Error patterns of Mandarin disyllabic tones by Japanese learners

Jung-Yueh Tu [1], Yuwen Hsiung [1], Min-Da Wu [1], & Yao-Ting Sung [1, 2]

[1] Center of Learning Technology for Chinese, National Taiwan Normal University
[2] Department of Educational Psychology and Counseling, National Taiwan Normal University

royaldu@ntnu.edu.tw, ywhsiung@ntnu.edu.tw, tarzan8720@yahoo.com.tw, & sungtc@ntnu.edu.tw

Abstract

Previous studies on Mandarin tone production indicate that there is no agreement on which tones are most difficult for L2 learners. Much of previous research on L2 learning of Mandarin tones has focused on monosyllables. In modern Mandarin, however, it is disyllabic words that dominate the vocabulary. This research investigates the production of Mandarin disyllabic tones by Japanese learners. In the current study, 25 Japanese learners of Mandarin were requested to produce 80 Mandarin disyllabic words with all tonal combinations (except for the neutral tone). The overall results showed a hierarchy of difficulty: Tone 3 > Tone 2 > Tone 1 = Tone 4. Most errors in the first syllable were found for Tone 2 and Tone 3 when followed by Tone 1 or Tone 4 (both start with a high pitch). In the second syllable, most errors were found for Tone 3 (misproduced as Tone 2). The findings are discussed in terms of the phonetic nature of Mandarin lexical tones and the interference from Japanese phonology.

Index Terms: Mandarin tones, disyllabic words, tonal production

1. Introduction

It is generally agreed that one of the most distinct features of Mandarin Chinese is lexical tone, which makes the mastery of the language quite challenging for most learners [1, 2]. Previous studies have suggested that language experience plays a crucial role in the processing of Mandarin tone [3]. It has been found that non-native processing of Mandarin tone reveals a somewhat different pattern from native speakers [4, 5]. In addition, L1 backgrounds of learners have great influence in the processing of Mandarin tone [3].

From the perspective of L2 speech, learning a tonal contrast involves both perception and production. While a majority of previous studies on the learning of Mandarin tones concern the processing of Mandarin tones in isolation, (i.e. monosyllabic words), this study investigates the production of Mandarin tone in disyllabic words. In Mandarin, it is the disyllabic words that dominate the modern vocabulary [1, 6, 7]. Also, disyllabic words can reflect the collaborative patterns of tones in speech, such as tone sandhi and tonal combination [1]. Thus, this study focuses on disyllabic words, intending to address the problem of tonal production in a larger linguistic units with contextual effects.

With an attempt to explore the effect of L1 on the tonal production in contexts, the current study examines the Japanese learners' production of Mandarin disyllabic tones.

1.1. Mandarin tone system

Mandarin is a tone language, which is typically described as having four lexical tones. Chao [9] used a numeric notation system on a 1-5 scale of pitch level (1 is the lowest and 5 is the highest of the speaker's pitch range) to distinguish among these tones: Tone 1, a high level tone (55); Tone 2, a mid- rising tone (35); Tone 3, a low-dipping tone (214); Tone 4, a high-falling tone (51). This describes the tone value for tones in isolation. It is noted that the rising tone (Tone 2) shows a dipping at the initial portion and then a moderately rising pattern in the final position [10]. In connected speech, it is found that Tone 3 always appears as a *half third tone*, with only the low-falling contour shape [11]. Tone 3 is changed into a rising tone similar to Tone 2 when this underlying Tone 3 is immediately followed by another Tone 3, which is so-called *third tone sandhi*. It has been found that in a Mandarin disyllabic word/phrase, the tone of the first syllable and the second syllable are compromised due to euphonious, and the pitch in the second syllable is usually lower than that in the first syllable [12].

1.2. Japanese pitch accent system

Japanese is classified as a pitch accent language. In a tone language, like Mandarin, each syllable of the word is associated with a separate tone, whereas a pitch-accent language uses pitch variation to give prominence to a syllable or mora within a word. Unlike Mandarin tone superimposed on monosyllabic words, Japanese pitch accent can be superimposed on disyllabic or multisyllabic words to make phonemic contrasts [13]. Mandarin is traditionally analyzed as syllable-based while Japanese is mora-based. While Mandarin has four tones with diversified pitches, standard Japanese (Tokyo dialect) has just two different pitches (H, L). Since Japanese is a pitch accent language, the location of the accent in a word is lexically indicated (i.e. marked in the lexicon), and given the location of the accent of a word, the pitch or tonal pattern of the entire word is predictable, unlike the case with tone languages [14, 15]. In Japanese, there can be only one peak with high pitch in a word. The pitches of the first and the second moras must be different.

1.3. Research questions

This study aims to investigate how Japanese learners produce Mandarin tones in disyllabic words. With an attempt to understand how and to what extent L1 experience with the prosodic features affect production of L2 tone categories, this study includes more participants and words than former studies [1, 8] to enrich empirical evidence as a ground for L2 speech research. The current study intends to answer two main research questions: a.)What are the general error patterns of Mandarin disyllabic tones by Japanese learners? b.)How does Japanese pitch accent influence tonal production of Mandarin?

2. Method

2.1. Participants

There were 25 Japanese learners participating in this study (15 females, 10 males; age range: 20-39 years old; mean age: 28.1

Copyright © 2014 ISCA

years old). (Originally 26 Japanese learners were recruited, but one participant had missing recording so that the data from the participant were not used in the analysis.) The participants all had less than 5 years of Mandarin learning experience before the experiment. They were from different regions of Japan and their native language was Japanese. None had difficulty in hearing and speaking.

2.2. Stimuli

The stimuli include 80 disyllabic Mandarin words with all tonal combinations, except for neutral tone (which is not the research target in this study). The stimuli cover four tones in the first syllable, four tones in the second syllable, and five disyllabic words in each tonal combination (4 tones * 4 tones * 5 disyllabic words = 80). The four tones were arranged into disyllabic words, then 16 tonal combinations can be retrieved. The numbers 1, 2, 3, 4 represent Tone 1, Tone 2, Tone 3, and Tone 4, respectively. Note that the tonal combination 3-3 should be pronounced as 2-3 due to the third tone sandhi. The design of word chart is illustrated below.

1-1, 2-1, 3-1, 4-1
1-2, 2-2, 3-2, 4-2
1-3, 2-3, 3-3, 4-3
1-4, 2-4, 3-4, 4-4

Since tonal production is dependent on participants' knowledge of the words (including their tonal specification), the disyllabic words in the stimuli were mostly taken from the *800 Chinese Words for Beginners* [16], which were familiar to participants. The order of the words were randomized to avoid participants' expectation of a pattern. Every word was presented with Mandarin phonetic symbols (Hanyu Pinyin) and Chinese characters.

2.3. Procedures

The production experiment was conducted in a quiet room. The utterances of the participants were recorded by a stand-alone microphone with a sampling frequency of 16 kHz and a resolution of 16 bit on a desktop. The recording was attained through using a recording program released by Speech and Machine Intelligence Laboratory of National Taiwan Normal University. Before the recording, the participants were familiarized with the practice section. During the recording, the participants saw a disyllabic word on the screen, and then clicked the recording icon when they were ready to produce the word. The participants were asked to read out these words with the correct lexical tones at a normal rate.

2.4. Analysis

From the recordings, the tonal errors made by the participants were judged by three phonetically trained native speakers of Mandarin. The 3 native speakers evaluated the recordings and labelled the tone of each syllable in the disyllabic words with a choice among the four lexical tones. When a production was too ambiguous and could not be categorized as any lexical tone, it would be labelled as the category 'other.' When there was any disagreement among them, the decision made by the majority was selected. Then the acoustic analysis was carried out using the software PRAAT with visual pitch contour to decide the label of the tone.

3. Results

3.1. Overall accuracy

One-way repeated-measures ANOVAs were conducted for the accuracy rate of tonal production. The result showed that the accuracy of tones was significantly different [$F(3,72) = 27.37$, $MSE = .012$, $p < .001$] (see Figure 1). The accuracy of Tone 3 was significantly lower than the other three tones, and that of Tone 2 was lower than Tone 1 and Tone 4.

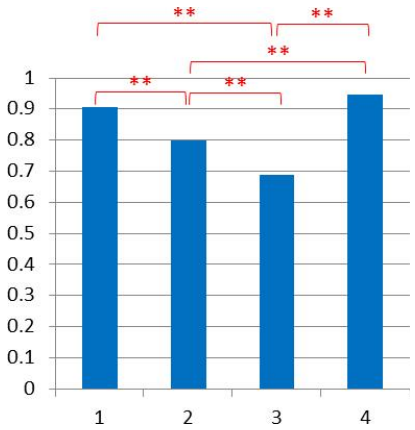

Figure 1. Accuracy of production for individual tone (* means $p < .05$; ** means $p < .01$).

3.2. Syllable effect

Paired Samples *t*-tests were conducted for the syllable effect on the accuracy of individual tone. The result showed that in Tone 2, the accuracy was significantly higher for the second syllable than the first syllable [$t(24) = 2.62$, $p < .05$]. For the other tones, the syllable effect on accuracy rate was not significant (all $p > .20$) (see Figure 2).

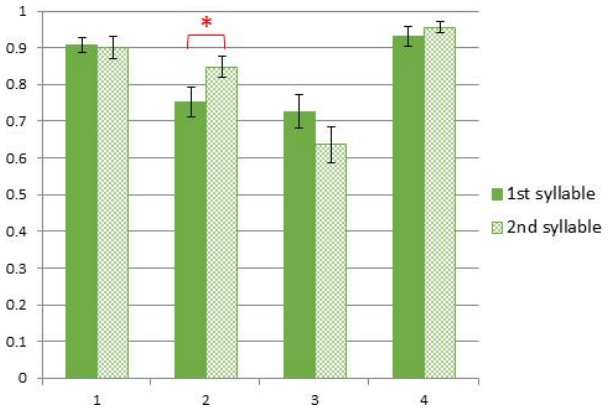

Figure 2. Comparison of accuracy between the two syllables for individual tone (* means $p < .05$; ** means $p < .01$).

3.3. First syllable

One-way repeated-measures ANOVAs were conducted for accuracy of tonal production in the first syllable. When Tone 1 in the first syllable is followed by different tones in the second syllable, its accuracy was significantly different [$F(3,72) = 3.14$,

MSE = .014, $p < .05$]. When Tone 2 in the first syllable was followed by different tones in the second syllable, its accuracy was significantly different [$F(3,72) = 11.06$, *MSE* = .058, $p < .001$]. When Tone 3 in the first syllable was followed by different tones in the second, its accuracy was significantly different [$F(3,72) = 4.69$, *MSE* = .014, $p < .01$]. When Tone 4 in the first syllable was followed by different tones in the second syllable, its accuracy was significantly different [$F(3,72) = 4.89$, *MSE* = .014, $p < .01$] (see Figure 3). As shown in Figure 3, the accuracy of production for Tone 2 followed by Tone 1 and Tone 4 was significantly lower than that for Tone 2 followed by other tones. In addition, the accuracy of production for Tone 3 in the first syllable was the lowest when Tone 3 was followed by Tone 4. It should be noted that due to the third tone sandhi rule, when Tone 3 was followed by Tone 3, the production of Tone 3 in the first syllable should be realized as a rising tone, similar to Tone 2. Thus, the accurate production for Tone 3 in the first syllable followed by Tone 3 was actually a rising tone, which was marked as T2 here.

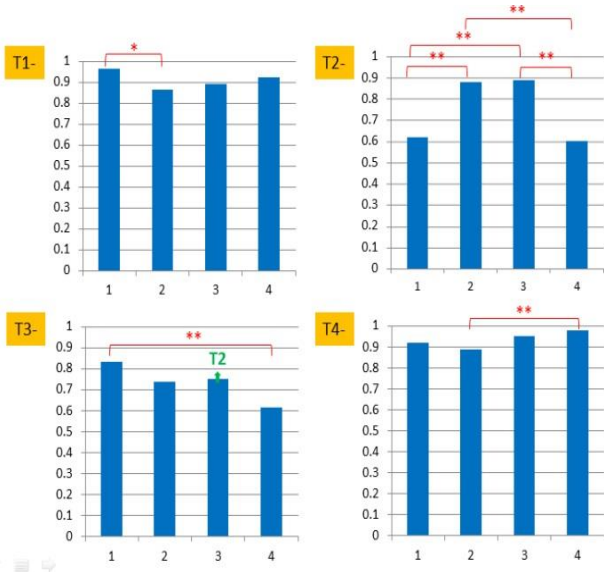

Figure 3. Accuracy of production for individual tone in the first syllable (* means $p < .05$; ** means $p < .01$).

3.4. Second syllable

One-way repeated-measures ANOVAs were conducted for accuracy of tonal production in the second syllable. When Tone 1 in the second syllable preceded by different tones in the first syllable, its accuracy was not significantly different [$F(3,72) = 0.57$, *MSE* = .019, $p = .64$]. When Tone 2 in the second syllable preceded by different tones in the first syllable, its accuracy was not significantly different [$F(3,72) = 1.65$, *MSE* = .029, $p = .19$]. When Tone 3 in the second syllable preceded by different tones in the first syllable, its accuracy was significantly different [$F(3,72) = 4.53$, *MSE* = .033, $p < .01$]. The accuracy of Tone 3 when preceded by Tone 3 was significantly higher than that preceded by Tone 4. When Tone 4 in the second syllable was preceded by different tones in the first syllable, its accuracy was not significantly different [$F(3,72) = 2.34$, *MSE* = .007, $p = .08$] (see Figure 4). As shown in Figure 4, the accuracy for Tone 3 in the second syllable was the lowest when Tone 3 was preceded by Tone 4.

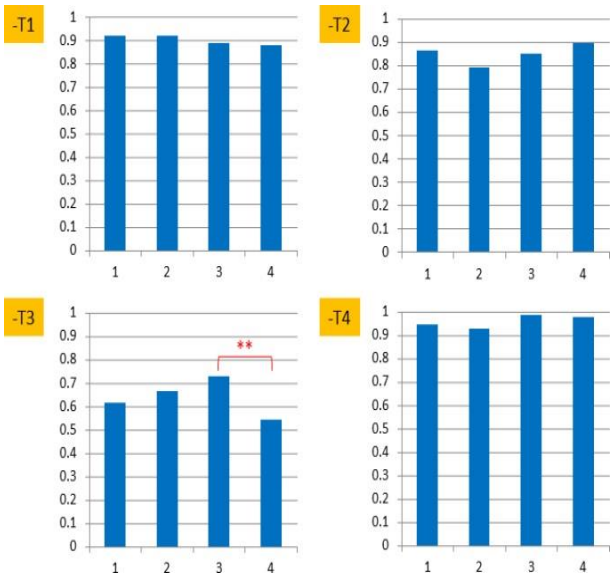

Figure 4. Accuracy of production for individual tone in the second syllable (* means $p < .05$; ** means $p < .01$).

3.5. Tonal error in each combination

The tonal error matrix for the first syllable and the second syllable in each combination are given in Table 1 and Table 2, respectively. (The correct response is highlighted in dark grey. When the percent of correct response is lower than 70%, the wrong response is highlighted in light grey.)

Table 1. Tonal error matrix- 1st syllable.

		1	2	3	4	Others
T1	11	96.0%	1.6%	2.4%	0%	0%
	12	86.4%	4.8%	1.6%	6.4%	0.8%
	13	88.8%	10.4%	0%	0.8%	0%
	14	92.8%	0.8%	2.4%	4.0%	0%

		1	2	3	4	Others
T2	21	6.4%	61.6%	30.4%	0.8%	0.8%
	22	9.6%	88.0%	1.6%	0%	0.8%
	23	7.2%	90.4%	0.8%	0.8%	0.8%
	24	5.6%	60.0%	32.8%	0%	1.6%

		1	2	3	4	Others
T3	31	1.6%	14.4%	83.2%	0%	0.8%
	32	4.0%	20.0%	76.0%	0%	0%
	33	13.6%	77.6%	4.0%	3.2%	1.6%
	34	6.4%	25.6%	61.6%	2.4%	4.0%

		1	2	3	4	Others
T4	41	6.4%	1.6%	0%	92.0%	0%
	42	9.6%	1.6%	0%	88.8%	0%
	43	3.2%	0%	0.8%	96.0%	0%
	44	2.4%	0%	0%	97.6%	0%

Table 2. Tonal error matrix- 2nd syllable.

		1	2	3	4	Others
T1	11	96.0%	1.6%	0%	2.4%	0%
	21	92.8%	3.2%	3.2%	0.8%	0%
	31	89.6%	5.6%	3.2%	1.6%	0%
	41	90.4%	1.6%	2.4%	5.6%	0%

		1	2	3	4	Others
T2	12	2.4%	86.4%	9.6%	1.6%	0%
	22	1.6%	79.2%	12.8%	0.8%	4.0%
	32	6.4%	84.8%	5.6%	0.8%	2.4%
	42	1.6%	89.6%	5.6%	2.4%	0.8%

		1	2	3	4	Others
T3	13	2.4%	24.0%	64.0%	0%	9.6%
	23	0%	22.4%	66.4%	0.8%	8%
	33	0%	19.2%	75.2%	2.4%	3.2%
	43	0.8%	35.2%	56.8%	0%	5.6%

		1	2	3	4	Others
T4	14	2.4%	0.8%	1.6%	94.4%	0%
	24	0.8%	2.4%	1.6%	92.8%	0%
	34	0%	0%	1.6%	98.4%	0%
	44	0.8%	0%	0.8%	97.6%	0%

4. Discussion

4.1. Tonal error patterns

The overall accuracy for individual tone showed the percent correct of Tone 3 is the lowest, followed by Tone 2, and then Tone 1 as well as Tone 4. Thus, it indicates the hierarchy of difficulty among the four lexical tones: Tone 3 > Tone 2 > Tone 1 and Tone 4.

With regard to the tonal combination, most errors in the first syllable were found in the tone pairs 2-1, 2-4, 3-4. That is, Tone 2 and Tone 3 were mostly misproduced when followed by Tone 1 or Tone 4. Tone 2 and Tone 3 were mostly misproduced as each other. It is probably because of the similar phonetic properties in Tone 2 and Tone 3, which both have a rising portion. Also, consider the phonological environment, the misproduced tone in the first syllable was followed by either Tone 1 or Tone 4, which both start with a high pitch. In the second syllable, most errors were found in the tone pair 4-3. Tone 3 was mostly misproduced as Tone 2, especially preceded by Tone 4.

The findings showed that confusion of Tone 2 and Tone 3 was a common error for Japanese learners. Tone 3 was the most difficult and mostly confused with Tone 2. Also, Tone 3 in the first syllable was usually not realized as a half third tone, but more as a rising tone, which might be perceived as Tone 2. In Mandarin speakers' pronunciation, Tone 3 in connected speech is usually realized as a half-third tone, but it is often produced as a full third tone in the first syllable by Japanese learners.

4.2. Interference from Japanese phonology

As mentioned earlier, Mandarin has four lexical tones while Japanese has only two pitches. From the production of Mandarin disyllabic tones by Japanese learners, it has been observed that Japanese learners generally have a narrower pitch range. In terms of tone register, Japanese learners often produced Tone 1 in Mandarin with a lower pitch. As for the tone contour, production of Tone 2 and Tone 3 by Japanese learners was somewhat different from that by Mandarin speakers. In the production of Tone 2, the pitch rising was not high enough; in the production of Tone 3, the dipping point was not apparent. Thus, the tonal production with narrower pitch range by Japanese learners may result in their tonal errors or foreign accent. The observation that Japanese speakers have a compressed pitch range is consistent with other studies on L2 speech sounds [17, 18, 19]. It seems that there is a general tendency for L2 speakers to compress pitch range, which needs further investigation.

In addition, the interference from Japanese language was revealed in the production of contour tones, i.e. Tone 2 and Tone 3. One type of errors was that Tone 2 was produced as a rising tone with an obviously level or a dipping part in the initial portion, which may sound like Tone 1 or Tone 3 plus Tone 2. A similar problem was also found in Tone 3 that it was produced as a low-dipping tone with a long rising part in the final portion, which probably sounds like Tone 3 plus Tone 2. This type of errors in contour tones is probably because tonal realization in Mandarin is syllable-based while in Japanese it is mora-based. Hence, the contour tones in Mandarin, which are usually longer in duration, were realized as bimoraic in the production of Japanese learners.

5. Conclusion

This study investigated the tonal production of Mandarin disyllabic tones by Japanese learners. The results showed that among the four lexical tones in Mandarin, Tone 3 was the most difficult and mostly misproduced as Tone 2. It was also observed that confusion of Tone 2 and Tone 3 was a common error for Japanese learners. Most errors in the first syllable were found for Tone 2 and Tone 3 when followed by Tone 1 or Tone 4. In the second syllable, most errors were found for Tone 3. Taken together, the findings suggested the effects of phonetic nature of Mandarin lexical tones and interference from Japanese phonology.

6. Acknowledgements

This research is partially supported by the "Aim for the Top University Project" and "Center of Learning Technology for Chinese" of National Taiwan Normal University (NTNU), sponsored by the Ministry of Education, Taiwan, R.O.C. and the "International Research-Intensive Center of Excellence Program" of NTNU and Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. NSC 103-2911-I-003-301.

7. References

[1] Chang, K.J., Chen, L.M, & Lee, N.C., "Tonal errors of Japanese students learning Chinese: A study of disyllabic words." Computational Linguistics and Chinese Language Processing, 11(3): 281-296, 2006.

[2] Guo, L & Tao, L., "Tone production in Mandarin Chinese by American students: A case study." In K.M. Chan and H.

Kang (Eds.), Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20), Vol. 1. (pp. 123-138). Columbus, Ohio: The Ohio State University, 2008.
[3] Ding, H., "Perception and production of Mandarin disyllabic tones by German learners." Speech Prosody, 2012.
[4] Jongman, A., Wang, Y., Moore, C., & Sereno, J.A., "Perception and production of Mandarin Chinese tones." In P. Li, E. Bates, L.H. Tan, & O. Tseng (Eds.), The Handbook of East Asian Psycholinguistics (pp.209-217). Cambridge University Press, 2006.
[5] Wang, Y., Jongman, A., & Sereno, J., "Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training." Journal of the Acoustical Society of America, 113: 1033-1043, 2003.
[6] Duanmu, S., "Stress and the development of disyllabic words in Chinese." Diachronica XVI, 1, 1-35, 1999.
[7] He, K. & Li, D., Xiandai Hanyu san qian changyong ci biao [Three thousand most commonly used words in modern Chinese]. Beijing: Beijing Shifan Daxue Chubanshe, 1987.
[8] Dong, Y., Tsubota, Y., & Dantsuji, M., "Difficulties in perception and pronunciation of Mandarin Chinese disyllabic word tone acquisition: A study of some Japanese university students." Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC 27), 143-152, 2013.
[9] Chao, Y. R., "A system of tone letters." La Maître phonétique 45, 24-27, 1930.
[10] Fon, J. & Chiang, W., "What does Chao have to say about tones?" Journal of Chinese Linguistics 27(1):13-37, 1999.
[11] Chao, Y.R., Mandarin Primer, Harvard University Press, 1955.
[12] Wu, Z. J., Introduction to the Phonetics of Modern Chinese, Peking: Chinese Language Teaching, 1992.
[13] Wu, X., Tu, J., & Wang, Y., "Native and nonnative processing of Japanese pitch accent." Applied Psycholinguistics, 33(3): 623-641, 2012
[14] Tsujimura, N., An Introduction to Japanese linguistics (2nd Edition), Oxford: Basil Blackwell, 2007.
[15] Tu, J., "Word Prosody in Loanword Phonology: Focus on Japanese borrowings into Taiwanese Southern Min." Doctoral dissertation, Indiana University, 2013.
[16] Steering Committee for the Test Of Proficiency-Huayu (SC-TOP), "800 Chinese Words for Beginners." Online: http://www.sc-top.org.tw/download/800Words_Beginners.pdf, accessed on 20 March, 2013.
[17] Ullakonoja, R., "Comparison of pitch range in Finnish (L1) and Russian (L2)." Proc. 16th International Congress of Phonetic Sciences (ICPhS XVI), Saarbrücken, 1701-1704, 2007.
[18] Traunmüller, H. & Eriksson, A., "F0-excursions in speech and their perceptual evaluation as evidenced in liveliness estimations." Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (PERILUS), 17, 1-34, 1993.
[19] Hincks, R., "Processing the prosody of oral presentations." Proc. InSTIL/ICALL2004 NLP and Speech Technologies in Advanced Language Learning, Venice, 63-66, 2004.