Production of Mandarin Disyllabic Tones by Vietnamese Speakers

Jung-Yueh Tu

National Chengchi University
jytu@nccu.edu.tw

ABSTRACT

This study investigates the tone production of Mandarin disyllabic words by Vietnamese speakers. In the study, there were 30 Vietnamese learners of Mandarin, who were requested to produce 80 (4 tones * 4 tones * 5 words of each tonal combination) disyllabic words in Mandarin. The overall results showed that Tone 4 was the most difficult. In the first syllable, most errors were found for Tone 3 when followed by another Tone 3 (the first Tone 3 should be pronounced as a rising tone but mispronounced as Tone 3), which indicated that Vietnamese speakers tend to underapply *Mandarin third tone sandhi*. In the second syllable, most errors were found for Tone 4 when preceded by Tone 4 (mispronounced as Tone 1). The findings will be discussed in terms of relevant L2 production models as well as effects of phonetic/phonological nature of Mandarin lexical tones and the interference from L1 phonology.

Keywords: Mandarin tones, L2 tonal production, disyllabic words, Vietnamese CSL learners

1. INTRODUCTION

The goal of this study to enrich the empirical evidence for the second language (L2) research. Models accounting for the production of L2 speech, particularly at the suprasegmental level, have been proposed. For example, *Prosodic Transfer Hypothesis* [1, 2] assumes that the types of prosodic representations that can be built in the L2 are restricted by the L1 prosodic constraints. Also, it hypothesizes that L2 learners fail to produce native-like speech because they are not able to construct prosodic representations required for the L2 speech production if those representations are disallowed in the L1. In addition, *Asymmetry Hypothesis* [3] proposes that the differences of prosodic domains and rule applications between L1 and L2 will affect L2 phonological acquisition. When the prosodic domain in L1 is smaller than L2, the learners will have positive evidence for the occurrence of alternation in a larger prosodic domain, in which the L2 acquisition will eventually take place. If the prosodic domain in the L1, however, is larger than L2, the learners will need negative evidence (e.g., forms and correction) for learning the proper domain setting of L2 alternation, in which L2 acquisition will not take place.

In terms of L2 speech, learning a tonal contrast involves both perception and production. Since most of previous studies on the learning of Mandarin tones concern the processing of monosyllabic tones, this study investigates the production of disyllabic tones. In modern Mandarin, it is the disyllabic words with a high percentage (69.8% of the total words) [4] that dominate the vocabulary [4, 5, 6]. Thus, this study aims to address the problem of tonal production in a larger linguistic unit with contextual effects, focusing on disyllabic words, which can reflect the collaborative patterns of tones in speech, such as tone sandhi and tonal combination.

In Mandarin Chinese, tonal acquisition is found to be difficult for L2 learners. How to master Mandarin tones is generally quite challenging for L2 learners. Building on the previous study regarding error patterns of Mandarin disyllabic words by Japanese and Korean learners [7, 8], the current study intends to expand our understanding of L2 tonal production by investigating the Vietnamese CSL learners with experience of another tonal language profile.

1.1. Mandarin tone system

As a tonal language, Mandarin has four lexical tones, including 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), and one neutral tone. The tonal system can be depicted either by a numerical system, with 5 indicating the highest point and 1 the lowest [9] or by a system of tonal features, H, M, and L, as presented in Table 1.

Table 1: Tonal system in Mandarin Chinese.

Tone categories	Tonal feature	Pitch value	Height
1	high-level (HH)	55	High
2	mid-rising (MH)	35	
3	low-dipping (MLH)	214	Mid
4	High-falling (HL)	51	Low

In connected speech, it is found that Tone 3 always appears as a *half third tone*, with only the low-falling contour shape [10]. A low-dipping Tone3 is converted to a rising Tone (similar to Tone 2) when followed by another Tone3, known as *third tone sandhi* [11].

1.2. Vietnamese tone system

Like Mandarin, Vietnamese is a tonal language. As for the number of tones in Vietnamese, most linguists ([11, 12, 13]) take there to be six, including thanh ngang 'flat,' a mid-level tone (33), thanh huyền 'deep,' a low-falling tone (21), thanh sắc 'sharp,'a mid-rising tone (35), thanh nặng 'heavy,' a mid-falling tone (3ˀ2ʔ), thanh hỏi 'asking,' a mid-falling(-rising) tone (313), and thanh ngã 'tumbling,' a mid-rising tone (3ˀ5). In Vietnamese, there are variations among speakers concerning how tone is realized phonetically. Also, there are differences among varieties of Vietnamese spoken in the major geographic areas (northern, central, southern) and among individuals [14]. The tone system in northern varieties is given in Table 2.

Table 2: Tonal system in (Northern) Vietnamese.

Tone ID	Tone name	Tonal feature	Pitch value
A1	ngang 'flat'	mid-level	33
A2	huyền 'deep'	low-falling (breathy)	21
B1	sắc 'sharp'	mid-rising, tense	35
B2	nặng 'heavy'	mid-falling, glottalized, heavy	3ˀ2ʔ
C1	hỏi 'asking'	mid falling(-rising), emphasis	313
C2	ngã 'tumbling'	mid rising, glottalized	3ˀ5

Revised from [11, 14]

Unlike Mandarin, Vietnamese tones do not rely only on pitch contour. Instead, Vietnamese often uses a register complex (which is a combination of phonation type, pitch, length, vowel quality and so on). The tonal feature of each tone is shown in Table 2.

1.3. Research questions

The current study aims to understand how native language experience with tones affects L2 speech production at the suprasegmental level by examining how Vietnamese speakers produce Mandarin tone in disyllabic words. The current study attempts to address three research questions: 1) What are the general error patterns of Mandarin disyllabic tones produced by Vietnamese CSL learners? 2) How do Vietnamese tone system and phonology influence tonal production of Mandarin? 3) How can the production of Mandarin disyllabic tones by Vietnamese speakers be accounted for by relevant L2 production models?

2. Method

2.1. Participants

There were 30 Vietnamese learners participating in the study (23 females, 7males; age range: 20-37 years old; mean age: 24.97 years old). All the participants have passed the self-made exam revised from TOCFL Band B test [15], which was corresponding to CEFR B1-B2 before the experiment. None had difficulty in hearing and speaking.

2.2. Stimuli

There were eighty disyllabic Mandarin words used in the stimuli list, which consisted of 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). Those words included all Mandarin tonal combinations, excluding the neutral tone. The four tones were arranged into disyllabic words, and then 16 tonal combinations can be retrieved (4 tones * 4 tones = 16 pairs). 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 similar to 2-3 due to the *third tone sandhi*. Consider that tonal production depends on participants' knowledge of the words with their tonal specifications, the disyllabic words were adopted from the stimuli used in [7], which were familiar to the participants. The presentation order of the words was randomized to avoid participants' expectation of a pattern. Every word was presented with Mandarin phonetic symbols (Hanyu Pinyin) and traditional Chinese characters.

2.3. Procedure

The experiment was conducted in a quiet room. The participants' utterances of the stimuli 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 by the software PRAAT. First, the participants were familiarized with the practice section. Later, they were requested to produce and record the stimuli. They saw a disyllabic word on the screen and then produce the word when they were ready. The participants were asked to read out those disyllabic words with the correct lexical tones at a normal rate.

2.4. Analysis

The participants' recordings of the stimuli were analyzed by three phonetically-trained native speakers of Mandarin. The recordings were judged by the 3 native speakers, who labelled the tone (with a choice among the four lexical tones) of each syllable of the disyllabic words. If the production was too ambiguous to be categorized as any lexical tone, it would be labelled as 'other.' When there was any disagreement on the tone labelling, 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 main effect of tone was significant in the study [MSE=.020, F=3.201, p=.026(<.05)]. The accuracy of Tone 4 (M=.825, SD=.187) was significantly lower than Tone 2 (M=.931, SD=.121) [SE=.036, p=.043(<.05)], and also lower than Tone 3 (M=.910, SD=.092) and Tone1 (M=.877, SD=.149). (see Figure 1)

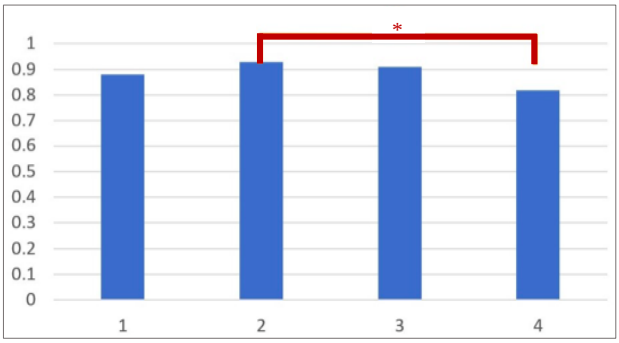

Figure 1. Accuracy of production for individual tone (*means p<.05).

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 1, the accuracy of the second syllable (M=.920, SD=.144) was significantly higher than the first syllable (M=.833, SD=.195) [t(29)=-2.802, p=.009(<.01)]. For the other tones, the syllable effect on accuracy was not significant (all p>.20). In Tone2, the accuracy of the second syllable (M=.945, SD=.138) was a little higher than the first syllable (M=.920, SD=.129). In Tone 3, the accuracy of the second syllable (M=.912, SD=.171) was also slightly higher than the first syllable (M=.908, SD=.126). In

Tone 4, the accuracy of the second syllable (M=.845, SD=.193) was also a little higher than the first syllable (M=.804, SD=.249). (see Figure 2)

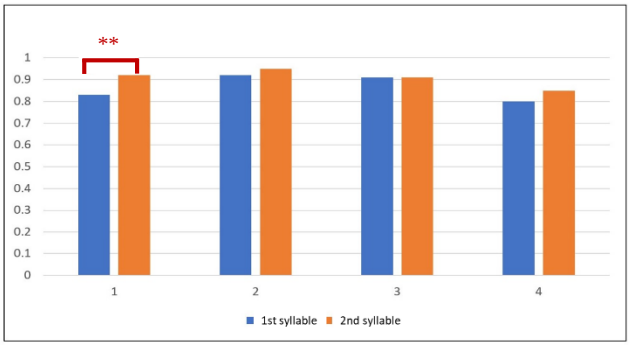

Figure 2. Accuracy for the four tones in the first and second syllables (**means p<.01)

3.3. First syllable

One-way repeated-measures ANOVAs were conducted for the accuracy of tonal production in the first syllable. The result (see Figure 3) showed that when Tone 1 in the first syllable was followed by different tones in the second syllable, its accuracy was not significantly different [MSE=.087, F=2.102, p=.104(>.05)]. When Tone 2 in the first syllable was followed by different tones in the second syllable, its accuracy was not significantly different [MSE=.034, F=.886, p=.451(>.05)]. When Tone 3 in the first syllable was followed by different tones in the second syllable, its accuracy was significantly different [MSE=.050, F=11.712, p<.001]. When Tone 4 in the first syllable was followed by different tones in the second syllable, its accuracy was not significantly different [MSE=.107, F=2.305, p=.081(>.05)].

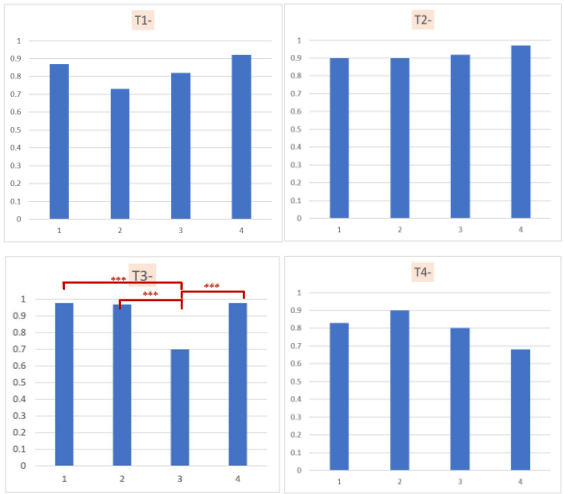

Figure 3. Accuracy of production for individual tone in the first syllable (***means p<.001)

As shown in Figure 3, the accuracy of production for Tone 3 followed by Tone 3 (M=.700, SD=.406) was significantly lower than Tone 3 followed by Tone 1 (M=.983, SD=.091), Tone 3 followed by Tone 2 (M=.966, SD=.126) and Tone 3 followed by Tone 4 (M=.983, SD=.091). Due to the *third tone sandhi* rule, when Tone 3 was followed by Tone 3, the first syllable should be pronounced as a rising tone, which is similar to Tone 2.

3.4 Second syllable

One-way repeated-measures ANOVAs were conducted for the accuracy of tonal production in the second syllable. The result showed that when Tone 1 in the second syllable was preceded by different tones in the first syllable, its accuracy was not significantly different [MSE=.049, F=2.442, p=.068(>.05)]. When Tone 2 in the second syllable was preceded by different tones in the first syllable, its accuracy was also not significantly different [MSE=.036 , F=2.392, p=.072(>.05)]. When Tone 3 in the second syllable was preceded by different tones in the first syllable, its accuracy was not significantly different as well [MSE=.53, F=0.978, p=.406(>.05)]. When Tone 4 in the second syllable was preceded by different tones in the first syllable, its accuracy was significantly different [MSE=.067, F=5.302, p=.002(<.01)].

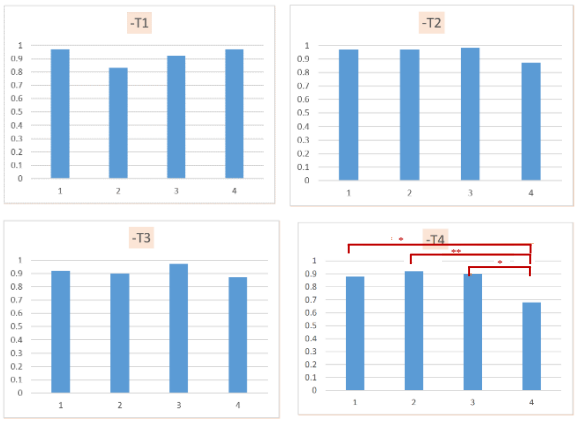

Figure 4. Accuracy of production for individual tone in the second syllable (**means p< .01, ***means p<.001)

As shown in Figure 4, the accuracy of production for Tone 4 preceded by Tone 4 (M=.683, SD=.352) was significantly lower than Tone 4 preceded by Tone 1 (M=.883, SD=.252), Tone 4 preceded by Tone 2 (M=.917, SD=.190) and Tone 4 preceded by Tone 3 (M=.9, SD=.359).

3.5 Error pattern in each tonal combination

The tonal error matrix in each combination for the first syllable and second syllable was shown in Table 3 and 4, respectively. Note that according to the *third tone sandhi* rule, a T3 followed by another T3 should be pronounced as a rising tone, similar to T2. Thus, in the tonal combination 3-3, the realization of T3 as T2 was presented as correct response. The correct response was highlighted in blue. If the accuracy rate of response was lower than 70%, the wrong response would be highlighted in pink.

Table 3. Tonal error matrix-1st syllable

		1	2	3	4	others
T1	11	0.88	0.00	0.00	0.07	0.05
	12	0.72	0.00	0.00	0.20	0.08
	13	0.82	0.02	0.00	0.15	0.02
	14	0.92	0.00	0.02	0.07	0.00
		1	2	3	4	others
T2	21	0.00	0.90	0.05	0.00	0.05
	22	0.00	0.90	0.07	0.00	0.03
	23	0.00	0.90	0.07	0.00	0.03
	24	0.00	0.97	0.02	0.00	0.02
		1	2	3	4	others
T3	31	0.00	0.02	0.98	0.00	0.00
	32	0.00	0.00	0.97	0.00	0.03
	33	0.00	0.70	0.22	0.00	0.08
	34	0.00	0.02	0.98	0.00	0.00
		1	2	3	4	others
T4	41	0.13	0.00	0.00	0.83	0.03
	42	0.08	0.00	0.00	0.90	0.02
	43	0.18	0.00	0.00	0.80	0.02
	44	0.27	0.00	0.00	0.68	0.05

Table 4. Tonal error matrix-2nd syllable

		1	2	3	4	others
T1	11	0.97	0.00	0.00	0.03	0.00
	21	0.83	0.00	0.00	0.12	0.05
	31	0.92	0.00	0.02	0.03	0.03
	41	0.97	0.00	0.00	0.03	0.00
		1	2	3	4	others
T2	12	0.00	0.97	0.03	0.00	0.00
	22	0.00	0.97	0.02	0.00	0.02
	32	0.00	0.98	0.02	0.00	0.00
	43	0.00	0.87	0.08	0.00	0.05
		1	2	3	4	others
T3	13	0.00	0.07	0.92	0.00	0.02
	23	0.00	0.05	0.90	0.00	0.05
	33	0.00	0.00	0.97	0.00	0.03
	43	0.00	0.08	0.87	0.00	0.05
		1	2	3	4	others
T4	14	0.10	0.00	0.00	0.88	0.02
	24	0.08	0.00	0.00	0.92	0.00
	34	0.05	0.00	0.00	0.90	0.05
	44	0.28	0.00	0.00	0.68	0.03

4. DISCUSSION

The overall accuracy rate for individual tone showed that the percent correct of T4 was significantly lower than T2. The result indicated a level of difficulty: T4> T1 & T3 > T2. That is, T4 is the most difficult and T2 is the least difficult. In addition, the syllable effect was found for Tone 1. That is, the accuracy of T1 in the second syllable was significantly higher than that in the first syllable, as indicated that the participants were better at producing Tone 1 correctly in the second syllable. Moreover, in the first syllable, the accuracy of production for Tone 3 followed by Tone 3 was significantly lower than that followed by the other tones. This result implied that the participants tend to *underapply* Mandarin *third tone sandhi*, which means they did not always pronounce the Tone 3 followed by another T3 as a rising tone (similar to T2) in the disyllabic words. In the second syllable, the accuracy of production for Tone 4 preceded by Tone 4 was significantly lower than that preceded by the other tones.

With regard to the error matrix in each tonal combination, most errors in the first syllable were found in the tone pairs 3-3 and 4-4. In the tone pair 3-3, Tone 3 in the first syllable was mostly mispronounced as Tone 3. In the tone pair 4-4, Tone 4 in the first syllable was mostly mispronounced as Tone 1. In addition, most errors in the second syllable were found in the tone pair 4-4, in which Tone 4 in the second syllable was mostly mispronounced as Tone 1. As previously mentioned, Tone 3 in the first syllable was mostly mispronounced as Tone 3 indicated that the participants tend to underapply the *third tone sandhi* rule. Some researchers document that in Vietnamese, sandhi occurs only in reduplicated constructions and that sandhi occurs only variably and does not apply to register, which must remain unchanging [16]. Thus, it is probably due to language transfer from Vietnamese phonology. Furthermore, the effects of Vietnamese phonology in the production of Mandarin tones were revealed in the tonal confusion between Tone 1 and Tone 4. As reviewed in 1.2., Vietnamese has six lexical tones. In Mandarin, Tone 1 (55) and Tone 4 (51) both start with a high pitch (5) while in Vietnamese, there is a mid-level tone (33) but no high-falling tone and the highest pitch of tones is (3). Therefore, the Vietnamese speakers tend to mispronounce Mandarin Tone 4 as Tone 1 for both the first syllable and second syllable in the disyllabic words.

The findings in this study suggest that the Mandarin tonal error patterns produced by Vietnamese learners were consistent with the assumptions in *Prosodic Transfer Hypothesis* and *Asymmetry Hypothesis*. The prosodic representations and rule applications in Mandarin bring about difficulties in L2 tonal acquisition for Vietnamese learners.

5. CONCLUSION

The current study investigated the tonal production of Mandarin disyllabic words by Vietnamese learners. The results showed that most errors were found in the tone pairs 3-3 and 4-4. The Vietnamese learners tend to underapply the *third tone sandhi* rule to Mandarin disyllabic words and mispronounce Tone 4 as Tone 1.

6.ACKNOWLEDGEMENTS

This study is funded by National Science and Technology Council, Taiwan (Grant No. MOST 110-2410-H-004-046). I thank Mei-chi Chao, Wei-ching Wang at National Chengchi University and Chih-yu Liao at University of Taipei for their help with partial data collection and analysis.

7.REFERENCES

[1] Goad, H., White, L. & Steele, J. 2003. Missing inflection in L2 acquisition: Defective syntax or L1-constrained prosodic representations? *Canadian Journal of linguistics, 48*, 243-263.

[2] Goad, H. & White, L. 2006. Prosodic transfer: L1 effetcs on the production L2 determiners. In: Bamman, D., Magnitskaia T., Zaller C. (eds), *Proceedings of the 30th Annual Boston University Conference on Language Development*, 231-224.

[3] Young-Scholten, M. 1994. On positive evidence and ultimate attainment in L2 phonology. *Second language research, 10*, 193-214.

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

[5] Duanmu, S.1999. Stress and the development of disyllabic words in Chinese. *Diachronica XVI, 1*, 1-35.

[6] He, K., Li, D.1987. *Xiandai Hanyu san qian changyong ci biao [Three thousand most commonly used words in modern Chinese]*. Beijing: Beijing Shifan Daxue Chubanshe.

[7] Tu, J., Hsiung, Y., Wu, M., Sung, Y. 2014. Error patterns of Mandarin disyllabic tones by Japanese learners. *Proceedings of the 15th Annual Conference of the International Speech Communication Association (Interspeech-2014)*, 2558-2562.

[8] Tu, J., Hsiung, Y., Cha, J., Wu, M. & Sung, Y. 2016. Tone production of Mandarin disyllabic words by Korean learners. *Proceedings of Speech Prosody 8*, 375-379.

[9] Chao, Y. R.1930. A system of tone letters. *La Maître phonétique 45*, 24-27.

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

[11] Doan, T.T. 2003. *A Concise Vietnamese Grammar*. Ha Noi: NXB The Gioi.

[12] Han, M. S. 1969. Vietnamese tones. *Studies in the Phonology of Asian Languages 8.* Acoustic Phonetics Research Laboratory, University of Southern California.

[13] Han, M.S., Kim, K.O. 1974. Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics 2*, 223-232.

[14] Vietnamese phonology https://en.wikipedia.org/wiki/Vietnamese_phonology, retrieved on December, 30, 2022.

[15] Test of Chinese as Foreign Language (TOCFL): https://tocfl.edu.tw/index.php/exam/test/page/1, retrieved on April, 16, 2023.

[16] Nguyen, A.-T. T., Ingram, J. C. L. 2007. Stress and tone sandhi in Vietnamese tone reduplications. *Mon Khmer Studies 37*, 15-39.