The Influence of L1 Tonal Background on the Acquisition of Cantonese Tones

Yu-Shen Liu
National Taiwan Normal University

Abstract

The present study looks into the perception of Cantonese tones by bilingual speakers of Mandarin and Taiwanese and investigates the extent to which L1 tonal backgrounds affect the acquisition of a new tonal system. The study recruited a group of Cantonese learners in Taiwan, who are all bilinguals of Mandarin and Taiwanese. The participants were asked to differentiate and recognize Cantonese tones in discrimination and identification tasks. The results show a discrepancy from previous findings that can be further explored and may serve as possible evidence that tone acquisition may be subject to the influence of multiple preexisting tone systems.

Keywords: tone, L2 acquisition, perception, Cantonese, Mandarin, Taiwanese

1. Introduction

1.1 Cantonese and Mandarin tones

Cantonese and Mandarin are both lexical tone languages, which apply tonal cues to contrast lexical meanings (Yip, 2002). There are four tones in Mandarin, traditionally described as High Level [55] (Tone 1), Rising [35] (Tone 2), Dipping [214] (Tone 3), and Falling [51] (Tone 4). As for Cantonese[1], there are six phonemic tones, including three level tones High Level [55] (Tone 1), Mid Level [33] (Tone 3), and Low Level [22] (Tone 6), two rising tones High Rising [25] (Tone 2) and Low Rising [23] (Tone 5), and one falling tone Low Falling [21] (Tone 4). Three extra tones, known as "entering tones", are often found in traditional description of Cantonese, designating Cantonese a language with nine tones. The three entering tones, High [5] (Tone 7), Mid [3] (Tone 8), and Low [2] (Tone 9), have similar f0 heights as the three level tones. They are independently categorized because they are characterized by having plosives at the end of the syllables and being relatively short in time span.

1.2 Theories of tone acquisition

[1] Hong Kong Cantonese.

Studies on phonetic learning have been focusing on the level of segmental features. Theories that have been proposed for tone acquisition are limited in number and mainly transplanted or extended from perception theories for segmental features. Some of these accounts are based on level of presentation, stating that speakers of non-tone languages cannot actually acquire lexical tones because their native grammar cannot perceive prosodic properties in a lexically contrastive way (Wayland & Guion, 2004). Others explain the process of acquisition from the perspective of category similarities. One example is Best's (1995) Perceptual Assimilation Model (PAM), which was originally developed to account for the acquisition of segmental features and later extended to the scope of suprasegmental features. The theory proposes that when exposed to foreign lexical tones, speakers of non-tone languages process these tones with reference to their native intonational categories while speakers of tone languages process the tones with reference to their native tone categories. Therefore, it is the degree of approximation/similarity between suprasegmental properties used in the two languages that may have an impact on the acquisition of foreign tones (Hallé et al., 2004)—whether the SLA learners' native languages are tonal or non-tonal is not the main concern here. Still others, as pointed in Wang et al. (2004), believe that the mere comparison between suprasegmental properties is insufficient and state that it is the degree to which acoustic features (average f0 or F0 contours) used to define tones that matters.

While it is plainly declared in Yip (2002) that so far little is known about tone acquisition, there have been an increasing number of studies focusing on the topic in the last decade. Among these studies, the comparison between performance of tonal and non-tonal language speakers has been of particular interest. It has been found that tonal background may facilitate tone acquisition. That is, speakers of tonal languages are more experienced than speakers of non-tonal languages in picking up tonal cues to contrast lexical meanings (Wayland & Guion, 2004). However, it has also been found that the existence of native tone knowledge is not necessarily advantageous in tone acquisition. In So & Best (2010), native speakers of Cantonese are reported to have greater difficulty distinguishing High Level and Falling Tones in Mandarin than English speakers, presumably because high level and falling tones are allotones of High Level Tone in Cantonese.

The acquisition of Cantonese tones is studied in Francis et al. (2008), which also compared the performance of tone language (Mandarin) and non-tone language (English) speakers. It is found that for Mandarin speakers, the High Level and High Falling Tones (Tone 1 and Tone 2) are most easily distinguished. The Mid-Level Tone (Tone 3) is also relatively easy, with 74.4% accuracy, even before training. The recognition of Low-Level Tone (Tone 6) seems to pose great difficulty to Mandarin speakers with 25.6% and 44.4% accuracy before and after training, respectively. Overall, there is a difficulty hierarchy of Low Level > Mid-Level > High Level of level tones in Cantonese for Mandarin speakers.

Based on previous findings, the present study goes further to look into the acquisition of a new tone system by bilingual speakers of two tonal languages, Mandarin and Taiwanese. Since Cantonese and Taiwanese both contain multiple level tones with varying f0 heights, it is possible that speakers of these languages are more sensitive to f0 height than those monolingual speakers of Mandarin.

2. Methods

Two tasks were conducted in this study: one discrimination task (Task 1), and one identification task with written cues (Task 2).

2.1 Subjects

Two native speakers of Mandarin and Taiwanese, both females, were recruited in the study. Subject 1, aged 25, has learnt Cantonese for 4 months. Subject 2, in mid-40, have learnt Cantonese for 3 months. The two subjects were classmates in a Cantonese course and received similar lecture. Neither of the two has received prior musical training, a factor that has been identified with regard to foreign tone perception[2].

2.2 Stimuli

Syllable /si/ and syllable /ji/ were used respectively in Task 1 and Task 2. The syllables were chosen out of some considerations. First, they can be found in most of

[2] Apart from L1 prosodic backgrounds, prior musical training has also been proposed as one of the possible confounding factors. It has been reported that, with regard to perception of foreign lexical tones, listeners who have received musical training in general perform better than those who have never received musical training (Gottfried & Riester, 2000).

the languages spoken in Taiwan and do not increase the loading of the subjects. Also, both syllables contain the same vowel /i/ to prevent any possible influence of alternative vowels on the perception of f0 heights. (Diphthongs are avoided for similar reasons.) The corresponding word lists in Cantonese can be found in Appendix 1.

The syllables were used in combination with the nine Cantonese tones, including six phonemic tones and three entering tones. In Task 1, each token is composed of a pair of two words. Both words have the same syllable /si/, bearing variant tones. A total of 45 combinations were made from the nine tones. The number was then multiplied by two with different ordering of the same two words in a pair, which makes a total of 90 tokens. A list of tokens can be found in Appendix 2.

For Task 2, both written cues and audio stimuli were provided. Each written cue contains one target syllable and one carrier sentence. Syllables were embedded in one of the two carrier sentences as shown below:

1. Sik^{1}	go^{3}	X	zi^{6}.	"[*Someone*] *knows the word X*"
識	個	X	字	
2. Hai^{6}	go^{3}	X	zi^{6}.	"[*It*] *is the word X*"
係	個	X	字	

The written cues were given only in Hong Kong Cantonese Romanization Scheme[3] and the Chinese characters are provided here for reference only. Chinese characters were avoided since they may prompt subjects with the Mandarin pronunciation of the characters and therefore hinder or even confound the subjects' performance. Audio inputs contain the same sentences as written cues, except that the target syllable may bear a different tone value. The target syllables in written cues were paired with audio inputs similarly as in Task 1, which makes a total of 90 tokens. All the tokens were randomly arranged in sequence.

It should be noted here that the carrier sentences were synthesized from recordings of single words and are not intended as natural stimuli with intonation. They are

[3] Also known as Jyutping, a Cantonese Romanization system that was designed and proposed by the Linguistic Society of Hong Kong in 1993.

provided solely as reference points of f0 heights. Since the perception of f0 height in tone recognition is relative in nature, one single syllable alone is insufficient to make judgments in an identification task. The case is especially true in Cantonese, which has several register tones in different f0 heights.

2.3 Procedure

As for procedures, Task 1 is an AX discrimination task in which subjects were asked to listen to a pair of two words and judge whether the two words have the same or different tone values. Each of the 90 tokens is separated by a clicking sound and there is a brief pause after every 30 tokens. The purpose of Task 1 is to evaluate subjects' on-line discrimination and see if the subjects are able to discriminate the tones in discussion.

In Task 2, an identification task, written cues were provided. Subjects were asked to listen to audio files and judge whether the tone value of the written cues meet the tone value of the audio input. They are instructed to answer "match" if they think the tone value of the written cues correspond to the tone value of the audio input and "mismatch" if they think the written cues mismatch the spoken cues. There are also 90 pairs of stimuli, each separated by a clicking sound, and a short break after every 6 pairs. The aim of Task 2 is to investigate the subjects' mental representation of the tones.

2.4 Results

2.4.1 Task 1

In general the accuracy is over 80%. If entering tones were considered as having the same f0 heights as level tones, then the accuracy is 86.67%. If entering tones were considered as having different f0 heights as level tones, that is, if subjects were allowed to take entering tones as different tones, the accuracy is even higher (93.33%). As shown in Table 1, subjects are able to distinguish one tone from another and tend to perceive entering tones as dissimilar from level tones.

Table 1. Overall accuracy of discrimination task in Experiment 1

	Subject 1	Subject 2	Average
Total no. of tokens	90	90	90

Total no. of correct answers for 6 tones	80	76	78
	(88.89%)	(84.44%)	(86.67%)
Total no. of correct answers for 9 tones	86	82	84
	(95.56%)	(91.11%)	(93.33%)

More details are revealed if we take a closer look to the confusions that the subjects have. The matching matrix in Table 2 was constructed based on the number of matching reports. Each time two sounds are reported as the same, it counts as one score. Hence, stimuli that have been reported as the same twice by both subjects respectively will gain 4 scores. Stimuli that have never been reported as the same gain 0 score and were omitted in the matrix.

Table 2. Matching matrix of discrimination task in Experiment 1

Word 1 (A)	Times of reporting as similar to Word 2 (X)								
	1	2	3	4	5	6	E1	E3	E6
1	4	-	1	-	-	-	-	-	-
2	-	4	-	-	1	-	-	-	-
3	-	-	4	-	-	-	-	-	-
4	-	-	-	4	-	1	-	1	-
5	-	2	-	-	4	-	-	-	-
6	-	-	-	-	1	4	-	-	-
E1	-	-	-	-	-	-	4	2	-
E3	-	-	-	-	-	1	1	4	-
E6	-	-	-	-	-	-	-	1	4

From Table 2, it is shown that all stimuli with two identical tones were correctly identified by both subjects. There seems to be virtually no difficulty if two similar sounds are identical.

As for confusions, there is a strong tendency to confuse High Rising Tone (Tone 2) and Low Rising Tone (Tone 5). The two contour tones have been reported as the same for three times (75%). Confusions between Low Falling Tone (Tone 4) and Low Level Tone (Tone 6) (25%), and Low Rising Tone (Tone 5) and Low Level Tone (Tone 6) (25%) were also reported. Since Tone 4 [21], Tone 5 [23] and Tone 6 [22] all have similar onset heights, the confusion may be accounted for as subjects using

the onset height as a cue to distinguish sounds and are less sensitive to the tone contours.

The number of confusions increases in the case of entering tones. High Entering Tone (E1) and Mid Entering Tone (E3) tend to be perceived as the same and have been reported so three times (75%). The confusion can be seen as analogous to that between High Level Tone (Tone 1) and Mid-Level Tone (Tone 3) (25%), the two level tones with similar f0 heights as E1 and E3. Mid Entering Tone (E3) was also confused with Low Entering Tone (E6), Low Falling Tone (Tone 4), and Low Level Tone (Tone 6). The increasing number of confusions for E3 than Mid-Level Tone (Tone 3) may be due to the shorter timer span of the entering tones, thus increasing the difficulty of perception and differentiation.

Also, it is found that none of the subjects consider entering tones as having the same tone value as their level tone counterparts. That is, the subjects do not consider High Entering Tone, Mid Entering Tone, and Low Entering Tone to be the same as High Level Tone, Mid-Level Tone, and Low Level Tone. And time sequence, that is, the W1 + W2 order does not show obvious influence.

2.4.2 Task 2

The overall accuracy lowered to 67.78% in the identification task. Differentiating entering tones from their level tone counterparts does not cause differences to the accuracy; the accuracy remains the same in both conditions.

Table 3. Overall accuracy of identification task in Experiment 1

	Subject 1	Subject 2	Average
Total no. of tokens	90	90	90
Total no. of correct answers for 6 tones	59	63	61
	(65.56%)	(70.00%)	(67.78%)
Total no. of correct answers for 9 tones	59	63	61
	(65.56%)	(70.00%)	(76.78%)

As shown in Table 4, High Level Tone (Tone 1) is mostly recognized. The accuracy is followed by High Rising Tone (Tone 2) and Low Level Tone (Tone 6). Low

Falling Tone (Tone 4) and Low Rising Tone (Tone 5) are poorly recognized with merely one correct report. Mid-Level Tone (Tone 3) is not recognized at all.

As for confusions, it seems that the confusion between High Rising Tone (Tone 2) and Low Rising Tone (Tone 5) that was found in Task 1 can also be found here. Audio input Tone 2 is identified as Tone 5 once and Tone 5 is identified more often as Tone 2 and Tone 3 than itself. While there is much confusion for almost all the audio inputs, it should be noted that the recognition of audio input for Mid-Level Tone (Tone 3) is particularly poor. The Mid-Level Tone is identified either as High Level or Low Level Tones and is not recognized in all cases. Low Level Tone (Tone 6) is at lease recognized in some cases.

Table 4. Matching matrix of identification task in Experiment 1

Audio input	Identified as					
	1	2	3	4	5	6
1	6	-	-	-	-	-
2	-	4	-	-	1	-
3	4	-	-	1	-	2
4	-	-	3	1	2	2
5	-	2	2	-	1	-
6	-	1	2	-	1	4
E1	3	-	-	1	-	-
E3	3	-	-	-	-	1
E6	1	-	1	1	-	2

3. Discussion

The present study is a preliminary study with limited number of subjects. Still, some ideas can be drawn from the results of the two experiments. Maybe the most obvious in the findings is not an answer to research questions but rather indication of confounding factors that need to be taken into consideration if one is going to gain a fuller picture of tone acquisition.

The results reported here correspond to previous findings in general and show influences from both L1s. It is found that the High Level Tone and High Rising Tone are the easiest tones in Cantonese to distinguish and acquire, which conforms to most

of previous reports (Francis et al., 2008). What was not found in previous findings was a relative ease of identifying Low-Level Tone in comparison with Mid-Level Tone. In Francis et al. (2008), it is reported that none of the Mandarin-speaking subjects know any other tone languages, while the Mandarin-speaking participants in the present study are capable of speaking more than one tonal language. Since Cantonese and Taiwanese both contain two or more register tones with contrasting f0 heights, speakers of these languages may be more sensitive to f0 heights than monolingual speakers of Mandarin who have no other dialect experience. Still, it is premature to claim bilingualism as a causal factor. It is unclear whether these pre-tuned perception weightings to certain tonal cues will overlap, counterbalance, or interfere the speakers' perception and acquisition of a foreign tonal system and the extent thereto. It could be worthwhile, however, to take the factor into account. The discrepancy from previous findings that was reported here could be further explored and may serve as possible evidence that tone acquisition may be subject to the influence of multiple preexisting tone systems.

For future research, it would be helpful to have some both monolingual and bilingual Mandarin speakers as subjects to further investigate the impact of bilingualism on tone acquisition. Also, apart from f0 heights, other types of tone features can be compared and analyzed cross-linguistically. Finally, pre-test training can be provided in the study design to ensure that all subjects have at least some level of metalinguisitc knowledge with regard to the tone systems in discussion.

References

Best, C. T. 1995. A direct realistic view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 171-204). Timonium, MD: York Press, Inc.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. 2008. Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics, 36,* 268-294

Gottfried, T.L., & Riester, D. N. 2000. Preliminary results from study of pitch perception and Mandarin tone identification. *Journal of the Acoustical Society of America, 108,* 2604.

Hallé, P., Chang, Y.-C., & Best, C. T. 2004. Identification and discrimination of Mandarin Chinese tones by Chinese versus French listeners. *Journal of Phonetics, 32,* 395-421.

So, C. K., & Best, C. T. 2010. Cross-language perception of non-native tonal contrasts: effects of native phonological and phonetic influences. *Language and Speech, 53(2),* 273-293

Wang, Y., Behne, D., Jongman, A., and Sereno, J.A. 2004. The role of linguistic experience in the hemispheric processing of lexical tone. *Journal of Applied Psycholinguistics, 25,* 449-466.

Wayland, R. P., & Guion, S. G. 2004. Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning, 54,* 681-712

Yip, M. 2002. *Tone*. Cambridge: Cambridge University Press.

Appendix 1: Word lists

	Tone 1 55	Tone 2 25	Tone 3 33	Tone 4 21	Tone 5 23	Tone 6 22	Tone 7 5	Tone 8 3	Tone 9 2
/si/	詩 si1 “poem”	史 si2 “history”	試 si3 “test”	時 si4 “time”	市 si5 “market”	事 si6 “thing”	色 sik1 “color”	洩 sik3 “leak”	食 sik6 “eat”
/ji/	醫 ji1 “cure”	椅 ji2 “chair”	意 ji3 “meaning”	姨 ji4 “aunt”	耳 ji5 “ear”	二 ji6 “two”	憶 jik1 “memory”	咽 jit3 “moan”	翼 jik6 “wing”

Appendix 2: List of tokens in Task 1, Experiment 1

Token 1-18		Token 19-36		Token 37-54		Token 55- 72		Token 72-90	
Si2	Si1	Si3	Si4	Si1	Si8	Si8	Si4	Si7	Si1
Si8	Si5	Si3	Si3	Si8	Si3	Si8	Si1	Si7	Si6
Si6	Si5	Si7	Si3	Si4	Si1	Si2	Si5	Si4	Si7
Si2	Si7	Si1	Si5	Si1	Si2	Si2	Si4	Si5	Si2
Si6	Si8	Si1	Si1	Si2	Si3	Si8	Si9	Si5	Si1
Si4	Si2	Si4	Si5	Si5	Si5	Si1	Si7	Si6	Si2
Si5	Si4	Si3	Si2	Si6	Si9	Si4	Si3	Si9	Si9
Si4	Si6	Si9	Si1	Si1	Si6	Si1	Si9	Si6	Si4

Si5	Si3	Si3	Si9	Si9	Si9	Si6	Si7	Si7	Si8
Si2	Si2	Si7	Si5	Si6	Si6	Si2	Si9	Si8	Si2
Si8	Si8	Si9	Si8	Si7	Si4	Si4	Si4	Si5	Si5
Si5	Si9	Si4	Si9	Si8	Si7	Si9	Si5	Si3	Si6
Si4	Si4	Si2	Si2	Si2	Si6	Si5	Si7	Si5	Si8
Si1	Si3	Si3	Si1	Si4	Si8	Si8	Si8	Si6	Si3
Si9	Si6	Si9	Si4	Si9	Si3	Si3	Si8	Si5	Si6
Si3	Si7	Si9	Si2	Si3	Si5	Si7	Si7	Si1	Si4
Si8	Si6	Si2	Si8	Si9	Si7	Si6	Si6	Si1	Si1
Si6	Si1	Si7	Si7	Si7	Si2	Si3	Si3	Si7	Si9

Department of English
National Taiwan Normal University
Taipei, TAIWAN
Yu-Shen, Liu: ysl@std.ntnu.edu.tw

聲調背景對粵語聲調習得之影響

劉又慎
國立臺灣師範大學

摘要

本文試探討國臺雙母語者對粵語聲調之感知判斷，以期了解母語聲調背景對聲調系統習得之影響。本試驗參與者為臺灣地區國臺雙母語之粵語學習者，受邀於試驗中區辨與辨識粵語聲調，其試驗結果與現有文獻綜合討論，或可作為日後聲調習得研究就聲調習得是否受原有聲調系統影響之參考。

關鍵字：聲調、外語習得、感知、粵語、華語、臺語