Tonal Aspects of Languages 2016
24-27 May 2016, Buffalo, New York

How Pitch Moves: Production of Cantonese Tones by Speakers with Different Tonal Experiences

Mengyue Wu [1], *Brett Baker* [1], *Janet Fletcher* [1], *Rikke Bundgaard-Nielsen* [2,3]

[1] School of Languages and Linguistics, The University of Melbourne
[2] La Trobe University, [3] MARCS Institute Western Sydney University

mengyuew@student.unimelb.edu.au

Abstract

This study investigates how native prosodic systems as well as L2 learning experience shape non-native tone production in terms of tone movement, a primary cue to tone identity. In an imitation task, the six Cantonese tones were produced by four speaker groups: native Mandarin speakers (tonal), native English speakers (non-tonal), native English speakers with Mandarin learning experience (L2 tonal) and native Cantonese speakers (control group). The results indicate that native prosodic systems influence non-native tone production: Mandarin speakers are more accurate on pitch contour than pitch height while English speakers perform better on level tones than contour ones. Furthermore, L2 tonal experience assists L3 tone production, as English-speaking Mandarin learners produce Cantonese tones in the most native-like shape, outperforming both Mandarin and English speakers.

Index Terms: non-native production; lexical tones; tone movements

1. Introduction

Over 70% of world languages are tonal. Tones in African and American languages tend to contrast by relative pitch height, while tones in Asian languages are distinguished from each other either by pitch contour alone (e.g. Mandarin) or by both pitch height and pitch contour (e.g. Cantonese) [1,3,6,7]. In "tone" languages, at least three major traits can differentiate lexical tones; these are: 1) pitch height, where fundamental frequency (F0) values are relatively constant for level tones and contour tones change from high to low, low to high; 2) direction, where each tone has an upward/downward direction along with the change of the height; 3) duration, measured from the syllable beginning till the end.

The target language here, Cantonese, has six contrastive tones: a high level tone (55), a high rising tone (25), a mid-level tone (33), a low falling tone (21), a low rising tone (23) and a low level tone (22). Mandarin, another tonal language, has a smaller tone system than Cantonese (see Table 1). Cantonese and Mandarin differ not only in the number of tones but also the nature of the tones. All Mandarin tones have different tonal contours, while Cantonese tones differ in terms of contour and also in terms of pitch height. In English, where tones are post-lexical, pitch accents also have F0 characteristics e.g. L* (low) and H* (high) pitch accents dock onto a rhythmically prominent syllable of an accented word. They have a pragmatic rather than lexically contrastive function in the language, unlike Cantonese and Mandarin. It is thus of particular interest to examine how speakers with these different native prosodic systems produce non-native tones.

Table 1. *A Comparison of Cantonese and Mandarin tones*

Tone Types	Cantonese	Mandarin
Level	High Level (55) HL	High Level (55)
	Mid Level (33) ML	
	Low Level (22) LL	
Rising	High Rising(25) HR	High Rising (35)
	Low Rising (23) LR	
Falling	Low Falling (21) LF	High Falling (41)
Falling Rising	N/A	Low Falling Rising (214)

The influence of a native prosodic system on the perception of intonation contours is indicated in several studies [9, 20]. English speakers (ES) focus on pitch height when perceiving non-native tones, while Cantonese speakers (CS) pay attention to both pitch height and pitch contour [7]. This can result in ES experiencing difficulty in perceiving tones with similar pitch height but different contours. Two studies have found that ES encounter difficulties with a Mandarin tone pair that has different tone contours but similar pitch heights [11, 18]. Comparable results are indicated for German listeners [5]. Previous studies also suggest that general psychoacoustic features universally influence speakers' perception, regardless of language background; for example, in the similarity and distance between the two L2 tones. Having a tonal language background does not automatically ensure that L2 perception of another tone language is easier, although the error patterns are more consistent [13]. It is also likely that listeners from non-tone language backgrounds will not perceive tones categorically (the way that L1 tone language speakers do), but rather in a 'psychoacoustical' way [10].

Second language (L2) tones are reported by L2 language learners to be difficult to produce [6, 16, 18], and incorrectly produced tones can negatively affect L2 comprehensibility [8, 10]. Such problems highlight the importance of determining what factors play a role in L2 tone acquisition. Does native language (L1) experience with tones help? Does it help to have acquired something similar before, even in an L2 setting? Will L1 and L2 abilities transfer to a new language? The current study addresses these questions in the production of Cantonese tones by speakers with different linguistic experiences. Previous studies have used a range of analytical methods. Two of the most common are 1) plots of F0 onsets and offsets; 2) trajectories of tone movements. This study focuses on tone movement, as part of a method which will enable us to test two of the three traits as mentioned above (pitch height and pitch contour). F0 movement is measured at every 10% time point, thus clearly recording how change of pitch differs between

http://dx.doi.org/10.21437/TAL.2016-29

speaker groups over the entire duration of a syllable bearing the tone.

We expect that the characteristics found in perception studies can be extended to production. Regarding predictions for the current study, Mandarin speakers (MS) might have more problems with pitch height, as they are used to relying on pitch contour in their native language [16, 21]. Regarding ES, they may be able to produce tones in a manner similar to producing intonation [7]. English listeners have experience with pitch via post-lexical accentuation and intonation, so it is possible they can categorise tones into their intonation system by interpreting them as post-lexical pitch accents and boundary tones, particularly for citation forms.

While it is established that L2 perception and production are related, the exact nature of this relationship needs to be determined. This study attempts to extend perception findings to production and seeks to uncover whether 1) tone production is influenced by L1 in the same way as in perception, and/or 2) if Mandarin learning experience assist Cantonese production by English speakers (L3 production in this sense)

2. Methodology

2.1. Participants

Three different speaker groups participated in the current study: 20 native Beijing MS (*M* age = 23.8; *SD* = 2.85); 20 native Australian English monolinguals (M age = 22.7; *SD* = 3.25) and 18 native Australian English speakers with intermediate Mandarin learning experience (EM) (*M* age = 24.3; *SD* = 3.72). The EM participants were all undergraduate students taking Chinese courses 3A/3B at the University of Melbourne. No participants in the groups reported previous experience with Cantonese or extensive musical training. In addition, a control group of 20 native Hong Kong CS (*M* age = 23.9; *SD* = 1.95) participated in the study.

2.2. Stimuli

Three syllables (/ba:p/ /bi:/ /bu:/) in six tones were recorded by a female native CS (aged 23). These tokens were chosen as none form a real word carrying six tones. Fifty-four tokens (3 syllables × 3 repetitions × 6 tones) were played randomly.

2.3. Procedures

An imitation task was conducted to investigate speakers' production of Cantonese tones. The whole design was made with E-prime 2.0—the participants saw a page displaying 'Listen' and heard the tone. They were then instructed to click and proceed to the 'Say' page by pressing any key. When finished, they could press any key to access the 'Listen' page for another tone. The three syllables were divided into three blocks; however, the order of the tones was not fixed, making the experiment more difficult.

2.4. Data Analysis

2.4.1. Normalisation

Some normalization procedures were undertaken to enable comparisons between speakers. We applied the enhanced pitch-synchronous overlap-and-add (PSOLA) technique to the data, which alters the duration without exerting changes on the pitch values, as in [15]: For the duration normalization, we first identified the longest F0 contour within each category and lengthened all other tokens to this duration. Though this approach constrains the investigation of duration somewhat, this procedure enables the possibility of linking observed perceptual patterns with the F0 dimension.

After duration normalization, F0 values were extracted with *Praat* 5.3 and *R* 2.15, using the autocorrelation method, with ranges set differently for female and male speakers (70-400Hz for female, 50-300Hz for male). In order to get a relative value for better comparison, each F0 value is converted from Hz to a logarithm-based T value, using the formula below:

$$T = \left[(lgX\text{-}lgL)/lgH\text{-}lgL\right] \times 5$$

In this formula, X is the F0 value at the given point, L is the lowest F0 value and H the highest produced by the speaker. This T value ranges from 0 to 5, corresponding to Chao's tone system. In the current formula, 0 represents the lowest pitch (when X=L) and 5 is the highest (when X=H). This way of mapping the F0 values to the five-step tone number system allows for easier comparison between tones produced by participants [4, 5, 9, 12, 13].

2.4.2. Tone Movement Plots

As pitch movement is an essential cue to tone identity in Cantonese, tonal contours have been plotted at every 10% time points. All of these analyses are based on equalized duration and normalized F0 values (T value).

3. Results

Tone trajectories by the four speaker groups are illustrated in Figures 1-4. The patterns of the production of Cantonese tones by native Cantonese speakers (Figure 1) were very similar to those found in previous studies. We note however, that the native speakers did not reach 5 in either T55 or T25, and that they produced small pitch movements, around 0.05. Some overlap can be found for T25, 21, 23, 22 before 20% into the syllable as they shared similar F0 onsets. The two rising tones 25 and 23 were almost identical before 30% of time, after then T25 rose higher. There was greater difference between T55 and T33 than between T33 and T22, though the latter two tones were still easy to separate. T21 started to drop from about the 20% time point. From offsets, we can see that T23 and T33 had a lot of overlap from 70% duration to the end. T25 and T55 had similar F0 offsets, which were around 4.5.

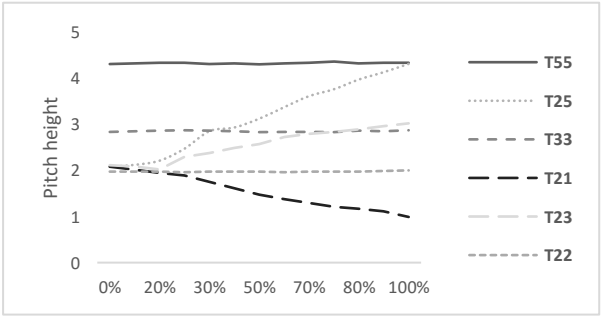

Figure 1: *Tonal Contour by native Cantonese Speakers*

Mandarin speakers (Figure 2), whose native language is tonal as well, had quite different production patterns. For example, T55 had a lower pitch level than native speakers and

the discrepancy between T55, T33 and T22 was much smaller compared to native speakers, especially for T22 and T33, which are fairly close to each other. Such a small difference could potentially cause perceptual confusion. The two rising tones T25 and T23 had different F0 onsets where they should be similar. But the Mandarin speakers clearly made the distinction between the rising slopes, though they were still different from native speakers' production. The falling contour had a less sheer fall before 70% and then dropped sharply from there until the end. However, it had a very high F0 onset, at around 3, possibly due to their native falling tone having quite a high onset (41). In general, MS were quite accurate in terms of the contour, but much less accurate in their sensitivity to pitch height.

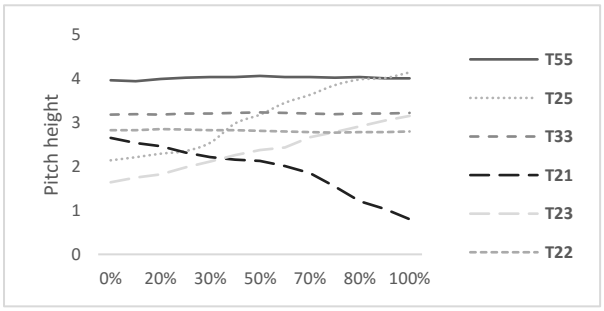

Figure 2: *Tonal Contour by Mandarin Speakers*

As shown in Figure 3, English speakers behaved differently and tended to produce every tone in a level shape – their production of the three contour tones: T25, 23 and 21 all had an F0 change ranges less than 2. However, the three levels were very native-like: they have similar level shape and pitch height, and the difference between T33 and T22 was still recognisable, with a discrepancy of about 1. But the onset area around 2 was very crowded – it was even difficult to tell T21, 23 and 22 apart before the 30% timepoint. The high rising tone T25 had a higher onset (about 3) however the low rising tone was produced in a more native-like way, partly due to the reason that the T23 had less F0 change itself. T21, interestingly, was produced with a sharp drop at around the 70% time points. In general, ES had more sensitivity to pitch height than did MS as they had great separation on the three level tones. However, their performance for contour tones was much poorer both in terms of pitch height and contour.

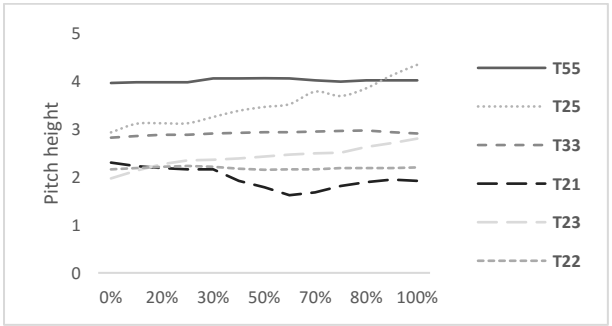

Figure 3: *Tonal Contour by English Speakers*

Figure 4 shows that English-speaking Mandarin learners had fewer problems than either Mandarin or English speakers. Surprisingly, they had the most native-like trajectory of the six tones. In terms of level tones, their high level tone was higher than the maximum of native Cantonese speakers. Their T33 was right above the 3 value and T22 was a bit lower than 2, which was quite native-like. Among the other three contour tones, T25 and T21 had similar F0 onsets but they proceeded in completely opposite directions. The other rising tone T23 had a lower onset but finally finished with the same offset as did T33. Roughly speaking, this production map was quite robust in terms of tone distinctions as each tone had clear path and little overlap with other tones, though the earlier parts of T23 and 22 were still very difficult to separate.

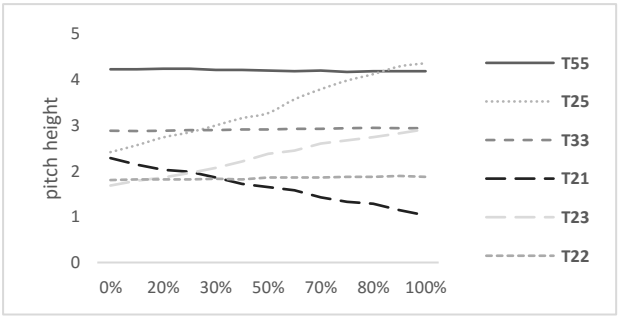

Figure 4: *Tonal Contour by English Speakers with Mandarin Learning Experience*

A two-way ANOVA was performed on all four groups' tone values at 10 timepoints – Timepoint is the within-subject factor while Group is the between-subject factor. The results revealed that for all tones, Group was a significant influencing factor (p <.001). For contour tones (25, 21 and 23), Timepoint was a significant influencing factor (p <.001), where significant tone movement is expected.

Further Tukey HSD tests revealed the difference between speaker groups when it comes to individual tones. For Tone 55, except from ES and MS, all other groups were significantly different from each other, p <.001. For Tone 25, MS and CS did not have significant difference between each other. The biggest difference can be found on English and Mandarin speakers where the p value is <.05. For Tone 33, except from EM and ES, all other groups are different from each other with Man and Can having the biggest difference 0.36. For Tone 21, ES was the only group having significant difference between CS: difference is 0.44 (p <.05). For Tone 23, significant differences can be found between all three non-native groups with the native speaker groups (p <.001). For Tone 21, significant difference was limited to CS and ES, CS and MS.

Another series of ANOVA were performed on each speaker group, with Timepoint and Tone type as two factors. For CS, MS and EM, Tone type has been significant factor: each tone is differently from each other. Timepoint, Tone Type and its interaction Tone × Type were all influencing factors (p < .001). For MS, F (5, 45) =40.575, 33 and 25, 22 and 25, 22 and 23 were not significantly different from each other. For ES, Timepoint was not a significant influencing factor, indicating that they failed to show significant pitch movement along time.

Tukey HSD tests indicated that for CS, all other tones are significantly different from each other (p <.001), except for Tone 22 and 21 (p <.05). The only indifferent pair is Tone 23 and 33 (p =.35). For MS, half tone pairs are significantly different from each other (p <.001), with most similar pairs being Tone 22 and 33 (p =.89), Tone 23 and 25 (p =.42). For ES, most tones are significantly different from each other (p <.01). For them, the most difficult pairs were Tone 22 and 25,

Tone 22 and 21 (p =.03), Tone 22 and 23. For EM, most tones can be differentiated (p <.001), though Tone 22 and 21, 22 and 23, 33 and 25 were slightly more difficult to pronounce for them.

Apart from the tone values at different time points, the pitch change range difference was analysed. The mean pitch change value was summarised in Table 2. The three level tones were clearly differentiated by CS: approximately 1.48 between T55 and 33, and 0.87 between T33 and 22. MS had quite small change ranges for all the three level tones, which were slightly larger than for CS, while their average F0 value is either too high (T33/22) or too low (T55). In particular, the difference between 33 and 22 produced by MS has a quite small difference, of 0.4. They had the biggest average pitch difference from the production of CS. ES had similar average pitch value with CS but the biggest pitch change ranges on the three level tones, followed by EM. EM, on the other hand, had almost the same pitch change range with MS production and similar average F0 values to ES.

For the two rising tones, native production had a wide pitch change range on T25 (2.23) and an average T value of 3.18. T23 had smaller pitch movement (1.09) and the average T value was around 2.56. MS had the most Cantonese-like pitch change range and average pitch height on T25. But they had a quite steep slope on T23, which should be very gradual. ES, however, had the smallest pitch change range on both T25 and 23. EM fell between the two groups on rising tones: they had a flatter shape on 23 than MS and wider pitch range on 25 than ES. ES had the smallest pitch change range on contour tones.

The low falling tone (21) had a native pitch movement of 1.09 and an average pitch value at 1.53. MS again had the most dramatic pitch change and a higher average pitch value. Similarly to rising tones, ES had the smallest pitch movement here. Interestingly, EM had the most native-like pitch change range and pitch height on this falling tone.

A two-way ANOVA was performed with Tone type as the within-subject factor and Group as the between-subject factor. It was found that both Group and Tone are significantly influencing the pitch change range (p <.001) but the interaction is not (p =.71). In Tukey-HSD tests, significant differences can be found between ES and CS, MS and CS (p <.001), CS and EM (p <.05) while MS and ES (p=.38) are not significantly different.

Table 2. *T-value change range*

	CS	**MS**	**ES**	**EM**
T55	0.02	0.05	0.06	0.04
T25	2.23	2.00	1.41	1.94
T33	0.03	0.04	0.09	0.05
T21	1.09	1.85	0.38	1.25
T23	0.99	1.51	0.83	1.23
T22	0.03	0.04	0.13	0.07

4. Discussion

Native as well as non-native experience with tonal languages exerts great impact on tone production. The results extend findings in L2/L3 perception to the production domain. Mandarin speakers are less accurate in their production of Cantonese tones that share the same tonal contour shape but where there are different heights for the native tone system compared to the non-native system. Mandarin Speakers tend to exaggerate pitch movement and have more problems with tones with medium pitch height. For the two rising tones, Mandarin speakers performed the best on T25, likely because the pitch range and pitch height are the closest to native production. T25 is more similar to Mandarin speakers' native rising tone which is a high rising 35. The low rising tone had a more dramatic rise than it should, which could be due to the fact that Mandarin speakers are most familiar with rising tones with large movements. The falling tone also shows a dramatic change for Mandarin speakers – their native falling tone starts from 4 and ends at 1, which may explain their production of Cantonese T21.

English Monolinguals, who have no experience with tonal languages, are quite sensitive to pitch height and have better performance on level tones than on contour ones. They tend to produce tones with less pitch movement. English speakers have small pitch movements on all tones, regardless of the tonal shapes. The possible reason is that they are much less sensitive to tonal contours thus they are less capable of producing them.

English-speaking Mandarin Learners, in contrast, are able combine their native sensitivity to pitch height and L2 experience with pitch contour. They have quite stable performance across all six tones: they are not as good as Mandarin Speakers on controlling pitch contour by stabilising level tones, but they are better than English Speakers on this aspect. Further, they outperform Mandarin Speakers on controlling pitch height. Mandarin Learners are the best speaker group when producing the more challenging tones, e.g. they have the most native-like production for the low falling tone.

5. Conclusions

The tone movement analyses in the present paper support the conclusion that native as well as non-native experience with tonal languages exerts great impact on tone production. In the present study, we show that Mandarin speakers are less accurate in their production of Cantonese tones that share the same tonal contour but where there are different heights for the native tone system compared to the non-native system. We also show that English speakers, who have no experience with tonal languages, are quite sensitive to pitch height and have better performance on level tones than on contour ones. English-speaking Mandarin learners, in contrast, are able combine their native sensitivity to pitch height and L2 experience with pitch contour.

This study contributes to the field of non-native tone production and the influence of tonal experiences on producing L2 and in addition, L3 tones. More research is in need to draw a solid conclusion on how native and L2 experiences tune L3 production at the same time.

6. References

[1] Abramson, A. S. (1962). *The vowels and tones of standard Thai: Acoustical measurements and experiments* (Vol. 20). Indiana University.

[2] Abramson, A. S. (1975). The tones of Central Thai: Some perceptual experiments. *Studies in Tai linguistics*, 1-16.

[3] Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology* (Vol. 102). Walter de Gruyter.

[4] Chuang, C. K., & Hiki, S. (1972). Acoustical features and perceptual cues of the four tones of standard colloquial Chinese. *The Journal of the Acoustical Society of America*, *52*(1A), 146-146.

[5] Dreher, J. J., & Lee, P. C. E. (1968). Instrumental investigation of single and paired Mandarin tonemes. *Monumenta serica*, 343-373.
[6] Gandour, J. (1977). On the interaction between tone and vowel length: Evidence from Thai dialects. *Phonetica*, *34*(1), 54-65.
[7] Gandour, J. (1983). Tone perception in far eastern-languages. *Journal of Phonetics*, *11*(2), 149-175.
[8] Howie, J. M. (1974). On the domain of tone in Mandarin. *Phonetica*, *30*(3), 129-148.
[9] Ladd,D.R.,Silverman,K.E.A.,Tolkmitt,F.,Bergmann,G.,&Scherer,K.R.(1985).Evidencefortheindependentfunctionofintonationcontourtype, voice quality, and F0 range in signaling speaker affect. *Journal of the Acoustical Society of America*, 78(2), 435–444.
[10] Lin, M., & Yan, J. (1988). The characteristic features of the final reduction in the neutral-tone syllable of Beijing Mandarin. *Phonetic Laboratory annual report of phonetic research*, *37*, 51.
[11] Ohala, J. J., & Ewan, W. G. (1973). Speed of pitch change. *The Journal of the Acoustical Society of America*, *53*(1), 345-345.
[12] Peabody, M., & Seneff, S. (2009). Annotation and features of non-native Mandarin tone quality. *Interspeech*, 460–463.
[13] Qin, Z., & Mok, P. P. (2011). Discrimination of Cantonese Tones by Speakers of Tone and Non-tone Languages. *Kansas Working Papers in Linguistics*, 34.
[14] Qin, Z., & Jongman, A. (2015). Does second language experience modulate perception of tones in a third language? *The Journal of the Acoustical Society of America*, *136*(4), 2107-2107.
[15] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.URL http://www.R-project.org/.
[16] Rose, P. (1987). Considerations in the normalization of the fundamental frequency of linguistic tone. *Speech Communication*, 6(4), 343–352.
[17] Repp, B. H., & Lin, H. B. (1989). Acoustic properties and perception of stop consonant release transients. *The Journal of the Acoustical Society of America*, *85*(1), 379-396.
[18] So, C. K. (2006). Perception of non-native tonal contrasts: Effects of native phonological and phonetic influence. *Proceedings of the 11thAustralian international conference on speech science & technology*, 438–443.
[19] Selkirk, E., & Shen, T. (1990). Prosodic domains in Shanghai Chinese. *The phonology-syntax connection*, *313*, 37.
[20] Ulbrich, C. (2008). Acquisition of regional pitch patterns in L2. *Speech Prosody*. 575-578.
[21] Wang, Y., Jongman, A., & Sereno, J. (2003). Acoustic and perceptual evaluation of Mandarin tone production before and after perceptual training. *Journal of the Acoustical Society of America*, *113*, 1033–1044.
[22] Wu, X., Munro, M. J., & Wang, Y. (2014). Tone assimilation by Mandarin and Thai listeners with and without L2 experience. *Journal of Phonetics*, *46*, 86-100.