Speech Prosody 2016
31 May - 3 June 2016, Boston, USA

Native Prosodic Systems and Learning Experience Shape Production of Non-native Tones

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

[1] School of Languages and Linguistics, The University of Melbourne
[2, 3] La Trobe University, MARCS Institute Western Sydney University
mengyuew@student.unimelb.edu.au

Abstract

This study investigates how native prosodic systems and second language (L2) learning experience shape non-native tone production. Speakers from tone language backgrounds (native Cantonese and Mandarin speakers [CS & MS]) and non-tone language backgrounds (English monolinguals [ES] and English speakers with Mandarin learning experience [EM]) produced the six Cantonese tones in an imitation task. The results suggest systematic effects of native prosodic systems on L2 tone production, regardless of tone or non-tone language backgrounds. MS have more problems with pitch height whereas ES tend to produce every tone in a level shape, which echoes the findings from previous perception studies. Further, MS's ability to integrate their native sensitivity to pitch height, along with their Mandarin training in pitch contour, contributes to their exceptional performance in producing the new tone language. Importantly, EM speakers performed better than MS speakers, suggesting that L1 experience with tone may be less helpful to learners than L2 tone acquisition experience, even when this L2 experience is with a different tone language (here Mandarin).

Index Terms: non-native speech production, lexical tone, linguistic experience.

1. Introduction

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.

In the current study, we examine the production of Cantonese tones by four different speaker groups who differ systematically in their native and non-native language experience with tone. These include: Cantonese speakers (L1 Cantonese tones); Mandarin speakers (L1 tone language with no Cantonese experience); English speakers (L1 non-tone language with no tone experience); and English speakers who are Mandarin learners (L1 non-tone with L2 tone experience). Such a design is crucial to our understanding of the factors that shape non-native tone production. A perception study [17] has found that both L1 and L2 experiences (English L1 and Mandarin learning as L2) contribute to L3 perception (Cantonese tones). It further proposes that when L2 and L3 belong to the same language typology, L2 modulates L3 perception. We extend this finding to production and investigate how L1 prosodic system and L2 experience influence non-native tone production.

Every language has its own prosodic system and uses fundamental frequency for different functions (e.g., for lexical contrast or for post-lexical functions). In tone languages, at least three major traits can differentiate lexical tones. These are 1) pitch height, 2) pitch contour and 3) duration. In English—a non-tone language—post-lexical pitch accents also have F0 characteristics (e.g., L* (low) and H* (high) pitch accents), which dock onto a rhythmically prominent syllable of an accented word. However, these have a pragmatic rather than lexical contrast function in the language. The target tone 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 tone language, has a smaller tone system than Cantonese. The comparison of this is presented in Table 1. Their tone systems differ from each other not only in quantity but also in the tones' nature. All Mandarin tones have different tonal contours, while Cantonese tones not only show difference in contour but also in pitch height.

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

Tone Types	Cantonese	Mandarin
Level	High Level (55)	High Level (55)
	Mid Level (33)	
	Low Level (22)	
Rising	High Rising (25)	High Rising (35)
	Low Rising (23)	
Falling	Low Falling (21)	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 by several studies [9, 20]. 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

10.21437/SpeechProsody.2016-120

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].

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 L2 experiences assist L3 production.

2. Methods

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 (/baap/ /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. Procedure

An imitation task was conducted to investigate speakers' production of Cantonese tones. The experiment was conducted in the MARCS Auditory recording studio, with a head-mounted microphone (Sennheiser SC230ML). 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. As each participant produced three repetitions, we were able to test the consistency of their production.

2.4. Data Analysis

2.4.1. Normalisation

Every speaker has a unique pitch range, ensuring it is impossible to have identical F0 patterns produced by different speakers. The high and low tones are all relative to each speaker's local pitch range; thus for better comparison, the raw data needs to be normalised. Here, the duration was normalised and then pitch values were modified into relative values adjusted within the speaker's unique pitch range, following [22].

For the duration normalisation, the longest F0 contour among each category was first identified. Others were then lengthened to this duration to preserve all F0 information [22]. The enhanced pitch-synchronous overlap-and-add (PSOLA) lengthening technique was adopted. This technique alters the duration without exerting changes on the pitch values. Although it restrains the investigation of duration, this procedure enables the linking of observed perceptual patterns with the F0 dimension.

After the duration normalisation, pitch 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) [4, 14]. To achieve a relative value for better comparison, each F0 value was converted from Hz to a logarithm-based T-value, using the following formula:

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

In this formula, X is the pitch value at the given point, L is the lowest pitch and H the highest produced by the speaker. The T-value ranged from 0 to 5, corresponding to Chao's (1930) 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 altering the pitch values into numbers enables easier comparison [12, 13, 14, 21].

2.4.2. Plots of F0 onsets and offsets

A common way to measure tone production is to plot F0 onset and offsets, with ellipses tagged for each tone type. We examine the tonal space in this fashion. Tone differentiation in each speaker groups' rising tones was expected to cluster closer to the y-axis, as they have higher offsets than onsets, whereas falling tones would cluster closer to the x-axis. An example of this is seen in the tone production of Cantonese by children with cochlear implants [1, 2]. In these studies, ellipses were drawn to visualise the space for each tone type. They were calculated to encompass 95% of the data points. The more discrete these ellipses were from each other, the better the production accuracy.

Three further analyses have determined how tones were differentiated within the tonal space and among tonemes, based on previous studies [1, 2]. The parameters calculated consisted of the axes' lengths, the areas of the tonal ellipses and the distances between the centre points of each ellipsis. It should be noted that the values on the x- and y- axes (0–5) corresponded to the maximum and minimum of the T-values, which echoed Chao's system. All other analyses were based on these values.

1) Measuring the tonal space

To perform this analysis, the three most differentiated tones were first identified. In Cantonese, the most

differentiated tones are Tones 1, 2 and 4 (55, 25 and 21 in Chao numbers respectively). If we drew a line to link the centre points of these, we would get a triangle roughly standing for the F0 range of the speaker. Thus, the tonal space we are comparing has been formed by these three points. The larger the number, the bigger the tonal space it stands for.

2) Measuring tonal differentiation within the tonal space

This index is based on the area of the triangle formed by the three most distant tones:

$$Index\ 1 = \frac{At}{Ae1,2,4}$$

Tonal differentiation across the tonal space is a function of the area of the ellipse space of each tone against the span of the triangle ($A_{e1,2,4}$) mentioned above. The bigger the target tone ellipse area, the more likely it is to overlap with other tones. Thus, the smaller the number, the more differentiated are the tones.

3) Measuring differentiation among tonemes

This index is based on the x- and y-axes values of the tones. Its function is the ratio of the distance of the ellipse centres against the average lengths of the two axes for all tones (Ave. Ax1+2):

$$Index\ 2 = \frac{Dist.}{Ave.\mathrm{Ax}1 + 2}$$

A series of t-tests on the differences between the non-native groups against the native group can show the observed differences between groups. This method makes no assumptions about whether speakers produced tones correctly or not and ensures that it is suitable to compare production by different groups of speakers, especially non-native speakers whose tonal production abilities are not known. Index 1 is more concerned with how much space the ellipse occupies within the whole tonal space, while Index 2 focuses on the centre distances between tones.

3. Results

A few tokens were eliminated due to 'creaky' or 'over-breathy' voice quality. The total token numbers for different speaker groups are presented in Table 2. The total number for CS, MS and ES is 360 (20 speakers × 6 tones × 3 repetitions); for EM, 326 (18 speakers × 6 tones × 3 repetitions).

Table 2. *Token Numbers for Different Speaker Groups*

	CS	MS	ES	EM
/a:/	353	342	342	314
/i:/	340	354	352	311
/u:/	339	351	337	318

The relative pitch values at the onset and offset time points were extracted from *Praat* and plotted with *R* [4, 14]; ellipses covered 95% of the data points. Figures 1 to 4 are illustrated with vowel /a/.

Figure 1 shows that even native CS have some within-tone category variation—they did not produce tones at exactly the same position every time. However, they produced hardly any overlap between the tone ellipses, except for a small portion between Tones 33 and 22, the mid- and low-level tones. In contrast, ellipses in Figures 2 and 3 indicated a great deal of overlap for MS and ES—it was extremely difficult to separate the six tone ellipses for ES's tone productions. Ellipses in Figure 4 (EM) were more separate than either ES or MS, although not as discrete as with native CS; however, most of the six tones were recognisable.

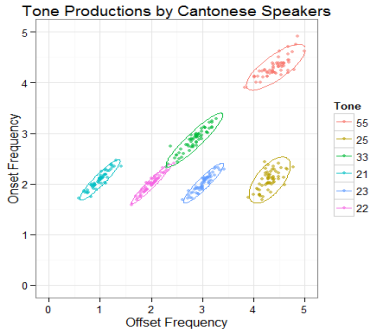

Figure 1. *Tone Production by Cantonese Speakers*

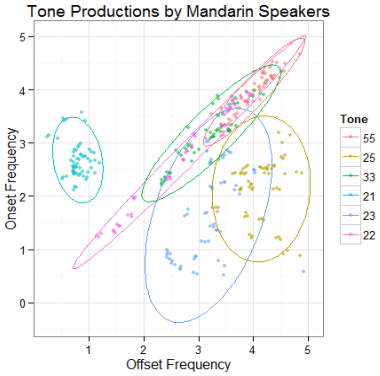

Figure 2. *Tone Production by Mandarin Speakers*

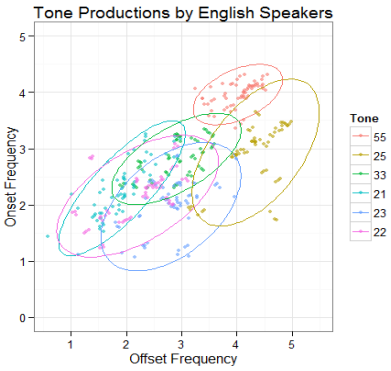

Figure 3. *Tone Production by English Speakers*

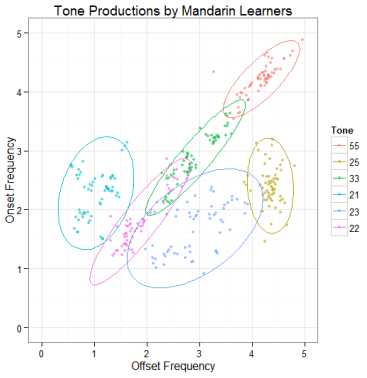

Figure 4. *Tone Production by Mandarin Learners*

1) Tonal space

The tonal spaces as defined earlier, formed by the three most distant tones (Tone 55, 25 & 21), were calculated on the basis of the relative onset and offset values of the centre points. CS tones had the largest tonal space of 3.89, followed by EM (3.06). This was slightly larger than MS (3.0), with the smallest tonal space by ES (1.35). Comparing the *t*-tests among the F0 onsets between all groups against native speakers, ES had the greatest difference (p=.008), followed by MS (p=.03). The difference between native speakers and EM was not significant (p=.86), nor was ES and MS (p=.45), while both ES and MS were significantly different from EM (p=.03 and .01 respectively). Interestingly, none of the three groups differed significantly from native speakers in off sets. In addition, the slope for each ellipse was compared with the regression model. The β coefficient for each tone type across speaker groups has been summarised as follows in Table 3: when the value is ≥ 0, the bigger the β value, the steeper the slope; when the value is < 0, then the slope is in the opposite direction and in this case, the smaller the β value, the steeper the slope.

Table 3. *Beta Values for Ellipse Slopes*

Groups	T55	T25	T33	T21	T23	T22
CS	.76	1.32	.93	.87	.84	.79
MS	1.06	1.88	.91	-1.46	1.69	.81
ES	.04	1.67	.56	.91	.72	.61
EM	.98	-1.75	1.02	1.54	.55	1.13

2) Tone differentiation within the tonal space

Table 4 presents the results of Index 1. Here, the smaller the number is, the more differentiated the tones are. Native speakers had the smallest number across all six tones, indicating that each tone took up quite a small part of the whole tonal space, leading to the least tonal confusion. In general, EM tones had smaller values than both MS and ES; except for T21, where MS had the smallest value. All the non-native speaker groups had the least tonal confusion on T55, which is probably because the high-level tone was the easiest tone to produce, regardless of language background. MS differentiated tones more effectively than ES across all categories except for T23, where ES outperformed MS.

Table 4. *Ellipse Areas against the Tonal Space*

Groups	T55	T25	T33	T21	T23	T22
CS	.35	.37	.29	.24	.23	.21
MS	.87	1.32	1.11	.65	2.03	1.74
ES	.73	.2.07	1.58	1.76	1.99	2.02
EM	.41	.57	.84	1.27	1.49	.99

3) Tone differentiation between tonemes

As shown in Figure 5, the larger the number for Index 2, the greater the difference between tonemes. This index shows the distances between ellipse centres: it is obvious that native speakers had the best tone differentiation. On this measure, the distances between ellipse centres for ES were fairly small, which means that a great deal of overlap between tones exists. However, if we look into their production plots (Figure 3) we can see that in general, the tones were realised at a reasonably correct pitch height, but with very different contours. For example, Tone 22 ranged from 1 to 3 in both offsets and onsets and Tone 33 was produced at a stable '3' level.

MS tones had the smallest distances between Tones 55, 33 and 22. This was especially so for 33 and 22, at 0.52. In addition, there was great overlap between the two rising tones (23 and 25), which should share similar onsets but which had completely different offsets. We can see from Figure 2 that the offsets for 23 ranged from 2 to 4 and that Tone 25 had a range from just over 3 to nearly 5. Thus, it was quite difficult to tell the two tones apart if both offsets overlapped. EM tones had slightly higher results than those for MS, representing better differentiation between tonemes. EM tones were better than ES at tonal contours and better than MS in terms of tonal height.

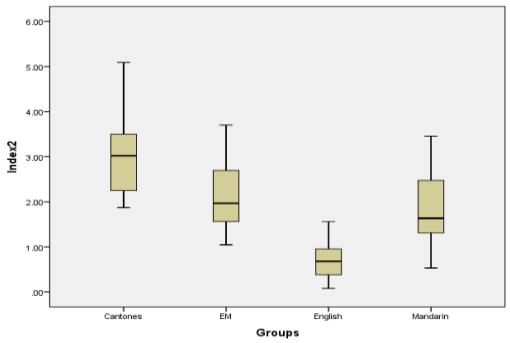

Figure 5. *Boxplot of Toneme Differentiation*

4. Conclusion

The most striking finding from this study is the fact that English speakers with L2 Mandarin outperform both L1 Mandarin and L1 English speakers. This indicates L2 learning experience shapes non-native production in a similar way to L1 experience does.

The current production results support the observations from perception studies that non-tone language speakers are more sensitive to pitch height and that tone-language speakers pay more attention to pitch contour. An L1 prosodic system greatly influences non-native tone production as well as perception. Regarding the three dimensions examined in this paper—tonal space, tone differentiation within tonal space and tone differentiation between tonemes—native CS are better than Mandarin learners of Cantonese, and better than MS. ES are least able to produce accurate Cantonese tone contrasts. MS are less sensitive to pitch contour, while ES pay more attention to pitch height, confirming previous findings [6, 7, 8, 10].

Speakers coming from a tone language background can produce tone contrasts more accurately than speakers with no prior tonal experience, given the evidence from tonal space and tone differentiation indices presented in this study. However, the fact that Mandarin learners perform better than both ES and MS, indicates that L2 experience can be transferred, along with native experience. Participants with native English experience are sensitive to pitch height; additionally, their L2 experience with Mandarin tones tunes their production abilities towards tone contour and possibly also to tone height. This is a topic for future investigation. Further analyses (e.g., F0 contours and acoustic duration) will be conducted to complement this preliminary analysis of Cantonese tone production by different speaker and learner groups. Future work shall extend the investigation to speakers with other language backgrounds (e.g., L2 English learners or L2 Thai learners).

5. References

[1] Barry, J. G., & Blamey, P. J. (2004). The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese. *The Journal of the Acoustical Society of America*, *116*(3), 1739–1748.

[2] Barry, J. G., Blamey, P. J., & Fletcher, J. (2006). Factors affecting the acquisition of vowel phonemes by pre-linguistically deafened cochlear implant users learning Cantonese. *Clinical Linguistics & Phonetics*, *20*(10), 761–780.

[3] Best, C. T., & Avery, R. A. (1999). Left-hemisphere advantage for click consonants is determined by linguistic significance and experience. *Psychological Science*, *10*(1), 65–70.

[4] Boersma, P. & Weenink, D. (2013):Praat: doing phonetics by computer [Computer program].Version 5.3.51, retrieved 2 June 2013 from http://www.praat.org/

[5] Ding, H., Hoffmann, R., & Jokisch, O. (2011). An Investigation of Tone Perception and Production in German Learners of Mandarin. *Archives of Acoustics*, *36*(3), 509–518.

[6] Francis, A., 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.

[7] Gandour, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11, 149–175.

[8] Gandour, J., Xu, Y., Wong, D., Dzemidzic, M., Lowe, M., Li, X., & Tong, Y. (2003). Neural correlates of segmental and tonal information in speech perception. *Human Brain Mapping*, *20*(4), 185–200.

[9] Grabe, M. E., Lang, A., & Zhao, X. (2003). News content and form implications for memory and audience evaluations. *Communication Research*, *30*(4), 387–413.

[10] Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421.

[11] Hao, Y.-C. (2011). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *40*(2), 269–279.

[12] Ladd,D.R.,Silverman,K.E.A.,Tolkmitt,F.,Bergmann,G.,&Scherer ,K.R.(1985).Evidencefortheindependentfunctionofintonationcont ourtype, voice quality, and F0 range in signaling speaker affect. *Journal of the Acoustical Society of America*, *78*(2), 435–444.

[13] Peabody, M., & Seneff, S. (2009). Annotation and features of non-native Mandarin tone quality. *Interspeech*, 460–463.

[14] 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/.

[15] Rose, P. (1987). Considerations in the normalization of the fundamental frequency of linguistic tone. *Speech Communication*, 6(4), 343–352.

[16] Qin, Z., & Mok, P. P. (2011). Discrimination of Cantonese Tones by Speakers of Tone and Non-tone Languages. *Kansas Working Papers in Linguistics*, 34.

[17] 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.

[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.