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

# The Acquisition of English Pitch Accents by Mandarin Chinese Speakers as Affected by Boundary Tones

*LIU Xing* [1], *CHEN Xiaoxiang*[2]

Department of English, Hunan University, P. R. China
[1]liuxing@hnu.edu.cn, [2]sophycxx@aliyun.com

## Abstract

This study investigates how Mandarin Chinese (MC) speakers learn English pitch accents under the influence of L1 prosody system. Based on Speech Learning Model (SLM), it is hypothesized that MC speakers will produce H* better than L* since MC speakers tend to associate stressed syllables with higher pitch. Another hypothesis is that MC speakers will produce pitch accents better in the sentence-final position due to tonal crowding effects and interrogative intonation features of MC. To test these hypotheses, 5 native English speakers and 12 MC speakers are invited to produce English declaratives and interrogatives so that different pitch accents can be elicited. A non-word pair of DAga and daGA only differing in stress position is embedded in sentence-medial and final positions. ToBI labeling shows how well MC speakers learn H* and L* respectively. Further statistical analyses of acoustic measures of pitch range and F0 change slope were conducted to triangulate the phonological data. The results show that MC speakers produce pitch accents as predicted. The study extends SLM in that learners encounter more difficulties in similar intonational categories as predicted.

**Index Terms**: pitch accents, intonation acquisition, boundary tones

## 1. Introduction

The influence of L1 on the acquisition of L2 sound system has long been regarded as an integral factor as evidenced by the ubiquity of “foreign accents” which are notoriously difficult to get rid of [1, 2]. However, to understand better how L1 influences function in L2 sound acquisition, there are 2 aspects in need of further research. First, more empirical evidence is needed to resolve the controversy over how L1-L2 similarities determine the difficulty of L2 acquisition. For example, while the tenets of Contrastive Analysis Hypothesis (CAH) [3] attribute the difficulty in L2 acquisition to L1-L2 dissimilarity, copious studies prove CAH wrong and many theoretical models are then formulated to predict that similarities pose more questions for learners. One of the typical theories is Speech Learning Model (SLM)[4], which hypothesizes that the “equivalence classification” based on L1-L2 similarities may block the establishment of a “new” category thus creating more difficulties for the learner.

Another aspect in L1 transfer is to extend L2 acquisition research from segmental features to suprasegmentals. When most transfer models focus on segmental aspects [4, 5], it is only until recently that a working model of L2 Intonation Learning theory (LILt) has been proposed by Mennen [6]. According to Mennen, L1 and L2 prosody systems can be compared from 4 dimensions (systematic, realisational, semantic and frequency) and hypotheses can then be formulated to predict the difficulty and process of L2 intonation acquisition.

This study, therefore, adopts this model to compare English and Mandarin Chinese (MC) prosodic systems and investigate how L1 prosodic features influence MC speakers in their acquisition of English pitch accents so that the issue of L1-L2 similarity/dissimilarity can be better understood in L2 prosody production.

### 1.1. Prosodic comparisons between English and MC

To understand how MC prosody system influence English pitch accent production, both English pitch accents and boundary tones are compared with their prosodic equivalents in MC based on LILt. The results from systematic and realisational levels are summarized in Table1.

| | English | MC |
|---|---|---|
| **Systematic** | Pitch accents (H*, L*) Boundary tones (L%, H%) | Pitch accents (?) Boundary tones (L%, H%) |
| **Realisational** | Higher pitch or lower pitch | Higher pitch, wider pitch range |

Table1. *Comparisons between English pitch accents and boundary tones and possible MC equivalents.*

English pitch accents are associated with prominence and they are an integral part of any intonational phrase. They can be both high-pitched and low-pitched, represented as H* and L* and their combinations in the Autosegmental-Metrical (AM) approach to intonation [7-10]. English pitch accents are anchored in lexically stressed syllables, which are acoustically more prominent with higher/lower pitch, longer duration, and less centralized vowel quality compared with unstressed ones. As for boundary tones, there are two types in English, H% and L%. Two typical combinations of tonal elements are H* L-L% used in declaratives and L* H-H% used in polar questions[11].

Unlike English, MC is a typical tone language with each syllable specified with a phonologically distinctive lexical tone [12]. There are controversies over the existence of pitch accents in MC intonation. For example, when Peng et al. included only stress levels in Mandarin intonation transcription[13], Lin argued that pitch accents are one of the two variables in Chinese intonation[14]. We follow Lin[14] in this study to assume there are pitch accents in Chinese so that English and MC pitch accents can be compared in the realisational dimension. However, since pitch accents in MC are associated with perceptually more salient syllables and realized with higher pitch contour and wider pitch range [14], there are no corresponding phonological element as L* in MC. Ou also found that MC speakers tend to perceive higher pitch syllables as being the locus of pitch accents more than they do

 10.21437/SpeechProsody.2016-196

with lower-pitched ones[15].

In MC, there are also boundary tones of L% and H% [13, 14], though, again, theoretical complications about the nature of boundary tones are found in literature[16]. Semantically, in MC, boundary tones carry the most important information for the listeners to decide if the sentence is a statement or question. Polar questions generally end with H%. An echo question without any interrogative particle tend to be perceived more as a statement if F0 of the last syllable is lowered [17].

Based on these results, it is worth investigating how Chinese speakers produce English L*, which finds no equivalence in MC. Especially, when it is aligned with the last syllable in a polar question, how will MC speakers produce L*, since it may be subject to the influence of tonal crowing [18, 19]? How will the overlapping of pitch accent L* and boundary tone H% influence their productions?

### 1.2. Hypotheses

Language acquisition models predict contrasting results based on cross-linguistic comparisons. For example, while CAH [3] predicts that the difficulty in acquisition lies mainly in the difference between L1 and L2, SLM[4] proposes 3 categories of different difficulty level: Equivalent, similar and different. The greatest difficulty comes from similar categories, while the different category being less difficult and identical features being the easiest.

Following SLM, our study will test the following hypotheses based on the comparison results from Section1.1.

1) MC learners of English can produce H* accurately because both English and MC have pitch accents and they are identically realized as high-pitched.

2) MC learners of English do not perform well in L* for they are similar but not equivalent categories.

3) Due to tonal crowding effects, MC speakers will produce L* better in the sentence-final position in the presence of H% brought by MC polar question intonation (the overlapping of L* and H%).

## 2. The study

### 2.1. Materials

To control for vowel context, we follow Shue et al. [18] to include a pair of non-words *DAga* and *daGA* differing only in the position of lexical stresses. Sentence types of statements and polar questions are designed to elicit the production of H* and L*. The 2 target words are embedded into sentence-medial and final positions so that the influence of boundary tones can be examined. 8 experiment sentences form 2 meaningful conversations with a picture showing the scenario. Table 2 gives an example of one of the conversations with target word *DAga*.

Table 2. *Conversation 1 of experiment materials.*

| | |
|---|---|
| Nancy: | Do you like ***Da**da* or ***Da**ga*? |
| Bob: | I like ***Da**ga*. |
| Nancy: (surprised) | Oh, did you say ***Da**ga*? |
| Bob: (bored) | Yes, I said ***Da**ga* just now. |
| Nancy: (unsure) | Really? Did you say ***Da**ga* just now? I thought you said ***Da**da*. |

### 2.2. Participants

5 native speakers (NS) (4 male and 1 female) are invited to read aloud the materials. They are native speakers of North American English from the mid-west United States, aged 22-34. Two learner groups are also recruited from first-year English major students at a key university in China. They are selected based on the class evaluation of "Pronunciation" course, which focus on both segment and intonation training of English pronunciation. Learners ranking as top 6 (3 female and 3 male) bottom 6 learners (2 female and 4 male) are included. They form 2 proficiency groups of advanced group (MC1) who are more native-like in pronunciation and beginner group (MC2) who need more practice in English pronunciation. All the learners have learned English as a foreign language for at least 10 years.

### 2.3. Recording procedure and data analysis

Recordings were made in a quiet office on campus. The experimenter first explained the purpose of recording to each speaker as to examine the acquisition of English lexical stresses. Real research purpose was not mentioned to ensure that participants can produce the intonation in a natural way. Speakers were then allowed enough time to practice and get familiarized with the materials. When they were ready, the experimenter recorded them with a digital recorder SONY ICD-SX713. After that, data were digitalized into WAV format on a MAC using Audacity 2.0.6 with 16k sampling rate. Data were then analyzed both qualitatively and quantitatively.

2 transcribers labeled data with MAE-ToBI. The inter-rater consistency is 86%. All learner data were compared with the production of native speakers. Deviations in pitch accent use were tallied and converted into percentage.

ProsodyPro[20] was then used to extract acoustic features of normalized F0 and F0 excursion of each pitch-accented syllable, as well as F0 change slope of target words[21]. ANOVA analyses of these acoustic measures were conducted with SPSS 22.0.

## 3. Results

### 3.1. Qualitative analysis

Native speakers demonstrate consistency of over 96% in pitch accent use with just one speaker using L+H* in a statement. We consider L+H* and H* as the same so that L2 production can be compared with that of native speakers.

L2 productions of pitch accents are compared with those in native speech. The cases of L2 deviation in all the experiment sentences are summarized in Table 3. Advanced learners in MC1 show fewer deviations of 12.5% from that of native speakers than the beginner group MC2 of 64.6%.

In statements to elicit H*, only 1 case of deviation is seen in advanced learner group. In MC2 group, although there are 11 cases of deviation, 6 of them are H*+L, which is not very distinct from H*.

The situation is much worse in polar questions that elicit L*. Much more deviations (25 cases) occur than in H*. In MC1 group, 5 deviations are found. In MC2 group, even more deviation cases occur. Instead of L*, the less experienced learners use H* or H*+L. One tendency worth noticing is that more deviations are found in sentence-medial than sentence-final positions in both learner groups. This may be related to tonal crowding effect (See Section 4.2 for further discussion).

| | | Non-final | Final | Total |
|---|---|---|---|---|
| MC1 | H* | 1 (2%) | 0 | 6 (12.5%) |
| | L* | 5 (10.4%) | 0 | |
| MC2 | H* | 6 (12.5%) | 5 (10.4%) | 31( 64.6%) |
| | L* | 10 (20.8%) | 10 20.8%) | |

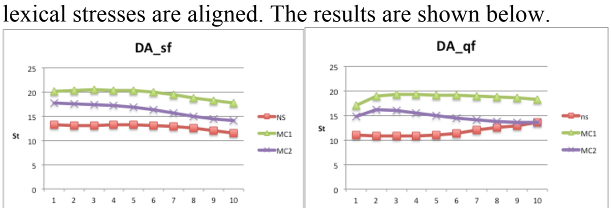
Table 3. *The number and percentage of L2 deviation in pitch accent use.*

### 3.2. Acoustic analysis

To examine how the phonological elements are realized phonetically and triangulate the qualitative data in the previous section, normalized F0 of target words in all productions (sentence type x position) of different groups were first converted into semitone and compared. Then the 2 measures $P_{F0}$ and slope were calculated to examine how pitch accents are realized.

#### 3.2.1. Normalized F0

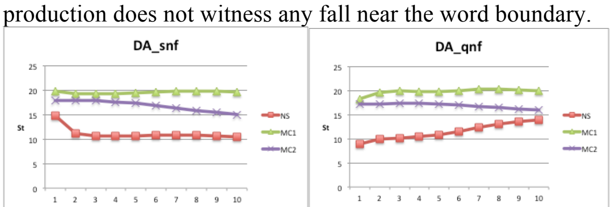
Normalized F0 values of pitch-accented words by different groups (NS, MC1 and MC2) were averaged respectively and compared. DA and GA indicate the syllables with which lexical stresses are aligned. The results are shown below.




1a) 1b)

Figure 1. *1a) F0 contour with DAga in final position in statements; 1b) F0 contour with DAga in final position in polar questions*.

As can be seen in Figure 1a), when DA is stressed lexically, MC1 and MC2 resemble NS performance in statements with DAga in sentence-final position by demonstrating H* and a fall near the word boundary, probably due to the boundary tone L% which aligns later with the syllable ga. However, as shown in 1b), for polar questions with DAga in sentence-final position, MC2 perform very differently from NS. Instead of demonstrating an L*, MC2 productions see an H* in the early part of the word and a fall by the end of the unstressed syllable ga, while MC1 production does not witness any fall near the word boundary.

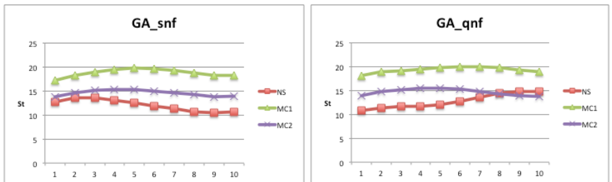



2a) 2b)

Figure 2. *2a) F0 contour with DAga in non-final position in statements; 2b) F0 contour with DAga in non-final position in questions*.

When producing DAga in non-final positions, although both learner groups perform like NS group by using H* in statements as shown in 2a), they both demonstrate a very different pattern from NS group by remaining high in polar questions as can be seen in 2b). However, NS start quite low from the beginning of the word and show a steady rise towards the word boundary.

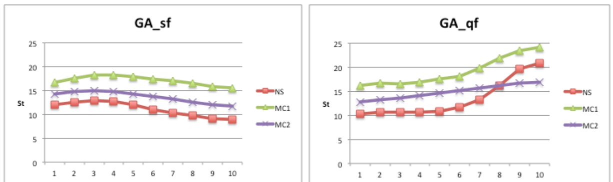
Figure 3 looks mostly the same as Figure2. With daGA in sentence-medial position, learner contour patterns do not change much with the only difference of lexical stress shifting from DA to GA, the second syllable. As can be seen from 3a) and 3b), both MC1 and MC2 groups demonstrate similar trends as in 2a) and 2b). The reason may be that no matter which syllable is lexically stressed, the pitch accent is not influenced by the boundary tone since the target word is not in sentence-final position.




3a) 3b)

Figure 3. *3a) F0 contour with daGA in non-final position in statements; 3b) F0 contour with daGA in non-final position in questions*.

However, as shown in Figure 4, the learner performance is significantly different from that of NS when daGA is in sentence-final situation, where the pitch accents overlap with boundary tones. 4a) shows F0 contour of GA when the target word *daGA is* in sentence-final position. Learner group performance resembles NS in demonstrating an H* but with an obvious fall possibly under the influence of the boundary tone L%. In 4b), both learner groups use a rising tone possibly under the influence of boundary tone H%, although it needs further explanation why MC2 do not demonstrate as much sharp a rise as the other 2 groups but a slight fall towards the end.




4a) 4b)

Figure 4. *4a) F0 contour with daGA in final position in statements; 4b) F0 contour with daGA in final position in questions*.

#### 3.2.2. $P_{F0}$

As mentioned above, $P_{F0}$ was calculated to understand the pitch range of each pitch-accented syllable. F0 excursion size of each target syllable (Excursion size 1) and the whole sentence (Excursion size 2) was first converted into semitone. $P_{F0}$ was then calculated by dividing excursion size 1 with excursion size 2.

$$P_{F0} = {Excursion\ size\ 1(St)}/{Excursion\ size\ 2\ (St)}$$

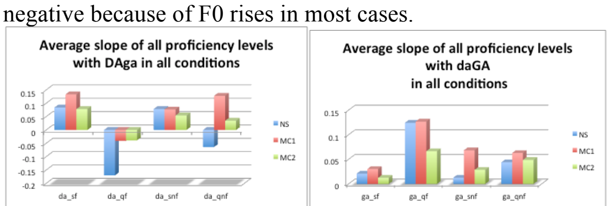
$P_{F0}$ of *DAga* and *daGA* in all experiment sentences was calculated respectively.

ANOVA analyses were then conducted with of group (3 levels), position (non-final vs. final) and sentence type (H* vs.

L*) as fixed factors. Results show that for DAga, none of the factors was significant. However, for daGA, the main effect of position was significant (Sig = .001). Mean values show that $P_{F0}$ in the final position is significantly higher, probably due to the effect of boundary tone.

### 3.2.3. *F0 change slope*

F0 change slope measures how fast F0 contour changes in the target word. In this study, we first calculated mean F0 difference (in HZ) between the first and second syllable in the target word, then divide the difference with the duration of the whole word (in MS). Figure 5 shows the average slope of F0 contour change for all three groups with DAga and daGA in sentence-medial and final positions in both statements and polar questions. For daGA (in Figure5a), the slope was mostly negative because of F0 rises in most cases.




5a) 5b)

Figure 5. *5a) Average slope of all 3 proficiency levels with DAga in all conditions; 5b) Average slope of all 3 proficiency levels with daGA in all conditions.*

ANOVA analyses were again conducted with group, position and sentence type as fixed factors. For DAga, sentence type (Sig = .000) is significant and proficiency is just above the significance level (Sig = .051). Post-hoc Tukey test showed that the major difference was between NS and MC1 (Sig = .005), but not between NS and MC2, which was a bit puzzling given the fact that MC1 are generally better in pronunciation. Further examination of data shows that mean value of NS is negative, and MC2 has the highest value. For daGA, however, the only significant factor is sentence type (Sig = .001), but not the other 2 factors. It is natural for slope to be steeper in questions than in statements.

## 4. Discussion

The experiment results largely confirm the 3 hypotheses to show the L1 MC prosody system does influence English pitch accent acquisition.

### 4.1. Pitch accent types

Qualitative data demonstrate that both high and low-level learners produce H* better than L* with much fewer deviations. It confirms Hypothesis 1 & 2 in that equivalent categories predicted by SLM can be easily acquired. In MC, phrase level prominence is realized with expansion of pitch range and lexical prominence is cued by higher pitch. However, MC does not employ low pitch to realize prominence as in English. Thus, although MC speakers produce H* with higher accuracy, L* still poses great difficulty for MC speakers. This can be explained by SLM that low-pitch accents can be considered a similar category, which poses the greatest difficulty for learners.

Phonetically, the tendency is also evidenced by the difference in MC1 and MC2 performance. Normalized F0 values show that for both H* and L*, MC1 performed more similarly to NS. However, for MC2, they produce L* much worse than H* by employing falls in both statements and polar questions. Less experienced learners may not have learned to cue English prominence with low pitch.

However, pitch range, a frequent MC prominence cue, has not been found significant in English pitch accent acquisition. This may need to be further explored by measuring F0 values of words before and after the ones aligned with pitch accents.

Pitch change slope has instead been found to be an important factor in pitch accent realization, since learners perform significantly from NS, although further study should be done to explore why in this study MC2 performance resembles that of NS, but not MC1. One likely solution is to find more advance MC learners of English.

### 4.2. Tonal crowding

With target words in final positions, results showed that learners resemble NS more when pitch accents are aligned with final syllables than with the penultimate ones. Hypothesis 3 is thus confirmed that learner performance resembles that of NS in the presence of H% boundary tone.

This may be a strong evidence for tonal crowding effect. In MC question intonation, the last syllable bears the most important role in deciding if the sentence is statement or interrogative. H% associated with the polar question conceals the poor L2 acquisition of L*. Therefore, with daGA in the final position of both statements and questions, it is natural for the learners to produce high tone due to the influence of H%.

However, when daGA is in sentence-medial positions where H% does not clash with pitch accents, learner performance is less similar to that of NS. In these cases, H* are employed because learners are subject to a negative transfer from MC to cue prominence with higher pitch rather than lower one. In addition, for MC2 to produce L% even in polar questions, it may be again the evidence of negative transfer from MC wh-questions, which are not phonetically cued by H%.

## 5. Conclusion

The study investigates the acquisition of H* and L* in English as influenced by boundary tones by examining both qualitatively and quantitatively data produced by MC speakers of English. In attempt to extend SLM into prosodic aspects, 3 hypotheses are formulated based on the comparative analysis of English and MC using LILt. It is found that MC speakers produce H* better than L* since MC prominence cued by higher pitch like English. However, English L* is not so well learned. In addition, tonal crowding has influence on MC speakers of English in that MC echo questions raise only the pitch of the last syllable with the rest of sentence remaining the same as statements. In this sense, the study provides empirical evidence to L2 acquisition of pitch accents and better explains the role of L1-L2 similarity/dissimilarity in crosslinguistic influences.

## 6. Acknowledgment

We would like to thank the support of National Social Science Research Fund (No. 14BYY144) and Hunan Province Social Science Fund (No. 14YBA078).

# 7. References


[1] T. M. Derwing and M. J. Munro, "Putting accent in its place: Rethinking obstacles to communication," *Language Teaching,* vol. 42, pp. 476-490, 2009.

[2] R. C. Major, *Foreign accent: The ontogeny and phylogeny of second language phonology*: Routledge, 2001.

[3] R. Lado, "Linguistics Across Cultures: Applied Linguistics for Language Teachers," 1957.

[4] J. E. Flege, "Second language speech learning: Theory, findings, and problems," in *Speech perception and linguistic experience: Issues in cross-language research*, ed, 1995, pp. 233-277.

[5] C. T. Best, "A Direct Realist View of Cross-Language Speech Perception," *Speech Perception and Linguistic Experience: Issues in Cross-Language Research,* 1995.

[6] I. Mennen, "Beyond Segments: Towards a L2 Intonation Learning Theory," in *Prosody and Language in Contact*, ed: Springer, 2015, pp. 171-188.

[7] D. R. Ladd, "Intonational universals and intonational typology," *Handbook on Language Typology and Language Universals. De Gruyter, Berlin,* 2001.

[8] D. R. Ladd, *Intonational phonology*: Cambridge University Press, 2008.

[9] M. E. Beckman and J. B. Pierrehumbert, "Intonational structure in Japanese and English," *Phonology,* vol. 3, pp. 255-309, 1986.

[10] J. Pierrehumbert and M. Beckman, *Japanese tone structure* vol. 15: MIT Press Cambridge, MA, 1988.

[11] A. Dainora, "Does intonational meaning come from tones or tunes? Evidence against a compositional approach," in *Speech Prosody 2002, International Conference*, 2002.

[12] X. S. Shen, "Relative duration as a perceptual cue to stress in Mandarin," *Language and Speech,* vol. 36, pp. 415-433, 1993.

[13] S.-h. Peng, M. K. M. Chan, C.-y. Tseng, T. Huang, O. J. Lee, and M. E. Beckman, "Towards a Pan-Mandarin system for prosodic transcription," *Prosodic typology: The phonology of intonation and phrasing,* pp. 230-270, 2005.

[14] M. Lin, "Chinese Intonation and Tone," *Applied Linguistics,* vol. 3, pp. 57-67, 2004.

[15] S.-C. Ou, "Taiwanese EFL learners' perception of English word stress," *Concentric: Studies in Linguistics,* vol. 36, pp. 1-23, 2010.

[16] P. J. A. Chen and A. Jiang, "Representation of Mandarin intonations: boundary tone revisited," in *Proc. 23rd North American Conference on Chinese Linguistics*, 2011, pp. 97-109.

[17] M. Lin, "Interrogative vs. declarative and the boundary tone in Standard Chinese," *Chinese Language,* vol. 4, pp. 364-376+384, 2006.

[18] Y.-L. Shue, S. Shattuck-Hufnagel, M. Iseli, S.-A. Jun, N. Veilleux, and A. Alwan, "On the acoustic correlates of high and low nuclear pitch accents in American English," *Speech Communication,* vol. 52, pp. 106-122, 2010.

[19] A. Arvaniti, D. R. Ladd, and I. Mennen, "Phonetic effects of focus and “tonal crowding” in intonation: Evidence from Greek polar questions," *Speech Communication,* vol. 48, pp. 667-696, 6// 2006.

[20] Y. Xu, "ProsodyPro—A tool for large-scale systematic prosody analysis," 2013.

[21] M. Ueyama and S.-A. Jun, "Focus realization of Japanese English and Korean English intonation," *UCLA Working Papers in Phonetics,* pp. 110-125, 1996.