

# Prosody Transfer and Suppression: Stages of Tone Acquisition

Chilin Shih and Hsin-Yi Dora Lu

Department of Linguistics  
Department of East Asian Languages and Cultures  
The Beckman Institute  
University of Illinois at Urbana-Champaign, U. S. A.  
cls@illinois.edu, hsinyilu@illinois.edu

## Abstract

This study investigates how native English speakers acquire Mandarin lexical tones, and whether they can express declarative and question intonation in a tone language. Production errors suggest three stages of tone learning. The first stage is characterized by a high error rate resulting from prosody transfer. The second stage shows moderate success in the suppression of L1 prosody especially in the utterance-initial positions. In the third stage, the error rate is low and the error patterns suggest that it is difficult to maintain tonal contrast in unstressed positions and with question intonation. Both male and female learners in general succeeded in suppressing declarative intonation but all female learners in the study failed to suppress question intonation. No learners in this study succeeded in using native-like declination for statements.

**Index Terms:** Tone acquisition, tone error pattern, prosody transfer, L1 prosody suppression

## 1. Introduction

The acquisition of lexical tone is a dynamic process where a second language (L2) learner adjusts to a situation where the fundamental frequency ( $f_0$ ) channel is used to encode lexical information in addition to paralinguistic information. Doing so requires the enhancement of the existing language faculty in the brain and the recruitment of additional neural and cortical areas [1, 2, 3, 4], which poses an interesting challenge to L2 learners.

Transferring selective properties of the first language (L1) into L2 is a strategy commonly employed by L2 speakers [5, 6], and subtle cues of L1 prosody may persist in the speech of fluent L2 learners, contributing to the perception of foreign accent [7]. Surprisingly, [8] did not find strong evidence of prosody transfer from English speakers learning Mandarin. They hypothesized that learners had limited attention span, and the high demand of lexical tone learning trumped the paralinguistic function of intonation. [9], in contrast, did find evidence of English-to-Mandarin prosody transfer.

The current study re-visits the issue of prosody transfer with attention to individual differences, investigating how native English speakers acquire Mandarin lexical tones, and whether they express declarative and question intonation while still producing correct tones.

Production errors suggest three stages of tone learning. The first stage is characterized by high error rate resulting from prosody transfer. The second stage shows moderate success in the suppression of L1 prosody especially in the utterance-initial positions. In the third stage, the error rate is low and the error patterns suggest that it is difficult to maintain tonal contrast in

unstressed positions and with question intonation. Both male and female learners in general succeeded in suppressing declarative intonation but all female learners in the study failed to suppress question intonation. No learners in this study succeeded in using native-like declination for statements.

## 2. Experiment Design

Digits and phone numbers in Mandarin tone 1, the high level (H) tone, were chosen as the experimental materials in consideration of ease of production, the learner's vocabulary size and their command of grammar, as well as whether the same text can be naturally expressed both in declarative and question intonation.

Among four Mandarin lexical tones, tone 1 is the easiest for L2 learners to master [10]. In terms of motor control, simulation models by [11] showed that a sequence of Mandarin tone 1 requires the least articulatory effort. Additionally, English speakers are capable of producing monotone English speech which has  $f_0$  contours that are similar to that of a Mandarin tone 1 sequence. Hence, tone errors in tone 1 sequences are more likely to be cognitively, rather than articulatorily based.

Telephone numbers were chosen to test utterance-level tone production because it can be used to convey declarative and question intonation without additional lexical items in both Mandarin and English. Furthermore, a ten-digit US telephone number in the format of ddd-ddd-dddd includes three phrases with at least three digits per phrase, allowing for the comparison of utterance initial, medial and final phrases, as well as phrase initial, medial and final syllables.

### 2.1. Subjects

Eight subjects were recruited to participate in the study. Two were native speakers of Chinese (1 male, 1 female). They served as the control group. Six L2 learners of Mandarin (3 male, 3 female) were recruited. They all spoke English natively. Among them, two (1 male, 1 female) were bilingual and had exposure to a tone language at home. One spoke Vietnamese and the other Cantonese. At the time of the recording, four learners were taking First Year Chinese (the second semester) and two were taking Second Year Chinese (the fourth semester).

### 2.2. Material

The experimental stimuli includes Chinese digits 1, 3, 7, 8 read in isolation and a digit string 338-811-3783 *san1 san1 bal-bal yil yil-san1 qil bal san1* in the format of a telephone number. This string was randomly generated from the tone 1 dig-

	Stage I	Stage II			Stage III	
Gender	F	F	M	M	F	M
Yr of Study	1	1	2	1	2	1
Language	E	E+V	E	E	E	E+C
% Error in Digit	56%	19%	20%	0%	0%	0%
% Error in Utt.	98%	45%	40%	21%	8.8%	7.5%
Utt. Dur (seconds)	5.73	4.81	5.00	4.96	5.14	4.35

Table 1: Background and tone production error rate from all learners.

its. Digit strings were marked with punctuations (a period or a question mark) and speakers were trained to produce statement and question intonation according to two written scenarios. All speakers successfully produced declarative/question distinction in their native language. Each subject recorded 4 repetitions of 10 digits in isolation first, then recorded 4 repetitions of 48 digit strings with various tonal combinations, followed by digit strings read in English. The recording order in each repetition was randomized. This study analyzed tone 1 production only. Each subject produced 16 tone 1 digits in isolation and 8 digit strings, yielding 80 tone 1 syllables in sentence production.

### 2.3. Experiment Procedure

The recordings were conducted in the sound proof room in the phonetics lab of UIUC using a Shure-SM58 microphone, Tascam digital audio recorder and an IBM T30 laptop. Subjects were seated with the microphone and a LCD monitor 1 foot in front of them. They were asked to read the digit or digit string from the monitor and made the recording at their own pace. They were encouraged to re-record the items as many times as needed.

After recording, tone production was graded by two native speakers in two rounds each. During the first round, each syllable/tone was graded as correct or incorrect. The inter-grader agreement is very good ( $\kappa=0.956$ ,  $z=23.8$ ). During the second round, the syllable/tone production was transcribed as 1, 2, 3, 4 if the tone production was perceived as being a good production of the corresponding Chinese lexical tones, 5 if the production was perceived as a Chinese neutral tone (a tone on an unstressed syllable), and 6 for none of the above. The inter-rater agreement is still very good ( $\kappa=0.87$ ,  $z=33.3$ ).

## 3. Analysis

Native speakers of Chinese did not make any mistakes. The averaged duration of their connected digit string production were 2.25 seconds and 2.62 seconds. That is about twice as fast as the learners.

All learners produced the correct tone of each digit some of the time, suggesting that they have learned Mandarin tone shapes and the lexical tones of the digits. Learners produced about twice as many errors when reading telephone numbers, compared to the error rate from digits produced in isolation. We expected more errors in sentence production, but not as many as that. One learner produced 56.2% tone error during single digit production, and the tones in digit string production were nearly completely wrong. Two learners made about 20% errors during single digit production, and their error rate increased to above

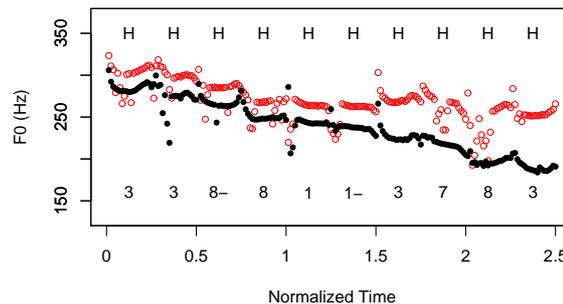


Figure 1: A female native speaker's production of a digit string in high level tones. The statement (filled circles) shows a strong declination effect. The question intonation (open circles) has a higher  $f_0$  than the statement for the entire utterance.

40% in digit strings. Three learners made no mistakes during single digit production, but had 7 to 21% of error during digit string production. Learners were grouped into Stage I to Stage III by their error rate in connected digit production.

Table 1 presents the background information of learners together with their error rate (percentage error) and speaking rate (seconds). *Yr of Study* shows whether the speaker was in the first or second year of Mandarin classes. *Language* provides language background. E is for students who grew up in a monolingual English environment. V and C are Vietnamese and Cantonese, respectively. *% Error in Digit* shows tone production error rate made when subjects were reading digits 1,3,7,8 in isolation. *% Error in Utt.* shows the error rate in connected digit, or utterance of telephone number production. *Utt. Dur* gives the averaged duration of the connected digit production for each learner.

The table suggests high correlations between monosyllabic error rate, digit string error and speaking rate. Learners with high error rate in monosyllabic production also have high error rate in connected digit utterances and with slow speaking rate. Year of study does not seem to be a strong predictor of tone production: one first year student performed well and one second year student lagged behind. Students with tone language background made tone production errors too, and do not necessarily perform better than monolingual English students. However, their speaking rate were faster compared to monolingual speakers with comparable error rates.

A logistic regression model followed by ANOVA was conducted predicting the tone production outcome (correct/incorrect) from sentence position (1 to 10), intonation type (declarative and question), gender and year of study and the interaction of position and gender. There were significant main effects of position ( $p < 0.0001$ ) and gender ( $p < 0.00001$ ). The effects of intonation type and year of study were not significant. There was a significant interaction of position and gender ( $p < 0.00001$ ).

## 4. Discussion

### 4.1. Production by native speakers

Native speaker's production is used as the gold standard against which L2 production is evaluated. Figure 1 shows averaged  $f_0$  contours from the female native speaker saying phone numbers

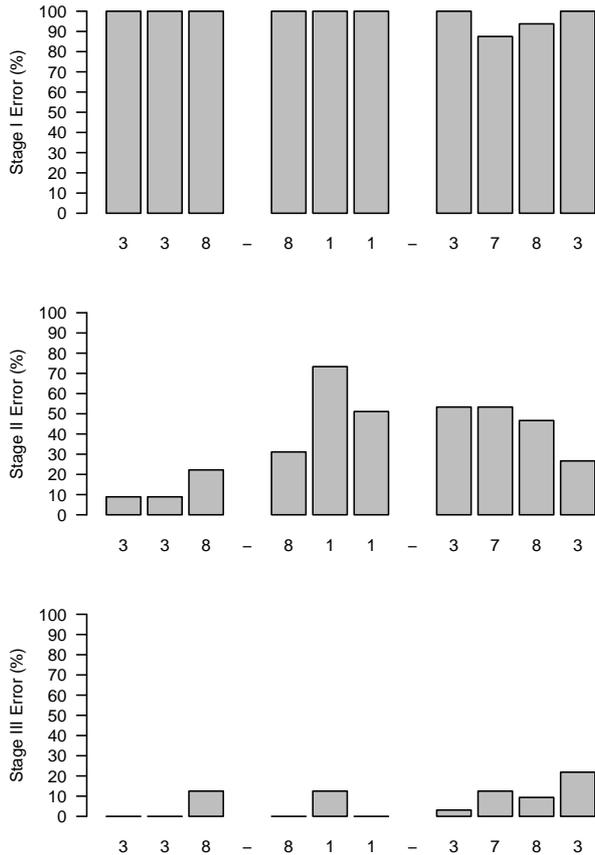


Figure 2: The dynamics of tone error reduction as learners advanced from Stage I to Stage III. Errors were plotted by the syllable positions in the utterance.

consisting of Mandarin tone 1. The statement is plotted with filled circles and question in open circles. The digit string is given at the bottom of the plot, and the expected tone sequence, a sequence of high (H) tones, is given at the top of the plot.

In a perception test, tone production of all syllables were judged as correct and were perceived as tone 1, even though the  $f_0$  contours may not be flat due to segmental effects and declination [12, 13].

Question intonation is signaled by a higher  $f_0$  values than for statements; the difference increases towards the end of the utterance. It may start as early as the beginning of the utterance as in Figure 1, or as late as the beginning of the last phrase. A statement can be signaled by a steep declination. The timing and amount of pitch manipulation vary by speaker and situation [14].

## 4.2. Production by learners

Learner's error rate is a good predictor of error type. One possible explanation is that learners go through predictable developmental stages of prosodic learning, and each stage is associated with characteristic error types. Naturally, early learning stages imply more errors, hence the correlation between error rate and error type. Table 1 sorts learners by error rate. Using the gaps in their error rate as the cut-off points, we combined learners into

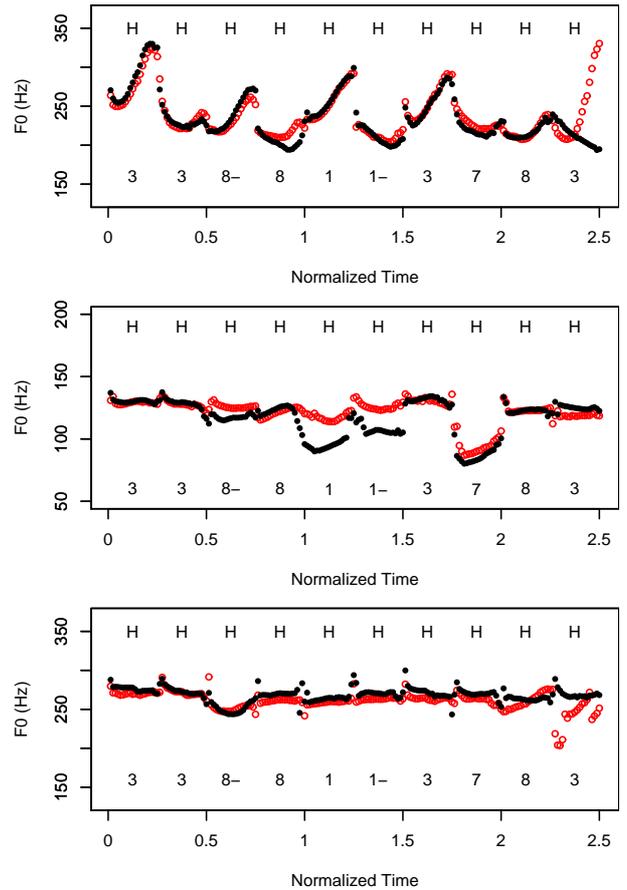


Figure 3: Averaged pitch tracks showing the tone production from three learners at different learning stages. Declarative sentences are represented by filled circles and question by open circles.

three groups as shown in Table 1, and report their error rate by sentence position in Figure 2 together with representative pitch tracks from each group in Figure 3, where averaged pitch tracks of the declarative sentence are plotted in filled circles and question in open circles.

### 4.2.1. Stage I: Prosody Transfer

Stage I represents the initial learning stage where tone error rate is high. Error analysis shows that the errors reflect L1 prosody transfer.

An example of Stage I production is shown in the top panels of Figures 2 and 3. Figure 2 shows 100% error in most sentence positions and Figure 3 shows that the digit string production is dominated by an English-like alternating pitch pattern. The speaker consistently used a falling pitch (H+L) and a rising pitch (L+H) to convey statement and question, respectively.

This speaker was able to produce correct tones for each of the Mandarin digits in isolation, demonstrating a knowledge of both tone shape and lexical tone association, though her performance was unstable even in monosyllabic production.

#### 4.2.2. Stage II: Suppressing L1 Prosody

In Stage II, error rate reduced considerably both in monosyllabic production and in sentence production. Learners were able to suppress L1 prosody occasionally and have better control of tone production early on in an utterance. However, they were not able to maintain consistent performance throughout the sentence even though in this experiment the target tones are all the same. The English-like alternating pitch pattern surfaced sporadically in the later part of the utterance.

The middle panel of Figures 2 combines data from three learners with sentence error rate from 20% to 45% and plot them by sentence position. The first two syllables of the digit string have the lowest error rate, while the 5th digit, being in the medial syllable of a medial phrase, has the highest error rate. Both positional dependent error patterns may be related to attention [15].

The middle panel of Figures 3 show averaged pitch tracks from a male speaker in this group. This speaker typically had correct tone production in the beginning and the end of the utterance, though he was unable to convey the distinction between declarative and question intonation.

#### 4.2.3. Stage III: Stress and Intonation Type

In Stage III, learners had good command of lexical tone shapes, as shown in the 0% tone production error in digits produced in isolation. Tone in sentence production was good too, with only occasional errors.

The bottom panel of Figures 2 combines data from two learners in this group. There were no errors in the first two syllables of the digit string. Error rates in other positions were low but the medial positions of phrases still posed problems, where the pitch pattern resembles English unstressed syllables. A possible explanation is that the learners were comfortable with the production of connected speech but were affected by the stressed/unstressed distinction in English, while speakers in Stages I and II were not fluent enough to express the stress/unstress contrast. Interestingly, the error rate in the utterance-final position did not show comparable improvement from Stage II. Error analysis show that most Stage III errors in this position came from female learners using the rising pitch movement (L+H) for question intonation. It appeared that female speakers pay more attention to sentence level prosody. In an attempt to convey question intonation but unable to do so with correct tones, they made mistakes by transferring question intonation from English and overwrote lexical tones. Male learners did not make tone errors in this position, but they also didn't differentiate question and statement intonation. No errors were found in the production of statement. There was no evidence that either female or male learners transferred English statement intonation to Mandarin.

## 5. Conclusions

This study investigated L2 Chinese learners productions of Mandarin high level tones in monosyllables and in utterances. The results identified several stages of prosody learning with characteristic error patterns. L2 errors reveal some aspects of English prosody that play a dominant role in L1 to L2 prosody transfer, as well as the capability to suppress L1 prosody in more advanced learners. The findings have implications to the teaching and learning of a second language, as proper prosody control is essential to effective speech communication.

Brain imaging research found a left hemispheric advantage

among native and bilingual Mandarin speakers processing Mandarin lexical tones, but not for nonnative speakers, though tone learning was associated with an expansion of neural process within existing language-related areas. It is possible that the three stages of prosody learning proposed here might be associated with different activations in the brain. We hypothesize that learners in different learning stages would exhibit a quantitative difference in the neural correlates and we plan to pursue this area of study in the future.

## 6. Acknowledgments

This project is based on work supported by the National Science Foundation under grant numbers IIS-0623805 and IIS-0534133. The opinions expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## 7. References

- [1] J. Gandour, D. Wong, L. Hsieh, B. Weinzapfel, D. Van Lancker, and G. Hutchins, "A crosslinguistic pet study of tone perception," *Journal of Cognitive Neuroscience*, vol. 12, pp. 207–222, 2000.
- [2] Y. Wang, J. A. Sereno, A. Jongman, and J. Hirsch, "fMRI evidence for cortical modification during learning of Mandarin lexical tone," *Journal of Cognitive Neuroscience*, vol. 15, no. 7, pp. 1019–1027, 2003.
- [3] Y. Wang, D. M. Behne, A. Jongman, and J. A. Sereno, "The role of linguistic experience in the hemispheric processing of lexical tone," *Applied Psycholinguistics*, vol. 25, pp. 449–466, 2004.
- [4] L. Liu, D. Peng, G. Ding, Z. Jin, L. Zhang, K. Li, and C. Chen, "Dissociation in the neural basis underlying chinese tone and vowel production," *Neuroimage*, vol. 29, pp. 515–523, 2006.
- [5] R. Ellis, *Understanding Second Language Acquisition*. Oxford University Press, 1985.
- [6] J. E. Flege, "The production of *new* and *similar* phones in a foreign language: evidence for the effect of equivalence classification," *Journal of Phonetics*, vol. 15, pp. 47–65, 1987.
- [7] M. Jilka, "The contribution of intonation to the perception of foreign accent," Ph.D. dissertation, The University of Stuttgart, Stuttgart, Germany, 2000.
- [8] T. Visceglia and J. D. Fodor, *Interfaces in Multilingualism: Acquisition and Representation*. Amsterdam: John Benjamins, 2006, ch. Fundamental frequency in Mandarin and English: Comparing first- and second-language speakers, pp. 27–59.
- [9] C. Pytlyk, "Interlanguage prosody: native english speakers production of mandarin yes-no questions." in *Proceedings of the 2008 Annual Conference of the Canadian Linguistic Association*, 2008.
- [10] M. Chan and H. Kang, Eds., *Tone Production in Mandarin Chinese by American Students: A Case Study*, The Ohio State University, 2008.
- [11] G. Kochanski and C. Shih, "Prosody modeling with soft templates," *Speech Communication*, vol. 39, no. 3–4, pp. 311–352, 2003.
- [12] C. Shih, "A declination model of Mandarin Chinese," in *Intonation: Analysis, Modeling and Technology*, B. A., Ed. Kluwer Academic Publishers, 2000, pp. 243–268.
- [13] C. X. Xu and Y. Xu, "Effects of consonant aspiration on Mandarin tones," *Journal of the International Phonetic Association*, vol. 33, no. 165–181, 2003.
- [14] J. Yuan, "Intonation in Mandarin Chinese acoustics, perception and computational modeling," Ph.D. dissertation, Cornell University, 2004.
- [15] P. Trofimovich and W. Baker, "Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech," *Studies in Second Language Acquisition*, vol. 28, pp. 1–30, 2006.