Perception of L2 Tones: L1 Lexical Tone Experience May Not Help

Xinchun Wang

Department of Linguistics
California State University, Fresno
xinw@csufresno.edu

Abstract

This study investigates whether adult L2 learners' experience with lexical tones and pitch accent in their first language facilitates the acquisition of L2 lexical tones. Three groups of beginning learners of Mandarin with different L1 prosodic experience: native Hmong (a tone language), native Japanese (a pitch and accent language), and native English (a non-tone, non-pitch accent language) speakers participated as listeners in a perception test on the four Mandarin tones. Results showed that native English listeners performed equally well as native Japanese listeners but native Hmong speakers performed significantly worse than the native Japanese and native English speakers in perceptual accuracy of Mandarin tones. The findings suggest that experience with lexical tones and pitch accent may not always facilitate learning. The lack of exact mapping of L2 tones onto L1 tones may interfere with the acquisition of nonnative tones especially at the initial stage of learning.

1. Introduction

The acquisition of nonnative lexical tones has proved to be difficult for adult second language learners [1, 2, 3]. Previous research on the effect of speakers' L1 prosodic experience on the acquisition of nonnative tonal systems has yielded conflicting results. For example, in a perceptual training study on Thai tones, Wayland and Guion [2] found that native Chinese speakers (Taiwanese and Mandarin) significantly outperformed native English speakers in discrimination and identification of the Thai mid level and low level tone contrast after a brief period of perceptual training. The authors concluded that prior experience with the tone system in one tone language might facilitate the acquisition of tone in another language. This is because the ability to track the change of F0 values, movement, and the direction of the movement at word level with one tone language may be transferable to the discrimination of tones in an unfamiliar tonal system.

However, conflicting results have also been reported. In a recent study of native Cantonese and Japanese speakers' perception of Mandarin tones, So [3] found that native Cantonese speakers consistently demonstrated greater difficulties in distinguishing the Mandarin tone 1- tone 4 and tone 2-tone 3 contrasts than native Japanese speakers before and after a brief period of training. Given the fact that Cantonese speakers have prior experience with lexical tone contrasts in their first language while Japanese speakers have only prosodic experience with pitch and accent at phrasal level in their L1, native Cantonese speakers' direct experience with a lexical tone system failed to facilitate learning, at least at the initial stage of learning. These inconsistent findings suggest that the relationship between learners' sensitivity to lexical tones or pitch accent as a result of first language experience and its effect on learning a new tonal system is not straightforward, at least at initial stage of learning.

The transfer of L1 experience to the acquisition of L2 speech has also been investigated at segmental level in previous studies. Some studies have explored whether the presence of certain features of place of articulation in the first language would assist its speakers in discriminating the non-native consonantal distinctions that share the same place feature but differ in manner. For example, Polka [4] examined whether Farsi speakers would have an advantage over English speakers in learning the Salish glottalized velar and uvular stop contrasts which form a non-native distinction for both native English and Farsi listeners. For the Farsi listeners, there is a phonemic feature distinction, velar versus uvular, occurring in different manners, voiced stops and voiceless fricatives. English, however, does not exploit the uvular place in its consonant system. The hypothesis was that if the Farsi speakers made use of the uvular place feature to identify the Salish distinction in a different manner of articulation, they would perform better than the native English listeners. However, the Farsi listeners as a group did not demonstrate a perceptual advantage because of their experience with the place contrast in another context [4]. The results suggested that the uvular place distinction in the L1 system was embedded in a certain manner of articulation for the particular consonant contrasts and did not transfer to a segment with a different manner of articulation.

Therefore, there is evidence from studies on both segments and lexical tones which suggests that positive transfer of L1 experience to the acquisition of L2 speech may not always occur. However, as there are conflicting results regarding the effect of L1 experience with lexical tones on the perception of L2 tones and both studies were based on two L1 speaking groups though a brief period of laboratory training [2, 3], future studies need to include more L1 groups to test the effect of different L1 prosodic experience on the acquisition of L2 tones in other learning environment such as formal L2 instruction in classroom setting.

This study investigates whether learners' experience with lexical tone and pitch accent contrast in their first language facilitates the acquisition of L2 lexical tones in formal classroom learning environment. Three groups of beginning learners of Mandarin (enrolled in a first semester Chinese course in a U.S. university) with different L1 prosodic experience: native Hmong (a tonal language), native Japanese (a pitch accent language), and native English (a non-tone, non-pitch accent language) speakers will be tested on their perceptual accuracy of Mandarin lexical tones. The two hypotheses to be tested are:

(1) If speakers' experience with the function of change in F0 height, direction and degree of movement that is used for contrasting lexical tones in their first language facilitates the learning of an unfamiliar tonal system though positive transfer, native Hmong

speakers will outperform native Japanese speakers and native English speakers in their perceptual identification of Mandarin tones.

(2) If speakers' L1 experience with lexical tones is a source of interference because of the confusion possibly caused by the interaction of the two tonal systems at initial stage of learning, causing native speakers of a tone language to less reliably discriminate tones from an unfamiliar language, native English and Japanese speakers will outperform native Hmong speakers in perceptual identification of Mandarin tones.

1.1. Mandarin and Hmong tones, Japanese pitch accent, and English Intonation

Mandarin Chinese contrasts four lexical tones, with Tone 1 having high-level pitch, Tone 2 high-rising pitch, Tone 3 low-dipping pitch, and Tone 4 high-falling pitch. The phonetic features of tones are manifested physically by the F0 values and contours. For the native Mandarin speakers, the primary cue for tone contrasts is F0 contour [5, 6]. Therefore, native Mandarin listeners attach more importance to "contour" than "height" dimensions of tones [7].

Hmong contrasts 7 lexical tones which are labelled as high level, high falling, mid rising, mid level, mid low, mid low breathy, and low glottalized [8].

The Japanese pitch accent rules function at phrasal level rather than at word level [9]. The change of pitch (intonation patterns) in English is not associated with the change of lexical meaning but functions at sentence and discourse levels.

2. Method

2.1. Participants

The participants were 37 Hmong (14 males, 23 females), 22 Japanese (10 males, 12 females), and 20 English speakers (13 males, 7 females) speakers residing in the U.S. at the time of the study. The Hmong speakers' average age was 18.5 years (range: 15-24). They were early Hmong-English bilinguals who were born in Hmong immigrant families in the U.S. A few were born in Thailand but moved to the U.S. with their parents as young children. The Hmong speakers all reported Hmong as their first language and began to learn English at elementary schools in the U.S. The native Japanese speakers were born and raised in Japan and were international students studying for undergraduate degrees in the U.S. The American English speakers were monolinguals with no L2 experience other than high school foreign language courses. None of the participants had resided in Mandarin speaking environment. All were enrolled in a first semester Chinese course in a U.S. university. Classroom teaching was mainly communicative emphasizing the four basic skills after an initial 3-4 weeks of concentration on the sound system and tones. Therefore, all participants were familiar with the four Mandarin tones and their function at lexical level by the mid-term when they took the perceptual tests. All had normal hearing by self-report.

2.2. Materials

The stimuli were produced by a male native Mandarin speaker who read a list of monosyllabic Mandarin words in Chinese characters presented with Pinyin. In order to ensure phonetic contextual variability, the words chosen for the perceptual test were 40 syllables (10 minimal quadruplets) combine various initial consonants and finals with different syllabic structures (i.e., V, CV, CVN, VN, CGlideV, CGVN). The stimuli were recorded in a quiet room on a PC computer using GoldWave software. The speaker read the list of words on a Shure SM 48 microphone connected to an M-Audio MobilePre USB preamplifier. The readings were recorded and saved at a sampling rate of 22050 Hz with 16-bit resolution. Each stimulus was normalized for peak amplitude for presentation. Before the test, the stimuli were screened for intelligibility by two native Mandarin speakers through four-way forced choice tasks. All the stimuli were 100 % accurately identified.

2.2. Procedure

The perception test was presented as a four-way forced choice identification task in a classroom. The participants first completed a brief language background form providing their age and first language experiences. The 40 stimuli were randomized and presented on a Macintosh computer through a built-in speaker system in the classroom. A printed out answer sheet in Pinyin was presented to the listeners before the test. The listeners heard each stimulus only once at an interstimulus interval (ISI) of seven seconds and circled the corresponding tone of the stimulus they heard on the answer sheet. A brief training session was performed to ensure the listeners understood the procedure before the real test began.

3. Results

3.1. Overall results

The mean percentage correct identification scores (% ID scores) by each group are presented in Figure 1. Overall, the mean correct identification scores were 61% for the Hmong group, 80% for the Japanese group, and 78% for the English group. A one-way ANOVA with Group as between subject factor revealed a significant effect of group [$F(2, 76) = 8.461$, $p = .000$]. Post hoc Tukey HSD test established significant differences ($\alpha < .01$) between the native Hmong and the native Japanese groups, between the native Hmong and the native English groups, but not between the native English and the native Japanese groups. These results indicate that, overall, the native Hmong group performed significantly worse than both the English and Japanese groups in perceptual accuracy of Mandarin tones while the native Japanese and Hmong speakers did not show any significant differences in their perceptual accuracy scores.

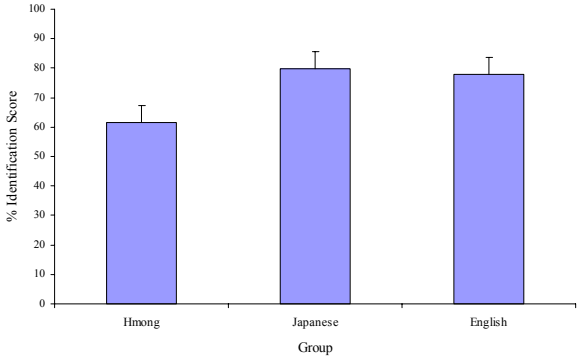

Figure 1. Mean percentage correct identification scores of Mandarin tones and standard errors for each group

3.2. Individual Tones

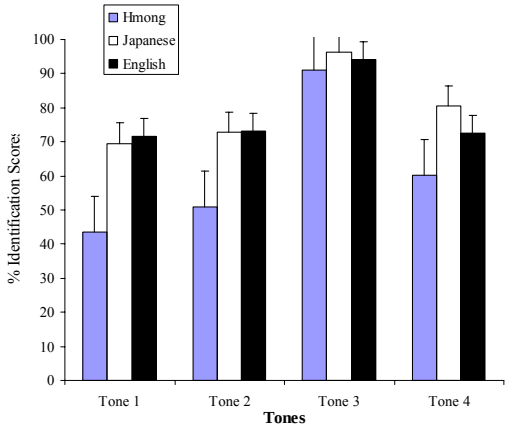

Figure 2. Mean percentage correct identification scores of the four Mandarin tones and standard errors for each group

To investigate listeners' performance on each tone, each listener's % ID scores were broken down to four individual tones and the mean scores for each group on each tone are presented in Figure 2.

A two-way repeated measures ANOVA with Group (Hmong, English, Japanese) as between subject factor and Tone (Tone1, Tone 2, Tone 3, Tone 4) as within subject factor yielded a significant effect of Group [$F(2, 76) = 1115.392$, $p = .000$], an effect of Tone [$F(3, 76) = 49.027$, $p = .000$] and an effect of Tone × Group interaction [$F(6, 76) = 3.355$, $p = .003$]. Follow-up multiple comparisons on Group established significant differences between the Hmong and the English groups ($p = .002$) and between the Hmong and the Japanese groups ($p = .000$) but not between the English and the Japanese groups ($p = .728$). Pair-wise comparisons on Tone revealed significant differences between Tone 3 and all the rest three tones ($p = .000$) and between Tone 1 and Tone 4 ($p = .001$) indicating that listeners across the groups perceived Tone 3 more accurately than the rest three tones.

To further investigate the effect of Group and Tone interaction, a series of one-way ANOVAs with Group as between subject factor on each tone revealed an effect of Group on Tone 1 [$F(2, 76) = 8.734$, $p = .000$], Tone 2 [$F(2, 76) = 6.917$, $p = .002$], and Tone 4 [$F(2, 76) = 4.437$, $p = .015$] but not on Tone 3 [$F(2, 76) = 1.267$, $p = .287$]. Post hoc Bonferroni tests ($\alpha = .01$) revealed no significant differences between the Japanese and the English groups on any of the 4 tones but established significant differences between the Hmong and the English groups on Tone 1 and Tone 2, between the Hmong and the Japanese groups on Tone 1, Tone 2, and Tone 4 indicating the native Hmong speakers had greater difficulties in perceptual identification of Tone 1 and Tone 2 than the other two groups. There was no significant difference among the groups on Tone 3.

Taken together, these results suggest that native Hmong speakers performed significantly worse than native English and Japanese speakers in perceptual accuracy on Mandarin tones, in particular on Tone 1 and Tone 2. The native Japanese and English speakers did not show differences in their performance on perception of Mandarin tones on any of the four tones.

4. Discussions and Conclusions

This study was carried out to investigate whether learners' first language experience with lexical tones and pitch accent facilitate their acquisition of L2 lexical tones. Revisiting the two hypotheses put forward earlier in the introduction, the results of the perceptual test do not appear to support hypothesis 1 which states that native Hmong speakers will outperform both native Japanese and English speakers in their perception of Mandarin tones due to positive transfer of their L1 experience with lexical tones to the perception of a new tonal system. In contrast to the prediction, the native Hmong group performed significantly worse than both native English and native Japanese speakers in perceptual accuracy of Mandarin tones. Obviously, the native Hmong speakers' experience with their first language lexical tones did not appear to give them any advantage over the native Japanese and English speakers whose L1 experience do not prepare them to associate the change of F0 in height and direction at word level.

The results appear to support hypothesis 2 which states that native English and Japanese speakers will outperform native Hmong speakers in perceptual identification of Mandarin tones due to the possibility of interference of Hmong speakers' L1 experience with lexical tones, causing them to less reliably identify tones from an unfamiliar language at least at the initial stage of learning. One might argue that the Hmong speakers' poor performance might be due to the fact that their perception of Mandarin tones was at phonemic level which was linguistically relevant while the native Japanese and English speakers were perceiving the tones at phonetic level. There is no evidence to support this argument from the experiment design in terms of using different ISI, which is often used to measure whether listeners' processing of lexical tone is at phonemic level (ISI = 1500 ms) or at phonetic level (ISI = 500 ms) in discrimination tasks in which the listeners judge the two tonal stimuli as the same of different. The current study used four-way forced choice identification task (listeners identify each stimulus as one of the four tones) with an ISI of 7s, a response time that is significantly longer than the 500 ms ISI at which phonetic level perception was believed to be possible in discrimination tasks [10]. Besides, by the time the participants took the test, they had already received eight weeks of instruction in Mandarin in which the initial 3-4 weeks were devoted to the sound system and tones, a test condition that was significantly different from previous studies in which the non-tone language speakers were tested without any prior exposure to lexical tones [1, 10].

It is not clear how exactly the Hmong speakers' experience with lexical tones interfere with the discrimination of L2 tones. One possibility might be that their poor performance is not due to the lack of experience with tones but rather to the lack of one-to-one mapping between the L1 and L2 tones. Such mismatch between the two tonal systems may cause confusion to the learners when the similar but not identical L1 tones interfere with the perceptual identification of L2 tones. Previous studies on L2 perception at segmental level suggest that listeners' L1 and L2 phonetic systems interact and influence with each during the process of learning [11, 12, 13].

Flege's [11, 12] Speech Learning Model (SLM) states that the success in the perception of L2 categories depends on whether a learner is able to establish phonetic categories for

the segments that exist in L2 but not in L1. According to the SLM, a new category fails to be established as an L2 speech sound in spite of the audible differences between the L2 sound and the closest L1 sound or between two closest L2 sounds if the learner fails to perceive such differences. Therefore, L1 and L2 speech sounds interact through a "category assimilation" mechanism. "By hypothesis, category formation will be blocked if instances of an L2 speech category continue to be identified as instances of an L1 category" (Flege et al. 2003, p. 469). The SLM also predicts that a "merged category" will develop overtime during the process of learning. Both the degree and direction of interaction of the two phonetic systems are related to the age of learning. Late L2 learners showed evidence of more unidirectional influence of L1 vowel system on L2 vowel system while early learners often show bidirectional influence of the two phonetic systems [13].

Interpreting the current data in terms of SLM, it is possible that the adult native Hmong speakers perceived Mandarin tones in terms of similar Hmong tones that are phonetically different from Mandarin tones blocking the formation of accurate Mandarin tone categories. Such speculation needs to be verified by how exactly the Hmong tones are mapped onto Mandarin tones though detailed acoustic analysis and cross linguistics perceptual tests. Future studies need to assess the category similarities and differences between Hmong and Mandarin tones, ideally, through direct assessment by having Hmong listeners identifying the Mandarin tones in terms of Hmong tone categories.

The current data also showed that native Japanese and English speakers' perceptual accuracy of Mandarin tones were not different and their performance was consistent across the four individual tones. The findings suggest that the presence of pitch accent feature in Japanese did not seem to give native Japanese speakers advantages or disadvantages over the native English speakers in learning the Mandarin tones. Based on the current findings, L1 experience with pitch and accent may not simply be viewed as extra source of help or interference in acquiring lexical tones. The findings are in agreement with results of studies on L2 speech perception at segmental level. As discussed earlier, the presence of certain features of place of articulation in the first language did not assist its speakers in discriminating the non-native consonantal distinctions that share the same place feature but differ in manner [4]. Therefore, there is evidence from studies on L2 segments and lexical tones which suggest that positive transfer of L1 experience to the acquisition of L2 speech may not occur.

Finally, according to the SLM, the ability to learn L2 speech categories is intact across the life span [11, 12] and learners' perception and production accuracy of nonnative speech categories will improve through long term exposure to L2 input or through perceptual training. It should be pointed out that the participants in the current study were still at very early stage of learning Mandarin as a foreign language. The test was conducted after only weeks of classroom instruction on Mandarin during a first semester Chinese course in a U.S. university. The participants had no exposure to any input other than the three-hour weekly classes. Moreover, due to the communicative nature of the instruction, the focus of teaching was not on tones or pronunciation of the sound system of Mandarin after the first 3-4 weeks of intensive pronunciation instruction. It is very likely that native Hmong speakers will gradually improve their perception accuracy of Mandarin tones through increased experience with Mandarin tones. A computer based training on perception and production of Mandarin tones was carried out with some of the participants reported in this study and results have shown that learners with different L1 experience all benefited from training [14]. Therefore, in addition to classroom instruction, laboratory training in both the perception and production mode will enhance learning.

5. References

[1] Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America, 106*, 3649-3658.

[2] Wayland, R. P., & Guion, S., G. (2004). Training English and Chinese listeners to perceive Thai Tones: A preliminary Report. *Language Learning, 54*(4), 681-712.

[3] So, C., K. (2005). The effect of L1 prosodic backgrounds of Cantonese and Japanese speakers on the perception of Mandarin tones after training. *The Journal of the Acoustical Society of America, 117*(4), 2427.

[4] Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America, 89*, 2961-2977.

[5] Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*, 61-83.

[6] Liu, S. & Samuel. A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech, 47*, 109-138.

[7] Wang, Y., Jongman, A., Sereno, J., (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after training. *Journal of the Acoustical Society of America 113*, 1033-1043.

[8] http://www.culturalorientation.net/hmong/hlang.html

[9] http://sp.cis.iwate-u.ac.jp/sp/lesson/j/doc/accent.html

[10] Burnham, D., & Francis, E. (1997). The role of linguistic experience in the perception of Thai tones. In A. S. Abramson (Ed.), *Southeast Asian Linguistic Studies in Honour of Vichin Panupong* (pp. 29-47). Bangkok: Chulalongkorn University Press.

[11] Flege, J. E. (1995). Second language speech learning: theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Baltimore: York Press.

[12] Flege, J., E., Schirru, C., & MacKay, I., R.A. (2003). Interation between the native and second language phonetic subsystems. *Speech Communication, 40*, 467-491.

[13] Baker, W., & Trofimovich, P. (2005). Interaction of native - and second language vowel system(s) in early and late bilinguals. *Language and Speech, 48*(1), 1-27.

[14] Wang, X. (2005). Training for learning Mandarin tones: A comparison of production and perceptual training. *The Journal of the Acoustical Society of America, 117*(4), 2425.