LEARNING MANDARIN TONES AT SENTENCE LEVEL THROUGH TRAINING: A PILOT STUDY

Xinchun Wang
Department of Linguistics, California State University, Fresno, 5245 N Backer Ave. M/S PB 92, Fresno, California 93740 USA

1. INTRODUCTION

Several previous studies have indicated that beginning level Mandarin learners improved significantly in identification and production of individual Mandarin tones in isolation after taking 2-3 weeks of training [1, 2, 3, 4]. However, perception and production of Mandarin tones on larger linguistic units such as phrases and sentences still poses significant challenges to learners who made progress on isolated words. No training studies on lexical tones have explored the effect of training beyond word level. This pilot study investigates the effect of perception with production training for learning Mandarin tones at sentence level. The research question is: will perception with production training using auditory and visual input be effective for learning Mandarin tones at sentence level?

2. METHOD

2.1. Participants

The participants were seven trainees (2 male, 5 female, mean age= 20) and 5 control subjects (3 male, 2 female, mean age = 20). All were beginning level Mandarin learners enrolled in a second semester Chinese course in a US public university. The participants' first languages (L1) include English (4), Hmong (4), and Japanese (4).

2.2. Stimuli

The stimuli used for pre- and post test were three sentences. Two of the sentences were statements and one was a simple question. The mean length of the sentence was 11 characters. The training stimuli consisted of 15 sentences and 48 phrases produced by four native Mandarin speakers (2 male, 2 female).

2.3. Procedure

For pretest, the participants read a list of 40 Chinese phrases and 10 sentences written in characters with pinyin (only 3 sentences were used for the current study). The recordings were made on a PC computer using GoldWave software through 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.

Individual training sessions were performed on a PC computer using Kay Elemetrics Sona Speech II Software as well as the same equipment mentioned in the above. During the training, the trainee opened and played back (through a pair of headphones) each training stimulus with real time display of the pitch contour in the top window of Screen A on the computer screen. The trainee then repeated the target phrase or sentence and recorded his/her own production of the target sentence by speaking into the microphone. The pitch contour of the trainee's production was instantly displayed in the bottom window of Screen B. The trainee could then compare his/her own production with the target phrase/sentence by playing them back repeatedly (auditory input). The trainee could also visually compare the tones by overlaying the pitch contour of the target sentence on that of his/her own production in different colors while alternately playing them back for auditory comparisons.

The training stimuli were blocked by speaker producing four training blocks each of which consisted of 15 sentences and 48 phrases. The trainees recycled the four speakers' phrases and sentences during the six hours of training that spread across 3-4 weeks. Immediately after the training, the trainees took the post test in which they repeated the tasks of the pretest. The control subjects took the pretest and post test at the same time interval but did not take the training.

2.4. Mandarin Listeners' Judgment

Four native Mandarin listeners (2 male, 2 female) residing in the US judged the speakers' productions of three sentences in a goodness rating task. The speakers' pretest and post test sentences were mixed and blocked by sentence. Two native Mandarin speakers' productions of the same sentences were also included in the rating stimuli. In each sentence block, 28 sentence stimuli were presented (1 sentence × 2 repetitions × 14 speakers). Individual rating tasks were performed on a Mac computer using custom designed software. The listeners rated each sentence along a continuum of 1 - 9 where 1 is labeled as "native like tones" and 9 as "very poor tones". They were told to pay attention to tones only while rating. Each listener had a trial session to learn the test procedure before the real rating tasks began. The order of the three sentences being rated was counter-balanced.

3. RESULTS

The mean rating for each of the Mandarin speakers' three sentences was 1.1. Not a single native production was rated above 2, indicating the raters were able to distinguish native from nonnative speech. As the research goal was not to compare native with nonnative production of tones, these Mandarin speakers' sentence data were not included in the analysis. Inter-rater reliability for the Trained and Control groups' productions were computed by sentence for the four raters by using Cronbach's α. The α values for the three sentences were .754, .702, and .831 respectively, which

were all acceptable. (In general, the cut-off point of Cronbach's α is .70, above which inter-rater reliability is acceptable.) Additionally, Pearson correlation tests revealed that inter- sentence correlations r ranged from .411 to .686, $p < .01$ for the three sentences. These results suggest that the four listeners had reasonable consensus on their ratings of the three sentences. Therefore, a single mean rating score for each speaker was obtained by averaging the four listeners' mean ratings of the 3 sentences. The Trained and Control groups' mean rating scores along with each individual speaker's data at pretest and post test are summarized in Table 1.

The mean score for the Trained group was 6 at pretest and 5.1 at post test. For the Control group, it was 6.8 and 6.7 at pre- and post test respectively. Paired t-tests revealed a significant difference between the mean rating scores at pretest and post test for the Trained group [t (13) = 5.987, p = .000] but not for the Control group [t (9) = .527, p = .611]. These results suggest that the trainees made significant improvement in their tone productions at sentence level that was not matched by the Control group.

As seen in Table 1, the overall mean ratings for the Trained group were lower (indicating more native-like tone) than the Control group at both pretest and post test. To investigate whether such differences were significant, two Independent-Samples t-tests (2-tailed) on the mean rating scores were performed. The results showed the difference between the groups at pretest [t (22) = 2.798, p =. 01] was significant. The difference was much more significant at post test [t (22) = 4.598, p = .000].

Table 1. Individual speakers' mean rating scores as judged by four native Mandarin listeners

ID	Trained Pre	Trained Post	ID	Control Pre	Control Post
T01	6.6	5.8	C01	6.6	6.7
T02	5.9	5.8	C02	7.3	7.1
T03	6.5	5.8	C03	6.0	6.4
T04	4.8	4.0	C04	7.4	6.8
T05	5.8	4.2	C05	6.5	6.6
T06	5.4	3.9			
T07	7.0	6.3			
Mean	**6.0**	**5.1**		**6.8**	**6.7**

The examination of individual speakers' scores showed that although each trainee's post test score was lower than his/her pretest score, only two of the seven trainees had a size of improvement that was beyond 1.5 along a scale of 9. None of their mean rating score fell within 2 standard deviations of the native Mandarin speakers' mean score, a standard that is often applied to measure whether the learners' performance reaches the level of native-like production in the literature.

4. DISCUSSION

This pilot study investigated the effect of training for learning Mandarin tones on a larger linguistic unit beyond the isolated tones. After taking 6 hours of training during a period of 3-4 weeks, the trainees' productions of Mandarin tones at sentence level were judged by native Mandarin listeners to be significantly better at post test than at pretest. The Trained group's improvement was not matched by the Control group. These results suggest that beginner level learners can improve their production of nonnative tones at larger linguistics units through perception with production training with both auditory and visual input. While the effect of training for learning nonnative tones is well documented in the literature, no previous studies, to my knowledge, have involved training on tones beyond isolated tones. Accurate productions of nonnative tones at phrase and sentence level is noticeably difficult for nonnative speakers due to different reasons. Therefore, it is encouraging to see that laboratory training is effective for learning tones at sentence level.

It is important to point out that although the trainees' improvement was significant, none of them reached the level of native-like production as judged by native Mandarin listeners. This may be due to the fact that six hours of training may not be sufficient for a more significant size of gain in production of nonnative tones at sentence level. More intensive training with more trainees is needed to test the ultimate effect of such training in future studies.

One limitation of the current study is the use of the rating score as the sole judgment of the production of Mandarin tones at sentence level. Although the listeners were instructed to pay attention to tones only, it is possible that their judgment might also be influenced by other factors such as segmental errors and speech rate. Future studies may include other assessment measures such as instrumental analyses of speech properties. Another limitation was the lack of strict control of the speakers' performance at pretest (not by design but due to lack of participants). It would be ideal to have same number of participants in both groups whose production scores were comparable at pretest.

REFERENCES

[1] Leather, J. (1997). Interrelation of perceptual and productive learning in the initial acquisition of second-language tone. In A. James & J. Leather (Eds.), *Second-language speech: Structure and process* (pp. 75-101). Berlin: Mouton de Gruyter.

[2] Wang, X. (2008). Training for learning Mandarin tones. In F. Zhang & B. Barber (Eds.), *Handbook of Research on Computer-enhanced Language Acquisition and Learning.* Hershey, PA: IGI Global.

[3] 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.

[4] Wang, Y., Jongman, A., Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after training. *Journal of the Acoustical Society of America 113,* 1033-104