Chun, D. M., Jiang, Y., & Ávila, N. (2013). Visualization of tone for learning Mandarin Chinese. In J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference.* Aug. 2012. (pp. 77-89). Ames, IA: Iowa State University.

VISUALIZATION OF TONE FOR LEARNING MANDARIN CHINESE

Dorothy M. Chun, University of California, Santa Barbara
Yan Jiang, University of California, Santa Barbara
Natalia Ávila, University of California, Santa Barbara

For a tonal language such as Mandarin Chinese, accurate pronunciation of tones is critical to meaning, and research suggests that computer-based programs that allow for visualization of pitch contours are helpful for improving learners' pronunciation (Chun, 1998; Hardison, 2004; Levis & Pickering, 2004). This paper reports on a pilot study using speech analysis software (*Praat*), which allowed L2 Chinese learners first to hear a native speaker of Mandarin say words and phrases while seeing a visual display of the native speaker's pitch curves, then to record themselves reading the same words and phrases, and later to compare their own pitch contours to those of the native speaker. Students in first-year Chinese were recorded reading words and phrases before and after two computer-based training sessions. Native speakers rated the words and phrases for accuracy of tones. Results indicate a ceiling effect for the pronunciation of mono- and disyllabic word tones, with 83.39% of tones produced correctly in the pre-test. Of the 16.41% of tones that were incorrect in the pre-test, almost 50% of them were pronounced correctly in the post-test. Students indicated in a post-study survey that seeing the pitch curves of both the native speakers and their own helped them improve their word tones.

L2 PRODUCTION OF MANDARIN CHINESE TONES

In Mandarin Chinese, there are 4 lexical tones manifested by different *pitch levels* (in acoustic phonetic terms, *fundamental frequency* or *F0*) and different *pitch contours*. As can be seen in Figure 1, which contains tones spoken by a native Mandarin speaker, tone 1 is high and level; tone 2 has rising pitch; tone 3 falls and then rises; and tone 4 starts high and falls sharply. A neutral tone that is unstressed and much weaker in intensity and shorter in duration also exists.

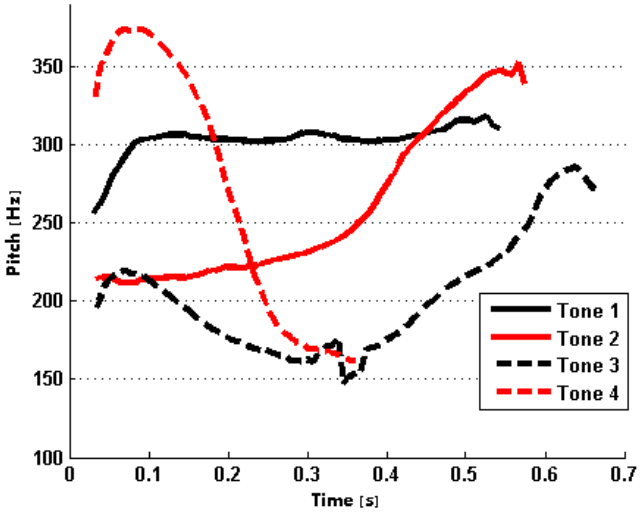

Figure 1. Acoustic Representation of Four Mandarin Chinese Tones

Research on the L2 acquisition of Mandarin Chinese tones has shown that learners whose L1 is a non-tonal language, e.g., English, have difficulties in perceiving and producing tones, but that training in both perception and production can be effective (Wang, Sereno, & Jongman, 2006). There is not universal agreement with regard to which tones are the most difficult. Shen (1989), for example, found that tonal errors made by American learners who had studied Chinese for four months ranged from 8.9% for tone 2 to 55.6% for tone 4, and that the mispronunciation of tones tends to lie in the pitch height (low, middle, or high) and not in the pitch contour (level, rising, dipping, or falling). For second-year American learners, Miracle (1989) found an overall error rate of 42.9%, with the errors relatively evenly divided among the tones. Chen (1997) found that tones 1 and 4 were easier than tones 2 and 3.

Molholt and Hwu (2008) advocated using acoustic analyses of speech (e.g., Figure 1) rather than simple descriptions or graphic representations that have traditionally been used in instructional materials (Figure 2). In line with this approach, other studies of L2 learning of non-tonal languages have demonstrated that when learners see acoustic visualizations of native speakers compared with their own pitch curves, the learners improve their pronunciation (Hardison, 2004).

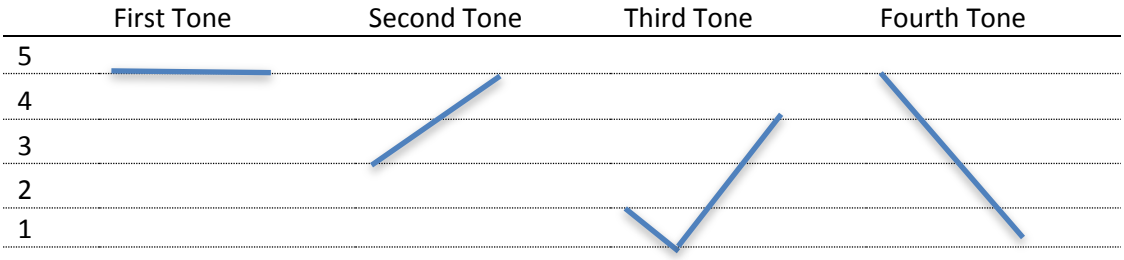

Figure 2. Typical Depiction of 4 Mandarin Tones.

PRESENT STUDY

This pilot study sought to determine whether providing L2 learners with visualizations of Mandarin native speakers' tonal pitch contours along with the learners' own pitch contours would improve L2 production of tones.

Participants

Sixteen students who were in their first-year (third-quarter) of studying Mandarin Chinese at a large state university in the western U.S. volunteered to participate. A list of the questions in the demographic survey can be found in Appendix A. None of them was a heritage speaker of Mandarin.

Materials and Procedures

The study was conducted in two one-hour sessions, a week apart. In the first session, students were given a pre-test: they were asked to read a list of 31 monosyllabic and disyllabic words or phrases and were recorded. They were then seated at a computer and shown the pitch curves of a Mandarin native speaker saying the 31 words or phrases (see Figure 3) and simultaneously heard the words spoken as they viewed the pitch curves that had been created with the acoustic software *Praat* (freely available at http://www.fon.hum.uva.nl/praat/). For 30 minutes, students

practiced repeating the words and phrases while hearing the native speaker and viewing the pitch curves.

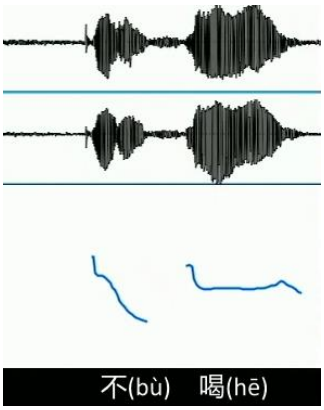

Figure 3. Native Speaker's Pitch Curves *bù hē* ('not drink').

A week later, students returned to the computer lab. The words that they had read the previous week had been pitch tracked by the researchers, and students could once again view the native speaker's pitch curves, but they could now also view their own production of these words and compare the two pitch curves (Figure 4).[1] After practicing again for 30 minutes, they were recorded reading the original 31 words and phrases. This recording would serve as the post-test; Figure 5 shows a student's pre- and post-test pronunciation of *bù hē*.

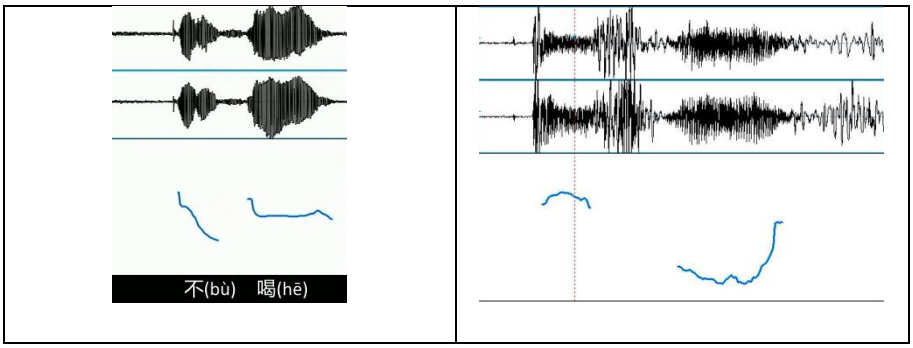

Figure 4. Native Speaker's *bù hē* (left) and Student's Pre-test *bù hē* (right).

[1] In other studies that used visualization of learners' pitch curves (e.g., Hardison, 2004), commercial software that provided a display of pitch curves immediately after learners' production was used. However, our study used the open source software *Praat*, which does not automatically provide pitch curves but rather requires student recordings to be processed individually "by hand." It is to be sure not as ideal as automatic, instantaneous feedback, but it does not require the purchase of costly software.

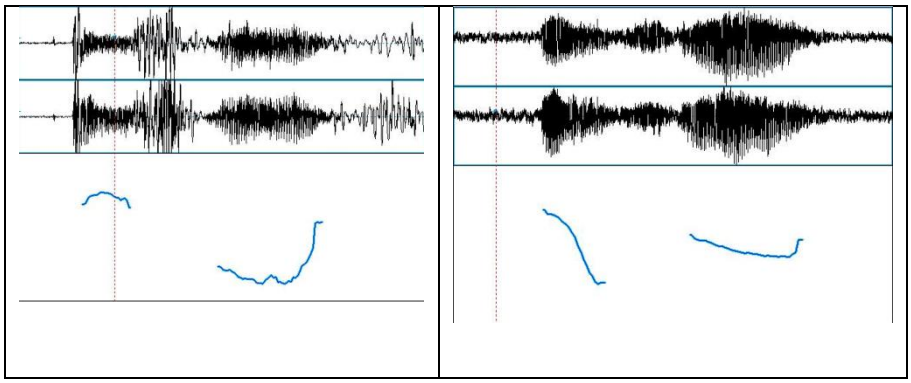

Figure 5. Student's Pre-test *bù hē* (left) and Student's Post-test *bù hē* (right).

Data Analysis

The pre- and post-tests were then assessed in two different ways in order to determine whether improvement had occurred. First, four native speakers of Mandarin Chinese (university Teaching Assistants) were presented with audio files of the 31 words spoken by the 16 learners. The words from both pre- and post-test were mixed together and randomized so that the raters would not know whether the recording was from the pre-test or the post-test. Raters assigned scores on a scale of 1-4, with 1=mostly correct and 4=mostly incorrect. Second, the same four native speakers were presented with pairs of words (the same words from the pre-test and post-test respectively) and asked to rate the improvement from pre-test to post-test on a scale of 1-5, with 1=great improvement; 2=some improvement; 3=only a little improvement; 4=no improvement or worse; 5=no change, both correct.

After the second session, students filled out a brief survey about the usefulness of seeing pitch curves, both of the native speakers and of their own.

RESULTS

Nineteen syllables in isolation or combined with others were selected for tone analysis based on a representative sampling of the four tones and tone combinations (Table 1). Among them were four 1st tone, five 2nd tone, six 3rd tone, and four 4th tone syllables. Within each tone group, at least one syllable belongs to each of the four categories: syllable in isolation, disyllabic tone + tone combination (t+t), disyllabic tone + neutral tone combination (t+n), and syllable in a sentence. These tone combinations are very common in Mandarin Chinese, and mastery of them is critical for learners to acquire correct pronunciation. Examining the students' accuracy of producing those syllables allowed us to determine which tones and tone combinations were most problematic for learners. It was also a systematic way to test the effectiveness of using pitch curves for tone training.

Table 1
Selected Syllables (Bolded) for Analysis

	Syllable in isolation	Disyllabic t+t	Disyllabic t+n	Syllable in a sentence
1st tone	***Shū*** ‘book’	*hǎo* ***shū*** ‘good book’	***māma*** ‘mother’	*zhè shì wǒ bàba* ***māma*** ‘this is my father, mother’
2nd tone	***chá*** ‘tea’	***nín*** *hǎo; hǎo* ***chá*** ‘how are you’; ‘good tea	***péng****you* ‘friend’	*zhè shì wǒ* ***péng****you* ‘this is my friend’
3rd tone	***nǐ; hǎo*** ‘you’; ‘good’	***nǐ*** *hǎo* ‘how are you’	***nǐ*** *men* ‘you’ (pl.)	***nǐ*** *hǎo;* ***nǐ****men hǎo* ‘how are you’; ‘how are you’ (pl.)
4th tone	***kàn*** ‘look’	*hǎo* ***kàn*** ‘good looking’	***bà****ba* ‘father’	*zhè shì wǒ* ***bà****ba māma* ‘this is my father, mother’

Pre-Test Results

The overall score statistics from the pre-test are displayed in Table 2, with the percentage of each score being the ratio of the score count over the total 304 syllables collected from the 16 participants. The learners’ pronunciation of tones was rated “1” (mostly correct) 79.61% to 87.17% of the time among the four raters. The percentage of tones that was rated “4” (mostly incorrect) ranged from 0.99% to 9.87% among the four raters. On average, 83.39% of the tones were rated “1”, and only 5.67% were rated “4.” In other words, raters perceived that the students mostly produced correct tones, but rarely incorrect tones. This suggests that a “ceiling effect” may exist, i.e., the training may not be effective for learners at this proficiency level, which may be attributed to the fact that they had been learning Chinese for at least six months before taking the test.

Table 2
Percentage (Pct.) of Scores in Pre-Test and Post-Test (1=mostly correct; 4=mostly incorrect)

Score	Avg. Pct. (%)	
	Pre-test	Post-test
1	83.39	85.23
2	7.24	6.93
3	3.70	3.14
4	5.67	4.70

We further broke down the ratings of tones to determine whether the students’ proficiency levels on the four tones were different. As shown in Figure 6, the 1st tone had the highest average percentage of rating “1” (90.23%) and the lowest average percentage of rating “4” (2.34%) among the four tones, indicating that the learners had the highest level of proficiency with 1st tone before training. On the other hand, the 4th tone had the highest average percentages of rating “4” (9.38%), suggesting the students made more mistakes when pronouncing 4th tone.

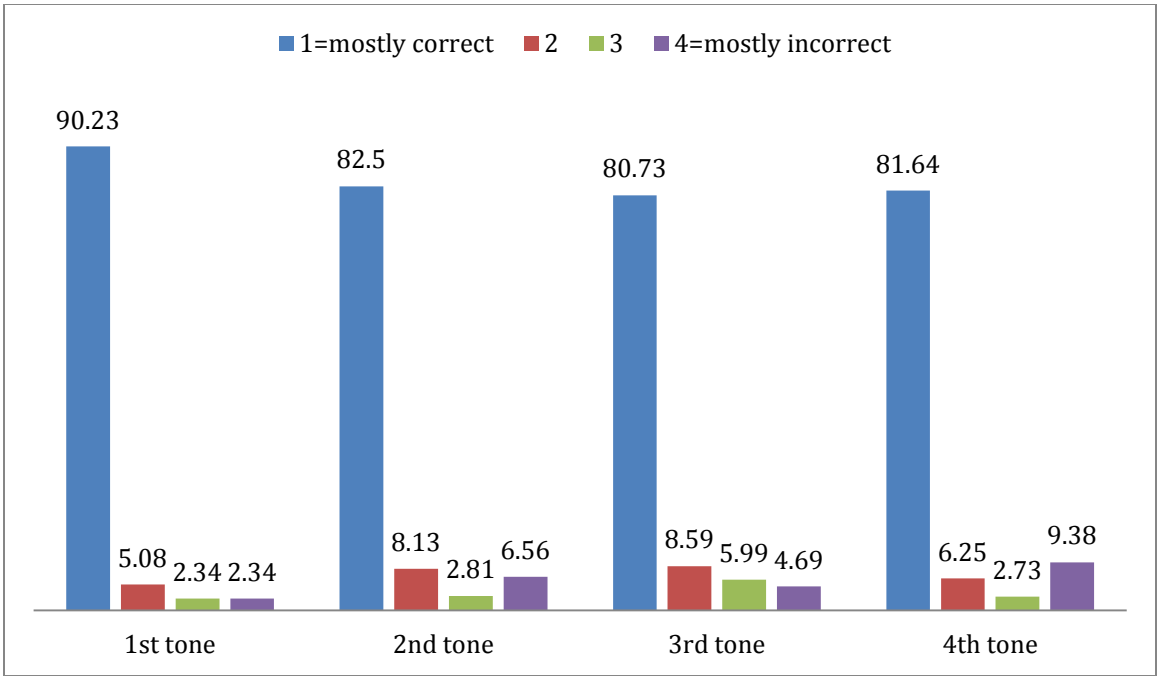

Figure 6. Average Percentages (%) of Each Score on Four Tones in Pre-test.

Post-Test Results

The first method used to determine whether students' tone production improved was to compare the percentage of scores in the post-test to the pre-test. Table 2 shows that the average percentage of score 1 increased by 1.84%, and average percentage of score 4 decreased by 0.97% from pre- to post-test. The change of scores suggests that the students produced more correct tones and fewer incorrect tones after training, although the difference in pre- and post-tests was limited due to the ceiling effect.

To further understand which tones showed improved production after training, we broke down the results by tones and computed the difference of the average percentages of correct tones in post- and pre-tests (see Figure 7). The higher average percentages of correct tones in the post-test indicate that learners improved on 2nd, 3rd and 4th tones after receiving the training. Notwithstanding the ceiling effect, learners improved their 4th tones by 5.42%. The reason might be that learners had a lower proficiency level on 4th tone before training, as suggested by the pre-test results.

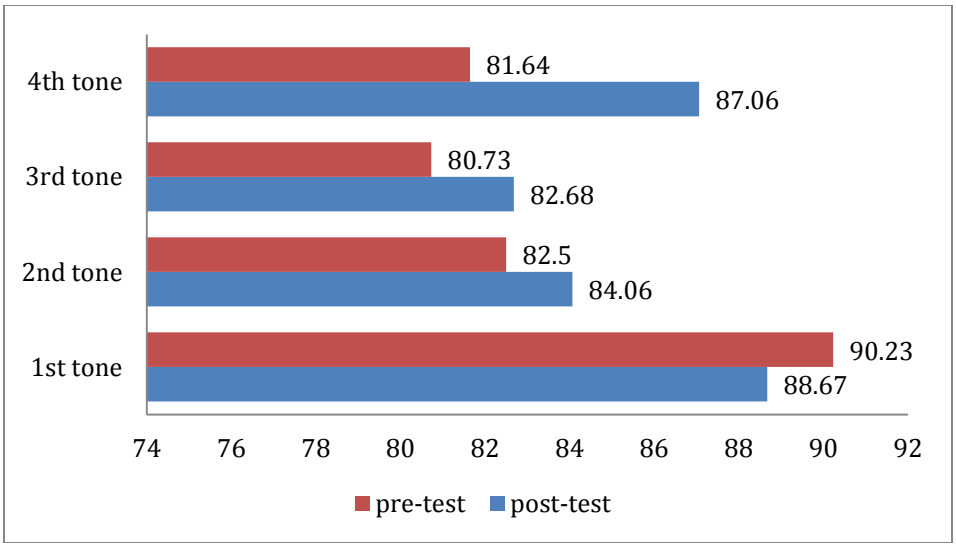

Figure 7. Percentages (%) of Correct Tones in Pre- and Post-tests.

Re-analysis of Pre- and Post-Tests

Acknowledging the unexpectedly high initial scores, we re-analyzed the pre- and post-test results and focused on the 16.61 % of the tones that were not produced correctly (i.e., did not receive a score of "1") in the pre-test (refer to Table 2, e.g., 1=mostly correct 83.39%; 4=mostly incorrect 5.67%). When comparing only the pre- and post-tests for the problematic tones (i.e., those rated "2," "3" or "4"), almost half of them, 49.5%, were rated "1" (mostly correct) in the post-test. The other 50.5% received ratings of "2," "3" or "4" in the post-test (see Figure 8).

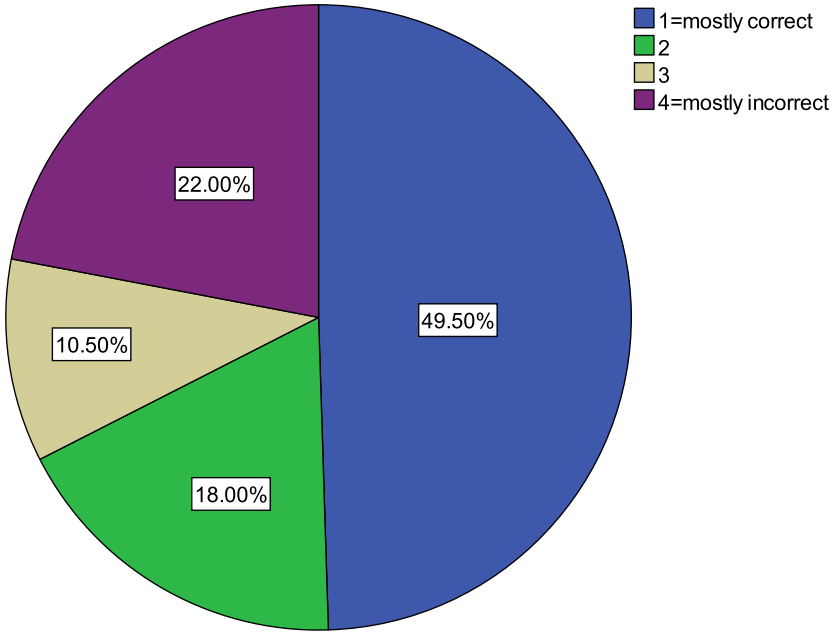

Figure 8. Distribution of Post-test Scores of the Incorrect Tones.

We also performed paired samples t-tests comparing the scores on the pre- vs. post-tests. In comparing all of the data, the mean of the pre-tests (M = 1.31, SD = 0.79) was not significantly different from the mean of the post-tests (M = 1.27, SD = 0.74), $t(1211) = 0.44$, $p > .05$. However, the mean scores differed significantly when comparing the ratings of the 16.61% of the tones that were incorrect in the pre-test with the ratings of these tones in the post-test, $t(199) = 10.25$, $p < .001$. The mean score of incorrect tones in the post-tests (M = 2.05, SD = 1.22) was .85 point lower than the mean score in the pre-tests (M = 2.90, SD = 0.88). The results showed that students made significant improvement in the incorrect tones with the aid of visualizing pitch curves.

Improvement Results

The second method of determining whether students' tone production improved from the pre- to post-test was to present the native speaker raters with the pre- and post-test recordings, one right after the other. Four native speakers then rated the pre- and post-test pairs from 1=great improvement to 4=no improvement or worse, and 5=no change, both correct. Table 3 shows the improvement results of the 16 students' scores. Due to the strong ceiling effect in the pre-test, the students produced both tones correctly, indicated by score 5, 67.39% of the time. The average of scores 1 to 3 is 17.01%, which means that students made a little, some or great improvement in the post-test approximately 17.01% of the time. Only 15.59% of the pre- and post-test pairs were rated "4," showing "no improvement or worse."

Table 3
Percentage (Pct.) of Scores in Improvement Analysis

Score	1	2	3	4	5
Avg. pct. (%)	5.78	4.36	6.87	15.59	67.39
Avg. of scores 1 to 3			17.01		

We further broke down the results by tones, and computed the means of percentages of scores 1, 2 and 3 (see Figure 9). The average percentages of improvement on 1st, 2nd, 3rd and 4th tone are, 20.24%, 14.29%, 17.91% and 15.87%. The results confirm that students achieved a certain degree of improvement on all four tones about 17% of the time. Tone 2 showed the least improvement and tone 1 the most improvement. Figure 10 is an example of a learner's pitch curves of the 4th tone syllable *kàn* ("to look") in pre- and post-tests. The learner appears to have made significant improvement with the aid of seeing pitch curves. However, Figure 11 is an example of another learner's pitch curves for the same word, but in this case, the pre- and post-tests are similar, both lacking the falling tone contour that is necessary for 4th tone syllables.

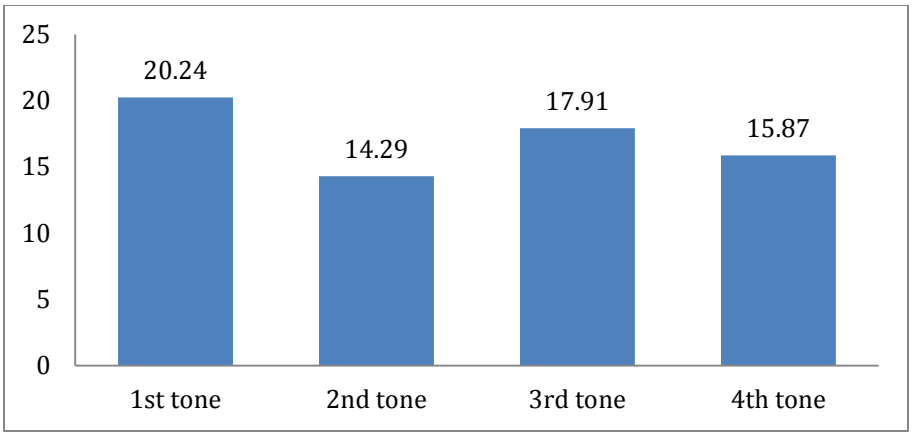

Figure 9. Average Percentage (%) of Improvement for Each Tone.

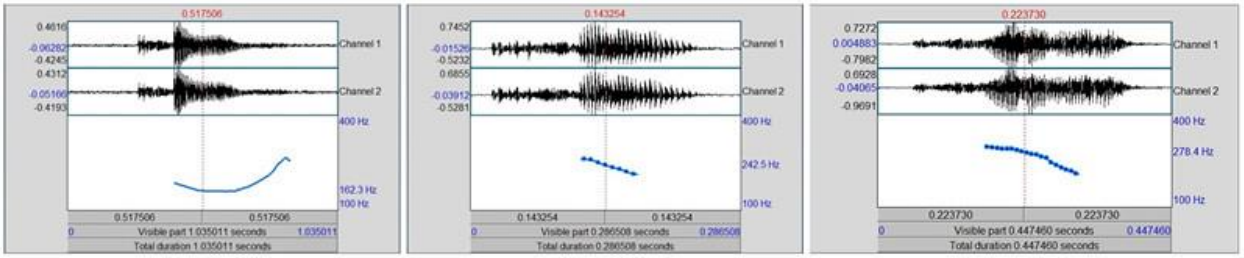

Figure 10. Pitch Curves of *kàn* ("to look") by Learner X in Pre-test (left), Post-test (middle) and by Native Speaker (right).

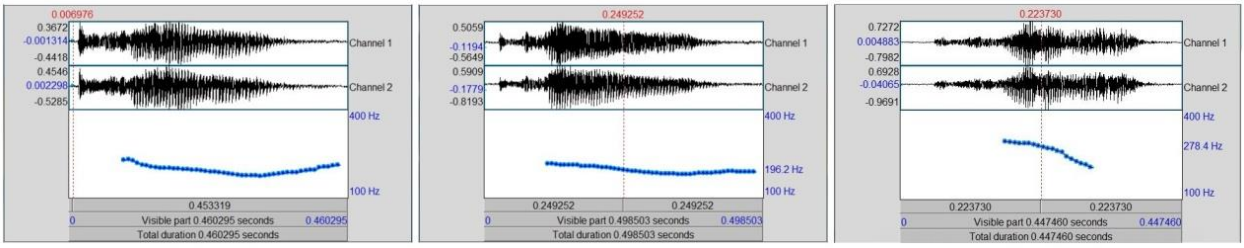

Figure 11. Pitch Curves of *kàn* ("to look") by Learner Y in Pre-test (left), Post-test (middle) and by Native Speaker (right).

DISCUSSION

The analysis of a representative sampling of the tone production data (19 words for each of the 16 participants, for a total of 304 samples) indicated, first of all, a marked ceiling effect in that 83.39% of the word tones were pronounced correctly or mostly correctly in the pre-test. This is likely due to the fact that these Chinese learners had been studying Mandarin for six months at the university level and that the selected words were all basic vocabulary items that learners were very familiar with.

In analyzing the production of individual tones, the data showed that 1st tones were pronounced correctly more often than any of the other tones, and that 4th tones were pronounced incorrectly more often than any of the other tones, corroborating earlier research by Shen (1989), who also found that 4th tones were most problematic for American learners of Mandarin.

Due to the ceiling effect and to the short duration of the training sessions, it is not possible to determine whether seeing the pitch curves had a direct effect on the learners' improvement from pre- to post-test, as found in other studies (Hardison, 2004). But the two types of comparison of pre- and post-test data both revealed some measure of improvement.

In the first rating method, raters scored randomized words from the pre- and post-tests, and a comparison of these ratings (averaged among all students and all raters) showed a marginal 1.84% improvement between pre- and post-test, with the greatest improvement of 5.42% in the most problematic 4th tone. However, a re-analysis of only the 16.61% of the tones that were incorrect in the pre-test showed that nearly half of them (49.50%) were pronounced correctly in the post-test. *T*-tests also confirmed that learners made statistically significant improvement in the incorrect tones between pre- and post-tests.

Using a second method to determine whether tone production improved, raters were presented with pre- and post-test recordings at the same time. A 17.01% improvement rate was found. In addition, a post-study survey indicated that the participants felt that seeing the pitch curves of the native speaker helped them improve their word tones (Figure 12) and that seeing their own pitch curves and practicing their pronunciation was also helpful (Figure 13).

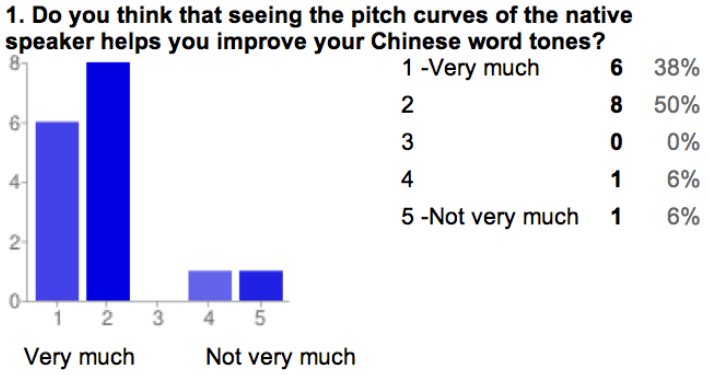

Figure 12. Student Survey Results on Helpfulness of Native Speaker Pitch Curves.

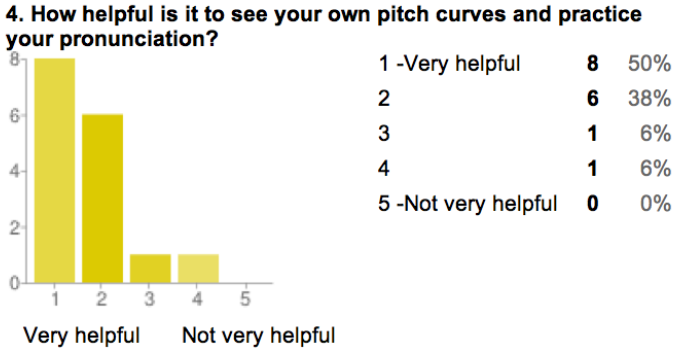

Figure 13. Student Survey Results on Helpfulness of Seeing Own Pitch Curves.

CONCLUSIONS

The purpose of this pilot study was to assess the usefulness of providing L2 Mandarin Chinese learners with visualizations of native speakers' pitch curves and the learners' own pitch curves while they tried to master Mandarin tones. The 16 learners were nearing the end of their first year of studying Chinese at the university level and proved to be very proficient at pronouncing the 19 basic, commonly used words and phrases, producing correct tones 83.39% of the time. Comparison of the learners' performance on pre- and post-tests did show improvement, with the amount of improvement varying, depending on the method of comparison. There was a statistically significant improvement in the 16.61% of the tones that were incorrect in the pre-test. The learners stated in a post-survey that they found the visualizations of the pitch curves to be helpful.

The results of this study will help in the design of a larger study with more students and not just volunteer participants. A longer, more systematic training program (i.e., more than two training sessions) will be implemented, and a more comprehensive list of words and phrases will be used, including words that might be less familiar to the learners. In addition, testing novice students who have not had as much exposure to Chinese may help reduce the ceiling effect. We anticipate that longer, sustained training with more words and phrases will provide more concrete and convincing evidence of whether visualizations of native speaker and learner pitch curves are effective for mastering Mandarin Chinese tones.

ACKNOWLEDGMENTS

We would like to acknowledge a generous Instructional Improvement Grant from UCSB and the assistance of Prof. Hsiao-Jung Yu, Dr. George Michaels, I-An Chang, Jingyu Huo, Hongran Peng, and Xiang Qiu, as well as all of our volunteer participants.

ABOUT THE AUTHORS

Dorothy M. Chun (Ph.D., UC Berkeley) is Professor of Applied Linguistics and Education at the University of California, Santa Barbara. Her research areas include: L2 phonology and intonation, L2 reading and vocabulary acquisition, computer-assisted language learning (CALL) and telecollaboration for intercultural learning. She has conducted studies on cognitive process in learning with multimedia and has authored CD-ROMs and websites for language and culture acquisition. Her publications include *Discourse Intonation in L2: From Theory and Research to*

Practice, 2002, John Benjamins and articles in *CALICO Journal*, *CALL*, *Foreign Language Annals*, *Journal of Educational Psychology*, *Modern Language Journal*, *System*, and the online journal *Language Learning and Technology*, which she has edited since 2000.

Yan Jiang is a Ph.D. student of Gevirtz Graduate School of Education at the University of California, Santa Barbara. Her research interests are: L2 teaching and learning and computer-assisted language learning (CALL). She received a B.A. in English from Tsinghua University, China, and a M.A. in Education from San Diego State University.

Natalia Ávila is a Ph.D. student of The Gevirtz Graduate School of Education at the University of California, Santa Barbara. She received a BA in Hispanic Language and Literature and in Education, and a MA in Linguistics from Pontifical Catholic University of Chile. She has authored an online composition course in Chile. Her research interests include composition, second language writing and digital literacies.

Dorothy Chun, Department of Education, UCSB, dchun@education.ucsb.edu

Yan Jiang, Department of Education, UCSB, yan_jiang@umail.ucsb.edu

Natalia Ávila Reyes, Department of Education, UCSB, naavila@umail.ucsb.edu

REFERENCES

Chen, Q-H. (1997). Toward a sequential approach for tonal error analysis. *Journal of the Chinese Language Teachers' Association, 32*(1), 21-39.

Chun, D. M. (1998). Signal analysis software for teaching discourse intonation. *Language Learning & Technology, 2*, 61-77.

Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology, 8*, 34-52.

Levis, J., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System 32*(4), 505-524.

Miracle, W. C. (1989). Tone production of American students of Chinese: A preliminary acoustic study. *Journal of the Chinese Language Teachers' Association, 24*, 49-65.

Molholt, G., & Hwu, F. (2008). Visualization of speech patterns for language learning. In V. M. Holland & F. P. Fisher (Eds.), *The path of speech technologies in computer assisted language learning* (pp. 91-122). NY: Routledge.

Shen, X-N. S. (1989). Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers' Association, 24*(3), 27-47.

Wang, Y., Sereno, J., and Jongman, A. (2006). Second language acquisition and processing of Mandarin tone. In Li, P., Tan, L.H., Bates, E., and Tzeng, O.J.L. (Eds.), *Handbook of East Asian Psycholinguistics* (Vol. 1: Chinese). Cambridge, UK: Cambridge University Press

Appendix A.

Student Demographic Questionnaire

Name:

Age:

Gender:

Year (freshman, sophomore, etc.)

Major (and Minor, if applicable):

Native/Home language:

Parents' native language:

Amount of time spent abroad and location:

Length of time studying Mandarin:

Do you speak Mandarin outside of class? With whom and how often?

Do you listen to Chinese songs or watch Chinese TV shows or movies? How often?

Other languages you speak and/or write: