Effects of F0 Feedback on the Learning of Chinese Tones by Native Speakers of English

Felicia Zhang¹ and Michael Wagner²

¹School of Languages and International Education
and
²Human-Computer Communication Laboratory, School of Information Sciences & Engineering
University of Canberra, Australia
(felicia.zhang, michael.wagner)@canberra.edu.au

Abstract
This paper reports on the effects of an interactive software tool on the learning of Chinese tones by English-L1 learners of Chinese. The software enables students to listen to model utterances by native speakers of Chinese and to record their own renditions of those utterances. Simultaneously the F0 contours of both the model utterances and the student utterances are displayed on the computer screen. The effect of the interactive software on the L2 tone production performance by a group of students is compared with that of a control group. The students who could make use of the software show a better approximation of native speakers in F0 mean and range.

1. Introduction

The subject Chinese Language and Culture Part 1 is read by beginning students of Chinese at the University of Canberra. Beginning in 2003, students in that subject had available to them an interactive computer program, which they could use at their own leisure to listen to course materials spoken by native speakers of Chinese. The program also displays waveforms and F0 contours and allows students to record their own rendition of the materials and compare the corresponding F0 contours. Four cohorts of students who were enrolled in the course in the years 2001 to 2004 were evaluated with respect to their performance in producing correct Chinese intonation after the first six weeks of the course. All of the students spoke English as their first language and were total beginners of Mandarin Chinese. Those who studied the course in 2001 and 2002 did not have the benefit of the interactive software support. The E-Group was also taught using the communicative teaching method with teaching materials on CD-ROMs, but without interactive software support. The E-Group was also taught using the communicative teaching method and it had available to it, and was encouraged to use, the interactive software tool.

2. F0 feedback tool

An interactive computer program Sptool was developed for beginning students of Chinese [3]. The program provides a graphical, as well as sound, user interface that allows students to play back Chinese phrases, spoken by native speakers, through the computer’s sound system and to simultaneously display the waveforms and F0 contours on the computer screen. By clicking on a screen icon, students are able to hear as many repetitions of the material as desired and to see the F0 contours belonging to the different syllables of the material.

By clicking on another screen icon, students are then able to record their own rendition of the material through the computer’s in-built microphone or through an attached external microphone. Subsequently, the students will see both the native speaker’s F0 contour more closely. Hence the purpose of Sptool is to enable students to practise the correct intonation of Chinese syllables, words and entire sentences by direct comparison of their F0 contours with those of the native model speaker, as is shown in Figure 1. The convenient interface allows students to repeat their recordings many times in quick succession, each time attempting to approximate the native model speaker’s F0 contour more closely. Hence the purpose of Sptool is to enable students to practise the correct intonation of Chinese syllables, words and entire sentences by direct comparison of their F0 patterns with those of native speakers, in their own time and in addition to the normal face-to-face contact with the language teacher.

3. Experimental methodology

3.1. Experimental data

After six weeks of instruction in the Chinese language, the students in both groups were given an assignment in
which pairs of students had to prepare a five- to ten-minute conversation in Chinese on one of a given list of topics. The conversations, which had to be conducted without written notes, were recorded using a desk microphone connected to a notebook computer at 44.1kHz and 16 bits/sample. The sides of the conversations were edited using CoolEdit 2000 and 60 seconds from every subject’s recording were chosen for analysis.

For comparison, similar conversational material was recorded for each subject in their native English. For further comparison, Chinese speech data from 13 male native speakers of Mandarin Chinese were taken from news broadcasts and other conversation materials. For all recordings, the F0 contours were extracted by means of the Praat speech analysis tool [2].

### 3.2. Effect of feedback on learners’ F0 range

The tonal productions of the 7 male subjects of the Control group and the 5 male subjects from the Experimental group were analysed using Praat. On a frame-by-frame basis, the fundamental frequency was determined for each voiced frame of each subject’s test data. From the collection of F0 values for each subject, the mean, standard deviation, minimum, maximum and range were determined for that subject. In turn, the group averages were determined for those 5 variables.

Table 1 shows those group statistics for the native Mandarin speakers (NMS), the E-Group subjects speaking Chinese and speaking English, and for the C-Group subjects speaking Chinese and speaking English, respectively.

Table 1 shows that average minimum fundamental frequency lies between 80Hz and 85Hz and hardly varies between the 5 cohorts. This is not surprising as the lower bound of F0 is strictly limited by the physiology of the human vocal cords. This is not the case to the same extent for the maximum F0, which is determined more by individual effort than by a strict physical limit. Accordingly, Table 1 shows a much larger variation of maximum F0. The native speakers have a higher average maximum F0 (268.19Hz) than both of the student groups speaking English (235.05Hz and 253.30Hz). This is clearly due to the tonal characteristic of Chinese where a majority of syllables is pronounced with tones that are at, or rise to or fall from, a high fundamental frequency.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Mean (Hz)</th>
<th>Std (Hz)</th>
<th>Min (Hz)</th>
<th>Max (Hz)</th>
<th>Range (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSM</td>
<td>131.52</td>
<td>30.83</td>
<td>80.23</td>
<td>268.19</td>
<td>187.97</td>
</tr>
<tr>
<td>E(G(Ch)</td>
<td>131.62</td>
<td>29.19</td>
<td>82.61</td>
<td>284.39</td>
<td>202.55</td>
</tr>
<tr>
<td>E(G(En)</td>
<td>115.26</td>
<td>22.39</td>
<td>81.84</td>
<td>253.30</td>
<td>171.46</td>
</tr>
<tr>
<td>C(G(Ch)</td>
<td>115.49</td>
<td>20.57</td>
<td>84.14</td>
<td>245.54</td>
<td>161.40</td>
</tr>
<tr>
<td>C(G(En)</td>
<td>110.16</td>
<td>16.68</td>
<td>80.92</td>
<td>235.05</td>
<td>218.37</td>
</tr>
</tbody>
</table>

Table 1. Group averages of speakers’ mean F0, minimum F0, maximum F0 and F0 range for 5 groups of male speakers

Table 1 shows further that both the C-Group and the E-Group raise their maximum F0 when they switch from English to Chinese. The C-Group raises its average maximum F0 by 10.49Hz to 245.54Hz, while the E-Group raises its average maximum F0 by 31.09Hz to 284.39Hz. Thereby, the C-Group falls well short of the maximum F0 of the native Chinese speakers while the E-Group raises their average maximum F0 higher than the native speaker group.
Accordingly, the average mean F0 for the native speakers is 131.52Hz, higher than the 2 groups, which have average mean F0s of 110.16Hz and 115.26Hz. When switching to Chinese, the E-Group raises its average mean F0 by 16.36Hz to 131.62Hz – almost the same average as the native Chinese speakers. By contrast, the C-Group only manages to raise their average mean F0 by 5.33Hz to 115.49Hz – well short of the native speakers.

The difference between E-Group and C-Group is shown even more clearly by the differences between the groups of F0 variance. For each subject, the F0 standard deviation (SD) is determined and those standard deviations are then averaged for each group. The native speakers have an average F0 SD of 30.83Hz, while the two groups speaking English have average F0 SDs of 16.68Hz and 22.39Hz. When switching to Chinese, the C-Group raises its average F0 SD by 3.89Hz to 20.57Hz, while the E-Group raises its average F0 SD by 6.80Hz to 29.19Hz, almost reaching the F0 variation of the native Mandarin speakers.

The results show that both groups of learners increased their mean F0 and their F0 variance when switching from English to Chinese, but that the group, which had access to F0 feedback using Sptool, achieved the necessary rise in F0 and extension of their F0 range to a much larger degree than the control group. Indeed, the results show that the former actually came very close to the F0 statistics of the native speakers while the control group fell considerably short.

Since the Experimental group of students obtained comparative feedback of their own F0 contours and those of model native speakers, a key hypothesis was that students learning using this approach would be able to ‘stretch’ their F0 range when speaking Mandarin or practising Mandarin. Acoustically, this means that their average F0 range should be similar to that of the native speakers. The standard deviation of the average pitch should also be similar to the native speakers’. Furthermore, the Experimental group’s mean SD, which measures the width of the pitch range, should be wider in Mandarin than when they are speaking their native language-English. One explanation for the significant difference between the E- and C-Groups is that the E-Group managed to significantly widen their pitch range when speaking Mandarin because they had the benefit of F0 feedback. The similarity in the mean F0 level between the E-Group and the native-speaker group means that students from the E-Group significantly adapted their F0 range towards the native models when speaking Mandarin. This is also evidenced by the comparison of the standard deviations of the average mean F0 of their English and Mandarin speech production. When speaking English, the mean standard deviation on the group average of mean F0 for the Control group is 20.57Hz. When speaking Mandarin, the mean standard deviation on the group average of mean F0 is 22.39Hz. When speaking Mandarin, the mean standard deviation on the group average of mean F0 for the male subjects is 29.19Hz. This represented an increase of 30.37%.

However, the increase in the Experimental group’s average means in Mandarin is not wide enough. Using Native speakers as a guide, further expansion of the E-Group’s mean standard deviation on the group average of mean F0 is required.

The significant differences between the Control group and the Native Speaker group suggest that the average F0 range of the C-Group of students did not undergo significant expansion by learning Mandarin without F0 feedback. Further proof comes from the comparison of the Group average of SD of the mean F0 of their English and Mandarin speech production. When speaking English, the mean standard deviation on the group average of mean F0 for the Control group is 16.68Hz. When speaking Mandarin, the mean standard deviation on the group average of mean F0 is 20.57 Hz. This represented an increase of 23.32%. The smaller increase in the width of the F0 range suggests that the lack of F0 feedback for the L2 students in the Control group disadvantaged those students in the process of learning near-native tone production in Chinese. Interview data with this group of students supports this explanation. It revealed that though the importance of tones was recognized, it was only recognized on an intellectual level. When it came to speaking Mandarin, students found it hard to incorporate tones in their speech under examination conditions or during their daily practice of Mandarin. Without the diacritics, they felt lost, as the only visual reminders of tones had been taken away.

### 3.3. Effect of feedback on learners’ L2 production

In the second experiment, the L2 production performance by the two learner groups – both male and female students – was judged by native Mandarin listeners. From the C-Group, oral performance data, collected for the students at the end of six weeks (30 hours) of face-to-face contact, were analysed through both auditory and acoustic tests, specifically with regard to tonal errors. Similarly, the Experimental group of students, the data of the students were analysed through both auditory and acoustic tests. The markers were native speakers of Mandarin from the University of Canberra. The materials were the spoken conversations from both groups of English native speakers. The markers did not know which group each student belonged to at the time of marking. Students were marked on a scale from 1 (not understandable) to 9 (near-native quality).

Significant differences were found in the speech production by the Control and Experimental groups. Listeners perceived the E-Group of students as more fluent than the C-Group of students. Table 2 shows the mean scores of listening tests.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>StdDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-Group</td>
<td>10</td>
<td>4.51</td>
<td>0.73</td>
</tr>
<tr>
<td>E-Group</td>
<td>11</td>
<td>5.33</td>
<td>0.82</td>
</tr>
</tbody>
</table>

**Table 2. Production quality scores for the 2 student groups.**

The E-Group scores were significantly higher (p=0.02) than the C-Group scores. To evaluate the reliability of the data obtained, the consistency of the markers’ perceptual ratings was also assessed. An inter-rater coefficient has been calculated [4] as:

\[
R_v = \frac{r_{n \times n}}{1 + (n - 1) r_{n \times n}}
\]

In this formula, \(r_{n \times n}\) stands for the reliability for all the markers’ ratings, \(n\) is the number of markers, and \(r_{n \times n}\) is the average correlation of the 9 markers. The inter-rater reliability rating obtained is 0.90, indicating very good agreement between the 9 markers.
3.4. Analysis of the learners’ tone production errors

An analysis of the tone production by the 2 student groups after 6 weeks of study was undertaken by 3 native listeners, who were also experienced instructors of Mandarin L2. They separately marked any deviant tonal production of a syllable with Chao’s tone letter system. A tone was considered mispronounced when judged as inaccurate by at least two of the listeners. Based on the auditory analysis, an acoustic study was carried out to verify the mispronounced tones.

The 10 students in the C-Group made 326 errors out of 1507 tokens, for an error rate of 22% from the oral input from each student in the C-Group. Since the oral production was not content-controlled, the distribution of different tones within each student’s sample varied. Figure 2a shows the tone error distribution over the C-Group.

Similarly, in the E-Group the 11 students made 674 errors out of 2536 tokens for an error rate of 26.58%. Figure 2b shows the tone error distribution over the E-Group.

![Distribution of Errors for Each Tone for the Baseline group](image)

![Distribution of Errors for Each Tone for the Experimental group](image)

Figure 2. Tone error distributions for (a) C-Group, (b) E-Group.

According to these results, we propose an order of tone difficulty for L2 learners with Tone 4 being the most difficult for learners, followed by Tone 3, Tone 1 and then Tone 2.

As the conversations were about oneself, family and hobbies, the data in both groups’ data contained many tone-3 pronouns such as ‘wo3’ and ‘ni3’. Due to the production of many questions involving the use of ‘sui4’ in the sentence-final position, tone-4 errors were also frequent. In fact, there was overwhelming evidence that the endings of sentences were a problem area for learners, especially when Tone 4 occurs in the final position of a sentence or question. This result is in agreement with [5] that large discrepancies exist between the tonal patterns of Tones 2, 3 and 4 before internal phrase boundaries and at the end of utterances. The contour errors made by the Experimental group consisted primarily of replacing the level contour with a falling contour.

The data on Tone 2 from both groups of students differ from native speakers’ production in a fundamental way. The register mistakes consisted primarily of beginning the tone too high in the tone space and producing Tone 2 with a down-trend rather than uptrend. According to [6], native speakers produce a low register dipping tone (212) in single tone/syllable words and either a falling or level contour in the low register when in combination with other syllables. However, complicated tone sandhi rules for Tone 3 mean that mistakes were frequently made with this tone. The contour errors made by the students consisted primarily of realizing Tone 3 as a rising contour or replacing a level tone with a falling or dipping contour. The confusion between the falling and level contours is understandable given this tone’s various realizations by the native speakers.

Students from both Control and Experimental groups realized most of the Tone 4 as the expected downward contour. However, two common problems remained, compared with the native speaker data. One is that students did not start Tone 4 high enough in the tone space. While most of the students in the Experimental group started Tone 4 above their average baseline F0, student 3 sometimes started above, sometimes below the mean F0 of his voice.

With respect to Tone 3, it has been suggested recently [5] that it should be taught as low-dipping. However, it is unlikely that this teaching strategy would solve every learner’s problems with this tone. As an alternative teaching method, using the materials covered in this study as an example, it could be better to group all the sentences starting with ‘wo3’ together and then ask the students to observe the behaviour of Tone 3 in these utterances. Out of 104 sentences, 47 sentences contained Tone 3 in various combinations. Out of the 14 sentences in Week 1’s materials, 10 sentences had Tone 3 in them. Therefore, the need to sensitize learners to the behaviour of Tone 3 in various combinations is important.

4. Conclusions

The experiments reported in this paper indicate that an interactive software tool which allows English-L1 students of Chinese to obtain audiovisual feedback including fundamental-frequency contours of both their own utterances and those of model native speakers, will improve the students’ production of correct Chinese tones and that the overall intelligibility of the students’ L2 production is significantly improved.

5. References