Real-time Display of Pitch-time Diagram as an e-Learning Tool for Acquisition of Mandarin Tones

Daphne Hui-Ying Hsieh (謝慧鶯)
School of Languages and Comparative Cultural Studies
The University of Queensland
Brisbane, Australia
d.hsieh@uq.edu.au

Mandarin tones are the changes in the fundamental frequency in the spoken words. With modern day personal computers it is possible to extract the fundamental frequency in a straightforward manner as pitch-time diagram. Real-time extraction of pitch-time diagrams can be used as visual feedback to assist the learner in identifying and correcting the learner’s tone errors. In this paper we describe the design and use of a pitch-time visualisation software as an e-learning tool for the acquisition of Mandarin tones. Initial experience shows that the visual comparison of students’ tone patterns against that of a reference pronunciation is very effective for a learner to self-evaluate and correct errors in his/her Mandarin tones.

Introduction

Mandarin tone is a suprasegmental feature of the language similar in its physical nature to intonation in non-tonal languages such as English. Functionally Mandarin tones serve to contrast lexical meanings. It is well known that it is difficult for adult speakers of a non-tone language to acquire near native proficiency in Mandarin tones in continuous speech. Incorrect use of Mandarin tones is a major problem in oral communication often leading to confusion due mainly to the rather limited number of Mandarin syllables. With increase in the importance of oral communication skills as a component of overall language proficiency, greater emphasis is placed on the acquisition of correct Mandarin tones.

The difficulty of Mandarin tone acquisition lies not in the production of the different tones per se as our experience indicates that most adult learners can articulate the four Mandarin tones accurately when pronounced in isolation. The difficulty lies in articulating the correct tones in continuous speech. Traditionally acquisition of Mandarin tones involves “listen and repeat” process. Our experience indicates that this is not an effective process: students often do not realise the errors in their speech by
aural comparison of their own pronunciation against a reference articulation by a teacher or on recorded media.

Visualization of Pitch-Time Diagram as a Feedback Tool

Visualization feedback scheme has been proposed since 1960’s as a mean for learning spoken languages [See Chun, 1989 for a description of the early development in this field]. In the case of tone feedback, this involves the use of the extracted tone or pitch patterns of spoken sentences. The perceived tone or pitch is simply change with time of the fundamental frequency, $F_0$, of the speech signal (Howie, 1974). $F_0$ arises from the vibrations of the vocal folds. $F_0$ can easily be extracted by pitch extraction software almost instantaneously. To be an effective tool, the feedback must be easy to use and easy to understand. The wide availability of personal computer (pc), pc-based sound recording software and pitch extraction software makes the use of pitch feedback possible. The fact that present day personal computers are operating at a speed sufficient to provide near to real-time pitch-time diagram is another factor making visual feedback a viable tool for computer-aided learning of spoken Chinese. A major requirement is that the software should allow a learner to visually compare his/her own tone pattern against that of a reference utterance, as well as the corresponding pinyin transcript, and thus identify his/her own errors and then apply the necessary correction. Of course the learner must have a basic understanding of the relation between the visual pitch patterns and the perceived tones. This is typically taught as part of the introduction to spoken Chinese.

In this paper we report on the development and the use of such a Mandarin tone visualization software to assist adult learners in their acquisition of Mandarin tones. The software is based on the widely available pc sound card, the multimedia control interface facility of the Microsoft Windows [Microsoft] and pitch extraction software from Summer Institute of Linguistics [Summer Institute of Linguistics]. It is simple to use and allows students to directly compare their own tone patterns of given sentences against those of a native speaker in “real time”. It is found that the visual comparison of learners’ tone patterns against those of the reference has been very effective in indicating the major tone errors in learners’ production and hence in assisting learners to correct their production. This is carried out at the learners’ own time and pace without requiring the help from a language teacher/tutor. The last point is of particular significance in view of the educational trends of self-directed acquisition of language skills through e-learning.

The effectiveness of this software is tested on a group of 15 first year university students who have studied spoken Chinese for one semester over a period of 3 months. Their assessment of the software and the effectiveness of the software in improving their tone production are also reported.

Tone Visualization Software

Our software is developed with Visual Basic language and utilises the Microsoft Windows multi-media interface facilities for both the recording and playback of
student’s utterances of Mandarin sentences (listed in pinyin transcription on the screen). The reference production of these sentences is stored as sound files in .wav format. In operation, a student selects a particular pinyin sentence to practice and plays the reference production; he/she then proceeds to record his/her own utterance of the same sentence. The student’s utterance is appended to a copy of the reference utterance as a temporary sound file. This file is then analysed by the Speech Analyzer and displayed as a pitch-time diagram (together with recorded sound waveform) which shows the pitch pattern of the reference production followed by that of the student’s own. The block diagram in Fig. 1 summarises the operation of the software.

Fig. 2 shows the computer screen when the Speech Analyzer is activated: on the left half of the screen is the list of pinyin sentences (with the selected sentence shown with a different background colour). On the right is the Speech Analyser Screen with the sound waveform on the top and the pitch-time diagram on the bottom. By comparing the pitch patterns of his/her own utterance against those of the reference, a student is able to determine for him/herself the tonal errors and repeat the exercise to achieve a better match between the two and hence more accurate tones. The pinyin sentence complete with tone marks also allows a separate check. The Speech Analyzer also provides the facility to playback a selected part of the sentence for more detailed analysis of a specific part of the sentence. The software also has the facility for student to check the pronunciation of specific Mandarin syllable (input by pinyin) in isolation in its various tones.

To demonstrate the effectiveness of visual feedback, we show in Fig. 3 the expanded view of the Analyzer window with the pitch patterns of the reference voice (female) followed by that of a female student’s production for the sentence: “Xiànzài chà shí fēn jiǔ diǎn le. Wǒ gāi zǒu le.” The pinyin sentence is superimposed on the figure to identify the pitch contour of the individual syllables. Comparison of the two pitch patterns show up the errors in the student’s utterance. For example the 4th syllable “shi” is in tone 2 and should have a rising pitch contour as is the case in the reference utterance. The student has however uttered it as a low dipping pitch resembling a half tone 3 contour. All together there are six errors in the student production and these have been marked on this figure.

**Evaluation of Software**

This software was tested for the first time in 2004 with 15 English-speaking adult students who have just commenced in their second course in Spoken Chinese in the University of Queensland. They have near identical language background – all have taken a 13-week first course in Spoken Chinese in the previous semester where they have been taught the nature of Mandarin tones and have practiced spoken Mandarin during that course. Prior to that, they have no experience with Chinese language or other tone languages. At the time of participation in this study, all these students have a good appreciation of the importance as well as the difficulties of acquiring accurate tones in continuous speech.
All participants were given a half-hour introduction and hands-on practice on the software so that they are confident in using the software as well as being able to detect their own tone errors by visually comparing their pitch-time diagram against that of a “reference”. They were also instructed on the use of other related facilities provided by Speech Analyzer, such as replaying part of a stored sentence. They are also reminded of instances of tone sandhi in Mandarin. For correcting the tone error in a particular syllable the participants were taught that if in doubt recite the four citation tones of the syllable in “mid sentence” to check which fits (the reference tone pattern). The actual study involves the students using the software in the manner described above to assist them in practising Mandarin tone production of sentences from Lesson 10 to 13 of Chinese For Today [Huang et al, 1996], the textbook for the course in Spoken Chinese the participants are enrolled in. Selected sentences are chosen from each lesson and recorded as sound files and tied to the corresponding pinyin sentences stored in the computer. Participants worked at the rate of one lesson per week as set by the rate of progress in the Spoken Chinese course. Beside this study where the participants aim at acquisition of correct Mandarin tones, they also attend to the other aspects of the course, such as vocabulary, grammar etc. with the rest of the class.

To assess the effectiveness of tone visualization in improving the tone habits, the participants were required to provide a recording of their reading of ten pinyin (including tone marks) sentences prior to and after four weeks of use of the software. These sentences, tabulated in Fig.4, are derived from the lessons in the students’ textbook. The emphasis is for students to articulate them based on the pinyin provided. Comparison of these two recordings is used to assess their improvement in Mandarin tones production. The assessment is concentrated on the quality of the tones of the individual syllables and not on the accuracy of the pronunciation of the initials and finals of these syllables. The assessment were carried out aurally by two native speakers (i.e. not using the tone extraction software) and a score of 1 is given for each syllable perceived to be uttered in the correct tone and a score of 0 for a tone perceived to be incorrect. The participants were scored in terms of the % error in tones made before and after using the software. Below is the summary of the outcome of the experiment:

<table>
<thead>
<tr>
<th>Tone Accuracy Before &amp; After Experiment</th>
<th>Average Correct No. of Syllables (%)</th>
<th>Std Deviation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total No. of Syllables =79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before using software</td>
<td>79.7</td>
<td>9.8</td>
</tr>
<tr>
<td>After using software</td>
<td>85.0</td>
<td>10.1</td>
</tr>
</tbody>
</table>

It is not clear how much of the improvement can be attributed to the use of the software as we did not have the opportunity to assess a control group of students who did not have access to the software. Nevertheless several participants showed that they have been able to rectify most of the glaring tone errors typically observed in students at
this stage of language acquisition. This is particularly encouraging in that the improvement has been achieved without the use of native language tutors. Participants have also commented on the ease of use of the software as well as being an effective way to detect tone errors.

On the operation of the software, it is found that participants have a tendency to be too soft in their voice when they were uttering the sentences. This often results in the pitch pattern not being extracted. Another operational problem is 50 Hz electrical mains induced hum in the recorded signal with some of the recordings due either to the poor quality of the microphone amplifier or poor grounding of the computer. Under this unsatisfactory condition, there is the tendency of Speech Analyser to home onto 50 Hz as the pitch frequency instead of the actual pitch frequency of the recorded voice signal. The fact that the microphones used have a low frequency gain roll-off at about 80 Hz has also reduced the low frequency components in the recorded signal, again leading to the loss of extracted pitch patterns at or below this frequency, this is particularly so with the tone 3 syllables uttered by male speakers. All these are technical problems which could be solved by appropriate changes in recording conditions.

Conclusions
We have developed a Microsoft Windows based software that allows learners of spoken Mandarin to extract the tone patterns of their spoken Mandarin and to evaluate the accuracy of these tones by comparison with those extracted from same sentences spoken by native speakers.

Initial evaluation of the software by a group of 15 adult English speakers who are taking first year Spoken Mandarin in the University of Queensland shows that the software is well accepted by the students as a learning tool. Assessment of their tone production of a set of short sentences before and after practicing with the software for four weeks indicates that the software has assisted the learners in detecting and hence removing typical gross tone errors observed in students with similar language background.

References
Fig. 1 Block diagram showing the main operations of the software

Recordings of 
reference 
utterances of 
sentences (as .wav 
files in computer)

Display of 
pinyin 
sentences

Selection of 
sentence & 
reference 
.wav file

Sound playback, 
extraction and 
output of pitch-
time diagram of 
reference + 
student’s input

Recording student’s 
utterance & 
appending to a copy 
of reference 
.wav file

Fig. 2. Computer screen when the Speech Analyzer is activated. Left hand side is the list of sentences to be selected (the selected sentence shown up with a different background colour). Right hand side is the Speech Analyzer screen: top half is the speech waveform and the bottom is the pitch-time diagram of the selected sentence.
Fig. 3. Speech Analyzer screen for the sentence “Xiàn zài chá shí fēn jiǔ diǎn le. Wǒ gāi zǒu le”. The main tone errors in the student’s production as seen in the pitch-time diagram are enclosed in dotted ellipses and the pinyin syllables underlined.

1. mā mā mā mā
   (Four tones of the syllable "ma")
2. Nǐ gěi wǒ mài jǐ ge xīn féng.
   (Please get me some envelopes)
   (It’s not mine. I’ve got mine here)
4. Yǐ ge xiǎoshí néng dào jīchāng ma?
   (Can we get to the airport in an hour?)
5. Tīng shuō zhè de tāngcūyú bǔcūo.
   (I’ve heard that the sweet and sour fish here is very good.)
6. Xiǎwǔ zánmen qù kàn diānyǐng, hào ma?
   (How about going to a movie with me in the afternoon?)
7. Xiànzhāi chá shí fēn jiǔ diǎn le. Wǒ gāi zǒu le.
   (It’s ten to nine now. I must be going.)
8. Lǎoxiānshēng chángcháng lái gōngyuán ba?
   (Sir, do you come to the park often?)
   (Mother Mā scolds the horse)
10. Shīshī shī shī shī.
   (History of Mr Shī eating a lion.)

Fig. 4. Pinyin sentences used to assess students’ tone production ability before and after using the pitch tracking tool.