A Chinese Tone Learning Platform Based on Advanced Speech Technologies

Chung-Hsien Yang ¹, Chin-Chin Tseng ², Yi-Jing Lin ¹

¹ L Labs Inc., Taipei, Taiwan
² Graduate Institute of Teaching Chinese as a Second Language, National Taiwan Normal University, Taiwan

ch.yang@llabs.com, tseng@ntnu.edu.tw, yj.lin@llabs.com

Abstract

This paper describes the tone-learning features of the MyET-MyCT software, which is a computer-assisted language learning platform based on advanced speech technologies. We describe the theories involved, the advantages it offered, and the challenges we faced when developing such a computer-assisted language learning system.

Index Terms: computer assisted language learning, automatic speech analysis, Chinese tones, Mandarin Chinese learning

1. Introduction

MyET-MyCT is a computer-assisted language learning (CALL) platform that focuses on the training of speaking English and Mandarin Chinese [1], [5], [6], [9], [13]. It is based on an advanced computer speech technology called “Automatic Speech Analysis System” (ASAS), which can analyze English or Chinese speech and pinpoint problems to individual sounds. As a result, MyET-MyCT acts like a one-on-one English tutor, providing users with detailed correction on aspects of pronunciation, pitch, rhythm, and emphasis.

Figure 1. The user interface of MyET-MyCT.

MyET works by having users speak into a microphone and mimic the teacher’s speech. Users receive scores on their speech, for immediate and quantified feedback. MyET then corrects the user’s accent by pinpointing within the sentence, which word, which phonemes need improvement. MyET subsequently offers remedy, through 3D mouth movement animation, to help users visualize how to correct the specific phonemes. MyET also points out where the user’s pitch is different from the teacher’s, and whether the user is speaking too fast or too slow. Finally, MyET also evaluates where the user is placing emphasis in the sentence. With MyET, users can practice the correct accent repeatedly until they develop the desired accent and the confidence to speak in public.

According to some independent research, the correlation between scores given by MyET-MyCT and scores given by human teachers is about 0.8, while the correlation of scores given by different human teachers is also around 0.8. This means the correctness of scores given by MyET-MyCT is similar to that of human teachers [2], [3].

MyET-MyCT was launched in 2002 as a commercial product. Since then, it has been downloaded and used by more than one million registered users and 300 universities all over the world. In order to support users from different countries, we provide user-interfaces in five different languages, including English, Japanese, Korean, Simplified Chinese and Traditional Chinese. Users may also choose between pinyin and zhuyin (a.k.a. “Bopomofo”) as the phonetic symbols to transcript Chinese texts.

And since the four tones has always been a challenge to foreign students learning Mandarin Chinese, MyET-MyCT has included some special features to evaluate the students’ performance in four tones and then provide feedbacks and suggestions for improvement [4], [8], [10].

2. Visual feedbacks to students

When students speak Mandarin Chinese into the microphone of MyET-MyCT, the pitch contour of each Chinese word is shown as a red line on the user-interface. Students can see and compare the pitch contours of their speech and those of the teachers’ speech immediately. Besides, if the tone of a certain word uttered by the student is different from that of the teacher, the waveform of that word will be highlighted in different color. Yellow indicates minor problems and red indicates major problems.

Figure 2. Visual Feedback to Students.
Such a computer-assisted tone learning system offers several advantages over the traditional tutoring offered by human teachers. First, with the tones shown as lines, foreign students may get visually feedbacks right away and adjust their pitch by try-and-error. According to the feedback from our users, this visual feedback is much easier to understand than verbal descriptions. Second, the software is more patient than human teachers. While human teachers may be impatient when the students make the same mistakes again and again, the MyET-MyCT software never get bored or tired. Third, students will not feel embarrassed when they make mistakes. They feel much more comfortable when trying different approaches to improve their tones.

The visual feedback is not only helpful to students; it can also help the teachers in understanding the nature of Chinese tones. For example, most Chinese teachers will be surprised to find that the pitch of their third-tones utterance usually does not fall and rise as described in the textbooks. Instead, the pitch of the third-tone in most Chinese conversations usually falls without rising again. This may explain why some foreign students are so confused when learning the third tone. The problem is not that students cannot imitate the fall-and-rise of the third-tone, but that the teachers do not explain it correctly in the first place.

![Figure 3. The pitch contours of some third-tone words.](image)

Although the scoring mechanism of ASAS may not be as sensitive as the ears of human teachers, it is much more consistent. Scores made by human teachers are prone to be inconsistent because teachers may get tired and have moods. Besides, different teachers may use different standards when grading. According to some independent study made by the academia, the scores given by ASAS are as good as those given by human teachers.

### 3. Technical challenges

When we developed the algorithms and visual feedback system of MyET-MyCT tone grader, we faced several technical challenges. The first challenge is to detect the pitch contour of each Chinese word correctly. Although there are many well-known algorithms that can determine the pitch of a speech segment by calculating the fundamental frequency [15], [16], this task is not as simple as it sounds. While fundamental frequency is almost always present in finals of Chinese words, it is usually missing or unstable in initials. Trying to calculate fundamental frequency in some Chinese initials (e.g. /b/, /p/, /f/, /d/, /t/) may result in incorrect pitch values and confusing visual feedback to students. Therefore, we must first decide the boundaries of initials and finals in each Chinese word, and then calculate fundamental frequencies only for finals and certain initials. And in order to find the boundaries of initials and finals, some speech-to-text (STT) algorithms must be used in the first place [18].

Even for final sounds, the calculation of pitch values is not trivial either. While human ears can usually detect pitch correctly in a noisy environment, existing algorithms do poorly when the background noise is present [17]. Besides, the harmonics in a speech voice may interfere with the calculation of the fundamental frequency. Sometimes the pitch calculated by a pitch-detection algorithm is actually the second harmonic instead of the fundamental frequency. The resulting scores and feedbacks given to a student will thus be incorrect.

In order to solve these problems, we use some “median smoothing” algorithms to fix the miscalculation of pitch caused by background noise and unstable vocal cord vibrations [11], [17]. We also added special checking rules in our software to detect the anomaly caused by harmonics.

The second challenge we faced is to give a correct tone-score to each Chinese word uttered by students, based on the pitch contour we detected. By “correct,” we mean that the scores given by MyET-MyCT should be similar to the scores given by a human teacher.

In order to give scores to individual Chinese words spoken by students, we must determine which speech segments correspond to which Chinese words. Again, some speech-to-text algorithms must be applied to locate the words. This is especially difficult when certain words are missing in the speech, or when certain words are badly spoken. Certain rules have to be inserted into the MyET-MyCT program to deal with these “exceptions” that are quite common among beginner language learners.

After locating each Chinese word in a speech signal, we must then decide what a correct pitch contour of a certain Chinese word should look like. As stated earlier, we found that the pitch contours of correctly pronounced Mandarin Chinese words may not always match the simple description we found in textbooks. The pitch contour of a second-tone word does not always rise significantly; the pitch contour of a third-tone word usually falls without rising again; and the pitch contour of a fourth-tone word sometimes looks very similar to that of a third-tone word. We found that while pitch-contour of the same Chinese word may vary significantly when surrounded by words of different tones in a sentence, a native Mandarin Chinese speaker can easily identify the original tone intended by the speaker. But this is not so easy for a computer program. Special rules must be added to deal with the exceptions caused by different tone sequences. For example, the tone sequence of “3-3” should be read as “2-3” (e.g. “雨傘”); and the tone sequence of “3-3-3” is sometimes read as “2-2-3” (e.g. “總統府”), and sometimes as “3-2-3” (e.g. “小雨傘”).

And after we decide the correct pitch contour of a given Chinese word, we can then give a score to the student’s utterance of the corresponding word by comparing the student’s pitch contour against the “correct pitch contour”. The length, average frequency, and slope of the pitch contour should all be considered in the calculation.

The third challenge we faced is to calculate the correct score of a whole sentence from the individual scores of the composing words. Again, by “correct,” we mean that the scores given should be similar to the scores given by human teachers. We found that the scores perceived by a human teacher is usually not the average of scores of individual words. For example, if there are ten words in a Chinese sentence and three of them are badly spoken in terms of tones, the tone-score given by a human teacher is usually 50 or lower instead of 70.

Giving scores to tone performance is very subjective. Different Mandarin Chinese teachers may give very different scores to the same sentence uttered by the same student. And while most teachers cannot describe the scores they give by a mathematical equation, they usually know what a “right score” is and what a
“wrong score” is. Under this circumstance, what MyET-MyCT can do is trying to mimic the grading behavior of human teachers. We collected hundreds of Mandarin Chinese sentences spoken by different foreign students, got them graded by about ten human teachers, and then tried to come up with an algorithm that can give similar scores to the same sentences. This process usually involves try-and-error.

4. Tone training courses

In terms of course contents, MyET-MyCT is an open platform instead of a closed system. New speaking-training courses can be added onto the platform relatively easily. Currently, the MyET-MyCT course database contains more than 20,000 English or Chinese conversations or paragraphs, which is equivalent to more than 500,000 sentences. Most of these contents come from famous publishers, universities, or language institutions in different countries. School teachers may also create their own MyET-MyCT courses to meet their special requirements [9], [12], [14].

MyET-MyCT has been working with the Graduate Institute of Teaching Chinese as a Second Language (TCSL) of National Taiwan Normal University to develop a course especially designed for the training of Chinese tones. This course contains sentences of different tone patterns. And in order to let the students focus on the four-tones, most of the sentences are combinations of Chinese finals in different tones and do not have any meaning.

![Figure 4. A tone-training lesson.](image)

Students using this training course may choose to practice individual sentences (i.e. individual tone patterns) or read through a whole lesson. When practicing individual sentences, tone scores of the sentence will be shown instantly. Students may also check the performance of individual Chinese word by clicking on the "Pitch" button. When a student chooses to read through a whole lesson, MyET-MyCT will give a tone score to the whole lesson. The scores are also stored on the MyET-MyCT server. Teachers may check the scores afterwards.

Besides, MyET-MyCT also monitors the long-term tone performance of a student over time. As shown in Figure 5, a diagnosis report on tone performance will be generated for each student. This report includes the error rate of each Chinese tone and provides suggestions for improvement. By reading this report, students will know which tones they are having problems, and a teacher may give individual tutoring based on this information.

![Figure 5. Diagnosis report on tone performance.](image)

5. Online speaking competitions

To encourage students using MyET-MyCT to improve their speaking abilities, we host speaking competitions over the Internet from time to time. Currently we are hosting competitions based on the tone-training courses created by TCSL of National Taiwan Normal University.

These competitions are open to the public. Any Internet user may join these competitions any time by downloading the free MyET-MyCT software and clicking a link on the web page. The competitors are asked to read through a list of sentences that consist of different tone patterns. A score will be given right away. And if the user’s score is high enough to enter the top-40 list, his or her name will be shown along with their national flags. Moreover, the national anthem of the champion will be played on the web page like in Olympic Games.

![Figure 6. An online competition of Chinese tone-patterns.](image)

A student may take the same competition multiple times. But only the highest score of a student is shown on the top-40 list. This is to encourage students to practice more. According to our experience, some students will read the same article in a MyET-MyCT online competition more than one hundred times. Even if they do not win some prizes, this turns out to be a very good practice and their speaking ability improved significantly.

These competitions also serve as a tool for us to collect the error patterns of students. By analyzing the scores stored on the competition database, we know which sentences (or tone-patterns)
are most difficult to foreign students. And by analyzing the
nationality of individual competitors, we will know how the
mother-tongues relate to the difficulties in pronouncing Chinese
tones.

6. Future work
Although the MyET-MyCT platform has been launched for almost
ten years, we have been focusing more on English-learning than on
Chinese-learning. And our research on the Chinese tones has just
begun. We are still looking for better algorithms to calculate the
pitch and to give correct tone scores to individual words and
sentences.

But with its huge user-base, the MyET-MyCT platform may serve as a test bed for different algorithms and linguistic theories.
And we are expecting rapid improvements. If MyET-MyCT gives
incorrect grades to certain sentences, we usually receive complains
from teachers and students within days or even within hours.
Therefore, we will know which algorithms work and which do not
right away. And we will know which theories must be mended to
meet the test of the real world.

Besides, we are especially interested in analyzing the
relationship between mother-tongues and error patterns in Chinese
tones. By understanding this relationship, we will be able to
suggest specific tone-pattern practices to a student according to his
or her mother-tongue. This information may also serves as the
basis for a teacher to design a Chinese course targeting a specific
country. But in order to achieve these goals, more data must be
collected from our users, and more analysis is required on the data
we already collected. We think some simple data-mining
 techniques might be required to accelerate this process.

And ideally, all these mechanisms should be linked together.
MyET-MyCT should be able to analyze the error patterns of a
student when he or she is practicing Chinese conversations, and
then suggests and provides remedy courses right away.

7. Conclusions
In this paper, we described the theories, advantages, and challenges
involve in designing the tone-training module of a computer
assisted language learning platform. There are plenty of rooms for
improvement, but the current result is rather encouraging. At least,
we proved that it is feasible to provide tone training on a large
scale with computer software based on advanced speech

8. References
[1] Tsai, P-Hua, "Bridging pedagogy and technology: User evaluation of
pronunciation oriented CALL software," Australasian Journal of
assessment: A preliminary study on the correlations between machine
[3] Chen, Hao-Jan, "Examining the Consistency of Evaluations Provided
by Three Automatic Speech Recognition Systems," Proceedings of
2006 International Conference on English Instruction and
Assessment, National Chung Cheng University, 2006.
[4] Li, Wei, "Speech Technologies in College Classroom: Their
Pedagogical Implications in Teaching/Learning Mandarin
Pronunciation and Tones," The Fifth International Conference and
Workshops on Technology and Chinese Language Teaching in the
Feedback on EFL College Students' Pronunciation Learning" ("語音
辨識軟體回饋對發音學習成效之影響研究"), Ph.D. Dissertation,
National Taiwan Normal University, 2004.
[6] Chiu, Shu-Zhen and Chen, Chun Li, "Result and Analysis of using
MyET in English Classes in a Primary School("『數位外籍教師』
(MyET) 輔助國小英語學習的成效與分析"), Selected Papers
from the Fourteenth International Symposium on English Teaching,
[7] Chen, Jung-Yi, "Evaluating the Learner-centeredness of Online
Courseware in Taiwan" ("台灣線上英語學習軟體其『學習者為中心』之特性評鑑研究"), Master thesis. Taipei University,
Taichung, 2007.
[8] Shu, Zhao-min, Lian, Yu-Ren, Lin, Chin-Hsi, and Li, Li-Chin, "Using
Online Speech Technology in Teaching Real-World Mandarin
[9] Chen, Li-Yuan, "NTU Chinese - Presenting and Sharing a Global e-
Learning Website" ("『台大華語』<NTU Chinese>全球數位學習
網站課程 展現與分享"), Proceedings of the 2007 Annual
Conference of the Association of Teaching Chinese as Second
[10] Wu, Fu-Ru, "Implementation and Results of Using Mandarin Courses
on the <NTU Chinese> e-Learning Platform" (《台大華語》數位
平台之課程實踐與結果報告—以教學成果之社會實驗課程的實
施與學習者回饋分析), Collected Essays of The Fifth International
Conference and Workshops on Technology and Chinese Teaching in
Important Element of ESP for Engineering Majors", Proceedings of
2008 National Conference on ESP; Kai Nan University, Taiwan, pp.
Mandarin Pronunciations" ("華語語音教學的新方法"), Proceedings of
[14] Chen, Li-Yuan, "Cutting-edge e-Learning Courses of the 21st
Century - Implementation and Deployment of the <NTU Chinese>
e-Learning Platform" ("21 世紀尖端網絡課程- 台大華語數位平
台的建置與實踐"), Collected Essays of The Fifth International
Conference and Workshops on Technology and Chinese Teaching in
[16] Boersma, P., "Accurate short-term analysis of the fundamental
autocorrelation method for measurement of the fundamental
[18] Rabiner, L.R., "A tutorial on hidden Markov models and selected
applications in speech recognition", Proc. IEEE, vol. 77, pp. 257-