Computer Assisted Language Learning in Teacher Education: Training of Tones and Stress Patterns in Asian Languages

S. Chen & M. Fu

University of Newcastle, Australia

Abstract

This paper investigates the application of speech recognition technology to assist English speaking teacher trainees of Asian languages to command tones and stress patterns in oral communication as well as the educational value of using computer assisted language learning techniques for effective teaching in teacher preparation. It will discuss some discoveries of a research project supported by Australian Research Council and demonstrate a new computer software developed by the University of Newcastle.

1. INTRODUCTION

The teaching of LOTE (Languages Other Than English) in Australian universities and schools is now highly placed on the list of national education priorities since understanding of and ability of using LOTE, especially Asian languages, have a significant bearing on the well-being of the nation in terms of its social, cultural and economic development. In 1992, the Council of Australian Governments (COAG) commissioned a working group to report on the position of Asian languages and cultures in Australian education. In 1994, the National Asian Languages and Studies in Australian Schools (NALSAS) Task force was set up to implement the aim to support enhanced and expanded Asian languages and Asian studies provision through all school systems with a view to 25% of Year 12 graduates having studied a language other than English from Year 3 and all students having Asian studies infused across all curriculum areas. Chinese (Mandarin), Japanese, Indonesian and Korean were targeted as four priority languages [1]. The teaching of Asian languages is expanding rapidly in schools all over the nation, creating a most demanding task in teacher preparation.

The desirability of greater integration of higher linguistic and cultural proficiency has a direct impact on language and teacher education. Improvement of the quality of language teaching and teacher training has become an urgent and crucial issue. To both teachers and students with a background of English or European languages, it is widely accepted that some Asian languages are particularly difficult to teach and to learn in terms of their phonetic systems and writing systems. First, most Asian languages have linguistic distinctive features such as tones (Chinese, Vietnamese, etc.) or stress patterns (Japanese, Korean, etc.). Second, the written forms of many Asian languages are not Romanization but a variety of scripts such as characters (Chinese, Japanese and Korean). A United States government research report has recommend that an average of 840 hours of study for European languages in comparison with 2400 hours for Chinese, Korean and Japanese to achieve a basic proficiency [2].

World-wide, computer technology has been employed to assist language learning mostly in European languages. Computer assisted language learning is regraded as an effective way of student-centred language learning [3], [4]. The initial efforts to use computer in Chinese language learning was made by

0-7803-3173-7/96/$5.00

Chinese language staff at the University of Illinois in 1970's [5]. In recent years, the incorporation of multimedia has enabled Chinese teachers to combine text, sound, and images in lesson presentation and more software have been developed [6], [7].

In Australia there are a few projects focusing on the writing systems of Asian languages, for instance, learning Chinese characters [8], [9], [10]. However, no research has been done in terms of using computers to teach tones or stress patterns of Asian languages because of the difficulties in applying speech recognition technology to language education.

Modern speech recognition techniques include Linear Predictive Coding [11], Hidden Markov Models [12], Spectral Analysis with Fast Fourier Transformation [13] and Neural Network [14]. Many successful speech recognition packages are available for specific applications such as automatic telephone operators, automatic billing service, voice data logging, voice-activated security systems and voice-activated keyboards. There is a potential feasibility to use speech recognition technology to distinguish tones and stress patterns for the purpose of language education.

The current research at the University of Newcastle is to investigate the application of speech recognition technology to assist English speaking teacher trainees of Asian languages to command tones and stress patterns in oral communication and to investigate the educational value of using computer assisted language learning techniques for effective teaching in teacher preparation.

This research is an interdisciplinary cooperative study by staff from language education and computer engineering. It is a pilot study focusing on teaching and learning tones of Mandarin Chinese phonetic system to English speakers. Specifically, the study attempts to seek answers to two crucial questions:

- To what extent can speech recognition technology be used to detect the inaccuracy of tones in language learners' speech and to correct the mistake automatically?
- To what extent can this application contribute to the teaching effectiveness in Chinese language training in teacher preparation?

2. PEDAGOGICAL PROBLEMS

The first step of this research is to identify the pedagogical problems of learning Asian languages. The key to the competency of spoken Mandarin Chinese is a good command of "Pinyin", a Romanization phonetic symbols including tones. The research was initiated through an investigation on problems of learning tones from 10 students of BEd (Languages/Asian Studies).

A range of methods have been employed for the investigation. Firstly, students' pronunciation of individual sounds by reading a designed passage covering the whole phonetic system was recorded and problems of tones was identified. Secondly, students' interactive conversations with investigators were recorded and problems of communication due to mispronunciation of tones were traced. Thirdly, students were interviewed to explicitly tell of their frustration of learning tones and their difficulties in commanding the Pinyin system. Their experiences were analysed.

The problems of mispronunciation and the main causes for those problems of learning tones have been identified as following:

(1) Mispronunciation occurs frequently in combination of some specific consonants and vowels. It is discovered that consonants could be divided into two sub-groups. Group A includes b, p, m, f, d, t, n, l, r, y, w. Group B includes j, q, x, z, c, s, zh, ch, sh. When consonants in Group B combine with any vowels, mispronunciation on tones occurs more than those in Group A if they combine with any vowels. This is probably due to the fact that there are no equivalents or similar sounds in English. The same reason could be used to explain why some diphthongs such as ia, ie iu, ua, ui, uo, etc, tend to be more easily mispronounced that others, for instance, ai, ao, etc.

(2) A number of last syllables in a sentence are often mispronounced. English speaking learners of Chinese have a habitual tendency of applying falling pitch to a statement and a rising pitch to a question. The change of tone at the end of each sentence affects the accuracy of tones of last syllabus which could

change the meaning. English speaking learners are normally not aware of this problem since the tone of each syllable is not a distinct feature in their mother tongue.

(3) The sound duration of each Chinese syllable pronounced by English speaking learners is apparently shorter that produced by Chinese native speakers. This is because English words are not only monosyllabic but also disyllabic and polysyllabic. Since Chinese characters follow the strict principle of " one character, one syllable", in the form of words which are most disyllabic, the average time duration for each syllable is much longer than that need in an English word. This habit of short duration by English speaking learners of Chinese hinders some of them to complete the full change of a pitch in a Chinese syllable. This is the major reason why they have difficulties to master the correct pronunciation of the third and fourth tones in particular.

(4) Compared with reading based on Pinyin, learners find more difficulties in pronunciation with tones when they read characters and communicate naturally. It is obvious that learners have to rely on memory since the characters do not explicitly indicate tones. For this reason, mechanic drills seem to play a significant role in the commanding of correct tones in spoken Chinese.

3. SYSTEM DESIGN

Based on the problems identified from an investigation on students' learning of spoken Chinese, a multimedia software is designed to apply speech recognition technology in order to assist students to overcome the learning difficulties. Consequently, the study of the acoustic mechanism of the tonal feature and mathematical algorithm for identification of tones become vitally important for the development of the intended software.

On the one hand, the study of the acoustic mechanism of the tone feature involves recording speech signals for a representative vocabulary (characters) with different tones and analyzing the common features of each tone. Methods used in the analysis are the linear predictive coding which is a time-domain method and spectral analysis which is a frequency domain method. Our work shows that tonal features are mainly reflected in the gradual variation of pitch frequencies although gradual volume changes are also noted. However, because the pitch frequency and volume vary from time to time, word to word, and speaker to speaker, great efforts are made to clearly identify the exact features of tones.

On the other hand, a thematical algorithm for identification of tones is also an arduous task. Once the main features of tones are found, a mathematical algorithm is established to identify the tones. The algorithm should not only be able to differentiate the tones, but also be able to provide a quantitative measure of the tone so that the result can be used to improve students pronunciation. To keep the algorithm fast and efficient, simple standard methods for detecting frequency and volume variations are used.

4. SOFTWARE DEVELOPMENT

In the process of developing the system software, it is discovered that

(1) The tonal information of the Chinese language is reflected in the change of the pitch frequency. For each word, it is found that such a change can be approximated by a second order (non-linear) time curve with sufficient accuracy. There are four tones in the Chinese language. The first tone gives a flat curve; the second one, an increasing curve; the third one, a parabolic curve which dips first and then shoots up; and the fourth one, a quickly decreasing curve.

(2) A method for estimating the pitch period of a Chinese sound is developed and tested. This is done in three steps: a) the speech signal is low-pass filtered to a sufficiently low signal bandwidth which maintains the pitch frequency; b) each word is divided into many sub words and their pitch frequencies are extracted; and c) the pitch frequencies of the sub words are best fitted with a second order time curve.

(3) Commonly used speech recognition algorithms are not adequate for tone recognition. This is because of two reasons: a) Most of speech recognition algorithms are developed for western languages and utilise a certain degree of linguistic features of the language which are not suitable for Chinese; b) Although some algorithms are shown to

be good even for Chinese speech recognition, they do not provide accurate information on the tonal variations of the speech which is essential for training of tones. Experiments have been carried out to confirm these points which justify the need for accurate computation of pitch frequency, as described in (1) and (2) above.

The initial investigation has paved the way to develop and implement an experimental system which can be used to recognise Chinese sounds and tones with sufficiently high accuracy. As a result, the standard algorithms for the English language is modified so that it can be used for Chinese. Through modifying the algorithm to incorporate the information of pitch frequency, tone recognition is achieved. Consequently, a numerically efficient algorithm is developed for pitch frequency computation before an experimental system is programed.

The hardware of the system consists of a high fidelity microphone, a 16-bit analog-to-digital (A/D) and digital-to-analog (D/A) board with the sampling rate no less than 22KHz and certain on-board digital signal processing functions, a 486 IBM (compatible) computer with 240MB hard disk, 16M RAM, 50MHz clock rate and a math coprocessor, and a simple audio amplifying equipment. The software consists of man-machine interface, recording of speech signals, preprocessing of signals, speech recognition module, and audio replay module.

The preprocessing of speech signals involves noise filtering, time-domain data compression and/or fast Fourier transform. The speech recognition module, which is the core of the software analyzes the preprocessed signals and computes a quantitative measure of the tone. This result is then visually displayed on the computer screen. The basic functions of the audio replay module are storing speech signals (teacher's pronunciations) and replay them (for students as models). Lab tests are carried out to check the quality of the experimental system. The results collected in the tests shall be used to upgrade the system.

5. SOFTWARE DEMONSTRATION

The modified system can be used on a 486 DX2-66 IBM compatible computer, with 16 megabytes of RAM and a 430 megabyte Hard Disk, equipped with a Sound Blaster card (for audio input and output) and DSP processor (for fast computation). The experimental system consists of (1) lesson planning, (2) designed training and (3) visualised learning. The system provides the teacher trainees flexibility to write their own program to meet the needs of students.

When the application is opened, directly from Windows 3.1, the initial screen appears with Lesson, Training, Learning and Help on the menu bar. To start a new lesson or to choose one from a number of lessons which have been planned before hand, the menu bar of Lesson should be clicked as shown in Fig. 1

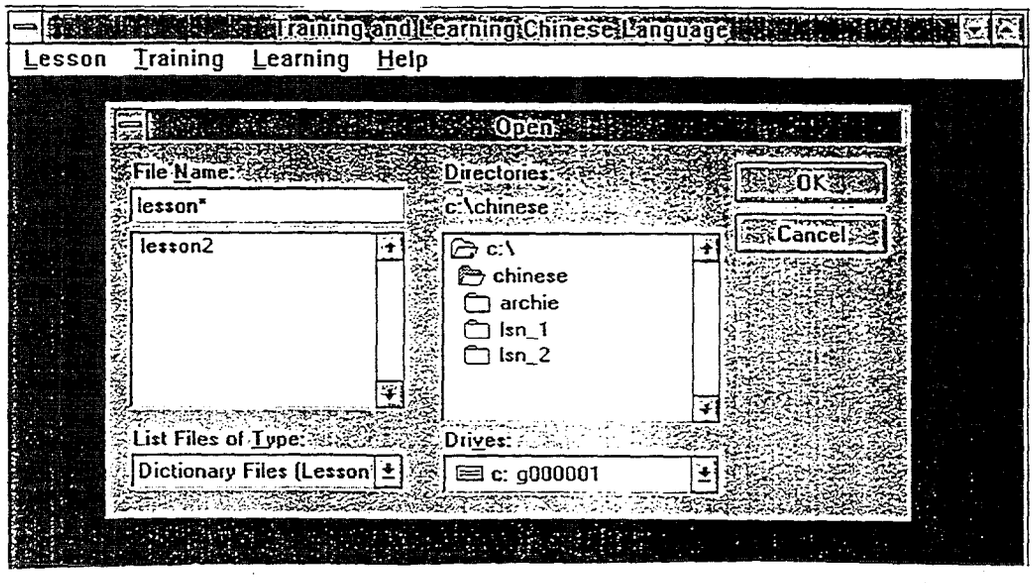

Fig. 1 The Screen of Chinese Software and Dialog Box

In the dialog box marked "File Name" there will be a number of lessons displayed, as "Lesson 1", "Lesson 2", etc. Double-click on the desired lesson, and the lesson will be opened. Suppose we are going to edit "Lesson 1" for an individual student or a group of students.We type "Lesson 1" in the box marked by "lesson*" and click "OK" before selecting "Begin" from the Training menu. The full range of words for that lesson is displayed on the screen (Fig. 2).Words can be removed or added by clicking "Delete" or "Add".

To remove a word, simply select it and click "Delete". To add a word, write the Pinyin in the space marked "Word" above the word selection panel, and click "Add". A new dialogue box appears as in Fig. 3. Click the "Record" button with the mouse and then clearly pronounce the desired sound. The duration of the pronunciation has a limit of two seconds. If the pronunciation is not successful, then click "Cancel" and redo it. In order to check whether the pronunciation is satisfactory, the recorded voice can be heard by clicking "Play". A satisfactory sound will be installed when the "OK" key is used.

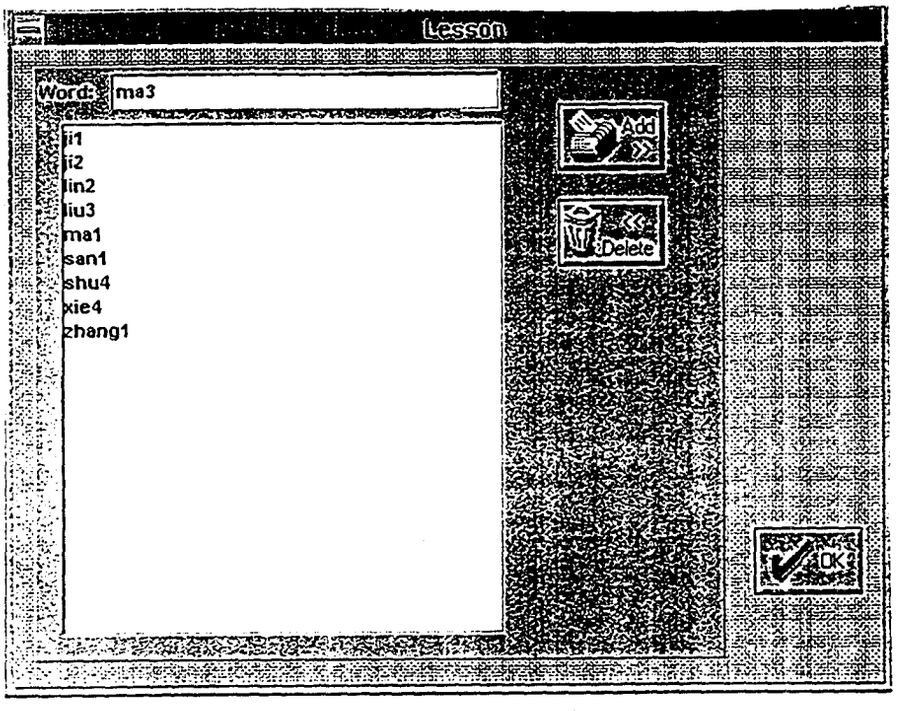

Fig. 2 Lesson Planning Screen

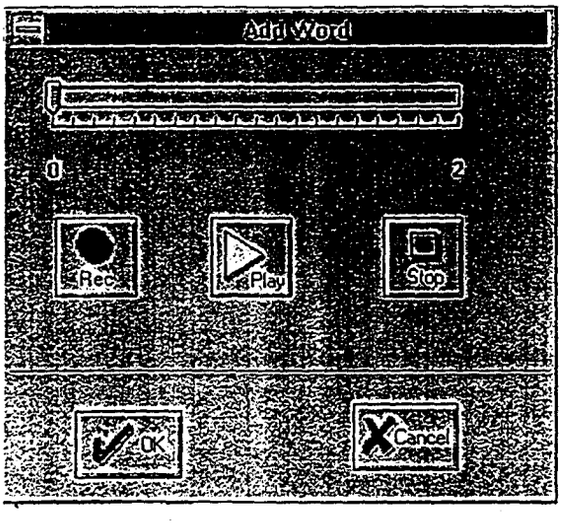

Fig. 3 Word-Adding and Recording

When the students are using this software, they select "Go" from the "Learning" menu after choosing a lesson. A new screen is shown. The selection of words in that lesson is displayed on the left-hand side of the screen (the rest of the screen temporarily blank), and they are then instructed to select a word. First they click "Play" to hear the teacher's voice for the word, and speech recognition analysis of the sound is displayed on the right-hand side of the screen (Fig. 4).

Then the students follow the example of the teacher's pronunciation. If they click "Record", their voice will be automatically recorded and the image of speech recognition analysis for the students' pronunciation is then shown in the middle of the screen. The spectrogram and pitch frequency of the student's and teacher's voice can be then compared. If they are similar (as in Fig. 5) then the student has pronounced the sound in the correct manner. If they differ significantly (as in Fig. 6), however, then the student has pronounced the sound wrongly.

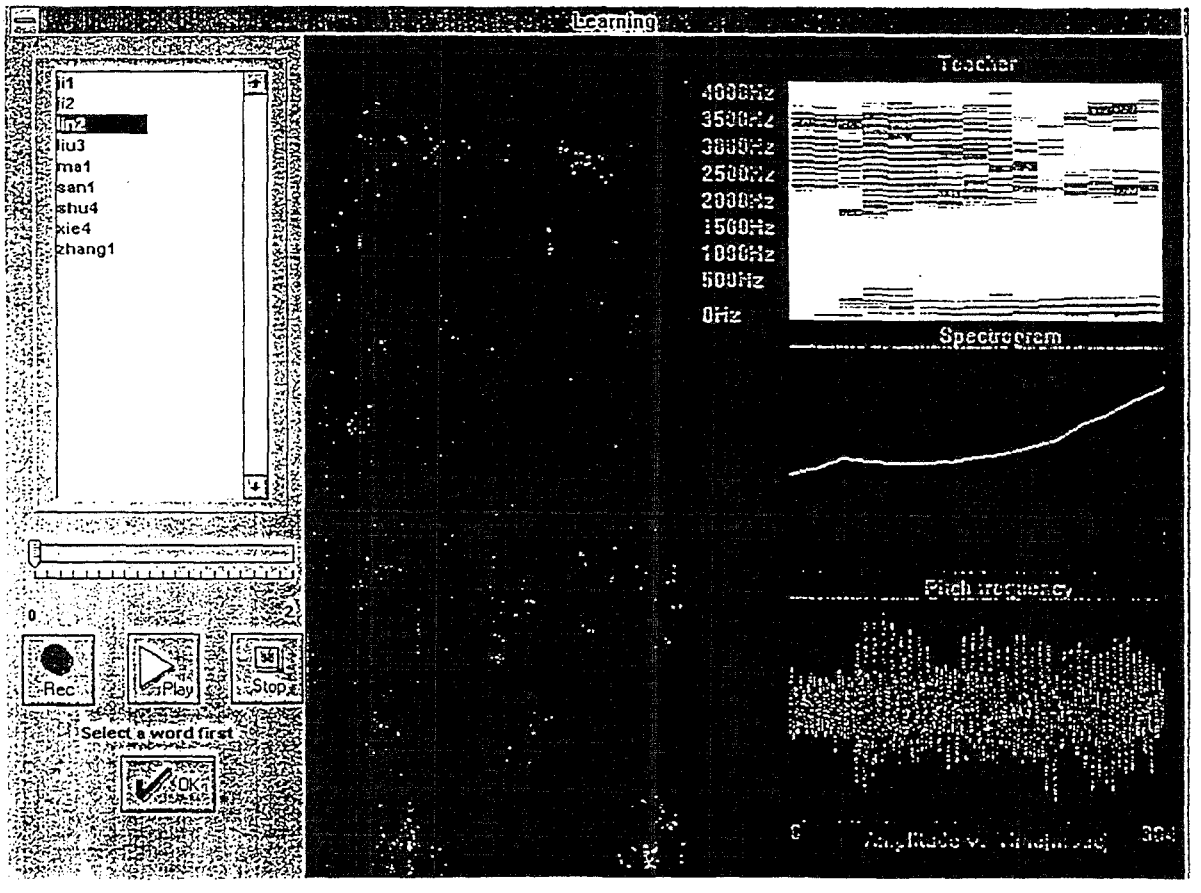

Fig.4 Demonstration of the Teacher's Pronunciation

Three separate pictures illustrate the spectrogram, pitch frequency and amplitude vs time. In particular, the curve of pitch frequency reflects the distinctive features of the four tones. Taking "Lin2" as an example, the number of 2 indicates the sound "Lin" is pronounced with the second tone, namely, the rising tone. The curve in the picture clearly demonstrates such a tendency.

In Fig. 5, since the student uses the same tone as the teacher, the curve shown in the pitch frequency is close to the teacher's example. In Fig. 6, we can see the curve derived from the student' pronunciation is totally different from teachers. The tendency of the curve shows that the students mispronounced the sound by using falling tone, in other words, the fourth tone instead of the second tone.

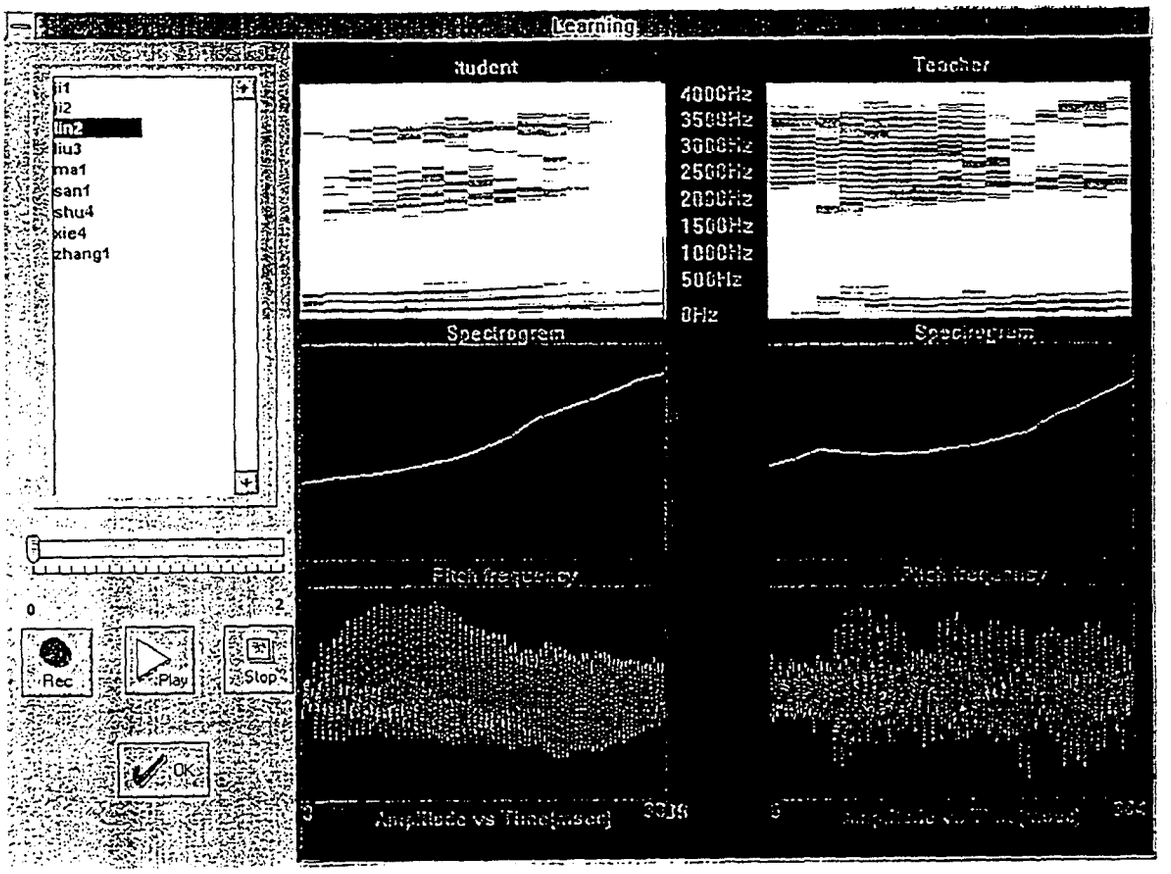

Fig. 5 Comparison of the Pronunciation: Correct

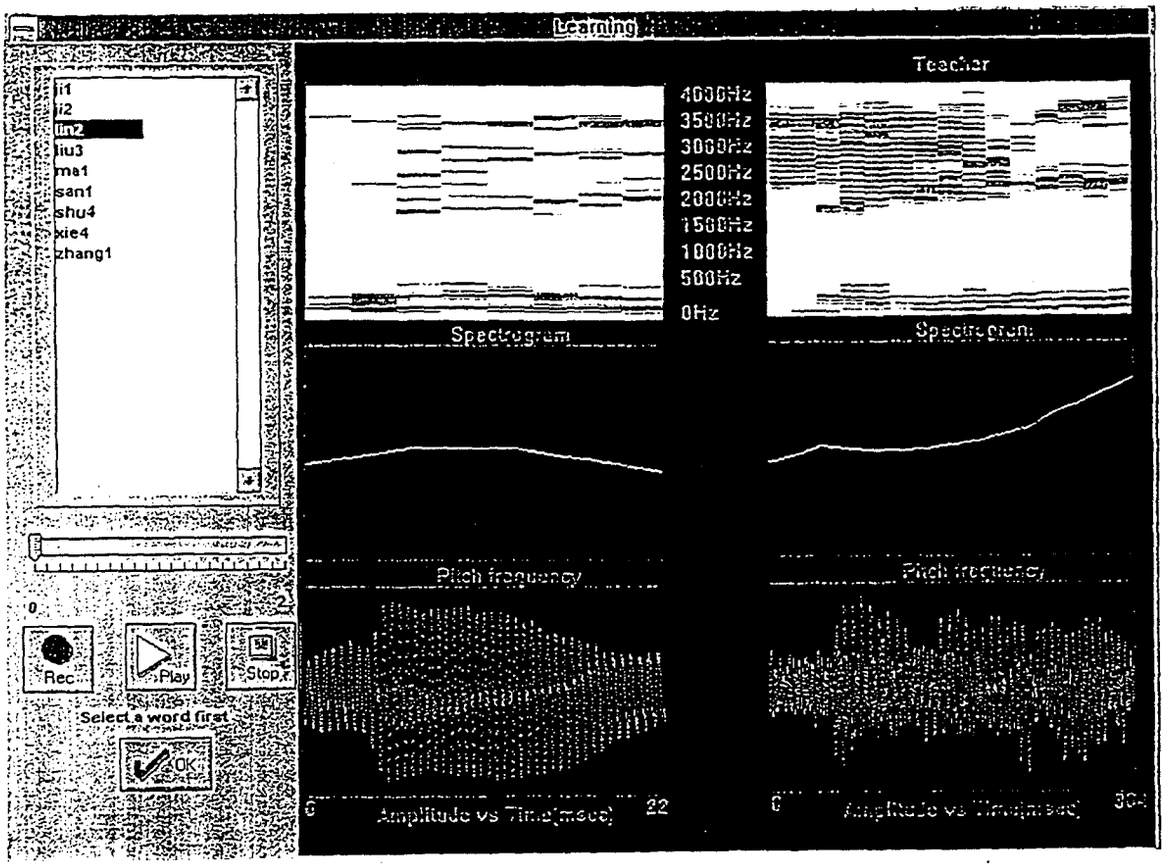

Fig. 6 Comparison of Pronunciation: Incorrect

Two advantages of the software can be found. First, the teacher's correct visualized tones are illustrated in the screen. Second, by comparing the curve, students will know whether their pronunciation is correct or not. If not, what caused the mispronunciation is also given.

6. FIELD EXPERIMENT

In order to assess the educational value of this teaching software, a field experiment is conducted in both university and school settings. Two teaching programs to teach Pinyin Romanization (the Phonetic system of Mandarin Chinese) are designed. One is for the teacher trainee to learn the four tones. The other is used by the teacher trainee to teach school pupils. Each teaching plan consists of a number of lessons focusing on the training of four tones. Standard sounds are recorded into the the educational system for comparison.

In the case of the study at the university, all the undergraduate students from BEd (languages) course are invited to use the software to learn Pinyin tones. The feedback was generally positive. The statistic data shows that 92% students regarded the software as a valid and effective instrument to learn the four tones. In particular, the software provides visual images of tones pronounced by the teacher and learner. In comparison with the standard curve, learners are able to identify what is wrong in pronunciation. Normally it is difficult for an English-background beginner of Chinese to have a clear understanding on the notion and difference of four tones. The software overcomes such a difficulty. In this sense, a classroom demonstration of the software will be most desirable and valuable to the English-background learners at the beginner's level.

It is also noticeable that most of the students (86%) agreed that the software is more useful for after-class practice rather than for classroom practice. A practice on a computer can be time consuming. Since the comparison of the curves is stimulating, students could spend hours on it without realizing the passing time. Sometimes it took longer time than expected to fulfil a planned learning task. Psychologically, students also felt much less stressed when their mistakes were picked up by a computer rather than a lecturer. Therefore, it would be ideal to make the software available for them to use at home.

In addition, a number of students (27%) thought that the software, especially the time control in the recording part was not easy to manage. The unexpected curves were not necessarily all caused by mispronunciations but by technic problem due to unfamiliar usage. A pre-training is needed before a learner starts to use this program confidently.

The software was also tried at Merewether High School in the Hunter Region, NSW. At this stage, the experiment is in progress. The initial response from both students and teachers was encouraging. In teachers' opinion, the software plays a new role of teaching four tones in the classroom although teachers' role is also indispensable. The software has a special function to detect learners' mistake automatically through the computer screen. However, they questioned the validity of self-learning by junior high school students because of the complexity of using the program. This point coincides with the comments made by some university students.

From the students' point of view, the program is interesting and stimulating. It tells the problems of pronunciation clearly. In the case of following a tape-recorder to practice, students can only be guided by a correct sound, but they cannot figure out whether the practice is correct not unless a teacher is present. By using the computer program, the students receive an instant feedback indicating the problems as well as some "reasons" behind them.The program does have a great potential for learner-centred language learning. On the other hand, the experiment has confirmed that supervision on self-learning is necessary. Computer software cannot substitute the role of a teacher in foreign language learning.

7. CONCLUSION

The tones and stress patterns in Asian language are often regarded as difficult aspects in the process of teaching and learning by English speakers. Tedious drills and constant corrections discourage learners' efforts on fluency and accuracy hence language teachers are facing a pedagogy problem. One way to arouse the learners' interest and to overcome the difficulties is to

design multiple links of sounds and images to allow for learner-centred discovery learning. Computer assisted language learning is recommended as an effective approach.

In order to develop a teaching software for learning Chinese tones, speech recognition techniques has been employed. The application of speech recognition techniques has made it possible to "visualize" the tones pronounced by teachers and learners through a multi-media IBM computer through which a comparison can be made to detect the problems of the tones. The system developed for teaching, although not be tested widely for various languages, has successfully illustrated its potential in an Asian language teacher training course.

Acknowledgement:

This research is supported by the Australian Research Council and the Research Management Committee at the University of University of Newcastle.

Reference:

[1] COAG *Asian Languages and Australia's Economic Future. A Report Prepared for the Council of Australia's Governments on a Proposed National Asian Languages/Studies strategy for Australian Schools.* Canberra: National Languages and Literacy Institute of Australia, 1994.

[2] D. Smith, B. Ng, K. Louie and C. Mackerras *Unlocking Australia's Language Potential--Vol.2 Chinese.* Canberra, NLLIA and DUET.

[3] D. Hardis and S. Windeat *CALL (Computer Assisted Language Learning)* Oxford University Press, 1989.

[4] J. Barker and R. N. Tucker (ed) *The Interactive Learning Revolution: Multimedia in Education and Training,* Kogan Page, London,1990.

[5] C.C. Cheng "Proactive Guidance in Computer-Assisted Language Learning." *Chinese Teaching in the World,* Vol. 3, 1995.

[6] C.C. Chen "Teaching Informational Signs Using Sound and Video Image." *Selected Papers from the Third International Symposium on Teaching of Chinese.* pp836-841 Beijing Language Institute Publishing House, 1991

[7] M. M. Yan *Active Chinese: a Multimedia Text.* Paper presented at the Regional Workshop for College-Level Chinese and Japanese Teachers: Integrating Technologies into the Classroom. Indiana University, 1993.

[8] P. Y. Lee *Chinese Character Tutor,* University of N.S.W. 1992.

[9] M. Sawer "Chinese Computer Assisted Language Learning at School and University" Conference paper presented at CSAA Conference 1993.

[10] N. Shaw "Flexible Delivery of the Chinese Course at Victoria University". Conference paper presented at CSAA Conference 1993.

[11] B.S. Atal "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," *J. Acoustic Soc. Am.*, 47, 65(A), 1970.

[12] S.E. Levinson and D.B. Roe "A perspective on Speech Recognition," *IEEE Communications Magazine*, vol. 28, no. 1, Jan., 1990.

[13] K.L. Lang and A.H. Waibel "A Time-delay Neural Network Architecture for Isolated Word Recognition," Neural networks, vol. 3, 1990.

[14] C.W. Loke "Speech Recognition System," Final Year Project Report, Dept. Electrical and Computer Engineering, The University of Newcastle, 1992. Supervised by Dr. M. Fu.

Dr. Shen Chen is a senior lecturer in the Faculty of Education at the University of Newcastle.

Dr. Minyue Fu is an Associate Professor in the Department of Electric and Computer Engineering at the University of Newcastle.