



Bridging pedagogy and technology: User evaluation of pronunciation oriented CALL software

Tsai Pi-Hua

China University of Technology, Taiwan

While it has become common to employ pronunciation oriented software to improve one's pronunciation in L2, both language teachers and L2 learners feel uncertain about choosing software to meet their purposes. Taking *MyET*, pronunciation oriented software written and highly praised in Taiwan, as a representative program, this study investigated its pedagogical usefulness through the viewpoints of nine junior college students (with three levels of English proficiency). The evaluation showed that *MyET* was able to differentiate between students at the beginning and intermediate levels, though its design for providing input and practice exercises has room for improvement. A questionnaire completed by the nine students indicated that they liked best the program's segment analysis and function of replaying target segments. Students enjoyed practising at their own pace and receiving individualised, immediate feedback from *MyET*, but considered the practice to be "mechanical." They expressed needs for more instruction on how to refine their pronunciation and for cumulative analyses of performances. This study contributes towards the design principles for pronunciation oriented software that can address users' language learning and practice needs. Implications for teaching pronunciation and selecting pronunciation oriented courseware are discussed.

Introduction

Pronunciation is an integral part of the communication process (Butler-Pascoe & Wiburg, 2003). Poor pronunciation (i.e. of phonetics and prosody) can distract the listener and make comprehension of the message difficult (Celce-Murcia & Goodwin, 1991), and this may result in negative social evaluation and discrimination (Lippi-Green, 1997; Munro, 2003). Learners are clearly aware that poor pronunciation represents a considerable barrier to their success in English and they give extremely high priority to mastering of pronunciation of the target language (Fraser, 1999; Nunan, 1988; Willing, 1988). Teachers need to know the best techniques for teaching pronunciation. Some teachers have complained that they do not know how to teach it though they try (Morley, 1994; Macdonald, 2002). According to Breikreutz, Derwing and Rossiter (2002), this arises because many teachers

have not received training in pronunciation instruction. Moreover, as many international language proficiency tests, such as the new TOEFL iBT (TOEFL, 1999) are beginning to include evaluation of examinees' oral ability, students seeking higher education have increased needs for instructional materials that can provide speaking practice.

Advances in technology have enabled automatic speech processing to be integrated into foreign language pronunciation training. The advantages of computer assisted pronunciation training (CAPT) software for improving English learners' pronunciation have been studied extensively (Molholt, 1988; 1990; Harless, Zier & Duncan, 1999; Holland, Kaplan & Sabol 1999; Kaplan, Sabol, Wisher & Seidel, 1998; LaRocca, Morgan & Bellinger, 1999; Eskenazi, 1999a, 1999b; Neri, Strik & Boves 2002; Butler-Pascoe & Wiburg, 2003; Kim, 2006). The untiring, non-judgmental nature of the computer allows students unlimited opportunities to review any part of the materials and receive additional assistance provided by the system. CAPT software enables students to study autonomously, choosing what function to use and how often they use it. On the other hand, teachers also benefit from employing CAPT software in their pronunciation classes as it can give students drilling practice, which teachers consider tedious and time consuming. Last but not least, CAPT systems offer an interactive learning environment in a range of modes: whole class, small group or pair, and teacher to student (Pennington, 1999).

CAPT software is not without its limitations. Most of the CAPT software has been criticised for being designed without a basis in pedagogical theory. Pennington (1999), for example, stated that most CAPT software laid overwhelming emphasis on the decontextualised mechanics of articulation. Breitzkreutz et al. (2002) commented that the most popular pronunciation software programs in Canadian classrooms focused exclusively on segments rather than prosody. The design of much of the CALL software has also been found to focus on the impressive multimedia capabilities of computers and to lack content that is linguistically and pedagogically sound (Chun, 1998; Derwing & Munro, 2005; Neri et al., 2002; Reeser, 2001). Chun (1998) and Neri et al. (2002), for instance, noted that though they look flashy to buyers, the graphical wave forms presented in software do not give meaningful feedback to users. Due to the above limitations, it has been suggested that more conclusive empirical evidence needs to be obtained for the pedagogical benefits of using computers in language classrooms (Chappelle, 1997; Dunkel, 1991; Salaberry, 1996).

The present study used *MyET* (LLabs, 2005), pronunciation oriented software designed by local engineers in Taiwan, as a representative program, to investigate the pedagogical usefulness of computer software for teaching English pronunciation. *MyET* received the "Digital Products

with Best Innovating Software” award in 2003 from the Industrial Development Bureau of Taiwan and the “Best Digital Publication” award in 2004 from the Government Information Office of Taiwan. Chen’s (2004) study on college students who used *MyET* found significant positive correlations between machine scorings and human graders. He suggested that subjects with different levels of language proficiency should be invited to further test the scoring validity of *MyET*.

To evaluate the feasibility of incorporating a CAPT system into pedagogy, one should investigate the learning environment the system provides, in addition to scoring validity. Some problems with its application can result from the human interface and input mode (microphones), rather than the speech recognition component *per se* (Ehsani & Knodt, 1998). The learner’s ability to interpret displays can also be a factor influencing the practicability of incorporating a CAPT system into pronunciation teaching (Chun, 1998). Furthermore, interaction between learners and technology should also be included in evaluations of computer assisted learning (Egbert, 2004; Pennington & Esling, 1996). Following a similar line, the analysis here will focus mainly upon user perspectives of the *MyET* environment.

The questions to be discussed in this paper are:

1. Is *MyET* able to differentiate between learners with different levels of English pronunciation proficiency?
2. What do learners feel about the usefulness of *MyET*’s pronunciation analysis and the functions the software offers? What type of pronunciation analysis do users see as the most informative?
3. Do learners feel that *MyET* is effective in improving pronunciation?
4. What implications does this present study have for English pronunciation teaching and pedagogical design of pronunciation oriented software?

The discussion in this paper is based on some theories of English language teaching and computer assisted language learning (CALL), presented in the next section. Following this review, a user evaluation of *MyET* will be presented. This study may help to improve the design of CAPT systems that can address learners’ language learning needs during practice sessions. It is also hoped that the analysis given here can help students or teachers understand which features need to be taken into consideration when choosing a pronunciation oriented software program to suit their needs.

Pronunciation teaching

Design of CAPT software needs to be based on contemporary pedagogy and the findings of research into second language acquisition (Pennington, 1999; Neri et al. 2002). Pronunciation teaching in second/foreign language

education has been found to emphasise prolonged and focused practice of a large number of linguistic items, such as individual vowel or consonant phonemes. In the late 1970s, the core of pronunciation teaching began to focus on learners' acquisition of English intonation, rhythm, connected speech, and voice quality setting (Celce-Murcia, Brinton & Goodwin, 2004). In addition to linguistic acquisition, the acquisition of communicative competence (e.g. the appropriate use of English in social contexts) was also emphasised at that time (Morley, 1994). It was believed that pronunciation learning and teaching needed to be placed in communicative contexts (Fraser, 1999; Wennerstrom, 1999). These changes meant that perfect or near native pronunciation was no longer the only goal of contemporary pronunciation instruction. The aim, instead, was to improve learners' intelligibility rather than to achieve total accuracy (Celce-Murcia, Brinton & Goodwin, 2004; Derwing & Munro, 2005; Jenkins, 2002; Morley, 1994).

The roles of learners and teachers are viewed differently in current pronunciation instruction. Learners are expected to be positively involved in their learning and to develop skills and strategies for monitoring their own speech production (Eskenazi, 1999b; Morley, 1991, 1994; Rypa & Price, 1999). Learners have their own learning styles, for example visual, auditory, kinesthetic, or tactile (Celce-Murcia, Brinton & Goodwin, 2004; Oxford & Anderson, 1995), so it has been suggested to teachers that learners should be presented with engaging input that accommodates these different styles. For example, phonetic input should be presented in both written and audiovisual forms so as to stimulate learners' interest.

Besides their roles as error correctors, teachers are expected to act also as facilitators, who offer various models, provide opportunities for practice, suggest specific techniques, and give encouragement and advice to the learner (Egbert, 2004). According to Dickerson (1994), providing various models is not enough to empower students in the area of pronunciation. He posited that when preparing a curriculum for pronunciation, teachers must consider three skills: prediction, perception, and production. Research has found a correlation between perception and production skills (Akahane-Yamada et al, 1996; Rochet, 1995), and ESL specialists have postulated that perception training and production training should go hand in hand in pronunciation instruction (Celce-Murcia, Brinton & Goodwin, 2004; Dickerson, 1994; Jones, 1997; Strevens, 1974). Dickerson (1994) believes that only after students are taught prediction skills are they able to pronounce the words they encounter. For example, the teacher can teach students how to apply rules, such as the following vowel prediction rules, to standard orthography to predict the pronunciation of words they have never seen before: a stressed VC+e predicts a long vowel; a stressed VC# predicts a short vowel; an unstressed VC+e predicts a reduced vowel (Morley, 1994, p.21). (V represents vowel, C consonant, and # the end of word position.)

Other techniques also can be used by teachers to stimulate their students' self improvement and help them become self instructors and lifelong learners. Pronunciation instruction should raise students' awareness of their own production, for example, through the use of student produced recordings (cf. Walker, 2005). Teachers should enable students to anticipate problems and errors before they actually occur (Kenworthy, 1987). Flege (1995) found that many L2 learners' errors could be attributed to unconscious interference from L1 phonological representations, so a contrastive analysis of the sound system of L1 and L2 may help give learners pertinent articulatory hints and help to avoid anticipated errors.

Theoretical framework of CALL

The CALL theoretical framework adopted in this study was based on the general guidelines Neri et al. (2002) set for developing pedagogical CAPT systems and on the principles that Pennington (1999) proposed for improving computer assisted pronunciation pedagogy. The conditions of constructing a superior CALL learning environment, as postulated by Egbert, Chao and Hanson-Smith (1999), will be used also to evaluate the design of the activities that *MyET* provides, though the conditions do not refer specifically to the design of an optimal environment for computer assisted pronunciation.

Neri et al. (2002) claimed that if the three most crucial factors that influence the acquisition of L2 pronunciation, input, output and feedback, are controlled well, then better pronunciation learning results can be obtained. *Input* refers to learners' amount of exposure to L2, which includes varied and meaningful materials and accommodates learners' needs and learning styles (Neri et al., 2002). The variety of language encountered must be sufficient for learners to continue to learn and improve (Egbert, et al., 1999), and one or more reference accents should be established (Pennington, 1999). When learning pronunciation with a CAPT program, some learners may set goals for intelligibility and others for accuracy. Thus, in order to help learners know if they have reached their goals for performance, Pennington (1999) suggested that developers of CAPT software decide clearly in the very beginning what performance is counted as having made progress towards or achieved a desired target.

Great exposure alone can not guarantee success in language learning. According to Neri et al. (2002), in second language acquisition it is necessary for learners to practise speaking the target language (i.e., *output*) to test their own hypotheses about L2 sounds. This process is also considered conducive to the development of self awareness in learners. As for practice activities, these cannot be decontextualised (Neri et al., 2002; Pennington, 1999). In personalised and real life contexts, tasks that learners

are asked to complete need to be authentic, and learners should be able to interact with an authentic audience (Egbert et al., 1999). To decrease learner anxiety during practice, it has been suggested that a CAPT program start from the easier stages and advance to more challenging ones (Pennington, 1999). An ideal CALL environment should give learners learning autonomy and enough time to finish tasks (Egbert et al., 1999).

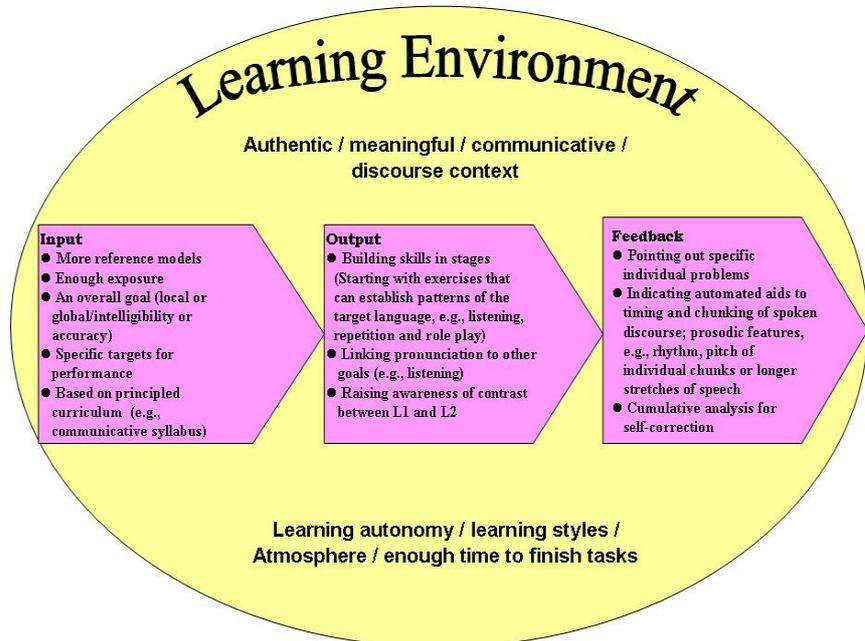


Figure 1: Basic features of an interactive pronunciation oriented software program

As for *feedback*, Neri et al. (2002) pointed out only certain pronunciation errors that may affect learners' intelligibility have to be highlighted. They stressed that feedback should focus on specific individual problems about segmental and suprasegmental errors, so that it can stimulate learners to attempt self improvement. Spectrograph and pronunciation models are employed in many automatic speech recognition (ASR) products to give learners feedback on their production. Pennington (1999) added that feedback should include automated aids for the timing and chunking of longer stretches of speech, such as displays of discourse intonation and comments like "too slow," "no linking," and "too many pauses." In addition to the presentation of learner's errors, some information should be provided for learners to improve their production, thereby enabling them to continuously reflect on their own learning (Egbert et al., 1999).

Pennington (1999) proposed that information, such as cumulative analysis and records of speech, be offered to help raise learners' awareness of the contrast between L2 and L1, and to develop their meta-analysis skills for self correction in the ensuing work. Based on the theoretical frameworks proposed by Neri et al. (2002), Pennington (1999) and Egbert et al. (1999), Figure 1 summarises the basic features of an interactive pronunciation oriented software program.

An introduction to *MyET*

MyET, a CAPT software program that uses ASAS (automatic speech analysis system), can identify words that are read aloud or spoken into any sound recording device. It displays the spectrum and contour of the user's utterance, and provides a scoring mechanism and key information that helps users to improve their pronunciation. Learners listen to utterances spoken by speakers from different parts of the world, who read from various sources: everyday conversations, English for specific purposes (e.g. business English and English for news), and excerpts from dialogues in movies. Then, learners record their own utterances. *MyET* claims it can explicitly pinpoint learners' pronunciation errors by giving one on one feedback that compares the learner's pronunciation with a model pronunciation. Figure 2 shows the interface of the pronunciation analysis that *MyET* provides.



Figure 2: Interface of pronunciation analysis for learners
 Figures 2-5 are Copyright *MyET*

On the right side of the interface is a scoring display of the learner's performance, including a display of overall score as well as individual scores for pronunciation (i.e. of segments), pitch, timing, and emphasis. Learners' spectrograms are displayed at the bottom of the interface below the model ones for visual inspection and comparison. Pronunciation errors are color coded to show the areas of the user's difficulty (see Figure 3).

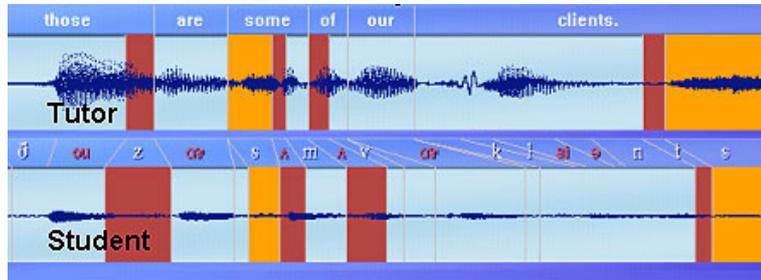


Figure 3: Display of pronunciation errors

Those learners who have difficulty pronouncing a particular English syllable or segment can just click on certain regions of the waveform of the model utterance or on the phonetic symbols (colored in orange) in the table of pronunciation analysis (see Figure 4) for model pronunciation of the target syllable or segment.

Pronunciation				
Expression	Comparison	Phoneme	Score	Remarks
Yes,	 	j	100	
		e	100	
		s	100	
those	 	ð	94	
		ou	100	
		z 	60	You can improve
are	 	ɑ:	100	
some	 	s 	68	You can improve
		A 	60	You can improve
		m	100	
Overall Pronunciation Score			87	

Figure 4: Pronunciation analysis of learners' production

Learners can also watch a 3D phonetic animation (see Figure 5) (i.e. the sagittal cross section of the vocal tract and the frontal lip view) illustrating how the target segment is pronounced. Underneath the animations, pronunciation tips for the target sound are offered.

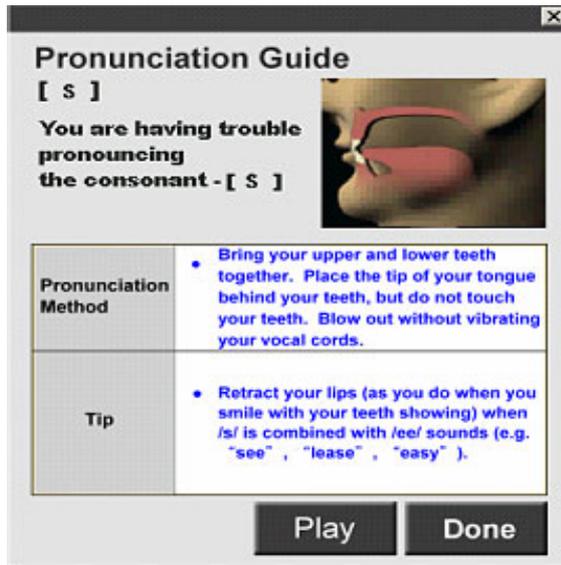


Figure 5: A 3-D animation with pronunciation tips

In the section titled "Test Yourself", a learning profile is created with a chart to show the learner's progress. In addition to autonomous oral practice, a "community website" is provided to enable learners to compare their scores with one another and to exchange ideas and opinions about English learning. Through the online community website, teachers or community leaders can also assign tests to their students or members and observe their performance and progress. More details about the features of *MyET* are available at <http://www.myet.com/en/Index.htm>

Methodology

Subjects

In the present study, the author recruited nine students (four girls and five boys) with Chinese as their L1, who were studying in a university of technology in Taiwan. The author had been teaching those nine students for three years before they were recruited to the experiment of this study. Considering the small size of the sample, the results of this study may not constitute a conclusive evaluation of the pedagogical effectiveness of *MyET*.

The nine students were divided into three groups with proficiency levels ranging from beginning to intermediate and advanced. This categorisation was based on the students' scores in English speaking proficiency for the past two years before the experiment and the interview the experimenter held, prior to the empirical experiment, to confirm the placement of their speaking proficiencies according to *ACTFL Proficiency Guidelines* (ACTFL, 1999). The beginning level students in this study were at the Novice-Mid level as defined in the *Guidelines*. They had difficulty producing even the simplest utterances and could hardly be understood. Those students categorised as intermediate had Intermediate-Low level speaking proficiency as defined in the *Guidelines*. They were able to handle a limited number of interactive, task oriented and social situations, such as asking directions and making purchases and with repetition. Although sympathetic interlocutors could understand their speech, linguistic inaccuracy was found. The advanced students defined in this study, whose speaking proficiency level approximated to that of Intermediate-High defined by the *Guidelines*, could produce connected discourse such as simple narration or description, converse with a number of strategies appropriate to a range of topics (though errors still existed in their speech) and generally, they could be understood by interlocutors.

Materials

The *MyET* software was the primary material used in this empirical study. A sign up sheet was designed for each student to make appointments for practice in a language laboratory at the school. A practice sheet (shown in Appendix A) was also provided for the purpose of keeping track of the students' progress and their reflections on using the system. Finally, a Chinese questionnaire (see the English translation in Appendix B) was used to elicit the students' perceptions of the pedagogical usefulness of the *MyET* interface (see Figures 2-5). Question 1 included five items, each with its own subcategories, pertaining to the usefulness of the recording function and *MyET's* pronunciation analysis, which includes segmental analysis and suprasegmental analysis (including pitch, timing, and emphasis). Question 2 asked the students if they were able to interpret the speech contours displayed by *MyET*. The second part of Question 2 inquired whether the visual phonetic displays could increase their awareness of pronunciation problems after the experimenter explained the spectrograms and waveforms that *MyET* produced. In Question 3, the subjects were asked if they thought the feedback *MyET* provided could boost their awareness of their pronunciation difficulties. Question 4 asked if the students considered *MyET* to be a good learning tool for improving and polishing their pronunciation. Finally, the students' overall comments on *MyET* were elicited in Question 5.

Experiment design and procedures

The methods employed in the present study for data collection included semi-structured and one on one interviews by the experimenter, and a survey using a questionnaire.

The experiment was conducted with the students one by one. In the beginning of the empirical experiment, each student was introduced to *MyET* and briefed on the purpose and procedure of the experiment. Each subject was told that he/she needed to sign up for practice sessions with *MyET* that would take place in a laboratory for 20 minutes three times a week for two weeks. Furthermore, the experimenter asked each student to keep track of one score as a representative score out of all the scores *MyET* gave for his/her utterances during each practice session and also to write down his/her reflections on *MyET*. After the introduction to *MyET* and the briefing on the experiment, the student started his/her first practice session with *MyET*. During each practice session with *MyET*, the experimenter distanced herself from the subject in the lab so that the student would not feel that he/she was monitored. At the end of the first meeting, each student was interviewed with semi-structured questions. First, he/she was encouraged to express his/her first impression of the design of *MyET*. Then, the subject was asked if he/she enjoyed practising with it and how he/she liked the design of *MyET*, such as the spectrum display. The student was told that both negative and positive comments on the system were welcome, and that his/her evaluation of *MyET* would not affect his/her grade for the school semester. The responses of each subject were taped and transcribed for subsequent analysis.

The experimenter did not provide any instructions on how to interpret the visual phonetic displays created by *MyET* until the beginning of each student's fourth practice session. These instructions were delayed to the fourth session in order to find out how much the students could understand from the displays without further instruction. At the end of the last practice session with *MyET*, each student filled out a questionnaire in Chinese on the overall design and effectiveness of the software.

Results and discussion

Questionnaire

Table 1 presents the means and standard deviations of all the items in Question 1: overall score, four independent phonetic analyses (segmental, pitch, timing and emphasis), and the recording, and (re)playing functions that the *MyET* program provides.

Table 1: Mean response scores on the effectiveness of items in Question 1

	Mean	SD
Overall score	3.67	.87
Segment analysis	4.22	.83
Phonetic symbol display and scores	3.89	1.27
Tips for pronunciation of segments	3.67	1.50
3 D animation of mouth	3.67	1.41
Contrast display of spectrogram with model utterance	3.22	.97
Comment	3.78	1.09
Pitch analysis	3.67	.87
Display of syllable and key	3.22	.97
Contrast display of intonation with model utterance	3.33	.87
Comment	3.11	1.17
Timing	3.78	.67
Display of syllable and speed	3.63	.74
Comment	3.63	.91
Emphasis	2.78	.97
Display of emphasis	3.00	.87
Comment	2.89	1.05
Recording function	4.67	.70
(Re)play function: click on certain regions of the waveform or the phonetic symbols for pronunciation of the target syllable or segment	4.33	.50

Note: 5 point scale (1 = not helpful, 5 = very helpful)

As shown in Table 1, among the four phonetic analyses *MyET* performs of each user's pronunciation performance, the segment analysis was found to be the most informative and helpful for pronunciation learning. The timing analysis was considered by the subjects to be the second most informative. As far as the students were concerned, the analysis of emphasis was not as helpful as the others. The results indicate that there is room for the improvement in the suprasegmental analysis *MyET* performs. Moreover, as one may find in Table 1, the subjects had a lower opinion of the visual phonetic displays. This echoes the comments made by Neri et al. (2002) on the graphical wave forms produced by CAPT software.

On the other hand, as shown in Table 1, the means for the "recording function" and "(re)play function" were the top two. This result is similar to that obtained by Walker (2005), who reported that the students enjoyed recording their own speech production because they could monitor their own pronunciation and measure their progress. This finding may inspire developers to embed more functions in their pronunciation oriented software program that will facilitate increased learning autonomy.

The answers to the first part of Q. 2 revealed that none of the subjects thought they could find out what was wrong with their pronunciation simply by looking at the spectrograms and waveforms displays *MyET* produced. However, in answer to the second part of Q. 2, five out of the

nine students reported that after receiving instruction from the experimenter, they could benefit from the visual phonetic displays and thereby better understand how to fine tune their pronunciation. In Q. 3, when asked if the feedback the software provided helped increase their awareness of the various aspects of their speech, all the subjects said "Yes." Moreover, the results of Q. 4 showed that if they were able to practise with *MyET* for a longer period of time, the students would consider *MyET* to be a helpful tool for improving their pronunciation in the future.

In their comments on *MyET* (Q. 5), most of the students used the word "cool" (in the sense of "good"), to express their feeling about the experience of learning English without being monitored. In their first trial, they were fascinated by the numerical score, spectrogram and waveform displays though they said they could not interpret them. The recording and replay functions also motivated them to keep practising and listening to the pronunciation of the models and their own. Above all, the students considered *MyET* to be a good tool for self study in that they benefited from the informative displays and tips on the pronunciation of segments.

While the students were amazed by the innovation *MyET* brought to their practice of pronunciation, they saw some room for improvement in *MyET*. With regard to *input*, they felt frustrated when they failed to match the speed of the model utterances, whose pitch and accent were hard for them to duplicate. They wished to have more choices of model utterances that matched their levels. Some students even suggested that non-native models of English utterances (e.g. English utterances produced by Chinese English teachers) be available for beginners to start with. According to some students, the paragraphs *MyET* provided for learners were either too short or too long (though some found the content interesting). Although Chinese translation was provided for each English paragraph, some students felt discouraged when they came across new vocabulary that they did not know. They indicated their need for a Chinese glossary of the difficult English vocabulary.

With regard to *output*, some students felt intimidated while interacting with the computer and felt that this type of practice and the computerised grading was too mechanical. Other students thought that more practice on certain utterances would not necessarily lead to higher scores. Others also indicated that sometimes the scores for their performance in vocabulary tasks were higher than those for their production of sentences. That is to say, they got higher scores for their vocabulary pronunciation than for their sentence pronunciation. Some of them attributed this discrepancy in their scores to the over sensibility of the microphone. As far as the *feedback* produced by *MyET* was concerned, some learners indicated that despite the visual displays produced by the program, they still did not know how

to fine tune their pronunciation. Besides the presentation of scores on production, some subjects wanted to see indications of their progress through each practice session so that they would know which specific kind of practice to focus on in subsequent sessions.

The scoring validity of *MyET*

As shown in Table 2, average scores for the students at the intermediate and advanced levels were higher than those for the students at the beginning level. This means *MyET* is able to distinguish between beginning and higher level learners. On the other hand, not much difference was found between the scores for intermediate and advanced learners. This result could be attributed to the fact that *MyET* focuses only on linguistic elements and not on overall communicative competence.

Table 2: Average scores of the performance of students at each level

	Advanced	Intermediate	Beginning
Student 1	87	86	88
Student 2	88	91	89
Student 3	90	89	72
Total	265	266	249

Suggestions for future improvement of CAPT software such as *MyET*

Based on the overall comments made by the subjects and the theoretical frameworks for this study (as summarised in Figure 1), some suggestions for the design of pronunciation oriented CALL software for English learners, such as *MyET*, will be made in the following section. Hopefully, these suggestions will be useful to teachers and students who consider using pronunciation oriented software programs to meet their teaching or learning needs.

Suggestions for the design of input

Considering that some students had difficulty producing utterances that matched the models spoken by teachers, *MyET* developers could provide model utterances at different speech rates and with various speaking styles for learners to practise according to their proficiency levels and learning styles. *Tell Me More* (Auralog, 2006) provides two programs levels, beginning and intermediate. Each one allows learners to alter the various elements of the program to match their individual levels closely. *Connected Speech* (Protea Textware, 2006) offers its users a choice between nine speakers with a range of accents and speaking styles. With such a design, lower level learners can hone their pronunciation and not become discouraged at the beginning. In addition, to help learners understand

which levels they belong to, it is also suggested that *MyET* provide diagnostic evaluation of learners' performance at the very beginning of each practice session. This may help learners decide on the pronunciation level they should start with.

To mitigate the mechanical feeling experienced by the subjects in this study, *MyET* developers could consider Purushotma's (2005) suggestion to incorporate materials from popular culture, such as voice navigated games or video games. For beginners, *MyET* could also offer a Chinese glossary of vocabulary to reduce their frustration with vocabulary they don't know.

Suggestions for the design of output

As for *output*, some subjects reported that though detailed analysis and some information about the pronunciation of English segments was provided by *MyET*, they still needed further instruction to improve their intonation and pronunciation of phonemes. This shows that it is desirable to provide perception and production exercises that users can do in stages, in order to first become aware of the differences between their native language and English, while also strengthening their knowledge of the sound patterns of English. For example, listening discrimination training could be provided to help learners understand the differences between the Chinese and English sound systems. This kind of practice may help learners avoid serious problems. Furthermore, the vowel prediction rules proposed by Dickerson (1994), which were introduced in the literature review in this paper, could be embedded into the program's exercises. Such exercises could help learners become familiar with learning strategies they can use to correct their own errors. Wei's review of pronunciation teaching strategies (Wei, 2006) is a great resource for CAPT developers who wish to design pedagogically correct exercises.

MyET developers may refer to *Connected Speech* (Protea Textware, 2006), *Pronunciation Power* (English Computerized Learning, 2006), and *Streaming Speech* (Speechinaction, 2006), which provide many perception and production exercises that help learners to become familiar with the sound patterns of English. *Streaming Speech*, for example, provides effective training in pitch and stress, and in the strategies people can use to achieve effective communication in real time (Levis, 2005).

As claimed by Greenspan and Lewis (2002), any language activity should be motivating enough to fuel learners' desire to develop receptive and expressive language. *MyET* developers could provide motivating exercises that help remove the mechanical feeling experienced by the students who tested *MyET* in this study. This is done by *Video Voice Speech Training System* (Micro Video Corporation, 2006), which provides a variety of games in

which learners manipulate their breath control, pitch, volume, and duration. For example, in a game named “Laser Blaster”, the learner destroys a target by vocalising at a specified volume and pitch.

Suggestions for the design of feedback

As for feedback, the learners in this study did not find *MyET*'s spectrogram and speech contour displays useful. To improve them, *MyET* developers may wish to refer to the notation system used in *Streaming Speech*. It incorporates notation that identifies stressed syllables (using uppercase letters or large circles) and is arranged in the shape of an intonation curve (for detailed description of the notation in *Streaming Speech*, see Lian, 2004). Such a notation scheme can draw the learner's attention to the prosodic features of a speech unit.

Some students reported that more practice did not necessarily result in higher scores, a result that may have been due to a speech recognition problem in *MyET*. To improve the program's speech recognition, *MyET* developers may refer to the suggestions made by Rypa and Price (1999). They recommended that software programs ask learners to repeat utterances that the programs don't recognise, rather than return them as errors, which can confuse learners. When the utterance of learners cannot be recognised, feedback could be provided in the form of a question such as “Pardon?” or “Can you repeat that?” so that learners will not feel discouraged and, above all, will learn how to make requests for clarification correctly.

To help learners understand how their utterances can be improved, *MyET* also needs to provide a feedback mechanism capable of directing the learner's attention to areas that need remedial practice (Ehsani & Knodt, 1998). Feedback like “*You have made wonderful performance on timing and more practice on segment pronunciation will be the goal of your next practice*” would be welcomed by learners because they would be told which specific areas needed more practice. Such feedback could also help learners feel that they are making good progress. Engwall (2006) considers such feedback to be the most important feature of a CAPT program because it increases learners' confidence in their pronunciation training. Furthermore, *MyET* should also keep track of each learner's performance during a practice session and, over a course of time, provide summary information about his or her performance. Such cumulative analyses can help learners to develop the meta-analysis skills that they need to perform self correction (Pennington & Esling, 1996).

To mitigate users' feeling that the computer grading mechanism is “mechanical,” the score displays could incorporate both visual and aural

media. As cited by Engwall (2006), the *European Ortho-Logo-Paedia* project uses visual maps, in which different target phonemes are placed and ASR is employed to show how close the user is to the different phonemes. When a student makes a serious error, the program shows a video clip of a teacher who uses pertinent gestures, eye contact, and a pleasant facial expression to give supportive feedback for correcting the error (Duncan, Bruno & Rice, 1995). Such a feedback display is necessary since it can help students learn a pragmatic skill, how to express support (Egber, 2004; Neri et al., 2002).

Suggestions for the design of the learning environment

Significant improvement could also be made in the design of the interactive *environment*. The students who tested *MyET* would not have considered the practice to be so mechanical if there had been more interaction between the computer and learner or among the learners. In the *Talk To Me* (Auralog, 2006) program, a user friendly interface enables learners to interact with the software by speaking. As a learner hears a question, it is simultaneously displayed on the screen. The learner is then supposed to reply with an answer he/she chooses from those provided by the software. Through speech recognition, the computer recognises the learner's utterance and accordingly moves on to the following conversation turn. Different choices lead the dialogue along different paths. Neri et al. (2002) commented that a program like this ensures a certain degree of realism since it simulates real life discourse.

Interaction with other English learners on the Internet can also increase the learner's use of English, make learning pronunciation more exciting and, hence, reduce the feeling of monotony resulting from listening or production practice. For example, the Internet community of *MyET* could use audio email software programs and structured audio chat so that learners could experience real communication, as recommended by Egbert (2004). An online community like this is a venue in which learners can receive feedback through peer and group interaction. According to Morris (2005), peer feedback can not only foster learners' increased awareness of language forms but also play an important role in their L2 development.

Conclusion

Student evaluation in this study indicated that the students benefited most from the functions (e.g. recording and playing) that made autonomous study easy. Additionally, *MyET* provides excellent feedback by showing the specific problems that learners have in segment pronunciation and by providing automated aids for the timing and chunking of spoken discourse. However, visual displays such as phonetic spectrograms were

not considered useful by the learners when they were not accompanied by instruction from a teacher. Moreover, the students had problems adjusting to the speeds of the models provided by *MyET*. The evaluation also showed that it would be desirable for *MyET* developers to design activities based on a pedagogical and communicative curriculum that could help learners become familiar with the English sound system before attempting production. The students also found the computerised grading system of *MyET* to be mechanical, and they wished the oral practice could be more interesting. Furthermore, they felt the need for cumulative analysis that could help them understand which type of practice they should focus on. The results of this study suggest that the developers of CAPT software in general should take learners' proficiency levels and learning styles into account. For learners with different proficiency levels and learning styles, a variety of materials and models should be presented at different speeds and in different speaking styles. For some students, additional support, such as a Chinese glossary of vocabulary, would help them make better progress in their practice with *MyET*.

Undeniably, this study has its limitations. The performance and viewpoints of nine subjects might not be sufficient for a thorough evaluation of *MyET*. A qualitative study like this aims to explore some issues that are related to learners themselves. An ideal, innovative CALL software program is one that makes language learning both enjoyable and productive. Thus, an evaluation done by learners can serve as a useful reference for improving the design of CAPT software. Moreover, the level of the advanced subjects in this study was equivalent to that of Intermediate-High speakers, according to the *ACTFL Proficiency Guidelines* (ACTFL, 1999). The results may have been different if the user evaluation had been conducted by advanced language learners whose English proficiency was sufficient for them to obtain high scores on the TOEFL test and the SPEAK (Speaking Proficiency Assessment Kit) test. As Bordonaro (2003) reported, advanced students enjoyed learning a language through interaction with native speakers rather than by using expensive language learning software to practice English. Above all, due to time limitations, this study did not evaluate the long term effects of incorporating *MyET* into pronunciation training. Hopefully, research in the future will probe the pedagogical effect of CAPT software programs like *MyET*.

This study has some implications for the teaching of English pronunciation and the pedagogical application of pronunciation oriented CALL software. In fact, the incorporation of CAPT technology into pronunciation teaching should be carefully considered by both English teachers and learners. Considering the number of pronunciation oriented CALL software programs available on the market, it is imperative that teachers understand the pros and cons of the programs and decide which software can meet the

needs of their students and help them improve their English pronunciation and communication skills.

To achieve this goal, Derwing and Munro (2005) suggested that teachers have a foundation in linguistics (such as phonology) and pronunciation research. Such a foundation will enable teachers to assess their students' pronunciation and help them develop their ability to introspect and draw comparison between L1 and L2. The same suggestion applies to computerised language instruction. It would be exciting to see more CAPT software that has a pedagogical basis. Only when technology dances with pedagogy can language learners sing like nightingales.

Acknowledgements

An earlier version of this paper was presented at the 21st International Conference on English Teaching and Learning in the Republic of China (ROC-TEFL), 2004. Special thanks go to three anonymous referees of AJET for their helpful and constructive suggestions and comments on this paper.

References

- ACTFL (American Council on the Teaching of Foreign Languages) (1999). *ACTFL Proficiency Guidelines - Speaking* (Revised 1999-PDF). [verified 13 Aug 2006] <http://www.actfl.org/files/public/Guidelinespeak.pdf>
- Akahane-Yamada, R., Tohkura, Y., Bradlow, A. & Pisoni, D. (1996). Does training in speech perception modify speech production? *Proceedings of the International Conference on Spoken Language Processing*. Philadelphia, PA.
- Auralog (2006). <http://www.auralog.com/>
- Bordonaro, K. (2003). Perceptions of technology and manifestations of language learner autonomy. *CALL-EJ Online*, 5(1), 1-21. [verified 10 Aug 2006] <http://www.tell.is.ritsumei.ac.jp/callejonline/journal/5-1/bordonaro.html>
- Breitkreutz, J., Derwing, T. M. & Rossiter, M. J. (2002). Pronunciation teaching practices in Canada, *TESL Canada Journal*, 19, 51-61.
- Butler-Pascoe, M. E. & Wiburg, K. M. (2003). *Technology and teaching English language learners*. MA: Pearson Education, Inc.
- Celce-Murcia, M. & Goodwin, J. (1991). Teaching pronunciation. In M. Celce-Murcia (Ed), *Teaching English as a second language* (pp. 136-153). NY: Heinle and Heinle.
- Celce-Murcia, M., Brinton, D. M. & Goodwin, J. M. (2004). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge: CUP.
- Chappelle, C. (1997). CALL in the year 2000: Still in search of research paradigms? *Language Learning and Technology*, 1(1), 19-43. [verified 10 Aug 2006] <http://llt.msu.edu/vol1num1/chappelle/>
- Chen, H. J. (2004). Automatic speech recognition and oral proficiency assessment. *Proceedings of International Conference on English Language Teaching Instruction and Assessment 2004*, 85-102. Taiwan: National Chung Cheng University.

- Chun, D. M. (1998). Signal analysis software for teaching discourse intonation. *Language Learning and Technology*, 2(1), 61-77. [verified 10 Aug 2006] <http://llt.msu.edu/vol2num1/article4/index.html>
- Derwing, T. M. & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39(3), 379-397.
- Dickerson, W. B. (1994). Empowering students with predicative skills. In J. Morley (Ed), *Pronunciation pedagogy and theory: New directions, new views* (pp.19-35). Alexandria, VA: TESOL Publications.
- Duncan, C., Bruno, C. & Rice, M. (1995). *Learn to speak Spanish: Text and workbook*. San Francisco: Hyperglot Software.
- Dunkel, P. (Ed) (1991). *Computer-assisted language learning and testing: Research issues and practice*. Philadelphia: Penn State University Press.
- Egbert, J., Chao, C. C. & Hanson-Smith, E. (1999). *CALL environments: Research, practice, and critical issues*. Illinois: Teachers of English to Speakers of Other Languages, Inc.
- Egbert, J. (2004). Review of *Connected Speech*. *Language Learning and Technology*, 8(1), 24-28. [verified 10 Aug 2006] <http://llt.msu.edu/vol8num1/review2/default.html>
- Ehsani, F. & Knodt, E. (1998). Speech technology in computer-aided language learning: Strengths and limitations of a new CALL paradigm. *Language Learning and Technology*, 2(1), 45-60. [verified 10 Aug 2006] <http://llt.msu.edu/vol2num1/article3/index.html>
- English Computerized Learning (2006). <http://www.englishlearning.com/en/>
- Engwall, O. (2006). Feedback strategies of human and virtual tutors in pronunciation training. *TMH-QPSR*, 48, 11-34.
- Eskenazi, M. (1999a). Using automatic speech processing for foreign language pronunciation tutor: Some issues and a prototype. *Language Learning and Technology*, 2(2), 62-76. [verified 10 Aug 2006] <http://llt.msu.edu/vol2num2/article3/index.html>
- Eskenazi, Maxine. (1999b). Using a computer in foreign language pronunciation training: What advantages? *CALICO Journal*, 16, 447-469.
- Flege, J. E. (1995). Second-language speech learning: Finding and problem. In W. Strange (Ed), *Speech Perception and Linguistic Experience: Theoretical and methodological issues* (pp. 233-273). Timonium, MD: York Press.
- Fraser, H. (1999). ESL pronunciation teaching: Could it be more effective? *Australian Language Matters*, 7(4), 7-8.
- Greenspan, S. I. & Lewis, D. (2002). *The Affect-Based Language Curriculum (ABLC): An intensive program for families, therapists, and teachers*. Bethesda, MD: Interdisciplinary Council on Developmental and Learning Disorders.
- Harless, W. G., Zier, M. A. & Duncan, R. C. (1999). Virtual dialogues with native speakers: The evaluation of an interactive multimedia method. *CALICO Journal*, 16, 313-337.
- Holland, V. M., Kaplan, J. D. & Sabol, M. A. (1999). Preliminary tests of language learning in a speech-interactive graphics microworld. *CALICO Journal*, 16, 339-359.

- Jenkins, J. (2002). A sociolinguistically-based, empirically-researched pronunciation syllabus for English as an international language. *Applied Linguistics*, 23, 83-103.
- Jones, R. H. (1997). Beyond 'Listen and Repeat': Pronunciation teaching materials and theories of second language acquisition. *System*, 25, 103-112.
- Kaplan, J. D., Sabol, M. A., Wisher, R. A. & Seidel, R. J. (1998). The military language tutor (MILT) program: An advanced authoring system. *Computer Assisted Language Learning*, 11, 265-287.
- Kenworthy, J. (1987). *Teaching English pronunciation*. New York: Longman.
- Kim, I.-S. (2006). Automatic speech recognition: Reliability and pedagogical implications for teaching pronunciation. *Educational Technology & Society*, 9(1), 322-334. http://www.ifets.info/journals/9_1/26.pdf
- LaRocca, S. A., Morgan, J. J. & Bellinger, S. M. (1999). On the path to 2X learning: Exploring the possibilities of advanced speech recognition. *CALICO Journal*, 16, 295-310.
- Levis, J. M. (2005). Software reviews: *Streaming Speech*: Listening and pronunciation for advanced learners of English. *TESOL Quarterly*, 39(3), 559-562.
- Lian, A. (2004). Review of *Streaming Speech*. *Language Learning and Technology*, 8(2), 23-32. [verified 10 Aug 2006] <http://llt.msu.edu/vol8num2/review2/default.html>
- Lippi-Green, R. (1997). *English with an accent: Language ideology and discrimination in the United States*. New York: Routledge.
- LLab (2005). <http://www.myet.com/en/Index.htm>
- Macdonald, S. (2002). Pronunciation - views and practices of reluctant teachers. *Prospect*, 17(3), 3-18.
- Micro Video Corporation (2006). <http://www.videovoice.com/>
- Molholt, Garry. (1988). Computer-assisted instruction in pronunciation for Chinese Speakers of American English. *TESOL Quarterly*, 22, 91-111.
- Molholt, Garry. (1990). Spectrographic analysis and pattern in pronunciation. *Computers and the Humanities*, 24, 81-92.
- Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, 25, 481-520.
- Morley, J. (Ed) (1994). *Pronunciation pedagogy and theory: New views, new directions*. Alexandria, VA: TESOL Publications.
- Morris, F. (2005). Child to child interaction and corrective in a computer mediated L2 class. *Language Learning and Technology*, 9(1), 29-45. [verified 10 Aug 2006] <http://llt.msu.edu/vol9num1/morris/default.html>
- Munro, M. J. (2003). A primer on accent discrimination in the Canadian context. *TESL Canada Journal*, 20(2), 38-51.
- Neri, A. Cucchiaroni, C., Strik, H. & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, 15, 441-467.
- Nunan, D. (1988). *The learner-centered curriculum*. Cambridge: CUP.
- Oxford, R. L. & Anderson, N. (1995). State of the art: A cross-cultural view of language learning styles. *Language Teaching* 28, 201-215.

Appendix B: Questionnaire

1. What features of this software program are helpful for improving your pronunciation?
(Please fill in a number to indicate the extent of helpfulness: 5 most helpful, 4 very helpful, 3 helpful, 2 a little bit helpful, 1 not helpful at all.)
 - ___ (a) Overall score
 - ___ (b) Correction of segment:
 - ___ display of phonetic symbols and scores
 - ___ tips for pronunciation of segments
 - ___ 3D animation of mouth
 - ___ contrast display of spectrogram with model utterance
 - ___ comments
 - ___ (c) pitch:
 - ___ display of syllable and key
 - ___ contrast display of pitch with model utterance
 - ___ comments
 - ___ (d) timing:
 - ___ display of syllable and speed
 - ___ comments
 - ___ (e) emphasis:
 - ___ display of emphasis
 - ___ comments
 - ___ (f) Recording design
 - ___ (g) Clicking on the phonetic symbols or certain regions of spectrograms to listen to target sounds
2. a. By studying the acoustic spectrum are you able to understand the weak points in your pronunciation? Yes, I can. ___ No, I can't. ___
 b. After the teacher interprets the spectrums, do you think studying and comparing the spectrums of the model utterance and your own can help improve your pronunciation? Yes, I do. ___ No, I don't. ___
3. After practicing several times, are you able to be aware of the correctness (or incorrectness) of your own pronunciation?
4. Given more time to practice, do you think this program could help you improve your pronunciation? Yes, I think it could. ___ No, I don't think so. ___
5. Would you comment on or provide an overall evaluation of this software program?

Tsai Pi-hua is a doctoral student in the TESOL program with the Department of English, National Chengchi University, Taiwan. She is also an instructor in English at China University of Technology, Taiwan, with fifteen years of teaching experience. Her research focuses on pronunciation teaching, discourse analysis and children's language acquisition.

Ms Tsai Pi-hua, Floor 6, No. 8, Lane 246, Yanji Street, Taipei 106, Taiwan.
 Email: tsaipihua@yahoo.com.tw