Computer-Assisted Pronunciation Learning in a Collaborative Context: A Case Study in Taiwan

Tsai Pi-hua

Holistic Education Center, Mackay Medical College, Taipei, Taiwan

tsaipihua@gmail.com

ABSTRACT

Computer-assisted pronunciation training (CAPT) software provides language learners with an individualized free environment where they can have access to unlimited input and repetitive practice pronunciation at their own pace. This study explores the impact of CAPT on 90 Taiwanese college students' pronunciation learning and examines if other kinds of mediation, such as peer support, could enhance its effect. It includes two experimental groups using *MyET*, a CAPT program designed in Taiwan, either independently (i.e., the Self-Access CAPT Group) or with peers (i.e., the Collaborative CAPT Group) while the control group only had access to MP3 files for practice. Though the quantitative results did not indicate group difference, the qualitative analysis showed that all three groups went through different learning processes. The Self-Access CAPT Group reported the most frequency in the category of *self-monitoring of language learning and production* while the Collaborative CAPT Group had the most frequency in the categories of *gains and strategies*. Lacking the mediation of peers and the feedback from *MyET*, the MP3 Group reported the highest frequency in *difficulties* and the least frequency of *gains and strategies* during the practice. Some pedagogical implications are also presented.

Keywords: Computer-assisted pronunciation training (CAPT), Collaborative learning, Strategy, Social cultural theory

INTRODUCTION

Pronunciation is an important factor in effective communication. Poor pronunciation may cause misunderstanding and therefore can become a barrier to communication. However, pronunciation instruction has long been ignored (Breitkreutz, Derwing, & Rossiter, 2001; Chun, 2012; Brown, 1991; Neri, Cucchiarini, & Strik, 2006). Learners' acquisition of English suprasegmentals did not receive much attention until the 1970s (Celce-Murcia, Brinton & Goodwin, 2004). Furthermore, in the history of L2 pedagogy, the core of pronunciation instruction stresses the importance of segments rather than suprasegmentals. A proficient L2 learner needs to have a good mastery of suprasegmentals as well as segmental pronunciation. Egbert (2004) highlights the fact that language learners are not able to speak and listen in a second language merely with phonemic correctness. Linguistic, syntactic and semantic information, according to Crystal (1981), can be more easily conveyed if a speaker can produce correct pitch variations in his/her own speech utterance, which consequently results in effective communication.

Prosody (i.e., stress, rhythm, pitches, and intonation) did not become a principal focus for pronunciation learning until the 1980s (Chun, 2002). Because of its importance in effective communication, prosody has been given high priority in the teaching of pronunciation (Dickerson, 1989; Gilbert, 1987; Hardison, 2004; Pennington & Richards, 1986). Instead of asking learners to speak as accurately and fluently as English native speakers, the goal of teaching prosody is to help learners achieve mutual intelligibility (Derwing & Munro, 2005; Jenkins, 2002). To achieve this goal, it is indispensable for teachers to include in their curriculum design suprasegmentals such as thought groups, stress, intonation, rhythm, reduced speech, and linking (Goodwin, 2001).

It was not until the late 1990s that the role of pronunciation pedagogy in English teaching was justified when many international language proficiency tests, such as the iBT TOEFL Test, began to include the evaluation of speaking ability. Similar to what has occurred in western countries, pronunciation (including prosody) learning

and teaching in Taiwan has been an area that does not receive enough attention. The reason that pronunciation teaching is not given a high priority is because most high-stakes examinations in Taiwan test students' reading and writing abilities instead of pronunciation. To cope with the shift in the testing trend, more and more pronunciation courses have been incorporated into English teaching in Taiwan. Importantly, technology has been employed to develop some pronunciation software to improve students' speech production. For example, LLabs Inc., a company located in Taiwan, uses advanced audio and visual technology in its product, *MyET* (MyET, 2012), to raise learners' awareness of their pronunciation problems. For its design of visualization of *MyET*, see Figure 1.

Because of the international shift in pronunciation instruction, there are more studies in Taiwan evaluating the efficacy of pronunciation software (cf. Chen, 2004; Chen, 2005; Chen & Chiu, 2005; Chen, M.W., 2006; Tsai, 2006). However, most of the studies mainly presented quantitative results of the learners' practice with pronunciation software. Therefore, it is necessary to probe empirically into the interaction between learners and pronunciation software and examine how they feel during the interaction. This present study investigates the impact of computer-assisted pronunciation training (CAPT) software, i.e., *MyET*, on learners' pronunciation learning including the difficulties the students had while using *MyET* as well as the strategies they developed from their interaction with the system of their peers. Moreover, this study aims to clarify the role of the CAPT system in pronunciation instruction and to investigate how other kinds of mediation can also reinforce its efficacy, such as the human element (peers or teachers).

To achieve the objectives stated above, two research questions await to be answered in this study:
- To what extent does CAPT software have an impact on students' pronunciation learning? Can practice sessions using the CAPT system result in a change in the students' performance in all the rating components or just in some of them?
- What are the learning processes of participants' pronunciation learning and their perceptions toward their learning through different mediations (i.e., CAPT software and peers)? What are the difficulties and challenges they encounter during the processes?

LITERATURE REVIEW
Much pedagogical instruction for foreign-language learners of English to practice pronunciation used to be based on the drill method (Spaai & Hermes, 1993). Rutherford (1987) and Schmidt (1990) suggest that instead of asking students to do rote imitation, teachers should try to promote discovery in their classes by focusing their learners' attention on specific targeted phonological forms in the input, and on the distance between the present level of their inter-language and their target form. However, language teachers seem to have difficulty in helping their students on an individual basis, detecting a distance due to his/her time limitation in class.

With the development of advanced technology, computer-assisted pronunciation training (CAPT) software has been developed to improve learner's pronunciation. Technology has made it possible to conduct a delicate and unbiased analysis of intonation, a difficult task for teachers who try to evaluate their students' performance based on human perception alone (Chapelle, 2004). CAPT software utilizes Automated Speech Recognition (ASR), a state-of-the-art technology that allows a computer both to recognize words that are read aloud or recorded and to compare a student's production of those words with that of a native speaker. Specifically, the software can provide learners with a visual feedback such as a contrast between pitch contours made by the learner and that of a model teacher. The feedback can then help learners understand how their speech deviates from that of the model utterance. Ideally, with constant practice and modification, learners will be able to narrow down the difference between their production and the model. The visual display can be conducive to the acquisition of correct intonation for learners of English because through the visual feedback, learners are made aware of their production differences from the model utterances and they can correct them in subsequent learning trials (Spaai & Hermes, 1993). Not only can CAPT software give students the autonomy to review any part of the materials as many times as they wish but it can also provide them with unlimited input, individualized feedback and additional assistance in a private, stress-free environment (Butler-Pascoe & Wiburg, 2003; Neri et al., 2002).

Though the integration of technology into pronunciation instruction has been found beneficial to pronunciation learning, Spaai and Hermes (1993) suggest that more conclusive empirical evidence is needed to verify the pedagogical benefits of using CAPT software in language classrooms. Some studies have explored the effect of speech technology (e.g., animation or visualization of pitch contours) on learners' learning of segments (Chen, 2012; Fatima Zaki Mohammad Al-Qudah, 2012; Hardison, 2003; Lively, Logan, & Pisoni, 1993; Motobashi-Saigo & Hardison, 2009; Wang & Munro, 2004). Others have examined its impact on users' learning of suprasegments such as intonation (Anderson-Hsieh, 1992 & 1994; de Bot, 1983; Delmonte, 2010; Hardison, 2004; Levis & Pickering, 2004). However, research related to its impact on learners' acquisition of the pattern of timing (i.e., tempo and rhythm) is limited. According to Chen (2006), timing is a crucial component that may influence native speakers' judgment of a foreign accent in learners of English speech patterns. Therefore, it is important to investigate the impact of practice with a CAPT system on EFL learners' performance of English timing pattern. Moreover, there is little research on computer-assisted pronunciation training that involves collaborative learning. The collaborative strategies are said to be the catalysts for peer interaction and are conducive to L2 learning (Oxford, 1997; Vandergrift 1997; Vygotsky, 1978).

Many studies have approached the effectiveness of speech technology solely by means of statistical evaluation. However, Lantolf and Thorne (2006) note that 'learner performance, despite its external manifestation, can have a very different underlying psychological status that changes over time' (p. 287). Because of the limitation of a quantitative study, a study needs to be done through cross-referencing of the learners' actual language gains and their reflections on their interaction with the technology. Advanced as modern technology appears to be, it is not uncommon to come across the fallacy that technology is the solution to all learning and teaching problems. To see through this fallacy, language teachers may need to know exactly to what extent the incorporation of technology into pronunciation instruction has influenced their students' learning. Apart from the employment of advanced speech technology, they may also want to use some other heuristic teaching methods in their pronunciation instruction. As Hirata (2004, p. 358) states, "The development of training techniques for L2 pronunciation is in its infancy, and there is much to be explored in assessing whether various methods of pronunciation training are effective in enabling subjects to accurately produce L2 contrasts."

While some teachers are advocating the use of technology in pronunciation pedagogy, others see collaboration learning as a method in improving students' learning of English pronunciation. Chela-Flores (2001) underscores that pronunciation pedagogy has to provide an opportunity for freer practice in which learners interact with peers in discourse situations that exemplify a variety of prosodic features. Such practice, she believed, can increase awareness of the communicative aspects of pronunciation. Apart from facilitating the language learning *per se*, collaboration learning was also found to be capable of raising the learner's awareness of some strategies that he/she alone would never think of applying in his/her learning of a language otherwise. Collaboration and social interaction are at the core of the Vygotskian socio-cultural theory.

A Vygotskian view of language acquisition is that it is essentially a social activity within a socio-cultural framework. According to Vygotsky (1978, p. 57), "Any function in the child's cultural development appears twice, or on two planes": first at the social plane (between people, i.e., inter-psychological) and then at an individual one (within the child, i.e., intra-psychological). Vygotsky underlines the importance of society because it determines human beings' behavior. He probed the causes and process of how interaction between two people in a dyad (i.e., on an interpersonal dimension) could lead to higher mental functioning (i.e., on an intrapersonal dimension). The Vygotskian sociocultural framework accentuates that knowledge is constructed through a process of collaboration, interaction, and communication among learners in social settings (Vygotsky, 1978, 1986).

There have been studies adopting the Vygotskian sociocultural framework to probe the myth of language proficiency. For instance, Swain, Brooks and Tocalli-Beller (2002) claim that peer collaborative dialogues can mediate second language learning. Jones (2006) also found that collaborative learning in a computer-based environment can support L2 learning. In a similar line, according to Chao (2007), a case study on community

learning reported that the assistance and guidance from teachers and peers as well as dialogic communication among them can have an invaluable impact on the participants' concept and attitudes toward the given task. According to Fang and Chen (2012), learners practicing pronunciation in two learning contexts (i.e., both computer-assisted pronunciation training and classroom-based pronunciation training) were able to develop more use of strategies. They assumed that the increase might result from the fact that the learners had more opportunities to observe and practice. This present study navigates how different mediations may impact students' pronunciation learning process.

METHODOLOGY
Subjects and experiment design
This study involved an experiment composed of 90 Taiwanese college English major students , who were divided into two CAPT groups (the experiment groups) and one non-CAPT group (the control group). The CAPT groups were the Self-Access CAPT Group and the Collaborative CAPT Group. The participants of the Self-Access CAPT Group studied independently using *MyET*, a CAPT system made in Taiwan, while those of the Collaborative CAPT Group practiced *MyET* with peers sharing one computer. In contrast, the non-CAPT group had only access to the written texts and the MP3 recording of the texts from *MyET*. For ten weeks these groups practiced the texts and the recordings of Part I of a play named *Cinderella* and Part I of *The Three Billy Goats* offered by *MyET,* which were featured with dramatic intonation. Appendix A (Excerpted from Tsai, 2006) presents a snapshot of the text for practice.

All the groups received the same pretest and posttest, i.e., an excerpt from the story, Cinderella. Each week after practice with the texts and listening to the recordings of the texts, the participants were asked to write down their learning reflections in their learning logs, on which their teacher would make her feedback every week. The teacher (i.e., the researcher of this study) played the role of facilitator, giving her students help only when they asked for it. It was not until the fourth week that the teacher/researcher of this study gave a ten-minute instruction on some basic concepts of English prosody to each group because quite a few participants had revealed that they had difficulties catching up with the reading speed of the model utterance. The teacher's instruction time was limited to 10 minutes because it was hoped that it would not influence the results of the experiment.

In the tenth week, a posttest was given to the students using the same test material and procedure as used in the pretest. The audio files of the participants' reading of Part I of *Cinderella* in both tests were collected for later analysis. The data collected were then managed through both statistical and qualitative analyses. A procedure sketch of this study is shown in Table 1.

Table 1: Procedure Sketch of this Study

Groups (N=30 students in each group)	Self-Access CAPT	Collaborative CAPT	MP3
Treatment (10 weeks)	Practicing *MyET* independently	Practicing *MyET* with peers	Practicing the handouts through MP3 player
Materials	1. *MyET* program with the texts and the recordings of Part I of *Cinderella* and Part I of *The Three Billy Goats*		1. Handouts with the texts and the recordings of Part I of *Cinderella* and Part I of *The Three Billy Goats*
	2. Learning logs		
Data Sets	1. The participants' recordings in the pretest and the posttest (i.e., Part I of *Cinderella*)		
	2. Learning logs		

Data analysis
All the audio files were then randomly uploaded to a database operating behind a rating website specifically designed and developed for this study. Each audio file was rated by four raters with a scale of 5 points. All the

sound files were presented to the raters without revealing the names of the participants and the order of the tests (i.e., pretest or posttest). Such arrangements prevented the raters from knowing whose production they were listening to and which productions had preceded or followed the treatment.

The rating criteria adopted in this study are similar to that of *MyET*, which categorizes the scores for a learner' production into four components: pronunciation, intonation, timing and intensity (i.e., loudness). Due to the fact that, the recording of the participants' speech might be affected by the operation of their microphones, the criteria this study set for the articulation rating did not include the category of intensity. The rating categories of this study, therefore, consist only of three components: pronunciation, intonation and timing. Additionally, a score for overall production was added to the rating items because, according to Chen's finding (2006), the subjective impression of human judgment may be more holistic than discrete.

As to the qualitative inquiry, in order to evaluate the learning processes each group went through and the performance differences across the groups during the process, the participants' weekly learning logs were compiled into a database for careful analysis. The analysis basically followed the naturalistic data processing procedure suggested by Lincoln and Guba (1985, pp. 336-356), a constant comparative content-analytic method. First, emerging themes related to the research questions were identified and were iteratively modified to reflect categories emerging from the data, such as those related to collaborative relationships or the role of technology in producing learning opportunities. Then the themes were categorized and interpreted, whose procedures are considered essential to all successful analyses of qualitative studies (Richards, 2003). Following that, the frequency of each category found in each group's learning logs was calculated and a summary table was made to display the pattern of learning progress each group had been through.

RESULTS AND DISCUSSION

This section presents the quantitative and qualitative analysis results of this present study. The former analysis is based on the scores given by the four raters of native speakers of English while the latter is grounded on the frequency counts the themes as emerged from the learning logs of the participants. First, the quantitative results are presented, which are followed by the qualitative ones.

Table 2 depicts the descriptive statistics on each group's progress in reading Part I of *Cinderella*. *P Progress* refers to the score mean difference between the segmental pronunciation rating from the pretest to the posttest for each group, *I Progress* to the change in the intonation rating, *T Progress* to that in the timing rating, and *O Progress* to that in the overall performance.

Table 2: Descriptive Statistics on the Three Groups' Progress

		N	Mean	Std. Deviation
P1progress	G1	30	0.16	0.45
	G2	30	0.21	0.44
	G3	30	0.21	0.48
	Total	90	0.20	0.45
I1progress	G1	30	*0.86*	0.60
	G2	30	*0.81*	0.63
	G3	30	*0.90*	0.72
	Total	90	*0.86*	0.64
T1progress	G1	30	0.70	0.49
	G2	30	0.59	0.60
	G3	30	0.63	0.55
	Total	90	0.64	0.55
O1progress	G1	30	0.53	0.44
	G2	30	0.51	0.52
	G3	30	0.55	0.49
	Total	90	0.53	0.48

Notes. Progress= the score mean difference from the pretest to the posttest

P1=pronunciation scores for the reading of Part 1 of *Cinderella*
I1=intonation scores for the reading of Part 1 of *Cinderella*
T1=timing scores for the reading of Part 1 of *Cinderella*
O1=scores for the overall performance of the reading of Part 1 of *Cinderella*
G1=Self-Access CAPT Group
G2=Collaborative CAPT Group
G3=MP3 Group

Table 3 shows that the degree of improvement in intonation precedes that of timing and segmental pronunciation. This result is encouraging for the integration of CAPT software into prosody instruction. As shown in Table 3 below, a one-way ANOVA involving Time (pretest, posttest) and mediation type (self-access, collaboration and MP3) did not reveal any significance of mediation in the practice of Part I of *Cinderella* (with all p-values > 0.05). That is, the employment of different kinds of tools for the practice did not result in a significant change in the students' reading performance in all the rating components in the posttest. This result is in accordance with Chen (2005), who reported that though the experiment group did make significant improvements as a result of the experiment, its improvement was not great enough to outperform the control group. One potential explanation for this result may be that the practice sessions were long enough for all the groups to improve themselves at the end of the experiment. Further study is needed to evaluate the accountability of this result.

Table 3: ANOVA on the Performance in Reading Part I of *Cinderella* across Groups

Progress	Groups	Sum of Squares	df	Mean Square	F	Sig.
P1	Between Groups	.05	2	.02	.11	**.889**
	Within Groups	18.47	87	.21		
	Total	18.52	89			
I1	Between Groups	.12	2	.06	.14	**.863**
	Within Groups	37.39	87	.43		
	Total	37.52	89			
T1	Between Groups	.17	2	.09	.29	**.749**
	Within Groups	26.82	87	.30		
	Total	27.00	89			
O1	Between Groups	.01	2	.00	.03	**.966**
	Within Groups	20.75	87	.23		
	Total	20.77	89			

Notes: Significance is set at p-value<0.05
Progress= the score mean difference from the pretest to the posttest
P1=pronunciation scores for the reading of Part 1 of *Cinderella*
I1=intonation scores for the reading of Part 1 of *Cinderella*
T1=timing scores for the reading of Part 1 of *Cinderella*
O1=scores for the overall performance of the reading of Part 1 of *Cinderella*

As for the qualitative results, four themes are generated from the reflections all the participants kept in their learning logs: *learning difficulties*, *gains*, and *strategies* used in their practice, and the *monitoring of one's own performance or language learning*. The participants in this study reported difficulties in comprehending the linguistic aspects of the material and in overcoming their psychological barriers during the learning process. The linguistic difficulties included the fast speed of model teachers' speech, and the variety of pitches manifested in the model teachers' speech depending on the context, i.e., sometimes low, high, lively, arrogant, sounding like an old man's or a young child's voice. As for the psychological aspects, the participants revealed that they were afraid that they might be overheard by the other classmates sitting next to them in the lab, so they were hesitant in practicing the texts loudly and felt shy in recording their reading.

The category of *gains* refers to the benefits the three groups reported to have had from their practice with the CAPT system (i.e., *MyET*) or the texts and the recordings of the fairy tales. Some of them even revealed in their learning logs that they could feel that they had made improvement in the first few weeks in terms of speech rate, intonation, segmental pronunciation, and in understanding of the vocabulary and sentences of the texts. Others claimed that their pronunciation practice helped in the development of skills other than speaking. The students also wrote that they enjoyed themselves during practice.

The students stated that they were able to deal with their learning difficulties with some strategies. Some would imitate the model teachers' reading and then practice the sentences over and over. The strategies that the students were found to use most often were listening and repeating, i.e., imitation, a result similar to that of Fang and Chen (2012). Others wrote that they would listen to the model utterance many times, read with variable intonations and emotions, resegment syllables (e.g., such as *once upon a time*), and control their pitches in reading.

The category of *monitoring their own or peers' learning and language production* refers to the students' evaluation of their own performance (such as *funny* and *too flat*) and the improvement that they made, getting to know the functions of intonation (e.g., to express different emotions), being aware of their incorrect pronunciation, observing the need to express feelings in their reading, and becoming aware of the need to control of the rhythm to make their reading sound more natural. The distribution of each theme category across the three groups is presented in Table 4.

Table 4: The Frequency of the Theme Categories Emerged from the Reflections

Categories	MP3 Group	Collaborative Group	Self-Access Group
D	**406**	192	219
G	203	**263**	243
S	104	**172**	152
MON	39	35	**58**

*Notes: **D** = difficulties, **G** = gains, **S** = strategies,*
***MON** = monitoring of one's own performance or language learning*

As indicated in Table 4, differences were found among different groups. The MP3 Group reported the highest frequency of difficulties (N=406). In contrast, the Collaborative Group had the lowest frequency in reporting their difficulties (N=192) and the highest frequency in describing their gains (N= 263), and strategy brainstorming (N=172). As for the Self-Access Group, it was discovered that they had produced the highest frequency of self-monitoring of language learning (N= 58). This might be because the members in the Self-Access Group had more time to continuously engage themselves in individual on-tasks. Their engagement might have facilitated their awareness of the learning progress during their practice with the system. Nevertheless, working alone, the students in the Self-Access CAPT Group reported more difficulties than the Collaborative Computer Group. Moreover, the former group did not develop as many strategies as the latter one. Though most students in the Self-Access CAPT Group had the potential to sustain their effort in practicing with the system, they might have been feeling lonely during practice. Some of them revealed their need to interact with a real person. Specifically, they were not able to have the support from their peers as experienced by the Collaborative CAPT Group.

Different from the Self-Access CAPT Group, the students in the Collaborative CAPT Group reported the highest frequency in gains and revealed that they had made much improvement in their fluency, their production of intonation, and segmental pronunciation. The collaboration between peers might have created some anesthetizing power, which made them feel that lots of improvement had been made. Explicitly, the students in this group might have received from each other some consolidating statements, such as praise, appreciation and

encouragement so that they thought that they had been performing very well. Metaphorically, the support the students received during the collaborative learning, like a pain killer, was able to alleviate the sense of difficulty and anxiety that might have arisen from their practice with the CAPT system otherwise.

The reflections of the Collaborative CAPT Group showed that the external assistance from their peers did facilitate their awareness of errors. Some students reported that they would revise their production upon receiving assistance from them. In other words, collaborative learning was also conducive to the self-regulation of their learning behavior. Moreover, with the mediated assistance from their peers, the collaborative group was also found to have generated the highest frequency of strategies to deal with their learning difficulty. Specifically, the students could brainstorm many strategies and give feedback to each other to improve their intonation performance. The quantitative results of this study attested the positive outcome of their collaborative learning in learning English prosody.

Like the Self-Access CAPT Group, the students in the MP3 group gained more time to spend on their learning since they did not have to work with other peers. With repetitive imitation, they could listen to the MP3 input more times and do more practice to achieve mastery. However, without peers' assistance and the feedback provided by *MyET*, they were not able to detect their production errors easily. This was the only group in which some of the students claimed that they did not have any difficulties. Their ignorance might have been induced by lack of mediating tools. Moreover, due to the lack of mediated assistance from peers and feedback from CAPT software, they were not able to generate as many strategies as the CAPT groups to tackle their own learning difficulties. As a consequence, the MP3 Group reported the highest frequency of difficulties and the least frequency of gains during the practice.

Finally, the participants' reflections showed that there is room for improvement of the feedback design of the CAPT system. First of all, some students using *MyET* complained about the fluctuation of scoring. They also expected to receive more constructive suggestions on how their production could be improved. For example, some students said though the feedback from the system had helped them modulate their pitch production somewhere in the sentences, they had no ideas how to modify their production to match that of the model utterances. Finally, some of the students considered practicing with the model teachers monotonous and mechanical after practicing for some time. To sum up, while the CAPT Groups (i.e., the Self-Access and Collaborative CAPT Groups) revealed that they benefited from the assistance from the CAPT system, there were times when they felt that their mediated learning through technology itself became a new problem to them.

As to the mediated learning through peers, the reflections of the Collaborative CAPT Group showed that the students were able to encourage and help each other. Some students even reported that the discourse help from their interlocutors (either more competent tutors or peers with equal status) enabled them to detect their gaps in the knowledge of the patterns of spoken English and to produce utterances they could not construct on their own. On the other hand, other students of the Collaborative CAPT Group revealed that they preferred autonomous learning methods because they considered it too tedious and time consuming to work out a way to collaborate with their peers. Some also revealed their concern about being laughed at in their production or not having peers competent enough to coach them for better production of pronunciation. Our finding here is in line with that of Swain, Brooks and Tocalli-Beller (2002), who reported that some learners might still rely more on teacher feedback than on peer feedback. Moreover, some students in this group reflected that working with their peers did take away the limited time they could have saved for their own individual practice.

IMPLICATIONS

The analysis of the students' learning logs verified that the software was able to specifically raise the students' awareness of the prosodic elements such as intonation, which was visually illustrated in various pitch contours. Nevertheless, new technology, such as some updated learning software, should be treated as a mediating tool to stimulate learning and thinking rather than as something that can replace the teacher's job. It was not unusual to hear that teachers can leave their teaching work to certain state-of-the-art software because of its specific

powerful features. As Warschauer (2005, p. 48) claims, "CALL advocates should not view the use of computers as an end in itself but as another tool to promote language learning [that] mediates and transforms human activities." Therefore, even if seemingly trivial instruction is rendered to a CAPT program such as *MyET*, teachers still need to scrutinize students' attitudes and responses to the computer-assisted learning program and provide immediate help for their students if the help is necessary.

The documentation of the learning process of the Collaborative CAPT Group was an example validating what socio-cultural theories have emphasized, i.e., the social nature of learning, and the interaction in cultural-historical contexts through which symbolic, physical, and mental spaces are mediated. The students collaborated with their peers to make and test their hypotheses of the patterns of English pronunciation and co-construct new and useful understandings of English pronunciation and thus developed their capability for learning it. As found in this study, the collaborative learning not only serves as a soothing function in relieving learners' anxiety and sense of difficulty but also boosts up more brainstorming for strategy use. Therefore, it is important for teachers to develop a curriculum that emphasizes interaction.

This study would like to suggest a staged instruction for English pronunciation or pronunciation teaching in general. At the beginning of the class, students can listen to the material and repeat after the model utterance for some time. Practice at this phase can help students internalize the text recording. In this way, the practice starts on an inter-psychological dimension (i.e., an interaction with the recording of texts) and then moves onto an intra-psychological plane (i.e., internalization through repetitive imitation). Then teachers can encourage students to work in pairs to practice some CAPT software. Practice at this stage again returns to the interactional dimension, in which peers help detect each other's errors, assist each other, and co-construct their knowledge of the features of the target language. Then, students can be left alone to work with the software to focus on their individual problems that have either been detected on their own or with their peer's assistance. The feedback that the CAPT system provides can further help identify the errors that cannot be easily or consistently detected by the human ear. Such practice enhances their learning on the intra-psychological dimension, that is, the mediated assistance from the CAPT system can facilitate the students' internalization of what they are learning. Above all, teachers need to invite their students to make reflections on their practice. As Swain (2000, p. 113) has claimed, "Through saying and reflecting on what is said, new knowledge is constructed." The proposed different stages for pronunciation teaching are illustrated in *Figure 1*.
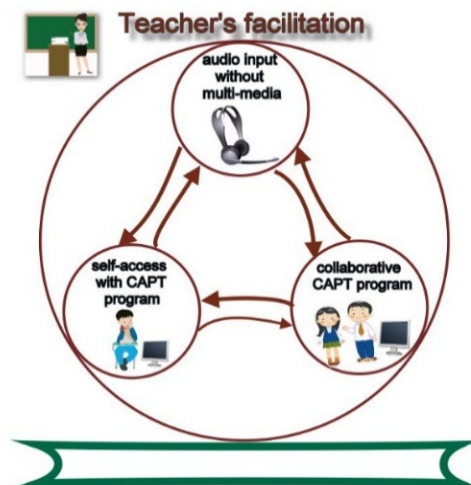

Figure 1. The Proposed Different Stages for Pronunciation Teaching

CONCLUSIONS

This study explores how various kinds of mediation (e.g., MP3, CAPT software and of peers) can support the pronunciation learning process of EFL learners in the colleges in Taiwan. The findings of this study present abundant evidence of the power of collaborative learning in mediating the learning process of the learners.

Moreover, this study sheds light on the difficulties and challenges of EFL learner during their computer-assisted pronunciation learning, and thereby becomes a good reference for both language instructors and CAPT developers. Based on their students' reflections, teachers can see clearly the learning process their students' progress through; and so teachers can further tailor the course to meet the students' needs. As for program designers, the results of this study also give insightful suggestions that may be beneficial for future product design. It is hoped that the program developers will be able to come up with a product that truly incorporates not only technology but also pedagogy.

Specifically, this study suggests a staged employment of various kinds of mediation for teachers to obtain a better outcome for their language teaching. Teachers should not expect that technology can solve all the students' learning problems. Instead, they should pay attention to the different roles assigned to technology and other kinds of mediation. If teachers can introduce various mediating tools to their students to facilitate their learning at different learning stages, they will be able to assist them to move to the next advanced learning stage.

If the peer-peer dialogues of the Collaborative Computer Group of this study could have been recorded, more substantial evidence could have been collected to illustrate the advantages of using CAPT software in a collaborative context. Future research can adopt the framework of Swain (2000) to examine the pattern of the dialogues between peers. According to Swain, collaborative dialogues can mirror the moments of language development. Further experiments involving more learners at different levels of proficiency and studies of retention would contribute to refining our understanding of the impact of different mediating tools on learners' pronunciation learning. As Lynch and Maclean (2001) reported, only less proficient learners were found to show improvements in phonology, syntax and lexis as a result of their incorporation of language from their interlocutors.

REFERENCES
Anderson-Hsieh, J. (1992). Using electronic visual feedback to teach suprasegmentals. *System, 20*(1), 51-62.
Anderson-Hsieh, J. (1994). Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction. *CALICO Journal, 11*(4), 5-22.
Breitkreutz, J., Derwing, T.M., & Rossiter, M.J. (2001). Pronunciation teaching practices in Canada. *TESL Canada Journal, 19*, 51-61.
Brown, A. (ed.) (1991). *Teaching pronunciation: A book of readings.* London & New York: Routledge.
Butler-Pascoe, M. E. & Wiburg, K.M. (2003). *Technology and teaching English language learners.* MA Pearson Education, Inc.
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.
Chao, C. C. (2007). Toward an understanding of computer-mediated EFL writing experience through Vygotskian perspectives. *Taiwan Journal of TESOL*, Monograph Series No.1. Taipei: National Chengchi University.
Chapelle, C. A. (2004). Technology and language learning: Expanding methods and agendas. *System*, *23*(4), 593-601.
Chela-Flores, B. (2001). Pronunciation and language learning: An integrative approach. *IRAL*, 39, 85–101.
Chen, C. L., & Chiu, S. J. (2005). The effectiveness and analysis of the use of *MyET* in assisting the English learning of students at Fuzhu Elementary School. *Selected papers from the 14th Int'l Symposium on English Teaching.* Vol. 2. (pp. 58-69). Taipei: Crane.
Chen, H. J. (2004). Automatic speech recognition and oral proficiency assessment. *Proceedings of International Conference on English Language Teaching Instruction and Assessment* (pp.85-102). Taiwan: National Chung Cheng University.
Chen, K. T. (2005). *An investigation on the impact of ASR software feedback on EFL college students' pronunciation learning.* Unpublished M.A. thesis. National Normal University. Taiwan: Republic of China.

Chen, M. W. (2006). *The impact of automatic speech technology on contrastive stress among adult EFL learners.* Unpublished M.A. thesis. Dayeh University. Taiwan: Republic of China.

Chen, Hsueh Chu. (2012). Hong Kong ESL learners' acquisition of English stress and assessment of an Online Tutoring Programme. *TESOL Journal*, 6, 45-62.

Chiu, T. L. (2005). *Effects of online conversation materials with the support of speech recognition technology on college EFL learners.* Unpublished M.A. thesis. National Tsing Hua University. Taiwan: Republic of China.

Chun, D. M. (2002). *Discourse intonation in L2.* Amsterdam: Benjamins.

Chun, D. M. (2012). Computer-Assisted pronunciation teaching. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Oxford: Wiley-Blackwell.

Crystal, D. (1981). *Clinical linguistics.* New York: Harper.

de Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech, 26* (4), 331-350.

Delmonte, R. (2010). Prosodic tools for language learning. *International Journal of Speech Technology, Online First.* [On-line]. *Available: http://dx.doi.org/10.1007/s10772-010-9065-1, (2014, 27 October).*

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. (1989). *Stress in the stream of speech: The rhythm of spoken English.* Urbana, IL: University of Illinois Press.

Egbert, J. (2004). Review of connected speech. *Language Learning & Technology*, January, *8(1),* 24-28. [On-line]. *Available: http://llt.msu.edu/vol8num1/review2/default.html, (2014, 20 August).*

Fang, T. & Lin, C. C. (2012). Taiwan EFL learners' pronunciation strategies in two learning contexts. *Journal of Language Teaching and Research*, 3 (5), 888-897. [On-line]. Available: *http://www.ojs.academypublisher.com/index.php/jltr/article/view/jltr0305888897, (2014, 23, July).*

Fatima Zaki Mohammad Al-Qudah. (2012). Improving English pronunciation through computer-assisted programs In Jordanian Universities. Journal of College Teaching & Learning, *9* (3), 201-208. [On-line]. Available: *http://journals.cluteonline.com/index.php/TLC/article/view/7085, (2014, 1 November).*

Gilbert, J. B. (1987). Pronunciation and listening comprehension. In J. Morley (ed.), *Current perspectives on pronunciation*, 29-39.

Goodwin, J. (2001). Teaching pronunciation. In M.Celece-Murcia (ed.), *Teaching English as a second or foreign language*, 3rd Ed. Boston: Heinle & Heinle Publishers.

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context and talker variability. *Applied Psycholinguistics, 24,* 495-522.

Hardison, D. M. (2004). Generalization of computer assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology, 8* (1), 34-5. [On-line]. Available: *http://llt.msu.edu/vol8num1/hardison/default.html, (2014, 30 October).*

Hirata, Y. (2004) Computer assisted pronunciation training for native English speakers learning Japanese pitch and duration contrasts. *Computer Assisted Language Learning 17*, 357–376.

Jenkins, J. (2002). A socio linguistically based, empirically researched pronunciation syllabus for English as an international language, *Applied Linguistics* 23(1), 83–103.

Jones, L. (2006). Effects of collaboration and multimedia annotations on vocabulary learning and listening comprehension. *CALICO Journal*, 24 (1), 33-58

Lantolf, J. P., & Thorne, S. L (2006). *Sociocultural theory and the genesis of L2 development.* Oxford: Oxford University Press.

Levis, J., Pickering, L. (2004) "Teaching intonation in discourse using speech visualization technology", *System 32* (4), 505-524.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage Publications, Inc.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94,* 1242-1255.

Lynch, T., & Maclean, J. (2001). A case of exercising: Effects of immediate task repetition on learner's performance. In M. Bygate, P. Skehan, & M. Swain (eds.), *Researching pedagogical tasks: Second language learning, teaching and testing* (141-162). London: Longman.

Motobashi-Saigo, M. & Hardison, D. M. (2009). Acquisition of L2 Japanese geminates training with waveform displays. *Language Learning & Technology, 13*(2), 29-47. Retrieved from http://llt.msu.edu/vol13num2/motohashisaigohardison.pdf.

MyET (My English Tutor), 2012. [On-line]. *Available: http://www.myet.com/MyETWeb/PersonalizedPage.aspx, (2014, 30 October).*

Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2002). The Pedagogy-technology Interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, *15,* 441-467.

Neri, A., Cucchiarini, C., & Strik, H. (2006). ASR-based corrective feedback on pronunciation: Does it really work? *Proceedings of International Conference on Spoken Language Processing 2006* (pp. 1982–1985). Pittsburgh, PA.

Oxford, R. L. (1997). Cooperative learning, collaborative learning, and interaction: Three communicative strands in the language classroom. *Modern Language Journal, 81* (4), 443-456.

Pennington, M. C., & Richards, J. (1986). Pronunciation revisited. *TESOL Quarterly, 20* (2), 207-225.

Richards, Keith. (2003). *Qualitative inquiry in TESOL.* NY: Palgrave.

Rutherford, W. (1987). *Second language grammar: Teaching and learning*. Harlow: Pearson Educational Trust.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129-158.

Spaai, G. W. G., & Hermes, D. J. (1993). A visual display for the teaching of intonation. *CALICO Journal*, *10* (3), 19-30.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (ed.), *sociocultural theory and second language learning* (pp. 97-114). Oxford: Oxford University Press.

Swain, M., Brooks, L., & Tocalli-Beller, A. (2002). Peer–peer dialogue as a means of second-language learning. *Annual Review of Applied Linguistics*, *22*, 171–185.

Tsai, P. H. (2006). Bridging pedagogy and technology: User evaluation of pronunciation oriented CALL software. *Australasian Journal of Educational Technology*, *22* (3), 375-397. [On-line]. *Available: http://www.ascilite.org.au/ajet/ajet22/tsai.html, (2014, 28 October).*

Vandergrift, L. (1997). The Cinderella of communication strategies: Reception strategies in interactive listening. *Modern Language Journal, 81* (4), 494-505.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press. Published originally in Russian in 1930.

Vygotsky, L. S. (1986). *Thought and language*. Boston: MIT Press.

Wang, X.- Munro, M.J. (2004) "Computer-based training for learning English vowel contrasts", *System, 32* (4), 539-552.

Warschauer, M. (2005). Sociocultural perspectives on CALL in J.L. Egbert and G.M. Petrie (Eds.). *CALL Research perspectives* (pp. 41–51). NY: Lawrence Erlbaum Associates.

APPENDIX A: A Screen Capture of a Learning Interface of *MyET*

(Excerpted from Tsai, 2006)