The effectiveness of integrating commercial pronunciation software into an ESL pronunciation class

by

Yang Liu

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

Major: Teaching English as a Second Language/Applied Linguistics

Program of Study Committee:
John Levis, Major Professor
Dan Douglas
Horabail Venkatagiri

Iowa State University
Ames, Iowa
2008

Copyright © Yang Liu, 2008. All rights reserved.
# TABLE OF CONTENTS

**LIST OF FIGURES**  iv

**LIST OF TABLES**  v

**ACKNOWLEDGEMENTS**  vi

**ABSTRACT**  vii

**CHAPTER 1. INTRODUCTION**  1
   1.1 Background  1
   1.2 The Current Study  3
   1.3 Research Questions  4
   1.4 Organization of the Study  5

**CHAPTER 2. LITERATURE REVIEW**  6
   2.1 The Effect of Pronunciation Instruction  6
   2.2 Computer-Assisted Pronunciation Teaching  10
      2.2.1 The role of computer technology in pronunciation instruction  10
      2.2.2 The effectiveness of CAPT  13
   2.3 Seferoğlu’s Study  15

**CHAPTER 3. METHODOLOGY**  19
   3.1 Participants  19
      3.1.1 Students  19
      3.1.2 Teacher  20
      3.1.3 Raters  20
   3.2 Materials  21
      3.2.1 Traditional classroom teaching material  21
      3.2.2 Pronunciation Power  22
      3.2.3 Pretest and posttest material  24
      3.2.4 The rating scale  24
      3.2.5 Post-class surveys  24
   3.3 Procedures  25
      3.3.1 Recruitment of students  25
      3.3.2 Pretest  26
      3.3.3 Pronunciation instruction  26
      3.3.4 Posttest  28
      3.3.5 Recording processing and rating  28
   3.4 Analysis  29
      3.4.1 Analysis of the rating sheets  29
      3.4.2 Analysis of the post-class surveys  30
      3.4.3 Identification of evidence for research questions  30

**CHAPTER 4. RESULTS**  31
   4.1 Research Question 1  31
   4.2 Research Question 2  33
   4.3 Research Question 3  38
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.</td>
<td>Distribution of comprehensibility scores of Group B</td>
<td>33</td>
</tr>
<tr>
<td>Figure 2.</td>
<td>Distribution of accentedness scores of Group B</td>
<td>33</td>
</tr>
<tr>
<td>Figure 3.</td>
<td>Distribution of the comprehensibility scores of the pretest</td>
<td>35</td>
</tr>
<tr>
<td>Figure 4.</td>
<td>Distribution of the accentedness scores of the pretest</td>
<td>35</td>
</tr>
<tr>
<td>Figure 5.</td>
<td>Distribution of the comprehensibility scores of the posttest</td>
<td>36</td>
</tr>
<tr>
<td>Figure 6.</td>
<td>Distribution of the comprehensibility scores of the posttest</td>
<td>37</td>
</tr>
<tr>
<td>Figure 7.</td>
<td>Example role play cards</td>
<td>51</td>
</tr>
<tr>
<td>Figure 8.</td>
<td>Example picture for picture-description task</td>
<td>52</td>
</tr>
<tr>
<td>Figure 9.</td>
<td>Front view of human demonstration</td>
<td>54</td>
</tr>
<tr>
<td>Figure 10.</td>
<td>Side view animation and front view of human demonstration</td>
<td>54</td>
</tr>
<tr>
<td>Figure 11.</td>
<td>Speech analysis</td>
<td>55</td>
</tr>
<tr>
<td>Figure 12.</td>
<td>Reading drill for sentence rhythm</td>
<td>56</td>
</tr>
<tr>
<td>Figure 13.</td>
<td>Reading drill for intonation</td>
<td>56</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1. Students in the study 21
Table 2. Independent samples t-test results for the comprehensibility scores of the pretest and posttest for Group B 32
Table 3. Independent samples t-test results for the accentedness scores of the pretest and posttest for Group B 32
Table 4. Independent samples t-test results for the comprehensibility scores of the pretest and posttest for Group B 34
Table 5. Independent samples t-test results for the accentedness scores of the pretest and posttest for Group B 34
Table 6. Independent samples t-test results for the comprehensibility scores of the posttest for Group A and B 35
Table 7. Independent samples t-test results for the accentedness scores of the posttest for Group A and B 36
Table 8. Students’ answers to the post-class survey Question 3 38
Table 9. Students’ ratings for the traditional class 39
Table 10. Group B’s ratings for the software 39
Table 11. Group B’s answers to the post-class survey Question 7 58
ACKNOWLEDGEMENTS

There are a lot of people I would like to thank for helping me with my thesis. I would like to thank my major professor, Dr. John Levis, who helped me a lot all the way through the research conduction and the thesis completion. I greatly appreciate his inspirational feedback and suggestions on my study as well as the generous encouragement he offered. I would like to thank my committee members Dr. Dan Douglas and Dr. Horabail Venkatagiri for their encouragement and great support of my work. I would like to thank Dr. Roberta Vann for supervising my teaching practicum and being supportive with my research. I would like to thank Felicity Douglas for helping me use the pronunciation software and the computer lab in the study.

At last, I want to dedicate this thesis to my dear parents, who have always been supportive for me no matter what. It was impossible to finish this thesis and my graduate education in America without their invaluable love and care.
This study examines the effectiveness of integrating commercial pronunciation software into an ESL pronunciation class of learners at university level. The study partially replicates Seferoğlu’s (2005) research design and seeks to confirm her findings through a revised methodology. Participants in this study were 18 international graduate students from various departments of the Iowa State University in the US. Students were assigned to two experimental groups which received traditional classroom pronunciation instruction and instruction that integrated the use of commercial pronunciation software, respectively, for six weeks. A pretest and a posttest using the same picture-description task were conducted in an attempt to find changes in the students overall pronunciation quality. As measured by the ratings of comprehensibility and accentedness from six native speaker raters, the group receiving software-integrated instruction did not show significant pronunciation improvement after the treatment. Neither did the two groups show significant difference in their pretest and posttest scores. Therefore the results did not confirm Seferoğlu’s findings. In addition, the students’ reflection on the instruction received was analyzed to explore which features of the two types of pronunciation instruction were considered most useful and least useful by the language learners.
CHAPTER 1. INTRODUCTION

1.1 Background

Pronunciation is important to second language learners because of its essential roles in oral communication, listener perception and speaker identity. A growing recognition of this fact has contributed to the exploration of new pronunciation instruction approaches. From the earliest Direct Method and the Naturalistic Approaches believing in purely imitation without explicit pronunciation instruction, to the currently dominant Communicative Approach emphasizing interaction as both the means and the ultimate goal of learning a language, practitioners have walked a long way in search of teaching effectiveness. Along with language teachers, researchers have made great effort to provide empirical evidence on the effect of pronunciation teaching methods and techniques (Elliot, 1995; Derwing, Munro & Wiebe, 1998; Derwing & Rossiter, 2003; Jamieson & Morosan, 1986; Yule, Hoffman & Damico, 1987; Macdonald, Yule & Power, 1994).

One prevailing issue in today’s pronunciation teaching and research is how to best teach pronunciation with the advantages of modern technological tools. Computer-assisted pronunciation teaching, or CAPT, has drawn great interest from teachers and researchers for several decades (Molholt, 1988). What computers can do seems quite promising for pronunciation teaching: computers are able to “provide learners individualized instruction, frequent practice through listening discrimination and focused repetition exercises, automatic visual support that demonstrates to learners how closely their own pronunciation approximates model utterances” (Levis, 2007, p. 184).
Previous research has studied the effectiveness of CAPT by focusing on a key feature shared by a certain group of software applications, for example, visual feedback (Anderson-Hsieh, 1994; Hardison, 2004; Hirata, 2004; Lambacher 1999; Neri, Cucchiarini & Strik, 2002a; Levis and Pickering, 2004 and Chun, 2007) and automatic speech recognition (ASR) feedback (Coniam, 1999; Derwing, Munro & Carbonaro, 2000; Kawai & Hirose, 2000; Kim, 2006; Neri, Cucchiarini, & Strik, 2006).

In terms of findings, past research has documented that CAPT could successfully promote certain aspects of pronunciation, for example, vowels (Wang & Munro, 2004); pitch accents and phonemic duration (Hirata, 2004, Kawai & Hirose, 2000); general segments (Neri, Cucchiarini, & Strik, 2006); pronunciation quality of individual words (Mich, Neri & Giuliani, 2006) and intonation (Hardison, 2004.)

However, few researchers have studied the general effects of using a comprehensive computer program with various features that allow a variety of learning activities on the learners’ overall pronunciation quality. The only one is Seferoğlu, who used the commercial English pronunciation teaching software- Pronunciation Power (2000). In her study (2005), Seferoğlu found that the group of students who followed software-based pronunciation instruction had significantly better posttest scores than the group who followed traditional classroom instruction, although the two groups had similar pretest scores, as shown by the results of an independent samples t-test.

The same software is one of the main computer programs offered by a small computer lab of the Graduate College at Iowa State University to their international students for oral English practice. I worked in the Graduate College with international students, mostly international teaching assistants, for one and a half years, used the software myself
sometimes, and had some experience of showing the newcomers how to use the computer
programs. I was always curious about how helpful those programs were for the students.

Seferoğlu has found larger pronunciation improvements in the students who used the
software than those who did not. However, several aspects of her research may have
weakened the validity of her findings. Therefore I conducted a study using the same software
but with a revised methodology.

1.2 The Current Study

The purpose of this study was to examine the effectiveness of using commercial
pronunciation software in an ESL pronunciation class of learners at university level.
Specifically, this study investigated whether the integration of the commercial pronunciation
software into an ESL pronunciation class generated positive effects on quality of learners’
pronunciation and if these effects were comparable to that under traditional teacher-led
classroom instruction. Furthermore, the study looked at the various features the traditional
classroom instruction and computer-based training have to see which features were likely to
have contributed more to the positive effects on learning.

This study partially replicates Seferoğlu’s 2005 study by 1) using two groups of
subjects for comparison between the learning results under traditional teacher-led classroom
instruction and computer-based training of pronunciation; 2) using the same pronunciation
training software; and 3) using a pre-test/post-test design to identify the improvements in
pronunciation quality.

One characteristic of this study is the attempt to reflect the authentic environment
where the target software is actually used at Iowa State University. In order to resemble the
real environment, I chose research subjects from a broad range of departments and colleges at ISU, representing a variety of nationalities and native languages. Also, the subjects in the experimental group worked in the computer room in the Graduate College and had the chance to work alone, which resembled the authentic situation in which individual students reserve the computer room for their own to practice.

1.3 Research Questions

This study examined the effectiveness of the integration of commercial pronunciation software in an ESL pronunciation class. It sought to confirm the findings of Seferoğlu’s (2005) through a revised methodology. This led to the first two research questions:

1. Does the integration of the commercial pronunciation software *Pronunciation Power* into an ESL pronunciation class result in improvements in the general quality of students’ pronunciation as measured by the ratings of comprehensibility and accentedness?

2. Does the group with exposure to the software-based pronunciation training improve differently from the group receiving traditional classroom instruction?

Along with the first two research questions, this study is also concerned with the various features of the traditional classroom instruction and computer-based training that were most likely to have positive effects on learning. This led to the third research question:

Which features of the commercial pronunciation software and the traditional classroom instruction were considered most useful or least useful by the students? Did the students prefer either of the two types of instruction?
1.4 Organization of the Study

Chapter 2 provides a general overview of the relevant literature and studies. Chapter 3 describes the methodology for this study, specifying the participants, the materials and procedures. It also describes the methods of data collection and analysis. Chapter 4 presents the results for each of the research questions. Chapter 5 discusses the possible explanations for the results, limitations of the study and implications for future researchers.
CHAPTER 2. LITERATURE REVIEW

The purpose of this study is to find out whether using commercial pronunciation software in an ESL pronunciation class will lead to improvement of the students’ pronunciation and whether the improvement is comparable to that under traditional classroom instruction. This study partially replicates Seferoğlu (2005), and seeks to confirm her findings through a revised methodology.

To contextualize the present study, this chapter will review the past literature on the effect of pronunciation instruction, computer-assisted pronunciation teaching (CAPT), and analyze Seferoğlu’s (2005) study in details.

2.1 The Effect of Pronunciation Instruction

The central issue of pronunciation teaching is how to maximize the beneficial effects of instruction for students. One aspect of this issue is what should be taught in class. Earlier teaching approaches focused on the segmental features of pronunciation, while more recent approaches have emphasized suprasegmental features such as sentence rhythm and intonation. However, this debate has never reached a conclusion. The growing trend in today’s pronunciation curriculum design is to integrate both the most important segmentals and suprasegmentals appropriately in one course. Another aspect of the issue is how to teach effectively. Along with the development of instruction approaches, traditional teaching methods and techniques have been advocated or challenged during different times while new methods and techniques have been created. A more balanced view today, as represented by
Celce-Murcia et al. (1996), is to use traditional pronunciation techniques as a starting point of class and later have students move to more communicative classroom tasks.

Empirical evidence on the effect of classroom pronunciation teaching practice has been provided by a number of studies (Elliot, 1995; Derwing, Munro & Wiebe, 1998; Derwing & Rossiter, 2003; Jamieson & Morosan, 1986; Macdonald, Yule & Power, 1994; Suter, 1976; Yule, Hoffman & Damico, 1987).

Positive effects have been documented for various teaching methods and techniques on various pronunciation features. For example, Jamieson and Morosan (1986) proposed a framework for successful pronunciation instruction and showed its effectiveness for the acquisition of a pair of contrastive English sounds by French learners. Elliot (1995) found that a multimodal teaching methodology effective in improving certain Spanish sounds in American SSL learners. Derwing, Munro, and Wiebe (1997) showed that long-term ESL individuals’ pronunciation improved significantly in terms of intelligibility, accentedness and comprehensibility in a program emphasizing global production skills.

However, some research has found few positive effects, or even a negative effect of pronunciation instruction under certain circumstances. Suter (1976) measured 61 non-native speakers of English on 20 variables suspected of displaying significant relationships to pronunciation accuracy using small interviews, questionnaires and psychological tests. The subjects’ oral performance was rated by 14 English native speakers for overall pronunciation quality. Surprisingly, Suter found the variable of total amount of formal classroom training in English had a negative correlation with the pronunciation scores. It appeared that the more total formal training on pronunciation a speaker had had, the less accurate the pronunciation tended to be. However, Suter argued that since all subjects had had some training before the
investigation, the variable actually measured the relationship of much versus little training instead of some versus none. Therefore he suggested that formal training under certain limits is essential, but beyond the limit is simply unproductive. Also, the study measured only the amount of training. The quality of training received was not considered and could have had a different correlation with pronunciation accuracy.

Yule, Hoffman and Damico (1987) studied some ESL students’ performance in a phoneme discrimination task before and after 8 weeks and 15 weeks of pronunciation training. They found that more than half of the subjects had actually decreased scores after 8 weeks of training, while their self-monitoring ability improved noticeably. The same subjects improved their performance after 15 weeks of pronunciation training. Therefore they argued that “there is a complex interaction over time between simply identifying a sound contrast and being confident that the identification is accurate.” (p. 768)

Macdonald, Yule and Powers (1994) compared learning outcomes under four instructional conditions- no instruction, teacher correction, self-study and interactive modification. The three groups of students’ oral productions were assessed before and after a single experimental treatment for the quality of some target words and phrases in the field of metrical systems. All four conditions yielded similar results. Therefore little positive effect was found to favor any type of instruction.

Derwing, Munro and Wiebe (1998) and its continued research by Derwing and Rossiter (2003) compared learning outcome of 48 ESL learners under different types of pronunciation instruction. Their focus was what to teach instead of how to teach. The three groups of subjects received segmental instruction, global (general speaking habits and prosodic factors) instruction and no specific pronunciation instruction, respectively. The
students’ oral production was rated by 48 native speakers of English on a 9-point scale for comprehensibility, accentedness and fluency. The results showed that, for a sentence reading task, both the segmental and global groups improved in terms of comprehensibility and accentedness. However, for a narrative task, only the global group improved in terms of comprehensibility and fluency. Although the segmental group made significantly fewer phonological errors than the global group at Time 2, it did not seem to have contributed to the raters’ judgment of comprehensibility of their pronunciation. In addition, the raters’ overall impressions for the speech samples showed that their attention was mostly drawn to prosodic problems instead of phonological errors, indicating the essential role of suprasegmentals in judgments of comprehensibility. Therefore they argued that prosody instruction should be more emphasized than segmental instruction.

In sum, the past literature has documented the likely effects of teacher-led classroom instruction on various aspects of pronunciation using various approaches. In the present study, the effects of classroom instruction are re-examined through a comparison with software-integrated instruction. Two groups of students were provided classroom instruction on both segmental and suprasegmental features, with a main focus on segmental features, through the classroom teaching methodology proposed by Celce-Murcia et al. (1996). However, one group of students received classroom instruction for half of the class time, while the other half was provided by software-based instruction. Comparison between the learning outcomes of the two groups provided evidence that both types of instruction were equivalent.

In the next section, I talk about the role of computer technology in today’s pronunciation teaching.
2.2 Computer-Assisted Pronunciation Teaching

Computer-assisted pronunciation teaching, or CAPT, has gained great attention from teachers and researchers since several decades ago (Molholt, 1988). What computer technology can bring into pronunciation teaching and the effectiveness of these contributions has been a main focus of research in the past decade or so.

In this section, I first review some research on the important features of pronunciation software applications. Then I review some of the various aspects of pronunciation found to be improvable with the help of computer technology.

2.2.1 The role of computer technology in pronunciation instruction

There are many things computers are able to do for pronunciation teaching: “computers can provide learners individualized instruction, frequent practice through listening discrimination and focused repetition exercises, automatic visual support that demonstrates to learners how closely their own pronunciation approximates model utterances” (Levis, 2007, p. 184). However, “the central issue in CAPT is the provision of adequate feedback…the most common methods are visualization and through automatic speech recognition” (p. 190).

Three kinds of well-known visual displays that have been advocated for a long time are spectrograms, waveforms, and pitch tracings, (Munro & Derwing, 1995; Munro et al., 1998; Derwing & Rossiter, 2003) although pitch tracings are considered more useful than the other two due to a relative ease in interpreting the displays (Levis, 2007).

Research using pitch display includes Hardison’s (2004) examination of the effectiveness of computer-assisted prosody training and its generalization to segmental
accuracy and lexical recall. She carried out two experiments in this study. In experiment 1, a real-time computerized pitch display is used in a 3-week training of French prosody training for some English speaking learners. One characteristic of the research is that native French speakers’ sentences were used as feedback to the subjects’ initial production. The subjects’ pre- and posttest productions were rated for the prosody and segmental accuracy and improvements were identified in both areas. Experiment 2 involved a memory recall task, and the result showed that the subjects’ lexical memory was also improved by their prosodic memory built through the training. Not only did pitch play was proved to be helpful to suprasegmental training in this study, but the positive effect of suprasegmental training on segmental and lexical was also identified.

Although visual displays such as spectrograms are argued to be less useful, their effectiveness is still supported by some research findings. For example, Coniam (2002) described a method for sensitizing trainee English-language teachers from Hong Kong to suprasegmental phonological features in English, particularly the concept of ‘stress timing’. The quasi-authentic spoken material of both Hong Kong English and American English drawn from a local TV show was analyzed through spectrograms to show the teachers the difference between the syllable-timed Hong Kong English and stress-timed American. The feedback from the teachers showed that they understood the relationship between a staccato rhythm and the concepts of stress and syllable timing better with the help of spectrograms. However, this study had a unique group of subjects (language teachers) instead of learners. It could be argued that this group had a different language knowledge background than the language learners CAPT is usually targeted at.
The other type of feedback-ASR has been even a more attractive topic for pronunciation teachers and researchers. “The central question in CAPT feedback is whether ASR can effectively provide immediate feedback that allows learners to know which parts of their pronunciation are correct and which are not” (Levis, 2007, p. 190).

Although the major concern with ASR is its accuracy for nonnative speech (Conian, 1999; Derwing, Munro, & Carbonaro, 2000), there have been findings that ASR can generate positive effects on second language pronunciation learning. For example, Neri, Cucchiarini & Strik (2006) studied a group of immigrants in the Netherlands assigned to three groups using three types of Dutch instruction: regular instruction with supplemental instruction from an ASR-based Dutch CAPT system, a CAPT system without feedback and no CAPT system. The students gave positive feedbacks to both the ASR-based CAPT systems and the CAPT system without ASR after the training. Also, the group with exposure to the ASR-based CAPT system showed the largest overall segmental improvements, followed by the group using CAPT system without ASR.

Mich, Neri and Giuliani (2006) tested the belief of many experts that a CALL (Computer-Assisted Language Learning) system based on ASR technology and on sound pedagogical guidelines could make positive contributions to classroom teaching, with the subjects being young Italian children learning English. They compared a group of pupils receiving teacher-led instruction with a group receiving ASR-based CALL for individual word pronunciation. The results showed that overall pronunciation quality of both general words and difficult/unknown words improved significantly for both groups of pupils. Therefore they argued that the system was equally effective in improving the pupils’ pronunciation.
Studies like the two described above actually used comprehensive CAPT programs in their experiments. However, the researchers’ focus was the key feature that they believed was contributing most to the efficiency of the whole CAPT system, instead of the combination of all the features of those programs that interacted with each other. There have been few studies on the general effects of using a comprehensive computer program with various features that allow a complete unit of pronunciation training activities (Seferoğlu, 2005). A detailed review of Seferoğlu’s study will be given in the last section of this chapter.

2.2.2 The effectiveness of CAPT

Past literature has documented improvements resulting from CAPT in various aspects of pronunciation. Wang and Munro (2004) trained Chinese speaking learners on three English vowel contrasts with identification tasks containing synthetic and natural speech stimuli two to three times per week for two months. The research design was unique in the way that the subjects in the experimental group determined the quantity and schedule of training. They found the subjects increased their perception performance on all the contrasts. Therefore they argued that “the techniques used in laboratory training can be successfully applied in settings in which learners participate according to their own preferences” (p. 550). However, the study examined the students’ perception only from their listening discrimination performance; it was unknown whether students would also try to make a difference when producing these contrasts due to their improved perception.

Hirata (2004) studied the efficacy of a pronunciation training program that provided fundamental frequency contours as visual feedback to English speaking learners acquiring Japanese pitch and durational contrasts. After 10 sessions’ training of Japanese sentences in
which words contrasted in pitch and duration, the subjects’ ability to perceive and produce Japanese pitch and contrasts was tested in two contexts, words in isolation and words in sentences. The results showed that the subjects improved significantly in both contexts, thus the effectiveness of the training program was proved.

Kawai and Hirose (2000) used an ASR-based CALL system to train Japanese learners on the pronunciation of Japanese double-mora phonemes. The system was able to assign the users intelligibility scores based on the users’ performance on a minimal-pair reading task and give corresponding feedback on whether a user should shorten or lengthen his pronunciation. The results showed the subjects improved significantly in making the short and long phonemes, with greater improvements on the short phonemes.

Hardison (2004) documented the effectiveness of computer-assisted prosody training and its generalization to segmental accuracy and lexical recall. Mich, Neri and Giuliani (2006) found the young learners’ overall pronunciation quality of both general words and difficult/unknown words improved significantly by an ASR-based CALL system.

However, few studies have studied the effect of CAPT programs on the general quality of the learners’ pronunciation (Seferoğlu, 2005). In the next section, I examine Seferoğlu’s (2005) study in details and discuss the need for a revised methodology to test the effectiveness of the same pronunciation software.
2.3 Seferoğlu’s Study

Seferoğlu (2005) conducted a research to study whether integrating commercial pronunciation software in an advanced English language class at the university level would result in improvements in students’ pronunciation at the segmental and suprasegmental levels.

Participants in the study were 40 students in the Department of Foreign Language Education at a Turkey university that were trained to be EFL teachers. They were assigned to two groups—the experimental group and the control group. For three weeks, the students in the experimental group used the target software Pronunciation Power in a multimedia language lab during class time while the control group followed traditional classroom instruction on the same course content. The course content included all the 52 English sounds and corresponding word or sentence level exercises available in the software. The experimental group was able to use the program on their own paces, but was required to cover everything the program offers.

The researcher used a pre- and posttest design to assess the participants’ change in pronunciation. At both tests, students were assigned different topics in relation to English teaching methodology and were required to give a ten-minute interactive presentation. The researcher rated each student’s performance while they were presenting, and another non-native speaker English teacher rated the video-recorded performance. They used a five-point scale to rate the level of communicative efficiency for each of the six elements of English sound system outlined by McDonough (1999, p. 265), including both segmental and suprasegmentals:

1. individual sounds: vowels and consonants
2. diphthongs and consonant clusters
3. linkage of sounds

4. stress pattern in polysyllabic words

5. sentence stress and rhythm, weak forms

6. intonation, and the use of varying pitch to formulate meaning and intention

The six scores were averaged to be the final score for a student. One thing to notice is that, after the pre-test the researcher met all students to discuss their performance and give specific feedback on their problems in both English segmentals and suprasegmentals.

When analyzing the data, the researcher ran an independent samples t-test for both pre- and posttest scores of the two groups. The results showed no significant differences between the two groups’ pretest scores, but showed significant differences between the two groups’ posttest scores, in favor of the experimental group.

The limitations of this study, discussed by the researcher herself, are as follows.

1. The students received individual feedback after their pretest. However, the experimental group had the chance to focus on their own problems since they were working on their own during the class; but the control group had to follow the same teacher-led instruction procedures with few chances to work on their own problems.

2. The researcher was both the teacher and one of the raters for the tests, and both the raters were non-native speakers of English. Results may have been different if the roles of teacher and rater were separated and if native speaker raters participated in the study.

3. The two raters judged data of different quality delivered through different media. This may have contributed to different results than that both raters were presented with data under the same conditions.
4. Students’ reactions to using the software were not sought in this study, which could have provided a further insight into the usefulness of the program and complemented the quantitative analysis with qualitative analysis.

5. The students involved in the study were being trained to be EFL teachers and had a high level of oral proficiency. It is necessary to conduct a research involving a different group of participants to see if the software was still effective.

However, I observed that there are some other limitations of the study. First, the study ensured the same content covered by both computer instruction and traditional teacher-led instruction, but did not provide a description of the traditional classroom teaching procedure or a discussion of the quality of the teacher instruction. There is a possibility that a disadvantaged teaching approach or dissatisfactory teaching ability of the instructor may have resulted in a poorer performance of students in the group.

Second, although the students’ performance was rated for each of the 6 elements of the English sound system, their final scores were averages of the scores for the 6 aspects. Therefore it was impossible to tell if they had only improved one aspect or all the aspects or improved to what extent for each aspect. Hence, the researcher should not have concluded that the students improved at both the segmental and suprasegmental levels, but should have only concluded there were improvements in general pronunciation quality.

Third, the results showed that the control group did not improve at all after receiving traditional pronunciation instruction. The researcher did not attempt to explain this phenomenon.

Fourth, although the researcher concluded with the general positive effect of the software on students’ pronunciation performance, she did not attempt to explore which
features of the software and the traditional classroom instruction were considered useful or not by the students.

Last, since the students in either group had no access to the other type of instruction at all, there was no way for a student to compare learning experience under the two different types of instruction.

Based on the discussion above, Seferoğlu’s (2005) study had limitations in the choices of the research subjects and pronunciation raters, the research procedures and the scope of the explanations for the results. Therefore, I would argue that her findings should be verified in another study using a revised methodology. In next chapter, I describe the methodology of the present study.
CHAPTER 3. METHODOLOGY

This chapter describes the methodology used to carry out the present study. There are four sections. The first section describes the participants in this study, including the students, the teacher and the raters. The second section presents the materials used in the study, including the traditional classroom teaching materials and the software used by the experimental group B, pretest and posttest, the rating scale used by the raters and post-class surveys. The third section describes the procedure for data collection. The last section explains the methods used to analyze data.

The whole study was integrated into my ESL teaching practicum during my study in the Master of TESL program at Iowa State University. Both experimental groups were taught by me for six weeks. Group A met twice per week in a regular classroom for 40-minute pronunciation classes; Group B met as a class only once per week, but in addition, each student had an individual 40-minute software-based pronunciation training session each week. The two groups studied exactly the same 14 segmental and 4 suprasegmental features of English pronunciation. The classroom instruction was supervised by one of the TESL professors in the department, who reviewed each lesson plan before my classes and observed the classroom teaching practice several times during the six weeks.

3.1 Participants

3.1.1 Students

There were originally 20 students in this study. These participants represented the population structure of the international graduate students at Iowa State University: their
native language background varied from Arabic to Vietnamese, however Chinese was the mother tongue of half of them; their age ranged from 24 to 35, with a mean of 30; half of them were male and half of them were female; and their academic majors varied from Agronomy to Statistics. All the students’ oral English proficiency was at about intermediate level, as assessed by the researcher subjectively from pre-research interviews. Based on their own wishes, these participants were put into the experimental group A, which followed the traditional pronunciation instruction only, or the experimental group B, which had exposure to the computer software, with an equal number of 10 participants in each group. However, after the third week of the research, 1 student from each group dropped out of the study. Therefore a total number of 18 students finished the whole process of research, 9 in each group. Information about students in each group is presented in Table 1.

3.1.2 Teacher

Both of the two groups were taught by the researcher. I am a non-native speaker of English. At the time of the study I was a second year M.A. student majoring in Teaching English as a Second Language (TESL) in the English Department of Iowa State University. Before teaching the two groups, I had taken 27 credits of TESL courses at graduate level. Teaching these two groups was also part of my teaching practicum, which was supervised by one of the TESL professors in the department.

3.1.3 Raters

The raters in this study were 6 graduate students from the Physics Department of Iowa State University, all of whom were native speakers of English. Two of them were females and the other four were males. Their age ranged from 24 to 29, with a mean of 25.
Table 1. Students in the study

<table>
<thead>
<tr>
<th>Group</th>
<th>ID</th>
<th>Gender</th>
<th>Age</th>
<th>Major</th>
<th>Native Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>C1</td>
<td>M</td>
<td>34</td>
<td>Math</td>
<td>Arabic</td>
</tr>
<tr>
<td>Control</td>
<td>C2</td>
<td>F</td>
<td>27</td>
<td>Statistics</td>
<td>Chinese</td>
</tr>
<tr>
<td>Control</td>
<td>C3</td>
<td>F</td>
<td>23</td>
<td>Statistics</td>
<td>Chinese</td>
</tr>
<tr>
<td>Control</td>
<td>C4</td>
<td>F</td>
<td>28</td>
<td>Physics</td>
<td>Chinese</td>
</tr>
<tr>
<td>Control</td>
<td>C5</td>
<td>M</td>
<td>33</td>
<td>Computer Science</td>
<td>Korean</td>
</tr>
<tr>
<td>Control</td>
<td>C6</td>
<td>M</td>
<td>40</td>
<td>Computer Science</td>
<td>Spanish</td>
</tr>
<tr>
<td>Control</td>
<td>C7</td>
<td>M</td>
<td>34</td>
<td>Computer Science</td>
<td>Spanish</td>
</tr>
<tr>
<td>Control</td>
<td>C8</td>
<td>M</td>
<td>32</td>
<td>Geology</td>
<td>Spanish</td>
</tr>
<tr>
<td>Control</td>
<td>C9</td>
<td>F</td>
<td>30</td>
<td>Agronomy</td>
<td>Vietnamese</td>
</tr>
<tr>
<td>Experimental</td>
<td>E1</td>
<td>M</td>
<td>27</td>
<td>VMPM</td>
<td>Arabic</td>
</tr>
<tr>
<td>Experimental</td>
<td>E2</td>
<td>F</td>
<td>26</td>
<td>Chemistry</td>
<td>Chinese</td>
</tr>
<tr>
<td>Experimental</td>
<td>E3</td>
<td>F</td>
<td>31</td>
<td>Genetics</td>
<td>Chinese</td>
</tr>
<tr>
<td>Experimental</td>
<td>E4</td>
<td>F</td>
<td>25</td>
<td>Material Science Engineering</td>
<td>Chinese</td>
</tr>
<tr>
<td>Experimental</td>
<td>E5</td>
<td>M</td>
<td>36</td>
<td>Chemistry</td>
<td>Chinese</td>
</tr>
<tr>
<td>Experimental</td>
<td>E6</td>
<td>M</td>
<td>30</td>
<td>Genetics</td>
<td>Chinese</td>
</tr>
<tr>
<td>Experimental</td>
<td>E7</td>
<td>M</td>
<td>24</td>
<td>Computer Science</td>
<td>Chinese</td>
</tr>
<tr>
<td>Experimental</td>
<td>E8</td>
<td>F</td>
<td>35</td>
<td>Civil Engineering</td>
<td>Thai</td>
</tr>
<tr>
<td>Experimental</td>
<td>E9</td>
<td>M</td>
<td>34</td>
<td>Agronomy</td>
<td>Thai</td>
</tr>
</tbody>
</table>

3.2 Materials

3.2.1 Traditional classroom teaching material

The words and sentences used in class for listening and reading were all from the software Pronunciation Power’s corresponding units. However, the format of paper-based exercises and the communicative games used in class were adopted mainly from three books:
Celce-Murcia, Goodwin and Brinton (1996), Lane (1993) and Hewing and Goldstein (1998). In addition, Novey and Cowin’s (1991) DVD set was used for presenting the animation of mouth movement and demonstration of example words and sentences of target sounds.

3.2.2 Pronunciation Power

The software used for Group B was Pronunciation Power (2000). The program contains 2 disks for different levels of oral English, Pronunciation Power 1 is for beginner to intermediate level students while Pronunciation Power 2 is for intermediate level to advanced level students. Both the disks cover the same topics, but Pronunciation Power 1 uses simpler words and sentences in exercises, and contains extra features such as picture-matching games. Considering the fact that they were all graduate students and their ages, I decided that Pronunciation Power 2 was more appropriate for them. It was the same choice as in Seferoğlu (2005).

Pronunciation Power 2 is organized by 52 individual sounds. For each sound, the program allowed students to do a complete unit of activities (sample screenshots are provided in Appendix A):

1. Watch the animation of tongue, jaw movement, and air flows for producing the target sound through side views of the human mouth. The animation could be watched step by step by clicking a button. Explanations of the symbols used in animation are also provided. Tips for pronunciation provided in an audio format with transcription and a glossary for the vocabulary used in the description of human mouth parts were also available.
2. Watch the demonstration through front views of the human mouth. Suggestions for pronunciation were provided in audio format with transcription.

3. View the waveform of the sound produced by a native speaker while listening to it. Record their own pronunciation and compare their own waveform with the native speaker’s in order to understand the difference.

4. Listen to sample words containing the target sound with transcription. Record their own pronunciation and compare that with the native speaker’s.

5. Listen to minimal pairs containing the target sound with transcription. Record their own pronunciation and compare with the native speaker’s.

6. Listen to sentences with transcription that embeds minimal pairs and identify which words are said.

7. Listen to sentences containing the target sound with a focus on suprasegmental features. Visual support along with the explanation of the visual symbols used was provided. Record their own reading and compare that with the native speaker’s. This feature is called S.T.A.I.R., which stands for Stress, Timing, Articulation, Intonation and Rhythm.

8. Listen to sentences containing a great amount of the target sound. Record their own reading of the sentences and compare with the native speaker’s.

One major limitation of the software material was that the suprasegmentals were not systematically put into individual units. Instead, they were built into the sentence exercises—the S.T.A.I.R. section for each sound without a rationale for associating a specific suprasegmental with a specific sound. For example, in the unit of the vowel sound /ey/, we would find exercises on sentence intonation; but in the unit of the vowel sound /iy/, we
would find exercises on sentence rhythm. It was impossible to predict where to find a certain type of suprasegmental practice.

3.2.3 Pretest and posttest material

The same picture-description task was used for both pretest and posttest. The picture used was adopted from the old testing material of SPEAK/TEACH tests, which are oral English proficiency tests that Iowa State University uses to assess the English speaking ability of its international teaching assistants. The picture was a five-frame cartoon of a humorous short story. See Appendix B for the content of the picture.

3.2.4 The rating scale

Munro and Derwing (1995), Munro et al (1998), and Derwing and Rossiter (2003) used a 9-point rating scale in their studies for rating several aspects of foreign-accented speech, especially comprehensibility and accentedness. For assessing comprehensibility, 1 point stands for the pronunciation very easy to understand, while 9 points means impossible to understand; for assessing accentedness, 1 point means no accent, while 9 points stands for very strong accent. The results of their studies have demonstrated the efficiency of the rating scales. In this study, in order to assess the students’ overall pronunciation quality, I used their 9-point scale for rating comprehensibility and accentedness to develop the rating sheets. Appendix C shows a sample of the rating sheets.

3.2.5 Post-class surveys

Post-class surveys were used to investigate the students’ reaction to the instruction. The survey for Group A contained ten questions. The questions mainly asked about three
aspects of the class they took: the helpfulness of the class as a whole, the helpfulness of each section/feature of the instruction and their comments on the advantages and disadvantages of the class.

The survey for Group B contained 12 questions. In addition to the same questions asked in Group A’s survey, they were also asked to compare the traditional classroom instruction with the software–based instruction and indicate their opinions about integrating commercial pronunciation software into traditional classrooms. Appendix D presents the two versions of post-class surveys used.

3.3 Procedures

3.3.1 Recruitment of students

The applicants for the study filled out an online registration form before having an informal interview with the researcher. In the registration form they provided basic information, indicated what English pronunciation problems they were having, and which class (Group A or B) they were interested in. The researcher then selected potential participants from the applicants so they would widely represent the international graduate students at Iowa State University and have individual interviews with them. In the informal interviews the researcher asked students about their oral English learning experience. The main purpose of the interviews, however, was to assess if the students’ oral English proficiency was at about intermediate level. After the interviews, the researched decided on the final list of participants for each group.
3.3.2 Pretest

After recruitment, the researcher set up times with each participant for individual pretest. All the individual pretest took place in the same classroom and followed the same procedure: First, the student signed the consent form with the researcher; second the researcher chatted with the student for 2-3 minutes as a warm-up activity before the pretest; third, the researcher presented the student the picture for description and gave instructions on what to do; fourth, the student was given 1 minute to prepare his/her speech; last, the researcher turned on a digital recorder, and had the student tell the story based on the picture for 2-3 minutes.

3.3.3 Pronunciation instruction

After the pretest, both Group A and B started their six-week pronunciation classes. The topics covered in the classes were chosen based on the students’ registration forms where they indicated what sounds they had trouble with, and the aspects of pronunciation Celce-Murcia, Goodwin and Brinton (1996) suggested to teach. The first four weeks covered the students’ most problematic sounds including 10 consonants and 4 vowels, while the last two weeks focused on suprasegmental aspects including word stress, sentence rhythm, linking speech and intonation. Appendix E presents the detailed syllabus.

Group A met every Monday and Thursday evening in a classroom for 40 minutes. The class procedures basically adopted the pronunciation class framework developed by Celce-Murcia, Goodwin and Brinton’s (1996): the students first received video or teacher instruction on the description of the target sounds/suprasegmentals and pronunciation rules; second, they did listening practice tasks; third, they did reading drill tasks; last, they played
various types of communication games. As introduced earlier, I was teaching the two groups also as part of my teaching practicum. The supervisor of my practicum reviewed every lesson plan before I taught a class, and she came three times respectively on the Thursday of the first, third and the fifth week to observe my control group’s class and give me feedback on various aspects of teaching. The Group A students had an average of 2.33 absences out of the 12 class sessions due to various reasons, with no student had more than 3 absences.

Group B met as a class every Monday after Group A in the same classroom for 40 minutes. The class content and procedures were exactly the same as that of Group A’s Monday class. In addition, each student in Group B had a computer-based pronunciation training session on either Wednesday or Thursday every week in lieu of a classroom session. They worked in the computer lab in the Graduate College where the target software was provided to students. Before they started to work, the teacher introduced them to the topic to work on and showed them where the unit was on the software program. For the very first time that the students used the program, they were instructed to go through every section of a unit to experience all the features of the program. However, after the first time, the students were told to use whatever feature they wanted to use in a unit. In the first four weeks, students were told not to use the feature S.T.A.I.R, because as discussed earlier, this feature practices suprasegmentals. In the last two weeks, students were told to use only this feature, and were told under which units they could find practice for a specific suprasegmental. After short teacher introduction, the students worked on their own in the computer lab for 30 minutes. After they had finished, the teacher answered whatever questions they had. The whole procedure of each computer session took around 40 minutes, just as for the regular class. The Group B students had an average of 0.56 absences out of the 12 class sessions
(including computer sessions) due to various reasons, with no student had more than 2 absences.

### 3.3.4 Posttest

On the Friday of the last week of the classes, individual students took the posttest in the same classroom as the pretest. The posttest followed exactly the same procedures as the pretest. After each student had finished their posttest, they were directed to a nearby classroom to fill out a post-class survey and leave it in an envelope.

### 3.3.5 Recording processing and rating

All the recordings in the study were saved as WMA files on a portable hardware. After the posttest, two recordings of two native speakers, 1 male and 1 female, performing the same task were added to the pool of the recordings from the pretest and the posttest. All these recordings were cut to from the beginning to be the same length of 25 seconds long, and then mixed and numbered randomly by the researcher. After putting down the number of each recording and the correct information about it, the names of the recordings were eliminated so it was impossible to tell if a recording was a pretest or posttest performance of a certain student without looking it up with the number.

Two weeks after the posttest, six raters listened to and rated all the recordings together in the library of the Physics Department of Iowa State University. Before the rating, the researcher briefly explained to them what they would listen to and what to rate. The pretest recordings of the two dropped-out students were played as examples to warm the raters up. Then after distributing the rating sheets, the researcher played the recordings one by one using an IMB laptop with external speakers. The researcher checked with the raters
before and after each recording had been played to ensure all the raters were working on the same recording. After rating, the researcher collected all the rating sheets.

3.4 Analysis

3.4.1 Analysis of the rating sheets

The researcher entered all the raw scores from the raters’ rating sheets into an EXCEL file. Using the simple Average Function, an average score from all the 6 raters for each student was calculated for comprehensibility and accentedness, respectively, and for both pretest and posttest, respectively.

With the processed rating scores, the researcher set up an appointment and met with a consultant from the Statistics Department at Iowa State University to discuss the appropriate statistical methods for comparing the score improvements of the two groups. With the help of the consultant, the following steps were made to analyze the ratings using the statistical software JMP:

1. The inter-rater reliability was calculated using a Pearson coefficient (r).

2. Because of the small sample size of the data and the purpose of this study, it was decided that a t-test was the most appropriate method to examine if there was any change in the students’ pronunciation. Therefore, a two-tailed T-test was first run on the data to compare the pretest scores and posttest scores for each group in an attempt to identify improvements.

3. However, because some data distributions seemed to have violated the normality assumption of t-test, it was decided a nonparametric test was also used. A Wilcoxon signed rank test was run on the data to confirm the results from the t-test.
3.4.2 Analysis of the post-class surveys

The post-class surveys contained three types of questions in terms of the format of the answers expected. The first type required the students to rate an aspect of their class on a 5-point rating scales; the second type required the students to select an answer from a list of choices provided; and the third type asked the students to write down their opinions in their own words.

For the first type of question, the ratings from the students were entered into an EXCEL file to calculate a group mean, using the same method used to process the ratings from the raters to generate average scores for each student.

For the second type of question, for each answer from an answer choice list, the number of times they were chosen by the students was counted, and the answer list was sorted in the order of the frequency of being chosen. In this way the degree of popularity of each answer was then clear.

For the third type of question, each student’s answer to each question was typed out and read to see if the students had similar opinions on the same topic. When similar answers were found, the frequency of the answer was noted down. In this way the degree of popularity of an opinion was then clear.

3.4.3 Identification of evidence for research questions

The raters’ rating results provided evidence for research question 1 and 2 from the listeners’ perspective, and the students’ self-evaluation from the surveys provided evidence for research question 1 and 2 from the students’ perspective. The students’ evaluation of the classes from the surveys provided evidence for research question 3.
CHAPTER 4. RESULTS

In this chapter, descriptive statistics of the raters’ ratings and summaries of the students’ survey answers are used to provide quantitative evidence for the first and second question investigating whether the pronunciation quality of the ESL learners changed after receiving traditional classroom or computer-software-integrated pronunciation instruction. Summaries and descriptions of the data in the post-class surveys are used to provide qualitative evidence for the first two questions as well as the third question which investigates the usefulness of each feature of the pronunciation instruction provided.

4.1 Research Question 1

Does the integration of the commercial pronunciation software Pronunciation Power into an ESL pronunciation class result in improvements in the general quality of students’ pronunciation as measured by the ratings of comprehensibility and accentedness?

In order to answer this question, the ratings of the pretest and posttest performance of the students in Group B were analyzed in seek of evidence. The inter-rater reliability for the ratings was calculated as Pearson coefficients (r) of 0.52 and 0.53 for the comprehensibility ratings and accentedness ratings, respectively. The mean scores for the native speakers’ recordings were 1.2 for both comprehensibility and accentedness.

The pretest scores and posttest scores of Group B were compared in an attempt to identify any difference. This comparison was first done by running a t-test on the comprehensibility scores and the accentedness scores respectively. Table 2 and Table 3 present the detailed results.
Table 2. Independent samples t-test results for the comprehensibility scores of the pretest and posttest for Group B

<table>
<thead>
<tr>
<th>Test</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>DF</th>
<th>Significance (two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>4.889</td>
<td>1.082</td>
<td>-0.883</td>
<td>16</td>
<td>0.390</td>
</tr>
<tr>
<td>Posttest</td>
<td>4.443</td>
<td>1.060</td>
<td></td>
<td></td>
<td>p&gt;0.05</td>
</tr>
</tbody>
</table>

Students in Group B had a slightly lower group mean score for comprehensibility after receiving instruction. Nevertheless, this improvement was not statistically significant. Similar to comprehensibility, although the students had a slightly lower group mean score for accentedness in the posttest, this improvement was not statistically significant either.

One concern about the data was that the distribution of the scores seemed to have violated the normality assumption of t-test. Figure 1 and 2 present the distributions of the comprehensibility and accentedness scores of Group B, respectively, in normal quantile plots). As shown in the figures, most of the residuals failed to fall closely along the straight line in the normal plots, which may have indicated non-normal distributions. In this case a nonparametric test may be a more powerful test. A Wilcoxon signed rank test was then run on the data to verify the t-test results. However, the results did not show significant difference between the pretest and posttest scores of Group B either: for comprehensibility, Z= -0.709, p = 0.479; for accentedness, Z= -0.401, p = 0.689.
In sum, the test results showed no significant improvement in the pronunciation of the group of ESL students who used the pronunciation software in the pronunciation instruction they received, as measured by the ratings of comprehensibility and accentedness.

4.2 Research Question 2

Does the group with exposure to software-based pronunciation training improve more than the group receiving traditional classroom instruction?
In order to answer this question, first, the pretest and posttest scores of Group A and B were compared in an attempt to identify any difference. The comparison was first done by running a t-test on the two groups’ comprehensibility and accentedness scores respectively. Table 4 and 5 present the results for the comparison between the pretest scores.

<table>
<thead>
<tr>
<th>Table 4. Independent samples t-test results for the comprehensibility scores of the pretest and posttest for Group B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5. Independent samples t-test results for the accentedness scores of the pretest and posttest for Group B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
</tbody>
</table>

As shown in Table 4 and 5, Group A and Group B did not have significant difference in either comprehensibility or accentedness in the pretest.

The concern that the distribution of the scores seemed to have violated the normality assumption of t-test was raised again (Figure 3 and 4 present the distributions of the two groups’ comprehensibility and accentedness pretest scores, respectively), which led to a re-analysis using a Wilcoxon signed rank test. The results confirmed the findings of t-test by showing no significant difference between the pretest scores of Group A and B: for comprehensibility, \( Z = 0.885, p = 0.376 \); for accentedness, \( Z = 0.044, p = 0.965 \).
Figure 3. Distribution of the comprehensibility scores of the pretest

Figure 4. Distribution of the accentedness scores of the pretest

Table 6 and 7 present the results for the comparison between the posttest scores of the two groups.

Table 6. Independent samples t-test results for the comprehensibility scores of the posttest for Group A and B

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>DF</th>
<th>Significance (two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4.833</td>
<td>1.110</td>
<td>-0.763</td>
<td>16</td>
<td>0.457</td>
</tr>
<tr>
<td>B</td>
<td>4.443</td>
<td>1.059</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

p>0.05
### Table 7. Independent samples t-test results for the accentedness scores of the posttest for Group A and B

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>DF</th>
<th>Significance (two-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5.537</td>
<td>0.498</td>
<td>-1.042</td>
<td>16</td>
<td>0.312</td>
</tr>
<tr>
<td>B</td>
<td>5.221</td>
<td>0.763</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(p>0.05\)

As shown by Table 6 and 7, there was no significant difference found in either comprehensibility or accentedness scores of the two groups in the posttest.

Similar to the pretest scores, the posttest scores also had the data normality issue (Figure 5 and 6 present the distributions of the two groups’ comprehensibility and accentedness posttest scores, respectively). Therefore, a Wilcoxon signed rank test was again run on the posttest scores to verify the t-test results. The results again confirmed the findings of t-test by showing no significant difference between the posttest scores of Group A and B: for comprehensibility, \(Z= -0.665, p = 0.506\); for accentedness, \(Z= -0.886, p = 0.376\).

![Figure 5. Distribution of the comprehensibility scores of the posttest](image-url)
In sum, there was no significant difference in either the pretest scores or the posttest scores of the two experimental groups, which means that the group with exposure to software-based pronunciation training did not improve differently from the group receiving traditional classroom instruction.

However, the students’ self-evaluation supplemented evidence from the raters’ perspective with evidence for the first research question from the students’ perspective. Question 3 in the post-surveys asked the students to choose answer from a list of statements that they felt true for them. The answers in the list described improvements in various aspects of pronunciation. The number of times each answer was chosen by each group of students was counted and is compared in Table 8.

As shown in Table 8, all the answers were chosen by multiple students in each group. In addition, 5 out of 9 control group students and 6 out of 9 experimental group students chose more than one answer, indicating that more than one aspect of their pronunciation was improved as a result of their self-assessment.
Table 8. Students’ answers to the post-class survey Question 3

<table>
<thead>
<tr>
<th>List of Answers</th>
<th>Frequency of being chosen</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group A</strong></td>
<td><strong>Group B</strong></td>
</tr>
<tr>
<td>a. I have more confidence in pronunciation than before.</td>
<td>2(22%)</td>
</tr>
<tr>
<td>b. I have learned things about pronunciation that I did not know before the class.</td>
<td>9(100%)</td>
</tr>
<tr>
<td>c. I can better distinguish between at least some of the English sounds in listening and speaking.</td>
<td>2(22%)</td>
</tr>
<tr>
<td>d. I can better pronounce at least some of the English sounds.</td>
<td>5(56%)</td>
</tr>
<tr>
<td>e. I have improved speaking in terms of word stress, sentence rhythm, linking words together or intonation.</td>
<td>2(22%)</td>
</tr>
</tbody>
</table>

However, out of the 5 types of improvements in terms of pronunciation, 4 had more occurrences in Group B than in Group A, with 3 of which occurred in Group B twice as many as in Group A. Therefore, I would suggest that the students seemed to like the software-integrated instruction more and consider the software-based learning experience more valuable than traditional classroom instruction.

4.3 Research Question 3

Which features of the commercial pronunciation software and the traditional classroom instruction are considered most useful or least useful by the students? Do the students prefer either of the two types of instruction?

To answer this question, I relied on the data from the students’ post-class surveys. In the post-class survey for both groups, Question 4 asked the students to rate each section of their class for helpfulness to them. For Group B, Question 4 also asked them to rate each feature of the computer software for helpfulness. All the ratings used a 5-point scale, on
which 1 = “not helpful at all” and 5 = “very helpful”. In the survey for Group B, Question 5 asked the students to compare their traditional class and the software-based training in terms of helpfulness.

For Question 4, Table 9 presents the students’ average ratings for each section of the traditional class on segmentals and suprasegmentals; Table 10 presents Group B’s average ratings for each feature of the target software.

### Table 9. Students’ ratings for the traditional class

<table>
<thead>
<tr>
<th>Class Sections</th>
<th>Average score</th>
<th>Class Sections</th>
<th>Average score</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. video of animation and human demonstration</td>
<td>4.06</td>
<td>a. teacher instruction on pronunciation rules</td>
<td>4.28</td>
</tr>
<tr>
<td>b. listening discrimination</td>
<td>4.22</td>
<td>b. paper-based small exercises</td>
<td>3.94</td>
</tr>
<tr>
<td>c. sentence reading drills</td>
<td>4.11</td>
<td>c. listening and reading drills</td>
<td>4.33</td>
</tr>
<tr>
<td>d. pair or team games</td>
<td>3.44</td>
<td>d. pair or team games</td>
<td>3.72</td>
</tr>
</tbody>
</table>

### Table 10. Group B’s ratings for the software

<table>
<thead>
<tr>
<th>Software Features</th>
<th>Average score</th>
<th>Software Features</th>
<th>Average score</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. animation</td>
<td>4.11</td>
<td>H. suprasegmental reading drills</td>
<td>4.33</td>
</tr>
<tr>
<td>b. human demonstration</td>
<td>3.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. speech analysis</td>
<td>3.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. sample words practice</td>
<td>4.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. minimal pairs practice</td>
<td>4.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. minimal pairs in sentences listening discrimination</td>
<td>4.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g. reading drills</td>
<td>4.11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
As shown in Table 9 and 10, all the features of both the traditional class instruction and the target software were rated higher than 3 points, which means they are all considered more than moderately useful by the students.

For instruction on segmentals, 3 out of the 4 (75%) sections of the traditional class were rated higher than 4 points while 5 out of the 7 (71%) features of the target software were rated higher than 4 points. The highest score was given to the listening discrimination tasks carried out in the traditional class, while the lowest score was also given to the traditional class for the communication games and activities.

For instruction on suprasegmentals, 2 out of the 4 (50%) sections of traditional class were rated higher than 4 points, while the only feature of the software that is focused on suprasegmentals was rated higher than 4 points. The highest score was given to the listening and reading drills in traditional class as well as the software’s reading drills. The communication games in traditional class were given the lowest score.

For Question 5, 4 out of the 9 (44%) students considered the traditional class and the software-based training the same in terms of helpfulness; 3 out of the 9 (33%) students thought that the traditional class was more helpful than the software-based training; 2 out of the 9 (22%) students believed that the software-based training was more helpful than the traditional class.

The main reasons the students thought the traditional class was helpful included the communicative interaction with the teacher and the classmates and the feedback from the teacher and the classmates. The main reasons the students thought the software-based training was helpful were the self-control of time and focus, and the technology enabled features such
as speech analysis, the possibility to repeat materials and to record and listen to oneself. Appendix F presents the answers from Group B students to Questions 5.

In sum, both of the traditional class and the software-based training have their own characteristics that the other type of instruction does not have, and all these characteristics were considered more than moderately helpful (above 3 points) by the students. There was no clear trend that the students preferred either type of instruction, but both types of instruction were given positive judgments from the students.
CHAPTER 5. DISCUSSION

The purpose of this study was to examine the effectiveness of integrating commercial pronunciation software into an ESL pronunciation class of learners at university level. The test results did not show any significant pronunciation improvement in either the group of students who received traditional classroom pronunciation instruction, or the group of students who had worked with the pronunciation software by themselves in half of their class time, after 6 weeks’ treatment. However, the students’ self-evaluation may have suggested that they considered the software-based learning experience more valuable than traditional classroom instruction. The analysis of the students’ post-class surveys also revealed that, although the features of the software were rated a slightly higher average score than the features of the traditional classroom, the students generally considered the traditional instruction and the software-based instruction equally helpful to them.

In this chapter, I first discuss the possible explanations for the findings, and then present the pedagogical implications of this study. Last, I discuss the limitations of this study and the suggestions for future researchers.

5.1 Explanations for the Results

This section first discusses the results for the first two primary research questions, then discusses the findings for the third research question.

5.1.1 The answers to the primary research questions

The comparison between the students’ pretest and posttest performance did not show any significant pronunciation improvement in the Group B students, who received
pronunciation instruction that integrated the use of commercial pronunciation software, Pronunciation Power. Also, the Group B students did not improve differently from the Group A students, who received traditional classroom instruction. This result did not confirm Seferoğlu’s (2005) findings using the same target pronunciation software: Seferoğlu found significant pronunciation improvement in the group of students who had received the software-based pronunciation training, as revealed by the results of a pretest and a posttest.

The main reason for the different findings may have been the different methodologies used to carry out the research:

1. The participants in this study were different from those in Seferoğlu’s study.
   a. The students in Seferoğlu’s study were EFL teachers in training and had a high level of oral English proficiency. This population was somewhat different from the general EFL/ESL learners in terms of their superior knowledge of language learning and the corresponding self-teaching/self-monitoring techniques they possessed. However, in my study, the students were all international graduate students from various academic fields with an intermediate level oral proficiency. These students had not received any special language-related training besides taking those regular ESL classes that most ESL learners are offered at school. Therefore I would argue that my students were a better representation of the general ESL learners.
   b. Seferoğlu as the researcher and the teacher in her study was also one of the two raters for the pretest and posttest. The failure to separate the roles of the teacher and the rater or to separate the roles of the researcher and the rater in a research could have led to great impact on the research results because i) the
teacher may give biased judgments on the students’ performance since he/she is already familiar with individual students’ problems and their learning progress, especially when the nature of the judgment is subjective; ii) the researcher may give biased judgments on the students’ performance since he/she has specific research goals and his/her own assumptions or hypotheses for the research. As a result, he/she may subconsciously give judgments that are consistent with his/her research hypotheses. Therefore I would argue that it was a more appropriate design to have had different raters than the teacher and the researcher in this study.

c. Seferoğlu’s study involved only two raters, both of whom were non-native speakers and EFL teachers. By contrast, the raters in this study were all native speakers who had limited exposure to accented speech and no professional training of languages. These differences in raters may have contributed to the difference in the ratings in two ways: i) because of the difference in their language learning experience, non-native speakers and native speakers of a language may have different perceptions of the communication quality of the target language regardless of their proficiency level. Therefore they may have different judgment of the same language production; ii) language teachers are trained to have more systematic knowledge and better understanding of languages than naïve listeners, therefore they may have better agreement on a speech production than naïve listeners, this is demonstrated by the high coefficient of 0.9 between the raters in Seferoğlu’s study versus the low coefficients of around 0.5 between the raters in the present study; in addition,
ESL/EFL teachers have much more exposure to foreign accented speech and hence have “better ears” for comprehending accented speech than naïve listeners. Therefore ESL/EFL teachers may have different perspective of a non-native speaker’s oral production than others. In sum, the difference in our raters may have contributed to the differences in the rating results of our studies.

2. The materials used in this study were different from that in Seferoğlu’s.
   
a. The measures of pronunciation quality used in this research were different from those in Seferoğlu’s study. Seferoğlu used a rating scale that broke down a language into segmental and suprasegmental features to assess each feature for communicative efficiency. By contrast, this study used the 9-point rating scales (Munro & Derwing, 1995; Munro et al., 1998; Derwing & Rossiter, 2003) for the general communication efficiency of the students’ speeches- the comprehensibility and accentedness perceived by the listeners. This difference in the rating objective and method may have led to the difference in the results.

b. The content and format of the pretest and posttest were different in the two studies. Seferoğlu had the students prepare a 10-minute presentation; while I had the students perform a 2-3 minutes narrative task. It did not describe clearly how much time the students in Seferoğlu’s study had to prepare their presentation and whether they used other resources to prepare; however, the students in this study had only 1 minute to go over the five frames in the picture and gave more spontaneous speech than in a prepared presentation. The difference in preparation time for the tests may have contributed to
different test results. Also, the students in Seferoğlu’s study were asked to present on a topic from their own academic field, while the students in the current study described a cartoon picture that they had never seen before. The different degree of familiarity with the topic and hence the vocabulary required by the topic may have contributed to the students’ difference in performance as well.

3. The procedures of this research were different from that of Seferoğlu’s study.
   a. The students in Seferoğlu’s study were given individual feedback after their pretest. However, the two groups could not have an equal chance to work on their own problems due to the different nature of the instruction they received. This fact may have impacted the results of Seferoğlu’s study. By contrast, the students in this study did not receive any feedback after the pretest.
   b. The treatment for the students lasted three weeks in Seferoğlu’s study, however, for how long they were trained per day was not clear in Seferoğlu’s description. In this study, students received 40-minute instruction only twice per week for 6 weeks. Therefore it was possible that the students in Seferoğlu’s study received a larger amount of instruction than the students in the current. Second, the way Seferoğlu integrated the software-based training was totally replace the traditional instruction with it for the experimental group. However, in this study I had a different interpretation of the “integration” of the software- my Group B students received software-based training for half of their total class time. Third, the quality of the traditional class instruction was also a concern in my study, while in Seferoğlu’s study
there was no discussion about instruction quality. As discussed in the third chapter, selection of teaching approach and materials was carefully conducted; regular feedback from the teaching supervisor was also used to improve my teaching process. In addition, my students were asked to reflect on the teacher’s teaching quality in the post-class surveys since I was concerned about the fact that I was just a teacher in training at the time of research. However, results showed that 13 out of 18 (72%) students considered my instruction quality similar to most of the ESL teachers they had had, while 5 out of 18 (28%) of them considered my instruction better than most of the teachers they had had; None of them thought the instruction was worse than the average level of teaching quality they perceived. Therefore, it was possible that an unsatisfactory quality of traditional classroom instruction had led to the result that the students receiving traditional classroom instruction had poorer posttest performance than the other group in Seferoğlu’s study. When the classroom instruction quality is ensured like in the current study, the learning outcome under the instruction may not be significantly worse than that under the software-integrated instruction.

c. The raters in Seferoğlu’s study rated data of different qualities delivered through different media-one rater rated real time presentations and the other rated the videotaped presentations. This may have led to a different result than when both raters were presented with data under the same conditions. In my study this problem was solved by having 6 raters rating at the same time
under the same condition. Therefore I would argue that my rating results were more reliable based on this difference in rating procedure.

However, my findings of no improvement in either group of students regardless of the different instruction they received was not unprecedented. In Seferoğlu’s study, she found no improvement for the group of students receiving traditional instruction as well, although she did not attempt to explain this phenomenon.

Similar results were also found in some other previous research on the effectiveness of pronunciation instruction. As reviewed earlier, Yule, Hoffman and Damico (1987) found that more than half of the subjects had actually decreased scores after 8 weeks of pronunciation training, with their self-monitoring ability improving noticeably. However, the same subjects improved after 15 weeks of pronunciation training.

Macdonald, Yule and Powers (1994) compared the immediate learning outcomes after one single treatment under four instructional conditions- no instruction, teacher correction, self-study and interactive modification. All four conditions yielded similar results. Therefore no positive effect was identified for any type of instruction.

Such findings of no improvement or even worse performance after immediate treatment may be explained by the Cognitive Theory. The concept of restructuring, defined by Rumelhart and Norman (1978) as a process that occurs “when new structures are devised for interpreting new information and imposing a new organization on that already stored”(p. 39), is very important in explaining the fluctuation in performance phenomenon in second language acquisition. As pointed out by Lightbown (1985), second language acquisition is not simply linear and cumulative, but also includes backsliding and loss of forms that seemed had been mastered. She explained this decline in performance by a process in which learners
have mastered some forms but then encounter new ones that require a restructuring of the whole system. Therefore one cannot assume that practice will directly and immediately accrue to skilled action, but rather accumulate as learners develop more efficient procedures (Kolers & Duchnicky, 1985). Hence, learners’ performance may be expected to “follow a U-shaped curve, declining as more complex internal representations replace less complex ones, and increasing again as skill becomes expertise” (MacLaughlin, 1987, p.152).

Although a delayed posttest in an attempt to find improvements after a period of time of instruction was missing from this study, the students’ self-evaluation in their post-class surveys may have indicated the positive effects of instruction on their internal knowledge system, which had not been reflected yet by their test performance.

The fact that the majority of the pronunciation topics covered in the six-week’s treatment was segmental may also be one main reason for no improvement in the two groups’ test results. As reviewed earlier, Derwing and Rossiter (2003) compared pronunciation teaching with differing pedagogical focus. They found that, on a narrative picture-description task, only the group receiving instruction solely on suprasegmentals improved in terms of comprehensibility and fluency. Although the group of students receiving segmental instruction made significantly fewer phonological errors than the suprasegmental group, it did not seem to have contributed to the raters’ judgment of comprehensibility of their pronunciation. In addition, the raters’ overall impressions for the speech samples showed that their attention was mostly drawn to prosodic problems instead of phonological errors, indicating the essential role of suprasegmentals in comprehensibility. Therefore, they argued that suprasegmental instruction should be more emphasized than segmental instruction in order to help students become more understandable. In the present
study, two thirds of instruction covered segmental features which the students had claimed to be their main pronunciation problems and the topics they hoped to be taught in class, with only one third of class time was spent on suprasegmental features. Therefore it was possible the results would have been different if more class time was spent on suprasegmentals instead of segmentals.

Also, as discussed earlier, different measures of pronunciation quality may have resulted in different findings. This study used the 9-point rating scales (Munro & Derwing, 1995; Munro et al., 1998; Derwing & Rossiter, 2003) for the general communication efficiency of the students’ speeches- the comprehensibility and accentedness perceived by the listeners. It is possible that after such a short time of instruction, improvements in general pronunciation quality are too small to be heard by the native speakers who were not familiar with accented speech. There is always a limit for scales, and it may be impossible to see a 5% improvement on a 9-point scale.

Besides the explanations explored above, one interesting observation at the posttest may also help explain the results to some extent. Some of the students actually appeared to be much more nervous in the posttest than in the pretest, although they were asked to perform exactly the same task in the two settings. As a result, they showed poorer fluency and made more pronunciation mistakes than their usual performance in class. I asked them for a reason, and was told that some of them were very anxious to demonstrate to themselves that they have improved their pronunciation through the efforts they had made in the 6-week class, and some others felt the pressure to demonstrate to me that they had learned and improved from the class, as I was sitting there listening to their oral performance. Therefore I
would argue that some interesting psychological issues may also have impacted the results in this study.

5.1.2 The answer to the secondary research question

The analysis of the students’ post-class surveys showed that most of the features of the both types of instruction received more than 4 out of 5 points, indicating a high level of usefulness to the students. However, there were some features that were rated below 4 points.

Surprisingly, the two lowest scores were given to the communication games played in the traditional class to practice segmental and suprasegmentals, respectively. The interactive communication tasks were highly promoted by English teachers and researchers such as Celce-Murcia et al. (1996). They suggest the use of activities such as role play, problem solving and information-gap tasks. These communication tasks from their book were used in the traditional class. Figure 7 shows an example role play activity using the two cards presented to practice the vowel sounds /iy/ and /I/ adopted from Celce-Murcia et al.’s book.

Figure 8 shows an example picture-description task for practicing linking speech.

<table>
<thead>
<tr>
<th><strong>Student A</strong></th>
<th><strong>Student B</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identity:</strong> Jean/Jim Green, student</td>
<td><strong>Identity:</strong> Mr./Mrs. Bean, physics teacher</td>
</tr>
<tr>
<td><strong>Situation:</strong> Your physics teacher, Mr./Mrs. Bean, has called you into his/her office because he/she suspects you of cheating on the quiz. You explain that you didn’t cheat. The heat was making you feel ill and you needed to take a pill.</td>
<td><strong>Situation:</strong> You have called your student Jean/Jim into your office because you suspect she/he was cheating on the physics quiz. You ask her/him to sit down and explain the situation.</td>
</tr>
</tbody>
</table>

Figure 7. Example role play cards
Here one particular part of the hypothesis that is currently being explained is examined and particular elements of that part are given careful scrutiny. However, students’ responses showed that at least some of the students considered these communication activities “fun” but not as useful as the other class sections. For example, one student commented that “we focused more on playing games itself rather than practicing pronunciation”; another student thought “playing games in class is good for young students … (because it can) attract their attention, but adults seem to dislike (such activities), or at least, think it is not very helpful, as I do.”

It was possible that the students simply did not know the purpose of these activities so they did not know what they should focus on during the activities. Although both the teacher and the supervisor of the teaching practicum were aware of the importance of the introduction of activity purpose, and made effort to practice it, it may still have not been enough to help raise the students’ awareness of the target language features during the communication process.

Another explanation for this finding may be that students were not given enough time in class to get comfortable with producing the target segmentals or suprasegmentals through
drilling practice before jumping into communication activities, hence they simply felt not ready to use these features in interaction. As suggested in the answers to the same question, the class time of 40 minutes as well as the time spent in class for oral practice was considered too short by some of the students.

Another two relatively low scores were given to the features of “human demonstration” and the speech analysis function of the pronunciation software. Figure 9 presents a screen shot for the “human demonstration” feature of the software for the consonant /f/ while Figure 10 presents the screen shot containing the features of “side view animation” and “front view of human demonstration” in one interface for the consonant /f/.

The “human demonstration” feature allowed the students to watch the video clip of sound demonstration through front views of a human mouth. Tips for pronunciation were provided in audio format with transcription through a click button above the video frame. However, as shown in Figure 10, this feature is presented with another feature—“side view animation of the sound production process” in the same interface. Sound demonstration and tips for pronunciation are also available in the animation section; in addition, the animation explained in details about lip, jaw, and tongue movement during the process of pronunciation with visual symbols representing abstract concepts of air flows and vocal vibration; it also allowed step-by-step view of the mouth movements. By contrast, the “human demonstration” feature seemed to repeat some of the functions of the animation with a lack of detailed description, which made it look unnecessary within the same interface. Therefore I guessed it was the competition of the two features with similar functions that led to the students’ devaluation of one of them.
Figure 9. Front view of human demonstration

Figure 10. Side view animation and front view of human demonstration

Figure 11 presents a screen shot for the speech analysis interface. The upper section presents a native speaker’s sound waveforms for the consonant /r/, while the lower section presents the waveforms self-recorded by the user of the same sound. Although the importance and efficiency of visual feedback have been proved in past literature (Anderson-Hsieh, 1994; Hardison, 2004; Hirata, 2004; Lambacher, 1999; Levis and Pickering, 2004; and Chun, 2007), as shown in the figure, little explanation or guidelines were provided by the software to help the users interpret the waveform displays and compare their own production.
with the native speakers. This may have made it hard for the students to understand the usefulness of this feature.

In contrast to the low scores, the two highest scores were given to the listening and reading drill section of traditional class instruction on suprasegmentals and the only feature that the software had for practicing suprasegmentals. Figure 12 presents an example of reading drill materials used in traditional class for sentence rhythm while Figure 13 presents a screen shot for the reading drills focusing on intonation.

The students’ high ratings for these two features suggest the importance and effectiveness of pronunciation teaching techniques with visual aids. It was interesting to notice that the “speech analysis” feature, which was also a type of visual feedback, was rated a low score, as discussed earlier. It could be observed that the example figures of the visual
aids for reading drills were transparent, especially with the explanation of symbols provided on the same interface in the software program. However, on the speech analysis interface, nothing more instructive than the simple guideline of “…compare the shape and sound of the waveform to the instructor’s waveform” was given. Therefore the learning potential of the speech analysis may not have been fully realized due to the lack of user guidelines for waveform reading.

Listen to the following sentences and repeat them. Pay attention to the stressed words.

. . . . . . .

1. You can see the house.
2. You can’t see the house.

Figure 12. Reading drill for sentence rhythm

It was also interesting to find no complaint about the software’s lack of additional function for practicing suprasegmentals, because before studying the post-class surveys, I
assumed the students would feel dissatisfied with the suprasegmental exercises provided by the software due to a lack of variety. However, the sole feature of the software for suprasegmental practice won the highest score from the students.

In sum, as discussed in the previous chapter, although the actual average score for each feature differed, all these characteristics of both the traditional instruction and the software-based instruction were considered helpful (above 3 points) by the students. There was no clear trend that the students preferred either type of instruction to the other, but both types of instruction gained positive judgments from the students.

5.2 Pedagogical Implications

Although the results of this study failed to provide statistically significant evidence for the effectiveness of using the target commercial pronunciation software in an ESL pronunciation class, positive feedback from the students on the software still implies the usefulness of integrating computer technology into pronunciation instruction.

In terms of how to integrate the computer programs into traditional pronunciation instruction, Question 7 in Group B’s post-class survey asked the students if they think computer software like the one they had used should be used in pronunciation class. Question 7 was a multiple choice question. The available answer choices presented to the students are listed in Table 11, in the order of number of times they were chosen by the students.

All the students believed that the software could be used to help pronunciation instruction in some ways. Most of the students believed that commercial pronunciation software could be used to replace some of the traditional classroom instruction on either segmentals or suprasegmentals, or could be used as outside classroom supplemental material.
for practicing either of these two aspects of pronunciation. Only 1 out of 9 students believed that software could be used to totally replace traditional classroom instruction.

Table 11. Group B’s answers to the post-class survey Question 7

<table>
<thead>
<tr>
<th>Available Answers</th>
<th>Frequency of being chosen</th>
</tr>
</thead>
<tbody>
<tr>
<td>f. It could be used to replace SOME of the regular classroom instruction of word stress, sentence rhythm, linking and intonation.</td>
<td>6 (67%)</td>
</tr>
<tr>
<td>b. It could be used to replace SOME of the regular classroom instruction of individual sounds</td>
<td>5 (56%)</td>
</tr>
<tr>
<td>h. It could be used as outside classroom supplemental material for word stress, sentence rhythm, linking and intonation.</td>
<td>5 (56%)</td>
</tr>
<tr>
<td>d. It could be used as outside classroom supplemental material for individual sounds</td>
<td>4 (44%)</td>
</tr>
<tr>
<td>c. It could be used to TOTALLY replace a human instructor for teaching individual sounds</td>
<td>1 (11%)</td>
</tr>
<tr>
<td>g. It could be used to TOTALLY replace a human instructor for teaching word stress, sentence rhythm, linking and intonation.</td>
<td>1 (11%)</td>
</tr>
<tr>
<td>a. It should NOT be used to replace any regular classroom instruction of individual sounds</td>
<td>0</td>
</tr>
<tr>
<td>e. It should NOT be used to replace any regular classroom instruction of word stress, sentence rhythm, linking and intonation.</td>
<td>0</td>
</tr>
<tr>
<td>i. It should NOT be used as outside classroom supplemental material at all.</td>
<td>0</td>
</tr>
</tbody>
</table>

This result supports the class design for Group B in this study- the target software was used to replace half of the regular classroom instruction. However, in authentic teaching environments, teachers should more flexibly take into account of the students’ needs and the characteristics of a particular computer program to determine when, how much and which part of the classroom instruction could be replaced by the computer program. Only in this way it is possible to maximize benefits for the learners from computer software maximized
and motivate the students fully. One evidence for this argument was that the Group B students had a much better attendance than the Group A students (0.56 absences of Group B vs. 2.33 absences of Group A), as they were allowed to reschedule the time of their computer session if they could not make the original appointment in a particular week.

This argument is also supported by Wang and Munro (2004). Their research design was different from other work examining CALL tools in that their subjects in the experimental group decided for themselves the quantity and schedule of training. Their subjects as a result enjoyed the process of scheduling and attending the training sessions, and successfully increased their perception performance on all the target vowel contrasts. These findings suggested that “the techniques used in laboratory training can be successfully applied in settings in which learners participate according to their own preferences” (p. 550).

The students’ ratings of the features of traditional classroom instruction as well as the computer software also have implications for teachers and software material developers:

1. Communicative activities should be carefully designed and carried out in pronunciation class only after the students have become familiar with the target language features through enough guided practice to ensure its efficiency.

2. Visual aids are important and effective in pronunciation training, and hence should continue to be promoted as a teaching technique.

The helpfulness of computer technology enable functions such as visual feedback could only be maximized when the users are well-trained to use the functions as well as to interpret the results of using them.
5.3 Limitations

This study included certain limitations. First, the small number of subjects in this study may have affected the generalizability of the findings. Second, the subjects were put into different groups based on their own wish, hence the uneven distribution of various demographic features of the subjects in the control and Group B may have affected the results. Third, the students only received 40-minute instruction twice per week for 6 weeks. Fourth, a delayed posttest to examine the time effect on the students’ learning was missing from the research. A delayed posttest is very important for exploring whether or not allowing an extra period of time for information processing would finally result in students’ pronunciation improvements. The results from a delayed posttest could provide empirical evidence for the cognitive concepts of restructuring and U-shape learning curve, and verify the findings from previous research on pronunciation instruction with similar results for immediate posttest such as Yule, Hoffman and Damico’s study (1987). Last, because the raters in the study were all ‘naïve’ native speakers of English who had limited exposure to accented speech and no professional language training, six might not have been an ideally large number of raters to ensure the inter-rater reliability. Also, the inter-rater reliability was calculated using a Pearson correlation coefficient (r), which averages the inter-rater reliability between each pair of raters without taking into account the magnitude of the differences between raters. It was possible to assess inter-rater reliability using a more appropriate statistic. In addition, the fact that the raters rated the recordings for both comprehensibility and accentedness together after listening to each recording only once may also have affected the results. The rating of one measure might have interfered with the rating of the other.
5.4 Suggestions for Future Research

As discussed in the previous section, this study included certain. Therefore future research on similar topic should overcome these limitations by:

1) Having a larger number of participants. As suggested by the statistical consultant who helped with this research, the number of students in each experimental group was not ideal for running statistical analyses because that the distribution of data may very likely to be abnormal. However, considering the necessity of keeping class size appropriate for pronunciation classes, I would suggest that future research organize more experimental groups instead of putting larger number of subjects in one group;

2) Having an even distribution of demographic features of the subjects between groups. Because of the number of applicants, it was very hard to ensure even distribution of demographic features of subjects between groups while satisfying the subjects’ requests for joining certain experimental group in this study. However, if possible, future research should try to have even demographic feature distribution among groups to minimize research variables between groups;

3) Providing the subjects a longer period of time and a larger amount of training. The pronunciation treatment in the present study was far from intensive training. Future research could provide more frequent than twice per week pronunciation instruction and have a longer than six weeks period of pronunciation course. It is possible for research with longer training time to find improvements in students’ pronunciation right after treatment as the students’ internal information system may have finished restructuring and become able to show the difference in their language output.
4) Using a delayed posttest to examine the time effect on the results. As discussed earlier, the results from a delayed posttest after a short-term pronunciation treatment like in the present study could provide empirical evidence for the cognitive concepts of restructuring and U-shape learning curve, and verify the findings from previous research on pronunciation instruction with similar results for immediate posttest. If a study could identify students’ pronunciation change right after a treatment, a delayed posttest could examine if this change is going to sustain over time, continue to change or fall back to the original state.

5) Having a larger number of raters if they had little background in language. Also, the rating of different measures of pronunciation could be carried out separately to minimize the interference with the each other. When having multiple raters, use a more appropriate statistic than Pearson correlation coefficient (r) to assess inter-rater reliability.

In addition, future research could examine a different computer program with more advanced features such as ASR. As discussed by Levis (2007), “the central question in CAPT feedback is whether ASR can effectively provide immediate feedback that allows learners to know which parts of their pronunciation are correct and which are not” (p. 192). Also, as suggested by Neri, Cucchiarini, Strik, and Boves (2002b), “ideal systems should always include an option to provide feedback by means of ASR technology, so that the user can receive immediate information on his/her performance” (p. 458). The software in this study was able to provide students some forms of visual feedback. However, the benefits that ASR technology could bring and have brought into CAPT were not reflected in the program used in this study. Therefore I suggest future research on a computer program with a larger variety of features.
APPENDIX A. SCREENSHOTS OF THE SOFTWARE FEATURES

1. Side view animation and front view of human demonstration

2. Speech analysis
3. Sample words practice

4. Minimal pairs practice

5. Minimal pairs in sentences
6. S.T.A.I.R.- Stress, Timing, Articulation, Intonation and Rhythm

7. Sentence reading drills
APPENDIX B. PICTURE USED IN THE PRETEST AND POSTTEST
APPENDIX C. SAMPLE PAGE OF THE RATING SHEETS

English Pronunciation Rating Sheet

Please indicate your gender and age:
Your gender:  M  F  Your age: ______

After you listen to each recording, please rate it for the overall comprehensibility of the pronunciation and the degree of accentedness on a scale of 9 points.

**Recording 1**
Comprehensibility:
(very easy to understand) 1 2 3 4 5 6 7 8 9 (impossible to understand)
Accentedness:
(no accent) 1 2 3 4 5 6 7 8 9 (very strong accent)

**Recording 2**
Comprehensibility:
(very easy to understand) 1 2 3 4 5 6 7 8 9 (impossible to understand)
Accentedness:
(no accent) 1 2 3 4 5 6 7 8 9 (very strong accent)

**Recording 3**
Comprehensibility:
(very easy to understand) 1 2 3 4 5 6 7 8 9 (impossible to understand)
Accentedness:
(no accent) 1 2 3 4 5 6 7 8 9 (very strong accent)

**Recording 4**
Comprehensibility:
(very easy to understand) 1 2 3 4 5 6 7 8 9 (impossible to understand)
Accentedness:
(no accent) 1 2 3 4 5 6 7 8 9 (very strong accent)

**Recording 5**
Comprehensibility:
(very easy to understand) 1 2 3 4 5 6 7 8 9 (impossible to understand)
Accentedness:
(no accent) 1 2 3 4 5 6 7 8 9 (very strong accent)

**Recording 6**
Comprehensibility:
(very easy to understand) 1 2 3 4 5 6 7 8 9 (impossible to understand)
Accentedness:
(no accent) 1 2 3 4 5 6 7 8 9 (very strong accent)
APPENDIX D. POST-CLASS SURVEYS

Post-class survey for Group A

Dear Students,

It has been such an enjoyable time studying English pronunciation together with you! Thank you very much for participating in the class and my research! Your answers to the following questions would greatly help me evaluate the class and complete my study, so please take your time to respond to each question with as many details as possible.

1. How many absences did you have? ______

2. Generally speaking, do you think the class was helpful to you? (highlight your choice of points)

   (Not at all) 1  2  3  4  5 (very much)

3. Please highlight any statements below that you feel true about you:
   a) I have more confidence in pronunciation than before.
   b) I have learned things about pronunciation that I did not know before the class.
   c) I can better distinguish between at least some of the English sounds in listening and speaking than before.
   d) I can better pronounce at least some of the English sounds than before.
   e) I have improved my speaking in terms of word stress, sentence rhythm, linking words together or intonation.

4. How do you rate each section of the class for helpfulness? (please highlight your choice, 1 for ‘not at all’, 5 for ‘very much’)

   Classes of Individual sounds
   a) Video demonstration and animation for individual sounds  1  2  3  4  5
   b) Listen to words with similar sounds and distinguishing between them  1  2  3  4  5
   c) Listen and repeating sentences with target sounds  1  2  3  4  5
   d) Pair or team activities and games  1  2  3  4  5
Classes of word stress, sentence rhythm, linking words together and intonation

a) Teacher’s explanation and illustration of the concepts and rules with the help of handouts

b) In class paper-based exercises such as marking stressed syllables

c) Listen to words and sentences and repeat them

d) Pair or team activities and games

5. How do you compare this instructor with other English teachers you have had before in terms of the efficiency of teaching? (highlight your choice)

a) Worse than most of the teachers I have had

b) Similar to most of the teachers I have had

c) Better than most of the teachers I have had

6. What do you think are some advantages of this class? (illustrate your answer with some examples)

7. What do you think should be improved about this class? (illustrate your answer with some examples)

8. Could you please give me some other comments on the class or suggestions for future class?
9. During this 6-week class period, were you doing any other oral English practice? What kind of practice and for how long each day? (Including review of class material by yourself outside classroom)

10. Usually, how many hours per day do you speak English? (excluding our class time)

This is the end of the survey.
Thank you again and enjoy the rest of the summer!
Post-class survey for Group B

Dear Students,

It has been such an enjoyable time studying English pronunciation together with you! Thank you very much for participating in the class and my research! Your answers to the following questions would greatly help me evaluate the class and complete my study, so please take your time to respond to each question with as many details as possible.

1. How many absences did you have? ______

2. Generally speaking, do you think the class was helpful to you? (highlight your choice of points)

   (Not at all) 1  2  3  4  5 (very much)

3. Please highlight any statements below that you feel true about you:
   a) I have more confidence in pronunciation than before.
   b) I have learned things about pronunciation that I did not know before the class.
   c) I can better distinguish between at least some of the English sounds in listening and speaking than before.
   d) I can better pronounce at least some of the English sounds than before.
   e) I have improved my speaking in terms of word stress, sentence rhythm, linking words together or intonation.

4. How do you rate each section of the class for helpfulness? (please highlight your choice, 1 for ‘not at all’, 5 for ‘very much’)

   **Regular class (Monday)**
   
   *Classes of Individual sounds*
   a) Video demonstration and animation for individual sounds 1 2 3 4 5
   b) Listen to words with similar sounds and distinguishing between them 1 2 3 4 5
   c) Listen and repeating sentences with target sounds 1 2 3 4 5
   d) Pair or team activities and games 1 2 3 4 5

   *Classes of word stress, sentence rhythm, linking words together and intonation*
   a) Teacher’s explanation and illustration of the concepts and rules with the help of
handouts

1 2 3 4 5

b) In class paper-based exercises such as marking stressed syllables

1 2 3 4 5

c) Listen to words and sentences and repeat them

1 2 3 4 5
d) Pair or team activities and games

1 2 3 4 5

Computer sessions (your own time)

Classes of Individual sounds (1-4 weeks)
a) Animation of the tongue, lip, jaw movements

1 2 3 4 5

b) Human demonstration

1 2 3 4 5
c) Speech analysis (where you compare your sound wave with the native speaker’s)

1 2 3 4 5
d) Exercise- samples words for listening, repeating and self-recording

1 2 3 4 5
e) Exercise- comparative word pairs for listening, repeating and self-recording

1 2 3 4 5
f) Exercise- listen to sentences and choose the word being said

1 2 3 4 5
g) Exercise- sentences for listening, repeating and self-recording

1 2 3 4 5

Classes of word stress, sentence rhythm, linking words together and intonation (5-6)
a) Exercise- sentences for listening, repeating and self-recording

1 2 3 4 5

5. Please compare the regular class and the computer session. (highlight your choice)

a) The regular class is more helpful
   
   Why do you think so? (illustrate your answer with examples)

b) The computer session is more helpful
   
   Why do you think so? (illustrate your answer with examples)
c) They are the same to me in terms of helpfulness. (please illustrate your answer with examples)

   Why do you think so? (illustrate your answer with examples)

6. How do you compare this instructor with other English teachers you have had before in terms of the efficiency of teaching? (highlight your choice)
   
a) Worse than most of the teachers I have had
   
b) Similar to most of the teachers I have had
   
c) Better than most of the teachers I have had

7. Do you think computer software like the one you have used should be used in pronunciation classes? (highlight your choice, could be multiple choices)
   
a) It should NOT be used to replace any regular classroom instruction of individual sounds
   
b) It could be used to replace SOME of the regular classroom instruction of individual sounds
   
c) It could be used to TOTALLY replace a human instructor for teaching individual sounds
   
d) It could be used as outside classroom supplemental material for individual sounds
   
e) It should NOT be used to replace any regular classroom instruction of word stress, sentence rhythm, linking words together and intonation
   
f) It could be used to replace SOME of the regular classroom instruction of word stress, sentence rhythm, linking words together and intonation.
   
g) It could be used to TOTALLY replace a human instructor for teaching word stress,
sentence rhythm, linking words together and intonation.

h) It could be used as outside classroom supplemental material for word stress, sentence rhythm, linking words together and intonation

i) It should NOT be used as outside classroom supplemental material at all.

8. What do you think are some advantages of this class? (illustrate your answer with some examples)

9. What do you think should be improved about this class? (illustrate your answer with some examples)

10. Could you please give me some other comments on the class or suggestions for future class?

11. During this 6-week class period, were you doing any other oral English practice? What kind of practice and for how long each day? (Including review of class material by
yourself outside classroom)

12. Usually, how many hours per day do you speak English? (excluding our class time) 

This is the end of the survey.
Thank you again and enjoy the rest of the summer!
## APPENDIX E. SYLLABUS FOR THE PRONUNCIATION INSTRUCTION

<table>
<thead>
<tr>
<th>Week</th>
<th>Monday Class</th>
<th>Thursday Class or Computer session</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Course introduction/overview</td>
<td>Vowel sounds /i/ &amp; /iy/</td>
</tr>
<tr>
<td>2</td>
<td>Vowel sounds /ɛ/ &amp; / ey/</td>
<td>Consonant sounds /r/ &amp; /l/</td>
</tr>
<tr>
<td>3</td>
<td>Consonant sounds /s/ &amp; /z/</td>
<td>Consonant sounds /θ/ &amp; /ð/</td>
</tr>
<tr>
<td>4</td>
<td>Consonant sounds /ʃ/ &amp; /tʃ/</td>
<td>Consonant sounds /ʒ/ &amp; /dʒ/</td>
</tr>
<tr>
<td>5</td>
<td>Word stress</td>
<td>Sentence rhythm</td>
</tr>
<tr>
<td>6</td>
<td>Linking speech</td>
<td>Intonation</td>
</tr>
</tbody>
</table>
APPENDIX F. ANSWERS TO SURVEY QUESTION 5 (GROUP B)

Q5 Comparison of regular class and computer session

a. The regular class is more helpful. (3 votes, main reason is interaction.)

  Comment 1: I think I can learn more from the regular class than the computer session. I just don’t like the software studying materials. If you pronounce one word in class, your classmate and teacher will notice if your pronunciation is correct or not right away and they will tell you. If you keep practicing with the computer, there is possibility that you did not realize you are wrong sometimes.

  Comment 2: It gives a chance to interact with the teacher and the students through activities and questions; it is helpful to have a lot of exercises with the answers. But also computer sessions play a great role to improve my pronunciation (different words, sentences).

  Comment 3: I think that the regular class is more helpful than the computer session, because I could ask and share my own idea with teacher and among students in the class, but the computer session not.

(3 votes, main reason is interaction.)

b. The computer session is more helpful. (2 votes)

  Comment 1: Computer uses wave images to compare your sound with standard one, which shows obvious difference. Computer can show the exact location of your tongue when you pronounce a consonant/vowel. That’s very clear.

  Comment 2: In the regular class, I can only read after what the teacher read, but she didn’t clarify whether I did it right or wrong. And in the computer aid session I can repeat what the computer read, and my improved accordingly, if any.

(2 votes)

c. The two types of classes are the same in terms of helpfulness. (4 votes, main reason is interaction in regular classes and self control of time and what to do in computer sessions)

  Comment 1: Although I think team or pair activities in regular classes are fun, but I think they are not much helpful as computer’s aid. However, in regular classes I can better distinguish pronunciation than computer classes, and much more ‘alive words’, though not be intended to be the textbook. Yet from computer session, I can read, listen, and compare more examples designed for specific pronunciation. So, I would like to conclude both are equal helpful to me.

  Comment 2: In the class, teacher gave explanation and students can interact with her if have some doubts. The computer session is helpful because students can repeat, record and listen to ourselves pronunciation to make sure we pronounce correctly.

  Comment 3: The teacher can explain for individuals’ specific questions while the computer session provides chances to repeat more.

  Comment 4: Facing the computer is less stressful than facing the instructor and classmates when one is awkward at certain pronunciation. The instructor can correct you when wrong pronunciation is made.
REFERENCES


