การปรับปรุงการออกเสียงภาษาอังกฤษของผู้เรียนชาวจีนที่เรียนภาษาอังกฤษในฐานะภาษาต่างประเทศโดยการบูรณาการการเรียนรู้ภาษาใช้คอมพิวเตอร์ช่วยกับเวอโบโทโนลิซึม

นางบี้ เคอ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศิลปศาสตรดุษฎีบัณฑิต

สาขาวิชาภาษาอังกฤษศึกษา

มหาวิทยาลัยเทคโนโลยีสุรนารี

ปีการศึกษา 2557

IMPROVING THE ENGLISH PRONUNCIATION OF CHINESE EFL LEARNERS THROUGH THE INTEGRATION OF CALL AND VERBOTONALISM

Bi He

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in English Language Studies

Suranaree University of Technology

Academic Year 2014

IMPROVING THE ENGLISH PRONUNCIATION OF CHINESE EFL LEARNERS THROUGH THE INTEGRATION OF CALL AND VERBOTONALISM

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

Thesis Examining Committee

_____

(Dr. Dhirawit Pinyonatthagan)

Chairperson

_____

(Assoc. Prof. Dr. Pannathon Sangarun)

Member (Thesis Advisor)

_____

(Prof. Dr. Andrew Lian)

Member

_____

(Asst. Prof. Dr. Harald Kraus)

Member

_____

(Asst. Prof. Dr. Arjuna Chaiyasena)

Member

_____

(Prof. Dr. Sukit Limpijumnong)

Vice-Rector for Academic Affairs and Innovation

_____

(Dr. Peerasak Siriyothin)

Dean of Institute of Social Technology

บี้ เคอ : การปรับปรุงการออกเสียงภาษาอังกฤษของผู้เรียนชาวจีนที่เรียนภาษาอังกฤษในฐานะภาษาต่างประเทศโดยการบูรณาการการเรียนรู้ภาษาใช้คอมพิวเตอร์ช่วยกับเวอโบโทโนลิซึม (IMPROVING THE ENGLISH PRONUNCIATION OF CHINESE EFL LEARNERS THROUGH THE INTEGRATION OF CALL AND VERBOTONALISM) อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร.ปัณณธร แสงอรุณ, 278 หน้า

งานวิจัยนี้ ศึกษาการเรียนรู้การออกเสียงโดยการบูรณาการ โปรแกรมคอมพิวเตอร์เพื่อการเรียนภาษาและระบบการเรียนรู้ด้วยตนเองแบบ verbotonal ซึ่งให้ข้อมูลทางด้านสัทศาสตร์ที่ถูกต้อง (Guberina, 1972; Lian, 1980; Guberina & Asp, 1981) (ซึ่งระบบมีชื่อว่า CALL-VT ในงานวิจัยนี้) งานวิจัยนี้ ศึกษาอิทธิพลของระบบ CALL-VT ที่มีต่อ (1) การเรียนรู้การออกเสียงของนักศึกษาจีนที่เรียนภาษาอังกฤษเป็นภาษาต่างประเทศ และ (2) การพัฒนาความสามารถในการเรียนรู้การออกเสียงภาษาอังกฤษด้วยตนเองของนักศึกษาดังกล่าว นอกจากนี้ งานวิจัยนี้ยังศึกษา ความคิดเห็นของนักศึกษา และครูผู้สอนที่มีต่อระบบ CALL-VT

นักศึกษาชาวจีน ชั้นปีที่ 1 วิชาเอกภาษาอังกฤษ จำนวน 96 คน จากห้องเรียนปรกติ 2 ห้องของ มหาวิทยาลัยชิงยี่ นอร์มอล ฟอร เนชันเนลลิตี้ เข้าร่วมการวิจัย หนึ่งในห้องเรียนทั้งสอง ถูกสุ่มตัวอย่างเพื่อเป็นกลุ่มควบคุม และ อีกห้องเรียนหนึ่งถูกสุ่มตัวอย่างเพื่อเป็น กลุ่มทดลอง งานวิจัยนี้ใช้ระเบียบวิธีวิจัยแบบผสม ระเบียบวิธีวิจัยเชิงปริมาณ ถูกใช้เพื่อประเมิน ความสามารถในการออกเสียงภาษาอังกฤษของนักศึกษา ก่อนเริ่มต้นการสอน และ หลังการสิ้นสุดการสอน และการพัฒนาความสามารถในการเรียนรู้การออกเสียงภาษาอังกฤษด้วยตนเองของนักศึกษาดังกล่าว ระเบียบวิธีวิจัยเชิงคุณภาพ ถูกใช้เพื่อวิเคราะห์ข้อมูลจากบันทึกของนักศึกษา ข้อมูลจากการสัมภาณษณ์นักศึกษาและครูผู้สอน

ผู้เชียวชาญทางด้านภาษาอังกฤษชาวจีน 4 คน และผู้พูดภาษาอังกฤษเป็นภาษาแม่ ซึ่งไม่ใช่ผู้เชี่ยวชาญทางด้านภาษาอังกฤษ 4 คน ให้คะแนนการออกเสียงของนักเรียนในกลุ่มทดลอง และกลุ่มควบคุมที่บันทึกเทปไว้ก่อนเริ่มต้นการสอน และหลังการสิ้นสุดการสอน ผลการให้คะแนนของผู้ให้คะแนนทั้งสองกลุ่มแสดงให้เห็นว่านักศึกษาในกลุ่มทดลองมีความสามารถสูงกว่านักศึกษาในกลุ่มควบคุมอย่างมีนัยยะสำคัญทางสถิติ ในทุกด้านที่ได้รับการทดสอบ คือ การออกเสียงหน่วยเสียง การอ่านออกเสียงคำ การอ่านออกเสียงบทความ และการออกเสียงประโยคในการสัมภาษณ์ รายละเอียดของผลการวิจัย จากการประเมินการอ่านออกเสียงบทความ และการออกเสียงประโยคในการสัมภาษณ์ โดยผู้ประเมินซึ่งพูดภาษาอังกฤษเป็นภาษาแม่ แสดงให้เห็นว่านักศึกษาในกลุ่มทดลองมีความสามารถสูงกว่า นักศึกษาในกลุ่มควบคุมอย่างมีนัยยะสำคัญทางสถิติ

ทางด้านการออกเสียง ความสามารถของผู้เรียนในการทำให้ผู้ฟังสามารถเข้าใจข้อความที่ผู้เรียนพูด และความคล่องแคล่วในการพูด

นอกจากนี้ ระบบ CALL-VT ยังน่าสนใจช่วยในการเรียนรู้ และทำให้การเรียนรู้สนุกสนาน และระบบนี้ ยังส่งเสริมการพัฒนาความสามารถในการเรียนรู้ด้วยตนเองของนักศึกษา

ผลการวิจัย ซึ่งประสบความสำเร็จอย่างสูง แสดงให้เห็นว่าระบบ CALL-VT มีประสิทธิภาพในการส่งเสริมการเรียนการออกเสียง ยิ่งกว่านั้น เมื่อเปรียบเทียบกับระบบการสอนแบบเดิม ระบบ CALL-VT เป็นระบบที่ดีกว่า เนื่องจากเป็นระบบที่ให้ประโยชน์ ไม่เพียงแต่ด้านการพัฒนาการออกเสียง แต่ยังให้ประโยชน์ต่อการพัฒนาทักษะด้านอื่นๆ ด้วย ผลการศึกษาของงานวิจัยนี้มีนัยสำคัญทั้งด้านทฤษฎีการสอนการออกเสียงภาษาอังกฤษ และการสอนภาษาอังกฤษเป็นภาษาต่างประเทศในประเทศสาธารณรัฐประชาชนจีนโดยรวม

สาขาวิชาภาษาต่างประเทศ ลายมือชื่อนักศึกษา_____

ปีการศึกษา 2557 ลายมือชื่ออาจารย์ที่ปรึกษา_____

BI HE : IMPROVING THE ENGLISH PRONUNCIATION OF CHINESE EFL LEARNERS THROUGH THE INTEGRATION OF CALL AND VERBOTONALISM. THESIS ADVISOR : ASSOC. PROF. PANNATHON SANGARUN, Ph.D., 278 PP.

PRONUNCIATION/CALL/VERBOTONALISM/AUTONOMOUS LEARNING/ CHINESE EFL LEARNERS/ RHIZOMATIC LEARNING

The present study investigated pronunciation learning by innovatively combining a CALL-based (Computer Assisted Language Learning) autonomous structure with the verbotonal system (Guberina, 1972; Lian, 1980; Guberina & Asp, 1981) of corrective phonetics (hence the name CALL-VT). It examined the effects of the CALL-VT system on Chinese EFL students in relation to pronunciation learning. In addition, it investigated the students' and teacher's perceptions towards the system and the development of students' autonomous learning.

Two intact class groups, 96 Chinese first-year English majors taking an English pronunciation course at Xingyi Normal University for Nationalities, participated in the study. One group was randomly designated as the control group and the other the experimental group. A mixed method design was employed: a quantitative framework was used to assess the students' English pronunciation at the beginning and the end of the pedagogical intervention as well as students' perceptions

and their autonomy development, while a qualitative framework was used to analyse the students' and teacher's interviews and student diaries.

Four Chinese experts on English and four naïve native speakers of English rated the recordings of both the experimental and control groups in the pre- and post-tests. Consistent rating results showed that the experimental group significantly outperformed the control group in all areas tested: phonemes, word-reading, passage-reading, and oral interview. More detailed findings from the native speakers in passage-reading and oral interviews revealed that the experimental group significantly outperformed the control group in pronunciation, comprehensibility and fluency.

In addition, the CALL-VT system was considered interesting, helpful, and enjoyable and it contributed to the development of students' autonomy.

The promising results indicate that the CALL-VT system was effective in promoting pronunciation learning. Furthermore, compared to the traditional approach, the CALL-VT system offers a better alternative since it benefited not only pronunciation but also other skills. The findings have theoretical and practical implications for the teaching of pronunciation and the development of comprehensibility and fluency in general and the teaching and learning of English as a Foreign Language in general contexts as well as in Chinese contexts.

School of Foreign Languages Student's Signature_____

Academic Year 2014 Advisor's Signature_____

ACKNOWLEDGEMENTS

This study would not have been completed without the contribution of numerous people. I would like to thank each of them here.

First and foremost, I wish to thank my advisers, Associate Professor Dr. Pannathon Sangarun and Professor Dr. Andrew Lian, for their intellectual support, encouragement, and enthusiasm. I had a countless number of consultation sessions with Professor Dr. Andrew Lian, which usually lasted several hours, sometimes deep into the night. It was him who ushered me into the field of Verbotonalism and his unique supervising style and painstaking efforts have enabled me to grow as a researcher.

I owe deepest thanks to my thesis examination committee: Dr. Dhirawit Pinyonattagarn (who served as chair), Assistant Professor Dr. Harald Kraus, and Assistant Professor Dr. Arjuna Chaiyasena. I'd also like to thank Associate Professor Dr. Kanit Khaimook for his help with statistical procedures. Without their valuable suggestions, my thesis would be of much poorer quality.

My grateful thanks also go to Associate Professor Dr. Anchalee Wannaruk (Chair of the School of Foreign Languages), Associate Professor Dr. Channarong Intraraprasert, Dr. Sirinthorn Seepho and many other faculty and staff members of the School of Foreign Languages, Suranaree University of Technology. They have always been kind and supportive during my study in Thailand.

Special thanks are also due to the Thai and Chinese experts and raters for their help with the tedious work of data collection and rating. They are Professor Dr. Xingbin

Tian, Mr. Kenneth Ray Wingate, Mr. Jonathan McClelland, Ms. Anegelina Loverde, Ms. Ting Chen, Ms. Jing Ai, and Ms. Lei Liu.

I would also like to thank the students in Xingyi Normal University for Nationalities. The experiment in this study could not have been conducted without their kind participation.

I also need to extend my deepest gratitude to my home university - Xingyi Normal University for Nationalities - for its financial support. Thanks also go to my colleagues – Professor Panshi Wei, Associate Professor Ankun He, and Associate Professor Meng Lou, for their support and encouragement, and those who shouldered my job obligations during my absence.

Special thanks also go to all my Chinese and international friends at Suranaree University of Technology for their help and encouragement. Especially, I would like to thank my former teacher Associate Professor Baoya Zhang for his language polishing work; Dr. Watinee Suntara for her friendship and abstract translation; Miss Suchada Chai for her help on many occasions; and Dr. Xuyang Liu and Mr. Fu Tian for their friendship. These friends and many others not listed here have made my life in SUT much easier and enjoyable.

Most of all, I feel deeply indebted to my father Mr. Jiaxin He, my mother Mrs. Jichun Yong, and my husband Mr. Longkun Yang. Very special thanks go to my cherished son Canning Yang for his love and bravery while his mother studied abroad. Without their sincere support, I would not have been able to complete my studies in Thailand.

Bi HE

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

LIST OF TABLES

Table **Page**

LIST OF TABLES (Continued)

Table **Page**

LIST OF TABLES (Continued)

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF ABBREVIATIONS

ANCOVA	Analysis of Covariance
CALL	Computer Assisted Language Learning
CALL-VT	Computer Assisted Language Learning-
Verbotonalism EFL	English as a Foreign Language
IELTS	International English Language Testing System
IOC	The Index of Item Objective Congruence
L1	First Language
L2	Second Language
MOE	Ministry of Education, China
PLE	Personal Learning Environment
SEA	Somatically Enhanced Approach
SPSS	Statistical Package for the Social Sciences
XNUN	Xingyi Normal University for Nationalities
QBMP	Qianxi'nan Buyi & Miao Autonomous Prefecture
VT	Verbotonalism

CHAPTER 1

INTRODUCTION

The present study aims at enhancing the pronunciation of Chinese EFL learners by applying a verbotonal approach to pronunciation improvement embedded in a self-regulating Computer Assisted Language Learning (CALL) environment. This chapter is an introduction to the thesis and it provides a background as well as a context for the present study. It also includes a statement of the problem, the rationale and significance of the study, the purpose of the study, research questions, the scope of the study and definitions of key terms.

1.1 Background of the study

The morphological, syntactic and lexical systems of any language (not just English) are made up of thousands of items. Of these, only a relatively small percentage is required at any one time for the purpose of communication. The reverse is true of the phonological system which usually consists of no more than two or three dozen significant units called phonemes all of which are required at all times in order for communication to occur (Renard, 1975). In other words, the phonological system, while relatively small in size, is extremely important for

communication through language (Gimson & Ramsaran, 1970). Among the four dominant macro-skills (listening, speaking, reading, and writing) of language learning, speaking has been considered the most challenging skill since it involves a complex process of representing meaning (Celce-Murcia & Olshtain, 2000). Moreover, speaking is ranked as one of the most required job skills for vocational college graduates in China (Wu, 2011). English majors in teacher training universities, who are trained to be English teachers in primary or middle schools, have ambitious goals in speaking, especially with regard to pronunciation. On this basis, pronunciation is considered an essential component in the range of all possible factors contributing to good spoken communicative competence. To put it another way, pronunciation occupies a central position in speaking (Pennington & Richards, 1986) because intelligible pronunciation is vital to successful communication (Levis & Grant, 2003). Garrigues (1999) also claims that good pronunciation is the foundation of effective spoken communication. Misunderstandings, or complete lack of communication, may occur when words or sentences are inappropriately pronounced or stressed. Be that as it may, Chinese EFL (English as a Foreign Language) learners are especially weak in speaking and pronunciation (Mak, 2011; Zheng, 2010). Therefore, there is an urgent need to improve Chinese students' pronunciation ability, especially the pronunciation ability of English majors in pre-service teacher training programmes.

The current curriculum for English majors in China was published by the Ministry of Education (MOE) in 2000 (MOE, 2000). This curriculum is a

programmematic document which offers overall guidance on teaching objectives, teaching materials, teaching hours and processes for all universities in China. The basic skills of the English language are identified explicitly in this curriculum. Phonetics, which focuses on pronunciation, is one of the compulsory courses in the first year of study. However, in order to meet the needs of universities in different contexts, the curriculum allows for specialized courses to occupy between 2,000 to 2,200 hours in total over 4 years. That is, each university can set appropriate teaching hours according to the conditions applicable to a specific cohort of students. Within this structure, many universities increase the teaching hours in phonetics owing to the importance of the mastery of good pronunciation. This is especially the case in teacher training universities, whose teaching hours for phonetics courses have increased from a suggested 36 hours to an actual 72 hours.

Xingyi Normal University for Nationalities (XNUN) is located in Qianxi'nan Buyi & Miao Autonomous Prefecture (QBMP), Guizhou Province, where 33 ethnic minority groups live in a community. There are 12 academic departments and one affiliated middle school. There are 23 majors offered, 13 of which are in pedagogic specialties. As of June, 27th, 2013, 6104 full time students were enrolled on campus. Of these students, 409 were enrolled as English majors, including 194 junior college students and 215 undergraduates. Phonetics is a compulsory course for both junior college and undergraduate English majors. As mentioned above, the number of teaching hours for phonetics was increased from the suggested number of 36 hours

per semester to 72 hours spread over two semesters. This decision was taken due to the poor pronunciation of the students. Because of their poor pronunciation background, all students need to work especially hard to pass the special phonetics test which has been conducted by the annual School of Foreign Languages for more than 12 years. This is a prerequisite for a qualified graduate majoring in English. The test was originally intended for English majors, but many non-English majors have chosen to take it in order to enhance their potential to secure a good job after graduation.

Before discussing current students' specific learning environment, a quick review of the features of language learning will be necessary and will help situate the research. It is widely believed that language learning is complex in nature both in terms of its context and in terms of the learners. Research studies in various areas such as learning strategies (A. D. Cohen, 1998; O'malley & Chamot, 1990), motivation (Z. Dörnyei, 2001) and cognition (Gass et al., 2003; Schmidt, 1990) have made us aware that many factors are involved in the process of learning and that language learning should be an important focus for research. Further, a shift from teacher-centred to learner-centred classrooms (Nunan, 1999) has made students more active in the learning process.

To put the present study in context, it is necessary to reflect on students' current learning environments. Given the increasing popularity and importance of

technology in language education, more contributions to technology-supported language learning are being made in the EFL context. In China, CALL (Computer Assisted Language Learning) is becoming a standard practice in the normalisation process (Bax, 2003; Chambers & Bax, 2006; Sun & Ye, 2006; Xiao, 2007) even in ethnically diverse areas like Qian Xi'nan Buyi and Miao Autonomous Prefecture (QBMP) though there are still some obstacles to be overcome (B. He et al., 2013). In other words, CALL, rather than being a miracle cure-all or something to be feared (Murray & Barnes, 1998), is becoming normal and a common, useful tool for EFL learners with students likely to become more independent and autonomous because of technology (St. Louis, 2006).

Beyond language learning, higher education has changed greatly in its nature with the advent of new technologies. Among these technology-motivated changes which include the educational experience, the research process, institutional expenditure and academic work, one of the newest developments seems to be the realization of the importance of personal learning environments (PLEs) based on technology (Archee, 2012). Even the possible consequences and effects of PLEs have not yet been imagined by most higher education administrators (Archee, 2012), and the coming into being of new learning environments cannot be denied.

The notion of personal learning environment (PLE) can be traced back to Johnson and Liber's (2008) work as they assert that the PLE came about in the UK

and the USA as a label which recognises the application of Web 2.0 social communication technologies to education. Martindale and Dowdy (2010) posit that PLEs are an outcome of tools like Web 2.0 and social media. Scholars (e.g., Dabbagh & Reo, 2010; Dron, 2007) have realized that students are able to create, organize, and share content with social media and PLEs are built on tools and services designed to help students aggregate and share resources, participate in collective knowledge generation, and manage their own meaning making. More evidence is found in Rubin (2010) and McGloughlin and Lee's (2010) work as they claim that PLEs allow students to select tools and resources to create, organize and package learning content to learn efficiently. PLEs are inherently self-directed and place the responsibility for organizing learning on the individual (Rubin, 2010). For instance, language learning, to a great extent, has stepped into a DIY (Do-it-yourself) era, which is constructed on a "just in time", "just enough", and "just for me" basis and is encouraged by emerging technologies (A.-P Lian, 2014a; A. B. Lian, 2014b). No doubt, there is a greater need and potential for self-regulated approaches to improving English pronunciation.

These new self-managed environments also provide more possibilities for learning theorists to raise their voices. Among those voices, we find Guattari and Deleuze (2000), French philosophers whose rhizomatic theory allows multiple, non-hierarchical entry and exit points in conceptualisation of the world, and which

have a place in the conceptualisation of language learning in a CALL environment. As Lian (2004) states, a rhizomatic structure may be seen as a self-regulating structure responsive to the learners' needs. These ideas are in harmony with a language learning environment, where learners can, to a great extent, be self-managed by accessing large number of both pedagogic and authentic resources organised in ways fit for learning. In this situation, students occupy a more central position in the learning process since they can build up their personal learning environments (PLEs) in personally relevant and effective ways (Valtonen et al., 2012). It is not difficult to see that the world is more information-rich and communication-rich today than at any time in its history as noted by Lian (2011). People can easily access various resources for obtaining information. If these resources are soundly organised then people will be able to manage their own learning on the basis of convenience and preference. Under these conditions, a self-regulated approach to learning pronunciation in PLEs could be an alternative and valuable possibility for English majors in Xingyi Normal University for Nationalities (XNUN).

1.2 Statement of the problem

The major problems existing in XNUN pronunciation teaching have been identified and summarized on the basis of three sources of data. One was from previous studies on pronunciation instruction at XNUN conducted by the researcher and others (B. He et al., 2014; Yan, 2008). Another was from a statistical analysis of

the pronunciation performances of all students enrolled in this pronunciation course in the past term at XNUN (see Table 1.1). The last was from interviews, conducted by the researcher, of teachers who have been teaching the phonetics courses and of the students enrolled in it.

Difficulties in learning and teaching pronunciation are identified as follows.

1) The first category of problems included:

- The teaching objectives of the pronunciation course are ambiguous. As Yan (2008) pointed out statements such as, "to train qualified professional teachers" and "to master appropriate knowledge on phonetics" are too broad and ambiguous as instructional objectives.
- The training given to the students in class could not be used in actual communication. To be specific, the students could pronounce single vowels, consonants, words, even sentences in class but could not use them in real communication with native speakers outside class.
- The intonation that students produced seemed to consist of "strange" or "mixed accents". Another study conducted by He et al. (2014) identified problems similar to those identified in Yan's (2008) study at XNUN. Problems in pronunciation teaching and learning included:
- The teacher is the major model of students' pronunciation imitation (only a few of them use models from CD and videos from the Internet).

Problems arise if the teacher makes a mistake or when he/she pronounces a word in a non-standard way.

- There is very limited application of technology even in after class activities. Then the resources for students to get suitable information and assistance seem to be limited.
- Students fail to transfer what they learned in class into real situations of English speaking.
- Teachers regard pronunciation instruction as challenging work since they believe that they themselves need to have a good pronunciation first and that pronunciation requires a great deal more work than teaching other subjects. So, only a small number of teachers are willing to teach this course.

2) The second category of problems was identified through a statistical analysis of pronunciation performances on the pronunciation examination for English majors in the second term of 2013 by the researcher. The results show that students' performances are not satisfactory (see Table 1.1).

In this test, the maximum score is 100. Students can receive academic credit only when they score 60 or above. Otherwise, they have to re-do the exam. In the worst case, they have to retake the course if they are unable to satisfy this requirement in the second exam. Occasionally scores have been inflated to avoid mass failure

and to help students satisfy the university course requirements. This is consistent with the data analysis from the semi-structured interview with teachers in this course. The following table (Table 1.1) summarizes results.

Table 1.1 Descriptive statistics of students' pronunciation test scores

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Test scores for class 1	48	20.00	88.00	59.48	15.02
Test scores for class 2	48	22.00	92.00	63.08	14.73
Valid N (listwise)	48				

3) The third category of problems was identified by interviewing students enrolled in the course and the teachers who had been teaching the phonetics course. The data analysis produced results in line with Kelly's (2000) statements of problems in pronunciation which he describes as a *paradox*. Although both teachers and students are able to realize the importance of the area and value it, pronunciation is often neglected in practice. The reason for this might rest on confusion in relation to how to teach and learn it. The same problems as outlined above such as inappropriate use of intonation and rhythm, failure to apply what was learned in class to real contexts, and teacher-centred learning environments were identified from the interviews.

The following procedure was followed in order to collect relevant data. An interview questionnaire was constructed by the researcher based on previous studies conducted in the university (B. He et al., 2014; Yan, 2008) as well as her experience of teaching. The interview questions were then modified according to a validity check performed with experts in the field. These experts rated the relevance of each item for the purpose of the questionnaire and the appropriateness of the content areas, and checked the evaluation form by using Item-Objective Congruence Index (IOC) calculations as a validation method for the relevance of the content and the objectives of the questionnaire. The result of IOC check for the semi-structured interview (revised version) for students was 84%, and that of the teachers was 82% (see Appendix E and Appendix F). The IOC results indicate that the interview questionnaires were valid and reliable. Data were collected and analysed thematically. Results were summarized as above.

As previously mentioned, there are 33 ethnic minority groups in QBMP. Consequently, Xingyi Normal University is very much ethnically diversified. Students' various linguistic backgrounds may also contribute to their difficulties in learning English pronunciation. In a study conducted by Xie and Liu (2004) in a similar context, ethnic and linguistic diversity was identified as a potential problem.

In order to address the problems of pronunciation instruction identified at XNUN, the present study attempted to improve students' pronunciation by raising

their awareness of pronunciation phenomena in a self-regulated learning environment. The focus of activities was on improving intelligibility of pronunciation as the final objective of the course was to enhance their communication ability (Trofimovich & Isaacs, 2012; Warren et al., 2009). Since they were provided with the opportunity to build their own learning environment with technology support in a CALL environment, it was anticipated that they would also benefit significantly in terms of learner autonomy improvement and even their lifelong learning.

1.3 Significance of the study

The primary significance of the study is to fill the gap in current pronunciation training systems and provide new evidence in research on pronunciation improvement in autonomous learning environments in China as no such empirical study has been conducted so far.

Second, the teaching approach proposed emphasizes the acquisition of suprasegmental features such as intonation, rhythm, stress, and pauses in whole utterances rather than single syllables. In this perspective segmental features such as consonants and vowels are expected to develop authentically in a manner more consistent with that of a native speaker's pronunciation, and individual phonemes will be set in place automatically as a result of the training programmeme. Therefore, the training on suprasegmental features of pronunciation is fundamental and will be

prioritised. Some studies have indicated that this focus is more effective than traditional approaches. Hence the teaching of pronunciation will be changed significantly from a fixation on the accurate production of individual sounds in isolation to concentrating on the broader, more holistic, and suprasegmental features of the communicative functions of speech (Brazil et al., 1980; Jones, 1997).

Third, the findings of the study will make a valuable contribution to the development of learner autonomy. More possibilities in students' self-regulation and self-management will be identified and thus a better approach to the learning of English pronunciation might emerge.

Fourth, the findings of the study will stimulate our thinking to reflect on how learning occurs as well as the roles of both teachers and learners in language learning in such an information-rich time.

Fifth, the present study, with its focus on tone (prosody), will be especially beneficial in the Chinese context where people have a tonal first language (L1).

Finally, the findings of the study may have some pedagogical implications for other skills in language learning, even in education more generally, both in China and other parts of the world.

1.4 Purpose of the study

In order to solve the problems in pronunciation teaching in XNUN, the purposes of this quasi-experimental study are as follows:

1) To develop an approach for pronunciation learning under a CALL environment based on the verbotonal theory of perception and phonetic correction (CALL-VT);
2) To investigate the effectiveness and efficiency of CALL-VT in pronunciation learning;
3) To compare the achievements of students who follow CALL-VT and those who receive pronunciation teaching in a traditional approach;
4) To investigate the students' opinions of CALL-VT in pronunciation learning ; and
5) To examine the students' development of learner autonomy after exposure to CALL-VT.

1.5 Research questions

In order to fulfill the research purposes of the study, the following research questions were proposed:

1) Is the CALL-VT system effective for pronunciation learning? If yes, in what ways?

2) Is there a significant difference in pronunciation improvement between the experimental and the control groups? If so, what is the nature of these differences?

3) What are the students' opinions of the CALL-VT pronunciation learning system?

4) What are the teachers' opinions of the CALL-VT pronunciation learning system?

5) Is student autonomy developed through the CALL-VT pronunciation learning system? If so, in what ways and to what extent?

1.6 Scope of the study (constraints)

1) The pronunciation ability focused on in this study refers to overall proficiency of English pronunciation including both segmental and suprasegmental features in terms of phonemes, words, sentences, short passages, and natural speech in daily life.

2) The target group consisted of 96 first year English majors at Xingyi Normal University of Nationalities, China. The participants were chosen on the basis of convenience and availability. Selection of students was not randomized and learners participated on the basis of their classroom enrollment. Consequently, there were not equal numbers of male and female students. Even though an equal number is not

required in studies considering gender (e.g., Young & Oxford, 1997), different numbers of male and female students still represents a limitation. Besides, the participants of this study were first-year undergraduate English majors. Other majors and levels were not included in this study. Because of this limitation, findings from this study should be treated with caution in making generalizations about pronunciation instruction of EFL learners.

1.7 Definitions of key terms

1) Computer Assisted Language Learning (CALL): CALL is an interdisciplinary term which may include a series of activities concerning language learning supported through the use of computers or any technology. Any research or study of the applications of computers or technology in language teaching and learning can be relevant to CALL. Furthermore, CALL is not a single idea, approach, or method, but any computer-based or technology enhanced procedure or process used to support language learning.

2) EFL learners: The term EFL learners in the present study refers to English major undergraduate students at Xingyi Normal University of Nationalities (XNUN), who, at the time of this study, had already learned English as a compulsory subject for at least six years in middle school and who were enrolled in their first year of university study.

3) Filtered language programme: Normal sentences in teaching materials are filtered through a low-pass filter using audio editing software such as *Audacity*. A low-pass filter removes all sound frequencies above the filter's cut-off frequency. In verbotonal filtering, the cutoff point is set at about the upper end of the speaker's fundamental frequency (F_0). In practice, this translates to cutoff points between approximately 320 Hz and 400 Hz. This procedure removes vowel and consonant sounds, liberates the intonation and rhythm patterns and makes them more salient (A.-P Lian, 1980). This enables learners to perceive the intonation patterns more effectively without requiring them to process the words. In this way, the students can focus entirely on the pattern which is presented optimally, the auditory processing load is lightened and that gives them more possibility of grasping the features of the intonation and rhythm of the sentences studied.

4) Personal Learning Environments (PLEs): In the present study, PLEs are described as a collection of various technical tools and software, together with people, in the learners' environment, to foster self-regulated and collaborative learning (Valtonen et al., 2012). PLEs can serve as platforms for both integrating formal and informal learning and fostering self-regulated learning in higher education contexts (Dabbagh & Kitsantas, 2012).

5) Pronunciation teaching: Pronunciation teaching refers to a compulsory course for English major undergraduates in China. In the present study,

pronunciation teaching includes basic theoretical phonetic knowledge together with teaching designed to improve English pronunciation of first year English majors at XNUN.

6) Self-managed, self-directed, self-regulated learning: In this thesis, all of these terms are used interchangeably and refer to the students' ability to take charge of their own learning.

7) Self-regulated Environment: Self-regulated Environment refers to the learning environment where students can learn on their own. They can choose the learning materials according to their needs and preference. They can also learn when they want to learn. Learning can happen anywhere, anytime and in any way.

8) Verbotonal Approach: The Verbotonal system is a comprehensive and effective phonetic correction approach applicable to any language and any type of learner. The Verbotonal system was initially used for the rehabilitation of people with severe communication problems and can be applied to various communication problems or situations, principally, deafness (C. W. Asp et al., 1981). Being different from conventional approaches to pronunciation instruction, the Verbotonal approach is designed to work in a global way by focusing on intonation, stress and rhythm. In so doing it seeks to act simultaneously on all language sounds hence, ostensibly, accelerating the learning of correct pronunciation and reducing the number of individual sounds requiring individual correction.

1.8 Summary

In this chapter, the researcher has given a description of the background of the study, including a preliminary research project, in order to position it in its context. The statement of the problem, the significance and rationale of the study, the research purposes and questions, the hypotheses of the research, the scope of the study, and the key terms frequently used in the study were briefly discussed. In the next chapter, a review of theories and research studies concerning pronunciation instruction for EFL learners will be presented and will lead to the construction of the study's theoretical framework.

CHAPTER 2

LITERATURE REVIEW

This chapter offers a review of the literature related to the present study. This review is in five sections. First, it discusses the notion, trends, and problems in previous studies on pronunciation instruction. The second section describes the theory of learning, including the understanding of pronunciation learning. In the third section, the theory of verbotonalism and its practice in language learning is reviewed. Following the third part, the notion and practices of autonomous learning and relevant studies are reviewed and discussed in the fourth section. Fifth, a review of Computer Assisted Language Learning (CALL) is provided together with the rhizomatic theory of learning and current learning environments. Finally, the theoretical framework of this study emerges on the basis of the reviews of the relevant literature and the purpose of the present study. A summary of the whole chapter is then presented in conclusion.

2.1 Pronunciation instruction

2.1.1 Pronunciation instruction from a historical perspective

Scholars have looked at the notion of pronunciation from many different perspectives. Richards and Schmidt define it as "*the way a certain sound or sounds are produced*"(J. Richards & Schmidt, 2002, p. 429). Reed and Michaud (2005) regard pronunciation as an integrated system, claiming that it is an integral component of language instruction. In their view, pronunciation consists of speaking and listening (or production and perception) (Reed & Michaud, 2005). There are 26 letters in the English alphabet but there are 44 speech sounds which might be differently pronounced in different contexts, hence the difficulty involved in the process of dealing with pronunciation (Reed & Michaud, 2005). They describe aspects of pronunciation as: two types of sounds (consonants and vowels), combinations of sounds, linkages of sounds, word stress, rhythm, weak forms, sentence stress, and intonation. They also recognize that every language in the world has different varieties and different accents and "*speaking with a foreign accent is only a 'problem' if it leads to breakdown in communication"* (Reed & Michaud, 2005, p. 12). From another angle, Dalton and Seidlhofer (1994) define pronunciation in two senses: pronunciation is the production and reception of sounds or speech; pronunciation refers to acts of speaking. Starting from the establishing of our identity as individuals in society, they interpret pronunciation from a social

context view, in which many aspects contribute to the identity of people. To describe what pronunciation is in an explicit way, Kelly (2000) uses a diagram to show a breakdown of its main features (see Figure 2.1). In his interpretation, phonemes, which are units of sound, are also known as segments. Suprasegmental features are the features of speech applied to groups of phonemes. "*The features which are important in English are stress, intonation, and how sounds change in connected speech*" (Kelly, 2000, p. 3).

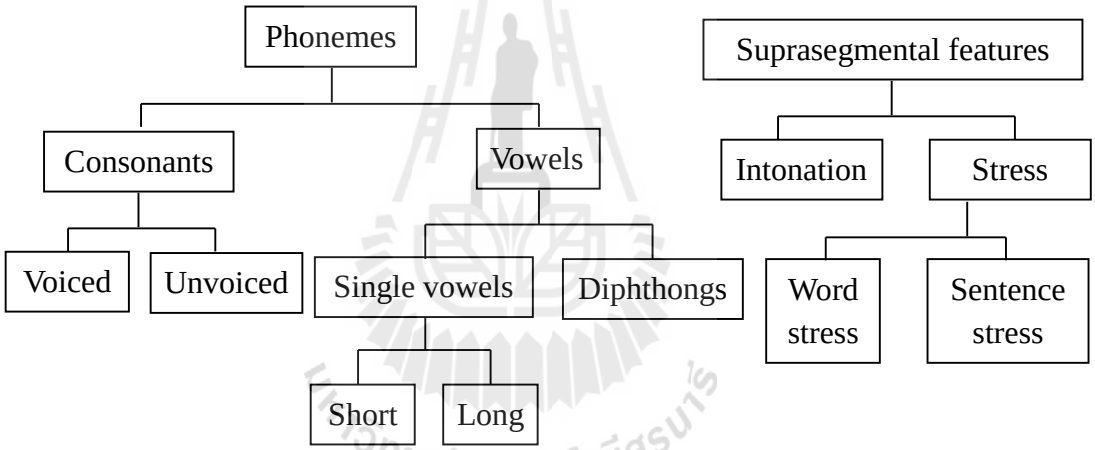


Figure 2.1 Main features of pronunciation (Kelly, 2000, p. 1)

In a different perspective, Trofimovich and Isaacs (2012) conducted a study to determine which aspects of second language speech are related to accent and which to comprehensibility. Results show that accent and comprehensibility (one measure of understanding) are overlapping yet distinct constructs; accent was uniquely related to aspects of phonology, including rhythm and segmental and syllable structure accuracy, while comprehensibility was chiefly linked to grammatical and lexical

richness. In spite of this, stress and rhythm are also closely related to comprehensibility. For instance, Munro and Derwing (1995a) noted that for a large number of listeners, comprehensibility ratings of ESL speakers' oral productions were correlated with the speakers' phonemic and prosodic errors.

Many scholars believe that pronunciation plays a crucial role in language use and language learning as the importance of pronunciation improvement in learning English has been recognized by many researchers (e.g., Renard, 1975; Warren et al., 2009) in the past decades. Furthermore, pronunciation is an important part of speaking because intelligible pronunciation is essential to successful communication (Levis & Grant, 2003). However, the position of pronunciation in language learning has either been put in a high position or simply ignored internationally (Fraser, 2000; Gilbert 2010; Levis, 2005). For example, in a key writing on foreign language teaching, Brown (2006) positioned pronunciation instruction as only a part of the teaching of listening and even worse, Su and Zhuang (1996) completely excluded pronunciation instruction from their book *Foreign Language Teaching: Theories, Practice and Approaches*. Another piece of evidence can be found in a study where the researcher recognised that pronunciation instruction was typically included as only a minor component in a speaking class (Breitkreutz et al., 2001). Pronunciation instruction has gone through ups and downs – neglected when the grammar translation approach dominated and popular in the heydays of the direct method and

audiolingualism (Jones, 1997). More evidence can be found in Celce-Murcia et al.'s (2010) book, in which they recognise that the teaching of pronunciation is largely irrelevant in some language teaching methods, such as Grammar Translation and the Reading-Based Approach. In such methods, oral communication is not a primary instructional objective. Going back to an earlier work, Kelly (1969) described pronunciation as the "Cinderella" area of foreign language teaching. He noted that Western philologists and linguists had put their eyes on grammar and vocabulary much earlier before pronunciation had been studied systematically.

Some key publications just simply omitted the issue of pronunciation, such as *The handbook of Second Language Acquisition* (Doughty & Long, 2003). Chinese scholars, Le and Han (2006)reviewed phonetic instruction from a historical perspective and commented on a teaching method named Jazz Chants. Phonetics instruction is heavily influenced by learning theories and approaches in second language learning. Under the influence of Behaviorism from the 1950s to the 1960s, pronunciation instruction emphasized differentiating phonemes. Students listened to phonemes and practiced their pronunciation intensively from smaller language units to bigger ones. The problem with this method is that students cannot produce the correct pronunciation even though they listen to the target pronunciation many times as they are not able to be aware of their mistakes: their perception is inadequate. From the 1960s to the 1970s, contrastive analysis was popular in pronunciation

instruction as the interference of the first language (L1) on the second language (L2) was widely recognized. A contrastive analysis of students' L1 and L2 compared consonants and vowels, sentence stress, liaison, rhythm, and intonation (Nunan, 2003). However, this kind of analysis of the comparison of mistakes seemed unnecessary since the interference between L1 and L 2 was limited and it was not surprising to encounter them in language acquisition (Kenworthy, 1987; Tarone, 1978). Then, communicative language teaching became dominant and has remained dominant. With these, linguists began to find that suprasegmental features were more important than segmental features for communication (e.g., Pennington & Richards, 1986; Xu, 2004). Hence, they began to take the social factors and context into consideration when discussing pronunciation instruction. Later on, theories like Focus on Form, Focus on Meaning, English as an International Language (EIL), and the Humanistic Approach influenced pronunciation instruction to some extent (Le & Han, 2006). Thus, students' emotional factors and self-concept began to be taken into consideration in pronunciation instruction (e.g.,Wen & Zhuang, 2005).

2.1.2 The goal of pronunciation instruction

What is the goal of pronunciation instruction in a foreign language? Numerous learners show their eagerness to have a native-like accent. As in Brown's (1992) research, many learners want to have a native-like accent, or at least a more authentic sounding one when they learn a language. Many other research studies

have demonstrated that learners wish to become native-like speakers whereas in fact results are not so optimistic. For example, Kenworthy (1987) says that the goal of good pronunciation may have been claimed as native-like pronunciation even though this goal will be achieved by relatively few people. Hence, the majority of learners have a very practical purpose for learning English which he terms as "comfortably intelligible" rather than pursuing a nearly impossible goal. Further, Reed and Michaud (2005) set the pronunciation goal as "let your speaking help your listening. Practice saying linked, reduced, deleted, altered, and contracted sounds in order to make sense of what you hear" (p. 15). However, they regard pronunciation as an integrated system and they emphasize error correction, which they believe to be effective in teaching pronunciation. Considerable research studies (e.g., Fathman, 1976; Krashen et al., 1982; Larsen-Freeman & Long, 1991; Oyama, 1976; Scarcella & Oxford, 1994; Seliger, 1975) have indicated that it is almost impossible for adults to attain a native-like pronunciation unless they are exposed to that language at a very young age (probably before 6). More recently, researchers (e.g., Celce-Murcia et al., 2010; Scarcella & Oxford, 1994) have come to believe that intelligibility is a more realistic objective than native-like pronunciation. This modest goal is to enable learners to surpass the threshold level so that their pronunciation will not detract from their communicative abilities. However, native-like pronunciation should not be discouraged, in particular for learners who want to have occupations like air traffic

controllers, telephone operators and English teachers. "High" goals are encouraged for them and appropriate pronunciation is necessary according to Kenworthy's (1987) understanding in his book *Teaching English Pronunciation.* However, as is evidenced in another wave of research, the teaching of pronunciation has moved substantially from emphasizing the accurate production of individual sounds to concentrating on the broader, more holistic, and suprasegmental features of the communicative function of speech (Brazil et al., 1980; Jones, 1997).

In the light of the above, the notion of *intelligibility* as a goal should be given careful consideration since it seems to be the ideal outcome for most EFL learners nowadays. When identifying the focus of instruction, it is helpful to distinguish between relevant terms such as accentedness, intelligibility, and comprehensibility (Munro & Derwing, 2006).

Kenworthy (1987) defines intelligibility as " being understood by a listener at a given time in a given situation", "the more words a listener is able to identify accurately when said by a particular speaker, the more intelligible that speaker is". It is about how closely a speaker's production approximates a native speaker's production. Derwing and Munro (1997) offered another definition as "the extent of the differences between native speaker and non-native speaker productions".

Comprehensibility refers to *"a listener's perception of how difficult it is to understand an utterance"* (Tracey M Derwing & Munro, 2005, p. 385). It is mainly

affected by the mean length of run and filled pauses (Kang, 2010). Non-native speakers' speech is less comprehensible when they speak slowly with a large number of hesitations. In addition to speech rate and pauses, sentence stress is another factor that affects comprehensibility.

Fluency, as used in this research, refers to what has been commonly referred to as the "narrow" definition of fluency (De Jong et al., 2013; Lennon, 1990). According to De Jong et al. (2013), fluency pertains to "smoothness and ease of oral linguistic delivery" (p. 395). It is possible to refine the concept of fluency further (Segalowitz, 2010) into cognitive fluency (to do with the speaker's abilities to efficiently plan and execute his/her speech (De Jong et al., 2013, p. 396)), utterance fluency (a construct with several aspects, e. g., breakdown fluency, speed fluency, and repair fluency (Skehan, 2003; Tavakoli & Skehan, 2005) and perceived fluency (the impression that listeners have of the fluency of a certain speech sample). Perceived fluency is our main focus as it is a measure of listener comfort – and our focus is on the listener. A more useful definition for perceived fluency may be found in Lennon (1990, p. 391) who defines fluency as "an impression on the listener's part that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently." While it is possible to distinguish between the three categories of fluency mentioned above, it is clear that there are strong associations between utterance fluency (which is related to cognitive fluency (De Jong et al., 2013)) and

perceived fluency (Rossiter, 2009). Thus, inevitably, the assessment of perceived fluency will be a reflection of the other two forms of fluency.

Accentedness is defined as "a listener's perception of how a speaker's accent is different from that of the L1 community" (Tracey M Derwing & Munro, 2005, p. 385). Accentedness is partially independent of comprehensibility and intelligibility (Munro & Derwing, 2006). Foreign accent affects communication as it reduces intelligibility. Accentedness is mainly determined by four variables (i.e., pitch range, stress, pause, and articulation rate). Among them, pitch is the most correlated feature, "the wider the pitch range, the less accented the NSs perceive the speaker to be" (Kang, 2010, p. 309). Kang suggests that stress plays an important role in accentedness: the more stressed words produced, the more accented the speech sounded to native listeners of English.

On the basis of the above discussion of the relationships between accentedness, intelligibility, and comprehensibility, Derwing and Munro (2006) assert that pronunciation instructors should concentrate on aspects of L2 phonology that affect intelligibility, comprehensibility and fluency rather than accentedness alone.

The present study adapts the notion of pronunciation ability offered by the new International English Language Testing System (IELTS). It consists of concrete and global features (Yates et al., 2011). Concrete features refer to segmental and prosodic features, including sounds of consonants and vowels, word stress, sentence

stress, chunking, rhythm, intonation, and speech rate, whereas global features including intelligibility, listener effort, and accentedness, are usually called oral proficiency (Tracey M Derwing et al., 2006). In this study, pronunciation ability refers to both segmental and suprasegmental features, as well as global oral proficiency which includes comprehensibility, fluency, and accentedness.

2.1.3 Methods of pronunciation instruction

Different forms of pronunciation instruction have appeared due to the different principles followed by those researching it. Pronunciation instruction has tended to be linked closely to the instructional method used in a particular era (Celce-Murcia et al., 1996). Pronunciation was almost irrelevant and ignored in the grammar-translation method. In the audio-lingual method, learners spent hours in the language lab listening to and repeating sounds. With the emergence of more holistic, communicative approaches to EFL instruction, pronunciation began to be addressed within the context of real communication (Morley, 1991).

A closer look at the approaches used in pronunciation instruction shows that they can be classified into 3 groups: articulatory; minimal pairs (binary oppositions); and technology assistance like machines, language labs, computer programmes etc. They all contribute to pronunciation learning to some extent, but none of them facilitates learners' perceptions, which is considered a key element in pronunciation learning. More evidence is revealed in the examples to be reviewed hereafter.

Celce-Murcia et al. (1996) summarize two general approaches to the teaching of pronunciation in the field of modern language teaching: the Intuitive-Imitative approach and the Analytic Approach.

In the Intuitive-Imitative approach, the availability of good models is presupposed. Sources like records, tape recording, language labs, audio- and video-cassettes, compact discs (CDs), and digital video discs (DVDs) have enhanced access to good models (Celce-Murcia et al., 2010). This approach depends on the learner's ability to listen to and imitate the rhythm and sounds of the target language. There is no intervention by way of explicit information. Problems arise when learners cannot distinguish between the sounds they hear. Logically, without this ability, they will not produce the appropriate sounds.

The Analytic Approach, on the other hand, *"utilizes information and tools such as a phonetic alphabet, articulatory descriptions, charts of the vocal apparatus, contrastive information and other aids to supplement listening, imitation, and production*" (Celce-Murcia et al., 2010, p. 2). In this approach, learners are explicitly provided with information and they focus attention on the sounds and rhythms of the target language. However, what if the learner does not know how to move his/her tongue or other articulatory organs? Such things are quite common in pronunciation learning since vocal organs are different physically from one person to another and a single articulatory description is always inadequate. Further, wrong

perceptions of the sounds will lead to serious problems in pronunciation. In the worst case, they will not be aware of their mistakes and keep pronouncing wrongly way as before.

As pointed out previously, 3 main approaches (articulatory; minimal pairs (binary appositions), and technology assistance) can provide an overall picture of all approaches in pronunciation instruction. For instance, in the Direct Method classroom, articulatory description is commonly used, the teacher (or any other resource like a recording) models a sound, a word, or an utterance, and the students just imitate or repeat. The teacher also uses a visual transcription system or charts to demonstrate the articulation of sounds (Celce-Murcia et al., 2010). And also, minimal-pair drills can be found in classroom teaching. Popular word drills like *sheep* and *ship*, *least* and *list* appear frequently in nearly every pronunciation book (e.g., Celce-Murcia et al., 2010; Christiane Dalton & Seidlhofer, 1994; Kelly, 2000; Reed & Michaud, 2005). Other forms of minimal-pair drills like syntagmatic drills (Don't sit in that seat/ Did you at least get the list?) and paradigmatic drills (Don't slip on the floor/Don't sleep on the floor) are also very popular in many approaches that Celce-Murcia et al.(2010) reviewed.

Another example can be found in Kelly's (2000) book entitled *How to Teach Pronunciation*. Kelly (2000) categorizes two key sides in pronunciation teaching, namely, the teaching of productive skills and the teaching of receptive skills. Drills

are regarded as a useful way to address both problems at once. Choral drilling, individual drilling, chaining (for sentences), open pair drilling, substitution drilling, minimal pair comparison, practicing pronunciation and spelling (e. g., homographs & homophones), listening activities, and reading activities are proposed to be carried out in an integrated pronunciation lesson (Kelly, 2000).

Even when machines or software are used, the above reviewed approaches failed to raise awareness of learners' perception. Even the followers of these approaches have realized this problem. For example, Gorjian et al. (2013) conducted a study to explore the effectiveness of a computer software programmeme named *Praat* in helping students to acquire prosodic features of the English language by visualising pitch contours. The result showed that this approach was more successful than the traditional non-CALL approaches in terms of learning prosodic features of pronunciation. However, they also became aware of what they could not do with this approach and they suggested that further work was needed to explore the effects that software-based approaches might have on language learners' perception. This idea finds earlier supportive evidence in Lian's (2004) review of *Streaming Speech,* a self-study computer package in training listening and speaking. As he concluded, *"the fundamental problem with mastering (or even competently approximating) the sounds of a foreign language depends largely on the ways in which the learner makes sense of the sound input"*. If all learners are exposed to the

same sound waves, no matter how much time they spend on this task, they may not interpret those sounds all correctly because they do not make sense of those sound waves in the same way as a native speaker. Thus, the trick to correct structuring, and therefore learning, is to make learners aware of what they have been previously unaware of. This is exactly what the CALL-VT system seeks to do in the present study.

As discussed in 2.1.2, and 2.1.3, both the goals and the methods of pronunciation teaching are beginning to pay more attention to suprasegmental features. To underpin this point, the following theoretical studies are reviewed in support of the notion of emphasizing suprasegmental features in the present research design.

Ken Lodge (2009) termed phonology as "the study of linguistic systems, especially the way in which sound represents differences of meaning in a language" (p. 8). Whereas segmental aspects of accent and pronunciation are individual sounds, also known as phonemes, (i.e., consonants and vowels), the term suprasegmental "refers to how speech sounds function and are affected at the sentence and discourse levels, i.e., over multiple sound segments that combine to make phrases, clauses, and sentences" (Celce-Murcia et al., 1996). The study of those suprasegmental features is often referred as the study of prosody, within which stress, rhythm and intonation patterns are regarded as important functions (Roach, 2010). To improve English pronunciation, teachers and practitioners in the field of English as a Foreign Language

(EFL) have paid much attention to finding effective ways of pronunciation learning. Drills and minimal pair exercises have been the magic solutions to teaching and improving learners' pronunciation for quite a long time over nearly the entire language teaching field. And this approach is still popular now among many educators. However, according to Molino's (2000) view, prosody is the first element one acquires when learning one's native language. Thus, a child will acquire the prosodic structures which include rhythmic and intonational characteristics before any other linguistic structures (phonemes, lexicon, syntax, etc.). This statement hints that prosody structures should play a more dominant role in instruction of pronunciation than they currently do.

Guberina's "Verbo-Tonal Approach" enhanced this view by using the prosodic structure of the target language as the "shell" for the improvement of pronunciation skills. When Lian (1980) stated the importance of intonation, he pointed out that different melodies in different languages "arrange" the vowel and consonant sounds of each language in a way which is specific to that language. Intonation also appears to be fundamental in other ways, such as the forming of one's own structured language and meaning systems in terms of feelings and pragmatic functions. Several studies (Brazil et al., 1980; G. Brown & Yule, 1983) are in line with the emphasis on suprasegmental features in spoken discourse by increasing awareness of their communicative function. Further, as Jones (1997) summarized,

many researchers called for a more “top-down” approach to pronunciation teaching. Here, the broader, more meaningful aspects of phonology in connected speech were emphasized rather than practicing with isolated sounds. Thus, the evidence summarised above leads to the conclusion that intonation training should come first, that it is fundamental and well-justifies in improving pronunciation (T. M. Derwing & Munro, 1997; Tracey M Derwing & Munro, 2005; Hahn, 2004) but that any remaining segmental difficulties should be given attention as well. This synthesis is consistent with the summaries about the current situation of pronunciation instruction by Celce-Murcia et al. (2010).

2.1.4 Research studies on pronunciation learning in China

Compared to other areas of language research, studies on pronunciation do not form a large percentage of research interests in language learning in China. One reason is that pronunciation is ignored as, internationally, it has been treated as a minor component in language learning (as discussed in 2.1.1). The other reason is that it is not easy to conduct an empirical study in this field. One piece of evidence for this is that not a single standard test for pronunciation can be found so far in China. Nevertheless, there have been many Chinese scholars conducting studies on pronunciation instruction (e.g., Fan & Nang, 2005; S. F. He, 1987; Luo & Zhang, 2002; Sui, 1997) over the past decades. This research ranges from teaching theories to practices in teaching methodology, from technology aids (H. X. Liu, 2012; Zhao,

2002) to comparisons between L1 and L2 (Gao, 2002). However, research on prosody and rhythm has become a new focus since the International Speech Rhythm Conference 2006 in Dresden University of Technology (Hua Chen, 2006). More and more researchers conduct studies to improve English pronunciation by emphasizing suprasegmental features in phonetics.

Most Chinese EFL pronunciation research studies have been conducted in the form of surveys (e.g., Fan & Nang, 2005; Y. Wang & Feng, 2010). They investigate learners' opinions towards pronunciation. And they survey the views of teachers to identify problems in teaching pronunciation. However, they could do no more than identify these problems. A few suggestions have been proposed and some action research has been done to improve students' pronunciation. Most of them result in some small improvement due to time on task. It is not surprising that there have been some empirical studies on how to improve English pronunciation. Among those studies, negative transfer of L1 (Chinese) to English pronunciation learning has been a hot issue for quite a long time (e.g., Wu, 2011). Some researchers have conducted studies on interference of various Chinese dialects in English pronunciation learning (e.g., Han, 2013). With the development of information technology, scholars have begun to study new methods of employing technology to assist pronunciation learning. They (e.g., Hong Chen, 2013; Sui, 1997; Zhao, 2002) made efforts to discuss ways of using technologies and their potential positive effects in

pronunciation instruction. For instance, Chen (2013) provided a background of pronunciation instruction and discussed the features of a software programmeme named *Praat*, which allows editing, annotation and transformation of speech sounds. The researcher came to the conclusion that *Praat*, with its pitch display, can help students improve their pronunciation by providing training on intonation. However, these studies did not originate with the nature of pronunciation learning (i.e., a theory of learning) and they seemed to use the "technology" as tool in teaching, on the basis of common sense assumptions rather than a coherent theory of learning.

To sum up, an insufficient number of empirical studies has been conducted on English pronunciation instruction in the Chinese context. Most studies focused on problems encountered by Chinese EFL learners, but few of them have solved those problems efficiently. None of them employed the verbotonal system and none of them situated pronunciation learning in a self-regulating CALL environment. The possible reasons for the failure of pronunciation instruction worldwide may rest in misconception in relation to learning theory. A review and theoretical discussion of learning theory is provided in the following section.

2.2 Theory of learning

Many theories of language learning have been in vogue over the years. After the popularity of the grammar-translation approach and the development of

behavioristic approaches such as the Audio-Lingual Method (ALM) (Andresen, 1990) language-teaching and learning approaches have continued to focus on language and its "ingestion" by learners. This gave rise to a strong tradition of contrastive studies with the accompanying belief that it was possible to identify problems and learning errors by contrasting the learner's mother tongue with the language to be learned. Theoretically, this list of predicted and predictable errors and problems could then be addressed in a kind of one-size-fits-all approach to foreign language learning. This was known as the strong form of contrastive analysis (Wardhaugh, 1970). It quickly became apparent that this approach simply did not work as students were producing idiosyncratic errors. This approach was then replaced by a weaker form of contrastive analysis, error analysis, which was based on observation and presented a more realistic approach to the problem of language learning. It no longer made theoretical predictions but observed and reported and constructed what later became known as "interlanguage"(Selinker & Rutherford, 2013). This purported to describe the state of linguistic knowledge of language learners. It too suffered somewhat from the same problems as the strong form of contrastive analysis in that linguists sought to identify a stable and staged model of linguistic knowledge for each linguistic group.

While these changes were occurring at the level of language study, attention started to turn towards learners as individuals. Jean Piaget's seminal work began to

capture attention (Piaget, 2013). In particular, Piaget's work argued that people were different and did not learn objects or "content". Instead, he argued that they constructed their own personal knowledge on the basis of past experience. This point of view gave rise to constructivism (Fosnot, 2013) and was reflected in language learning in the works of scholars such as Dulay, Burt and Krashen (1982). They sought to demonstrate that language acquisition was a process of "creative construction".

The intellectual basis of this research work is rooted in the notion of Piaget's view of construction of personal knowledge and verbotonalism, though developing in parallel with Piaget's studies, reflects the same concerns. In a sense, both Piaget and Guberina (verbotonalism) were postmodern scholars before the advent of postmodernism as such and no doubt contributed to postmodern thought, at least in the educational area. These principles will now be developed in an intellectual framework which will lay the foundation for the rest of this study.

2.2.1 Notion of learning

What is learning? How does learning occur? In the researcher's view, one's research interests should originate from how we understand learning and how learning occurs. No matter how one understands the notion of learning, there is a general common definition that acquiring knowledge means "acquiring new information". This kind of understanding does not take sufficient account of how knowledge is

constructed and it ignores the interaction between the self and the outside world. It reifies knowledge, i.e., knowledge is seen as an object to be internalised. Our theoretical view of learning is greatly influenced by Pierre Bourdieu's concept of Habitus (Bourdieu, 2005). In Bourdieu's view, a human being constructs him/herself in every aspect of life on the basis of their personal history. Personal history is not static since people interact with the outside world (which also keeps changing) every second. How they make sense of the world and of themselves constitutes their knowledge. And also, this is how we learn: something stimulates us, and we react to it on the basis of what makes sense to us and we change. When that happens, our constitution of knowledge changes and this is our moment of learning. That is to say, one can only learn when one can make some sense (not necessarily the correct sense) of what one is experiencing. The difficulty lies in how to raise awareness of what is not already in one's history. This is critically important because we recognize and value what is "in" our "history", whereas we exclude and ignore that which is not.

These ideas are consistent with Ania Lian and Andrew Lian's interpretation of learning. They argued that language learning required a rethink about differences between people for their unpredicted needs (A.-P Lian & Lian, 1997). In their argument, language learning is essentially about the management of meanings. Meaning is not objectively present but the result of each person's interactions with the

world. Also, by considering the interactions between people and even within oneself, the process of language learning is essentially a process of accumulating history and establishing relationships (A.-P Lian, 2014a). Thus, learning seems to be a practice-oriented and sociohistorically conditioned process. Once students' awareness is raised (something becomes meaningful), their learning can begin to occur, in a process involving complex interactions between students and the world, and also within themselves. This view is consistent with a postmodern approach to learning.

2.2.2 A postmodern view of learning

Many scholars such as Derrida, Lyotard, Baudrillard, Nietzsche, and Foucault (see Boje et al., 1995) define postmodernism from their own perspectives and those definitions make postmodernism multiple and complicated. To pick up some pieces of definitions, postmodernism is regarded as a term which "must be the most over and under-defined" (Hutcheon, 2003). It is seen in different, even conflicting ways in a great number of fields (Cova, 1996). Hutcheon (2003) argues that postmodernism is a contradictory phenomenon, one that uses and abuses, installs and then subverts. From the perspective of language learning, postmodernism differs from most approaches to learning in two fundamental ways: 1) rationality and logic are not important to attaining knowledge; 2) knowledge can be contradictory. Because of the contextual nature of knowledge, individuals can hold two completely incongruent views of one subject at the same time (Kilgore, 2001). To sum up,

postmodernism is a philosophy that advocates a complete rethink of learning, including pronunciation improvement.

An example can be found in a study on postmodernism and the Web as the researchers conclude that postmodernism illuminates thinking in the new information medium (Berthon et al., 2000). The postmodern approach to learning is founded upon the assertion that there is not one kind of learner, not one particular goal for learning, not one way in which learning takes place, nor one particular environment where learning occurs (Kilgore, 2001). In the present study, students need to be able to set their own goals and choose learning materials to improve their pronunciation in different ways in addition to classroom study because they will all construct their personal knowledge differently. This is in line with Kilgore's (2001) assertions about the postmodern view of knowledge :

> *Knowledge is tentative, fragmented, multifaceted and not necessarily rational.*
> *Knowledge is socially constructed and takes form in the eyes of the knower.*
> *Knowledge is contextual rather than "out there" waiting to be discovered.*

Hence, knowledge can shift as quickly as the context shifts, the perspective of the knower shifts, or as events overtake us. Therefore, the matter of learning is concerned with meaning making (i.e., the meaning we/the learners make). To raise awareness of the students becomes the key point in learning (pronunciation learning in our case).

In the present study, filtered language training (in the experimental group) will help to give a sense of the intonation and rhythm of English language by by-passing blockages set in place by past experience. The general pedagogic assumptions governing the work to be performed here can be summarized as follows.

- Learning and knowledge construction are essentially individual and dependent on each learner's internal logical and representational systems which, in turn, are based in part, if not entirely, on each learner's personal history (A.-P Lian, 2004).
- Learners' logical and representational systems constrain the ways in which learners are able to interpret and construct relevant new knowledge and processes, including linguistic, semiotic and communicative knowledge and processes.
- Critical to changing students' logical and representational systems is awareness-raising (sometimes also referred to in the research literature as noticing) (A.-P Lian, 1987, 2004, 2011; Schmidt, 1990). Awareness-raising can lead to a modification of learners' perceptual systems and, as a consequence, a reorganization of their logical and representational systems (i. e., their personal histories). This will be the function of digital filtering and the gesture work incorporated into CALL-VT.

In this perspective, the essential function of teaching is to help modify learners' logical and representational systems. These systems are particularly resistant to change. In this experiment, both awareness-raising and the strengthening of newly-acquired phenomena are to be achieved, inter alia, through a variety of strategies, both conscious/reflective as well as unconscious.

In order to achieve the research purposes on the basis of the researcher's understandings of the above theory of learning, a verbotonal approach embedded in a autonomous CALL environment is proposed and described in the following section.

2.3 The Verbotonal System

Verbotonalism finds its origins in an intellectual framework similar to that described above. However, verbotonalism has a special advantage in relation to this study in that it explicitly combines a postmodern view of learning with the learning of pronunciation.

In light of problems discussed in Chapter 1, the review of pronunciation difficulties encountered by EFL learners, and the above review of learning theory, it is proposed to trial an approach named Computer Assisted Language Learning-Verbotonalism (CALL-VT). Before doing so, however, a review of the theory and practice of the verbotonal system is useful.

2.3.1 Verbotonalism (General principles)

The verbotonal system finds its origins in the pioneering work of Petar Guberina who developed a new method for correcting pronunciation known as the Verbotonal system of Phonetic Correction (Guberina, 1972; Renard, 1975) in the 1950s. The Verbotonal system was initially used for the rehabilitation of people with severe communication problems, principally, deafness (C. W. Asp et al., 1981). Different from the conventional approach, this approach seeks to teach the deaf by concentrating on using people's residual hearing to hear and to speak. As Renard (1975) asserted, the problem is to teach the pronunciation of a new language to people who already have a tool for oral communication: their mother tongue. This language effectively acts as a blockage which filters any possible incoming signals. Below is a short synthesis of the features of the verbotonal system.

Firstly, the key principle and also a starting point of the verbotonalism is the notion of meaning making. The verbotonal approach (VT) considers speech production from a holistic perspective. From a global view, VT treats pronunciation as a process of several procedures. People perceive the sounds on the basis of their history which does not have to do only with their first language but many other "things" that the person's history may contain. We don't hear sounds as they are. We give those sounds the meaning we know by reconstructing them in our perception. Then, we produce the sounds by means of oral articulation.

Secondly, to underpin what is argued in the first point, each speaker is both a producer and a perceiver of speech. In introducing the VT approach, Guberina and Asp (1981) take speech as a social event. In their interpretation, when we want to say something, the "meaning" of what we speak out is transmitted "*not only by linguistic elements but also by the auditory and visual information present in the rhythm, the intonation, the loudness, the tempo, the pauses, the tension, and the gestures of the speaker*." (Guberina & Asp, 1981, p. 1). That is, the auditory and visual information in a person's production reflects how he/she perceives speech and people make sense of the world by negotiation with others. They perceive the world in their own way and they adjust themselves to the world in which they are living in order to survive. The same is true of sounds. When students perceive and produce sounds in the target language, they tend to do so in their mother tongue system and they have to react to the interlocutor by adjusting themselves to the way a native speaker of the target language might perceive. Therefore, we must change the speaker's perception if we want to change his/her production. According to the metaphor of the "phonological sieve" developed by Trubetzkoy (1939), each person has the phonological system of their mother tongue. When they hear the mother language spoken, they will intuitively use the language system of their mother language to analyse what has been said, just like a "phonological sieve", which accepts mother tongue sounds and rejects the others. These mistakes and

misinterpretations happen when this sieve is not suited to the foreign language. In the present study, how students perceive a sound is of great concern and the process of both perceiving and producing is embedded in the experimental design.

Thirdly, the verbotonal procedures *"follow the pattern of language development observed in babies who have normal hearing"*(Guberina & Asp, 1981, p. 1), and, oral ability or spoken language should be taught first. A new born baby produces sounds by crying, babbling, and cooing when his/her speech organs are not sophisticated enough to allow him/her to speak. In this stage, a baby's vocal activity is not a response to his/her sense of hearing but a response to his/her proprioceptive sense. During the process of maturation, rhythm and intonation patterns, rhythmic motor activities, a baby's vestibular, tactile, and proprioceptive sense all contribute to his/her language development. Thus, when a baby produces the first meaningful word, he/she has already learned how to manipulate rhythm and intonation to assign different meanings to the world. That is, rhythm and intonation transmit meanings before a baby really speaks. This is also the case in adult spoken language (Guberina & Asp, 1981). This perspective can explain why Guberina stressed the importance of rhythm and intonation in producing perceived speech. More important, it is the starting point of the focus on rhythm and intonation in the present study which involves both theory and practice in English pronunciation improvement. Evidence of the same understanding is echoed by other scholars' descriptions of how

people learn to speak (e.g., Christiane Dalton & Seidlhofer, 1994). More evidence is found in some scholars' (e.g., Leppänen et al., 2010; Tincoff & Jusczyk, 1999) work. They provide empirical evidence to support the idea that early auditory abilities impact on later outcomes of language development in normal infants. Mueller, et al. (2012) conducted a study to test whether and how basic auditory processes are related to online learning of a linguistic rule in infants and adults. They conclude that "*the ability to extract linguistic rules develops in early infancy and is tightly linked to functional aspects of basic auditory mechanisms"*. They believe that the perceptual ability form the gateway to spoken language. This ability may be an important determinant of language learning process.

Fourthly, speaking is a whole-body experience and body movement is of great importance and will support learning of pronunciation. Body movement is a common but important technology in language use and social interaction. And it has always been an indispensable part of the rhythm i. e. "dance" of any culture. Birdwhistell (cited in A.-P Lian, 1980) finds that 65% communication is nonverbal and only 35% can be attributed to the verbal element. This finding implies the great importance of movement and gesture in communication. When people speak, it would appear that their body moves both consciously and unconsciously in such a way as to emphasise the stressed syllables of the utterances (A.-P Lian, 1980). This is also consistent with the behaviour of a baby expressing their meaning at a very

young age. A strong supporting point emerges from the work of Condon and Ogston on synchrony (Condon & Ogston, 1966). They discovered that, in normal behaviour, speech and body motions were precisely and rhythmically coordinated:

> *Intensive analysis revealed harmonious or synchronous organizations of change between body motion and speech in both intra-individual and interactional behaviour. Thus the body of the speaker dances in time with his speech. Further, the body of the listener dances in rhythm with that of the speaker!* (Condon & Ogston, 1966, p. 338).

This discovery is cited in many disciplines ranging from experimental psychology to social anthropology and linguistics (Gassin, 1990) and it is obviously of great importance for the understanding of second language learning. Gassin (1990) presented Condon and Ogston's major findings concerning synchrony and emphasized their relevance with respect to second language learning. These two discoveries were termed self-synchrony and interactional synchrony. Another way of talking about this is to say that when we speak we dance in synchrony with our speech (this is self-synchrony). When we speak to others, others dance too, in rhythm with that of the speaker (this is interactional synchrony). Body movements are thus to be rhythmically integrated, co-occurring components of total speech behaviour. We are writing our rhythms on each other. Therefore, it is both natural and necessary to consider body movements and gestures when teaching spoken language. This provides a good explanation of why practitioners of pronunciation

instruction do exercises on body movements and gestures like Lian (1980) and Zhang (2006). Furthermore, this is also closely linked to the emphasis on rhythm and intonation since a speaker's body and gestures are synchronically related to their own rhythm and to their interlocutor's speech rhythms, at least when speaking in their mother tongue. Learning a foreign language may lead to a failure of self-synchrony because of the possible dislocation between gestures, rhythms and the holistic rhythm of the new language. Interestingly, Dalton and Seidlhofer's (1994)understanding of pronunciation is in line with this notion: "pronunciation is never an end in itself but a means to negotiate meaning in discourse..."(p. 1). And they use this understanding as a guide to select the aspects covered in their book. In our view, this is an essential and inseparable piece of competent language use for students to learn when we are teaching pronunciation. On the basis of strategies used in verbotonalism and of the findings from movement researchers, Asp, Kline, and Koike, in their essay (2012), described body movements as one of the treatment tools to restructure the child's brain. Their research indicated that body movements based on seven parameters could both stimulate and correct errors of listening through spoken language, and could be applied to all communication problems. Other studies like Dalton and Seidlhofer's (1994) are also in line with this idea as they state that "we use our bodies to communicate with others".

Fifthly, coherently related to the above, prosodic (or suprasegmental) features of spoken language are emphasized. To put it another way, importance is given to intonation and rhythm in the verbotonal system. Guberina insisted on the importance of suprasegmental (or prosodic) features of spoken language: rhythm, pitch variation and stress. Further studies reveal that the acquisition of a language's rhythmic structures is recognized as fundamental to the successful acquisition of the language including its structures as Zhang (2006) synthesized in her Somatically Enhanced Approach. Other scholars, Gilbert, for instance, believe that instruction should focus on learner production of suprasegmental aspects of English pronunciation (Gilbert, 1987, 2008; Gilbert 2010). However, she argues for a systematic teaching and insisted on training students through meaningful pair drills which focus on generating extensive student production of specific points of suprasegmentals (e. g., thought groups, stressed words in sentences). Gilbert's perspective of classroom teaching is different from Guberina and Zhang's beliefs. Furthermore, the idea of systematic teaching is different from the instructional procedures in the present study although both studies highlight the suprasegmental features.

2.3.2 The verbotonal approach to phonetic correction

The verbotonal system is a comprehensive and effective phonetic correction method which works with any language and any type of learners. Many studies have

been conducted to demonstrate its effectiveness. For example, Hu and Uno (2005) carried out a study to examine the effectiveness for Japanese beginners learning the voice tones of the Chinese Language through a newly developed teaching method based on the verbotonal method. The results supported the hypothesis that body movements were useful for distinguishing and learning to pronounce 4 different kinds of voice tones in Chinese. The participants were 35 university undergraduates in a basic Chinese class. After 7 weeks (2 sessions a week, 90 min per session) of training, the subjects could do voice-tone learning more effectively and appropriately in a shorter period of time and obtained high scores on almost all tasks. Another trial of verbotonal approach was conducted by Hang (2012) in Japanese teaching to Chinese students. Her study started with a description of the current Japanese teaching, proposing a verbotonal approach to teaching Japanese sounds on the basis of analysing reasons for current difficulties in teaching Japanese and the differences in Chinese and Japanese language. The study showed a positive result in applying the verbotonal method to teaching Japanese sounds, suggesting that the method was a good trial for language teachers. However, there is no empirical research in which the verbotonal approach is applied in teaching English pronunciation in the Chinese context, especially, in a self-regulating embedded CALL environment. Thus, the present study intends to fill this research gap.

2.3.3 Pedagogic considerations on the basis of the CALL-VT approach

As discussed earlier, EFL learners hear the target language through their personal histories which include their "mother tongue sieve" (Trubetzkoy, 1939) and other filters. In other words, EFL students will encounter difficulties in recognizing English sounds since the language system of their mother tongue, possible other languages and other linguistic experiences will prevent them from giving meaning to the frequencies of specific sounds as a native speaker might. Being unable to perceive English sounds, they then produce a distorted version of that sound which matches the ways that they perceive the sounds in question. In this context then, the ear seems to be "deaf" to the L2 sounds. Appreciating the problem of acquiring sound, Lian (1980) noted that there were two necessary phases in the learning of pronunciation: it was necessary, a) to defeat the students' "deafness" to the sounds by enhancing their perceptions of the sounds to be heard and, after good perception had occurred, b) to practice intensively the perceptual and articulatory patterns of the sounds in question. The pedagogic design of the present study relies heavily on these two important phases in pronunciation learning (and will be developed in Chapter 3).

In the first phase (to defeat students' "deafness" to sounds), a sensitization session is suggested in order to raise their awareness of the target language speech phenomena. This awareness-raising will be facilitated through a load-lightening

process. The learning of pronunciation involves both perception and production. Pronunciation also involves the participation of the body as a whole: it is a form of behaviour (A.-P Lian, 1980). Verbotonalism further postulates that general bodily tensions, including posture and gesture are necessarily connected with the tensions of phonetic articulation. This is because our body functions as a system where all gestures (including articulatory gesture) are connected with one another and mirror one another in an act of self-synchrony. In turn, these general body tensions become embedded in the culture of the groups or communities that speak a specific language in a specific way. In order to pronounce a language the same way, people need to produce the same or similar bodily tensions. When learning an L2 , the self-synchrony of the L1 becomes disrupted as the physical body is habituated to move in a particular way while the L2 seeks to impose a new set of physical tensions, postures and body rhythms. A form of potentially counter-productive dislocation between speech and gesture can then set in as the voice goes one way (under conscious control) and the body goes another way (under automatic control). It will be necessary to enhance multi-sensorial perception of the articulation of the L2, to facilitate pronunciation learning and therefore restore the required self-synchrony for optimal production of the suprasegmental and segmental components of the L2.

In this context, the use of relaxation techniques for the body as a whole will help to reduce the muscular tension and movement habits likely to be antagonistic to

the learning of the pronunciation of the L2, (for supporting evidence, consult Fidelman, 1993a, 1993b; Wylie, 1977). Thus the act of pronunciation takes on a holistic physical dimension which can be exploited through natural and corrective gesture to support good pronunciation. Relaxation of the body will bring about a lowering of conscious and unconscious resistance to the imposition of a new set of physical body tensions and rhythms which the learning of a foreign/second language seeks to impose. Following that logic, before any corrective intervention, the students need to be relaxed both physically and psychologically. To assist in this objective a relaxation phase may be of value.

The relaxation phase adopted in this study is adapted from the success of relaxation techniques used in language learning in Suggestopaedia the Lozanov method of language learning. It was claimed that this method could speed up learning by some 50 times. As the classroom is also a social site of learning, lowering the learners' level of inhibition can also make the learners' egos more permeable (Guiora et al., 1972). There are two steps in this relaxation-based procedure. Step 1: Ask students to sit in their preferred positions and prepare their bodies for new stimuli by engaging in mind-calming exercises. Step 2: Students repeatedly listen to the rhythm of low-pass filtered sentences. Low-pass filtering facilitates perception of the rhythms and intonations of language and reduces the processing load on the learner. Filtering is a form of load-lightening (or

simplification) where learners need only process the intonation/rhythm but none of the words. During this period, as much auditory stimulation as possible is given to the students to let them "feel" the intonation patterns of the sentences. Low-pass filtering preserves low-frequency sounds. The human body is very sensitive to low frequencies (think of the beating of a drum which "hits" you "in the stomach"). Thus the learners hear the intonation, or sentences, not only through their ears but also their body as a whole. Therefore, either consciously or unconsciously, students will perceive much more through their whole body rather than just through their ears and they will also perceive/receive intonation through their muscular system. Relaxation techniques appear to be an effective way of reducing, if not eliminating, the muscular tensions and movements of students' mother language, Chinese (in this case) and making them more receptive. For this reason, the relaxation phase of the approach is extremely important and needs to be highly systematic to ensure that the intonation patterns are internalised as much as possible through a process of intensive awareness-raising. Awareness-raising is further enhanced by teacher intervention. For instance, the teacher will trace out, with the hand, the shape of the intonation curve and beat out the rhythm of the sentence. Students watch and copy as closely as possible. They should also walk about, their feet coming down on every stressed syllable as Lian (1980) and Zhang (2006) did for their students learning French and Chinese intonations.

A reinforcement session then follows. This consists of a number of steps including extending perception by providing variations in the intonation pattern being studied, strengthening perceptual mechanisms by offering discrimination and other exercises as well as further developing articulation.

These considerations will form the core of the CALL-VT approach developed in this study. They will be supplemented by an information technology infrastructure to be described later.

2.4 Autonomous learning environment

Autonomous learning has long been discussed and there has been an increasing tendency among researchers in the field of Second Language Acquisition (SLA), especially, with the shift from teacher-centred to student-centred learning environment (Benson, 2007; Godwin-Jones, 2011; Holec, 1981; Little, 1995). Yet, implementing autonomy remains problematic (Gremmo & Riley, 1995; Ho & Crookall, 1995) though computer technology has contributed much to facilitating autonomous learning (Godwin-Jones, 2011; Hayta & Yaprak, 2013). As Lian (2011) noted, the world is more information-rich and communication-rich today than at any time in its history. People can easily access various resources for obtaining information and they are increasingly heading to manage their own learning on the basis of convenience and preference.

2.4.1 Definitions of autonomy in language learning

The concept of autonomy in language-learning originated from a project housed in the "Centre de Recherche et d'Applications Pédagogiques En Langues" (CRAPEL) at the University of Nancy, France. The founder of the CRAPEL was Yves Chalon who was considered by many to be the father of autonomy in language learning. Chalon died in 1972 and Henri Holec, a prominent figure within the field of autonomy, was given the leadership of CRAPEL.

Many scholars have sought, and continue to seek to define and understand the concept of autonomy from different angles. In his project report to the Council of Europe in 1981, which was regarded as the first contribution to "autonomy" in a document (Benson, 2001). Holec defined autonomy as "the ability to take charge of one's learning". Such responsibility includes the responsibility for making decisions on determining the objectives, defining the content, type and rate of progressions, selecting methods and techniques to be used, and evaluating what has been acquired (Holec, 1981). In short, for him, autonomy was the learner's sense of responsibility.

In Little's terms, learner autonomy is "essentially a matter of the learner's psychological relation to the process and content of learning, a capacity for detachment, critical reflection, decision-making, and independent action" (Little, 1991, p. 4). In Little's description, autonomy is not limited to learning without a teacher. In a classroom context, it is not a matter of letting the learners get on with things as

best they can. Moreover, autonomy is not a steady state achieved by learners (Little, 1990). From Little's point of view, the capacity is concerned more with taking control over the cognitive processes involved in effective self-management of learning. As Benson (2001) states, Little's definition of autonomy adds a vital psychological dimension that is often absent in other definitions of autonomy.

In his book *Self-instruction in Language Learning*, Dickinson (1987) defines autonomy as a particular attitude to the learning task, where "the learner accepts responsibility for all the decisions concerned with his learning but does not necessarily undertake the implementation of those decisions". This is called "full autonomy". A notion of semi-autonomy is also described as "this conveniently labels the stage at which learners are preparing for autonomy". He believes that there are various levels of autonomy and he suggests recognizing a scale of degree of autonomy. In another research article by Dickinson (1995), autonomy is seen as a capacity for active, independent learning.

Benson understands autonomy as "people taking more control over their lives". To him, autonomy in learning is concerned with learners taking more control over their learning both in and outside the classroom, including the purposes of learning and ways of how learning occurs. In another book, *Autonomy & Independence in Language Learning*, Benson and Voller follow Holec's 1981 definition. They also state that the word autonomy has been used in five different

ways: learners study entirely on their own; such skills can be learned and applied in self-directed learning; capacity is suppressed by institutional education; learners are responsible for their own learning; and learners have the right to determine the direction of their own learning (Benson & Voller, 1997). According to them, philosophy and psychology are concerned with individual autonomy, which "implies both responsibility and freedom from constraint " (Benson & Voller, 1997).

Since autonomy is defined as the capacity to take charge of, or responsibility for, one's own learning, what "taking charge" or "taking responsibility" mean should be a concern (Benson, 2001). By taking a close look at Holec's and Little's definitions of autonomy, Benson (2001) argued that the value of three levels of learner control should be recognized. These three levels of control are: learning management, cognitive progresses and learning content. They are clearly interdependent and it can be helpful to consider each dimension separately when we tend to study autonomy in detail. In another piece of work, Benson and Voller (1997) outlined correspondences between three versions of autonomy which are technical, psychological and political autonomy. According to them, there are three approaches to issues of knowledge and learning: positivism, constructivism and critical theory. This work described the concept of autonomy in regard to philosophy and politics (Benson & Voller, 1997).

Keeping the above in mind, it seems that autonomy is not accepted as an absolute concept. It encompasses concepts from different domains such as politics and education, philosophy and psychology (Blin, 2004). And also, the literature of autonomy displays studies which have been conducted to discover the processes and factors affecting autonomy. For example, Benson (2001) argued that there are some misconceptions about the nature of autonomy and its implementation. It is often assumed that autonomy implies learning in isolation, without a teacher or outside the classroom (Benson, 2001). But research studies on autonomy in recent years have emphasized that the development of autonomy necessarily implies collaboration and interdependence. Another example is the study entitled "Learner autonomy in the language classroom: from teacher dependency to learner independence" conducted by Egel (2009). As the author stated, some researchers have claimed that the origins of autonomy were rooted in the European continent. On the contrary, there were researchers claiming that the very idea of autonomy has deep historical roots in Eastern philosophies. Through the above representations of learner autonomy found in the earlier language teaching literature, a more comprehensive understanding of learner autonomy can be drawn out. That is, autonomous language learners take responsibility for their own learning and can do so without a teacher's intervention.

Ambiguities or tensions often characterize the implementation of language teaching approaches (Benson & Voller, 1997) and these could explain the variety of

terms and concepts associated with the notion of autonomy. For example, Dickinson (1987) included terms such as “self-instruction”, “self-direction”, “self-access materials” and “individualized instruction”. A decade later, Dam and Little (1998) include such terms as humanistic language teaching, collaborative learning, experiential learning and learner-centered classroom. However, most people now agree that autonomy and autonomous learning are not synonyms of “self-instruction”, “self-access”, “self-study”, ”self-education”, “out-of-class learning” or “distance learning”. These terms basically describe various ways and degrees of learning by oneself, whereas autonomy refers to abilities and attitudes. The point is, then, that learning by oneself is not the same thing as having the capacity to learn by oneself and to control one’s path to learning success. Further, autonomous learners may, in fact, also be better than others at learning by themselves, but they do not necessarily have to learn by themselves.

2.4.2 Why does autonomy matter?

Why does autonomy matter in the field of language education? Why do scholars keep talking about and doing research on learner autonomy? As Benson (2001) comments, the concept of autonomy has become part of the mainstream of research and practice within the field of language education. This is partially due to previous successful projects associated with autonomy and the efforts of those practitioners in autonomy. Moreover, the basic ideas of autonomy have also been in line with major innovations in language teaching theory and methodology. The

innovations have shifted from the behaviorist assumptions towards more communicative approaches. And the idea that language learning should focus on how to communicate also supports the notion of learner-centeredness. To be brief, communicative teaching, learner-centeredness and autonomy share a focus on the learner as the key agent in the learning process (Benson, 2001).

Coincidentally, when Klimova and Semradova (2012) described present trends in the teaching of foreign languages, "the concept of autonomous learning is being implemented" was included as one of the nine most recent principal trends. Learners are seen as individuals who can and should be autonomous. Moreover, autonomous education helps to aid a student in his/her independent learning (Klimova & Semradova, 2012).

Why is autonomy good for both language educators and learners? In spite of the fact that it fits into the mainstream of SLA mentioned above, it is also consistent with the characteristics of current (21st century) learning environments (Andrew Churches. 21st century) According to Joshi's (2011) study, achieving the conditions of autonomous learning partially relies on the factors of self-esteem and motivation. Thus, students can actively be involved in activities like setting goals, choosing materials, evaluating process, choosing learning strategies, and so on (Cotterall, 2000). Especially in a technology enhanced learning environment, flexible, autonomous learning is essential to success (M Warschauer et al., 2000).

Most of all though, the simplest and most compelling reason is that learners, because of their different personal histories, will encounter unpredicted and unpredictable needs which have to be met quickly and effectively in personally relevant ways as and when they appear. That is when autonomy is necessary. In the end, we are alone in the act of learning. We might as well know how to help ourselves: the act of learning is necessarily solitary.

From the above review of the understandings and value of autonomy, we can see that autonomy has a long and respected tradition in educational, psychological and philosophical thought. The concept of autonomy in language education is well researched both in theory and practice and continues to be researched. More importantly, the role of technology in achieving autonomous learning cannot be underestimated. In this context, the use of modern technology to support autonomous learning is critically important and is an integral part of the current research.

2.5 Computer Assisted Language Learning (CALL)

2.5.1 What is CALL?

In the *Longman Dictionary of Language Teaching and Applied Linguistics*, Computer-assisted language learning (CALL) was defined as "the use of a computer in the teaching or learning of a second or foreign language"(J. C. Richards & Schmidt, p. 101). CALL can take the form of activities which see learning through other media but which use the facilities of the computer, activities which are extensions or

adaptation of print-based (or class-room based) activities and activities unique to CALL (J. C. Richards & Schmidt). Levy (1997, p. 1) defined CALL as "the search for and study of applications of the computer in language teaching and learning". The subject CALL is interdisciplinary in nature, and it has evolved from early efforts to find ways of using the computer for teaching or for instructional purposes across a wide variety of subject areas. Moreover, as Beatty (2003) stated, Computer-assisted language learning, CALL, was a young branch of applied linguistics and was still establishing its directions. To accommodate the changing nature of CALL, Beatty (2003) defined CALL as "any process in which a learner uses a computer and, as a result, improves his or her language" (p.7). As this definition implies, CALL covers a wide range of activities which makes it difficult to describe as a single idea or simple research field. Issues of materials design, technologies, pedagogic theories and models of instruction have come to be encompassed in the study and research of CALL and continue to develop, grow and diversify.

2.5.2 History of CALL

Computers have been used in language learning and teaching for more than 40 years. Several scholars divided the history of CALL into phases or pedagogical approaches according to their own theory.

Warschauer and Healey (1998) divided CALL roughly into three main stages, namely, behavioristic CALL, communicative CALL, and integrative CALL.

Each stage parallels a certain technology to some extent as well as a certain pedagogical approach. Behavioristic CALL, according to Warschauer and Healey (1998), conceived in the 1950s and implemented in the 1960s and 1970s, could be considered a sub-element of the broader field of computer-assisted instruction. Likewise, Warschauer and Healey (1998) claimed that communicative CALL, the next stage of CALL, emerged in the late 1970s and the early 1980s. However, criticism came to communicative CALL as well in late 1980s to early 1990s. Many teachers were paying increasing attention to a more social or socio-cognitive view of teaching than a cognitive view. Approaches such as task-based, project-based, and content-based learning all tried to integrate learners in authentic environments, and also to integrate the various skills of language learning and use. This led to a new perspective, namely, integrative CALL, which seeks both to integrate various skills and also integrate technology fully into the language learning process. With these integrative approaches, learners learn to use technological tools as an ongoing process of language learning and use.

Briefly speaking, as Warschauer and Healey (1998) concluded, if the mainframe was the technology of behavioristic CALL, and the PC (personal computers) the technology of communicative CALL, the multimedia networked computer was the technology of integrative CALL: today's environment which is now augmented with mobile technology.

Scholars have different classifications of the history of CALL. Beatty (2003) said that early statements concerning the history of CALL easily became irrelevant because of new advances in technology. Nevertheless, she (2003) tried to play safe and divided the history of CALL into three phases chronologically: CALL in the 1950s and 1960s; CALL in the 1970s and 1980s; and CALL in the 1990s.

Another scholar, Bax (2003) gave his own ideas of grouping the phases of the development of CALL. He stated that Warschauer and Healey's discussions of the phases of CALL showed significant differences in different publications and these inconsistencies were peculiar and avoidable though they were very important in themselves. Another doubt is that those so-called phases were ambiguous and the terming paradigms or perspectives added conceptual confusion. The third point was the unclear criteria of the three categories: Behaviouristic CALL (perhaps the most plausible category), Communicative CALL and Integrative CALL (two categories far less satisfactory).

After he analysed Warschauer and Healey's three phases of CALL, Bax (2003) used the more general terms of "approaches" instead of "phases". He called the first approach "Restricted CALL" which dominated from 1960s until about 1980 because the term "restricted" was more comprehensive, more flexible and therefore more satisfactory as a descriptor. He termed the second approach "Open CALL" which has lasted from the 1980s until today, with some Restricted CALL

manifestations still valuable in their place, since it was relatively open in all dimensions — from the feedback given to students, to the software types, to the role of the teacher. The third approach, Integrated CALL, only exists in a few places and a few dimensions, but is far from common. The content of this approach is, as he named it Integrated CALL and Integrated language skills work. As Bax (2003) claimed, his above classification is more accurate as a description of what happened in the past and is happening now. This framework might allow us to define our practice in CALL in some details.

2.5.3 Self-directed learning in CALL environments

As described above, CALL is an interdisciplinary term. This idea is reflected from another perspective as Lian (2011) pointed out that even the universe is interdisciplinary in nature and that disciplines are actually human categorization. This realization is having an impact on the educational world and on language learning. Traditional artificial academic boundaries and categorizations are weakening (Lian, 2011) and as a consequence, new attitudes are appearing in a highly information-rich world largely accelerated by the development of modern information technology. Greater connections are now being established between traditionally distinct areas of study and research, previously closed educational disciplines are now opening and facing changes and challenges of many kinds. Technology has also empowered the intellectual fringe-dwellers and more voices, both dominant and less

dominant now have a chance of being heard. These changes have empowered not only theoreticians but potential learners who have the possibility to solve learning problems for themselves through the availability of tools ranging from Google to MOOCs. It is interesting to note how eagerly people have been availing themselves of the opportunities for education with behaviour patterns not unlike those of the proletarian autodidacts of the late 19th and early 20th centuries. From the perspective of language learning theory and practice, these technological developments have provided opportunities for thinking differently, perhaps more interestingly and, clearly, with greater and richer intellectual and social diversity creating a new evolutionary climate for new models of learning to emerge.

Creativity can now come from both students and teachers. And learners can have more power in their learning activities. One such evolutionary model can be found in the development of rhizomatic systems as proposed by language and non-language scholars (e.g., A.-P Lian, 1996; A.-P Lian, 2004; A-P. Lian & Gonzalez, 2008; Pineda, 2014). These systems have a strong focus on self-directed, self-regulated, self-adjusted, and self-managed learning which we will take as being roughly synonymous.

Originally, the rhizome is a philosophical concept developed by Gilles Deleuze and Félix Guattari in their Capitalism and Schizophrenia project (Deleuze & Guattari, 1980). Deleuze and Guattari use the term “rhizome” and “rhizomatic” to

describe theory and research that allows for multiple, non-hierarchical entry and exit points in data representation and interpretation. They outlined four principles: 1) any point of a rhizome can be connected to any other, and must be ; 2) only when the multiple is effectively treated as a substantive, "multiplicity" that it ceases to have any relation to the one; 3) a rhizome may be broken, but it will start up again on one of its old lines, or on new lines ; and 4) a rhizome is not amenable to any structural or generative model; it is a "map and not a tracing".

A rhizome is generated by the connections and flows of our past and our present (including our choices and imposed actions, e. g., by teachers) to guide us into our future. In some sense, it is the "natural", organic way of learning, consciously or unconsciously, formally or informally. Pineda (2014) understands rhizomatic learning as "an organic, developed system of habits, attitudes and personal practices of discovery, meaning making and validation of what we perceive as knowledge". As she pointed out, not everyone perceives in this manner right away and not everyone is ideally self-regulated. Therefore, it is important to identify how rhizomatic learning takes place. An earlier statement on rhizome and rhizomatic learning is found in Lian's work as he argues that the rhizome is the path we follow to construct personal knowledge. It may include the path we follow through representations of knowledge (not knowledge as such) since meaning cannot be transferred/communicated from one person to another. It can only be represented through language or other semiotic systems.

Rhizomatic systems are essentially needs-based systems and are designed to meet the needs of learners as and when these needs are elicited. They take as their point of departure the perceptions and performances of learners as they actually happen and not as a course designer might imagine them to happen or how a statistical model might predict they would happen. Both course designers and statistical models are necessarily flawed as they cannot possibly take account of all the variables involved. These principles are in total in harmony with the principles that govern verbotonalism, autonomy and computer enhanced language learning systems.

2.6 Theoretical framework

Given the coherence of the above-mentioned systems, we are now able to determine the characteristics of an individualized, computer supported, self-regulated rhizomatic system for pronunciation improvement based on verbotonal theory of perception.

While the focus of contributions in this research is on pronunciation improvement by applying a verbotonal approach to pronunciation embedded in a self-regulating CALL environment, the importance of the over-arching theoretical framework cannot be over-looked. The starting point of this study is verbotonalism (VT), a way of correcting pronunciation for foreign language learners on the basis of a theory of perception. It argues that one's ear will only hear what it expects. The

ear tends to select and recognize the sounds which are meaningful to it. Thus, the present study starts from a training of students' perceptual mechanisms to enable them to have better pronunciation by raising their awareness.

Secondly, in a background of increasing popularity and importance of technology in language education, the theoretical framework of this study also finds it origins in the works of Deleuze and Guattari (1987), French philosophers whose rhizomatic theory allows multiple learning networks. As Lian (2004) argues, a rhizomatic structure can be seen as a self-regulating structure responsive to the learners' needs. These ideas are in harmony with today's world where students can learn to be self-managed by accessing numerous resources as and when they need them under autonomous control.

Thirdly, also informing the study and following from the above is the notion of autonomy as proposed by Holec (1981) and which emphasizes "internal evaluation" by the student. It reminds us to pay attention to aspects of students' self-awareness. In spite of the differences in defining the term "learner autonomy", there is a general agreement that one of the crucial components of autonomy is the ability for managing one's own learning (Benson, 2007; Cotterall, 1995a; Dickinson, 1987; Hedge, 2001; Holec, 1981; Little, 1991; Littlewood, 1996). This kind of ability tends to be developed and enhanced in models such as the DIY (Do-it-yourself) model encouraged by emerging technologies and practiced by countless individuals

nowadays (A.-P Lian, 2014a; A. B. Lian, 2014b). Since there is evidence to encourage us to develop learner autonomy, it is interesting to provide a place to let it happen to benefit students in language learning. There will be much more space in the development of self-regulated learning in this study. And this is also a key focus of the study.

Last but not least, Sugata Mitra's (2003) experiments with children's pronunciation improvement using, on their own, a speech to text software called DRAGON encouraged us to explore the possibilities to develop learners' self-regulating ability so as to build up an effective personal learning environment. Furthermore, this study also encouraged us to believe that not all learning needs to be based on systematic carefully prepared materials. This is especially the case with adults, who, as opposed to children, may have more ways and strategies to develop their self-managed learning since they have stronger motivation to work at their own pace and preference.

The above ideas are coherent with a postmodern view of our perceptions of language learning, and our cognition of the world. To sum up, verbotonaliam, rhizomatic theory, theories on prosody, the notion of learner autonomy, the features of some current approaches to language learning environments and modern technology constitute the conceptual framework for the present study. It provides the over-arching framework for the project which is summarized in Figure 2.2.

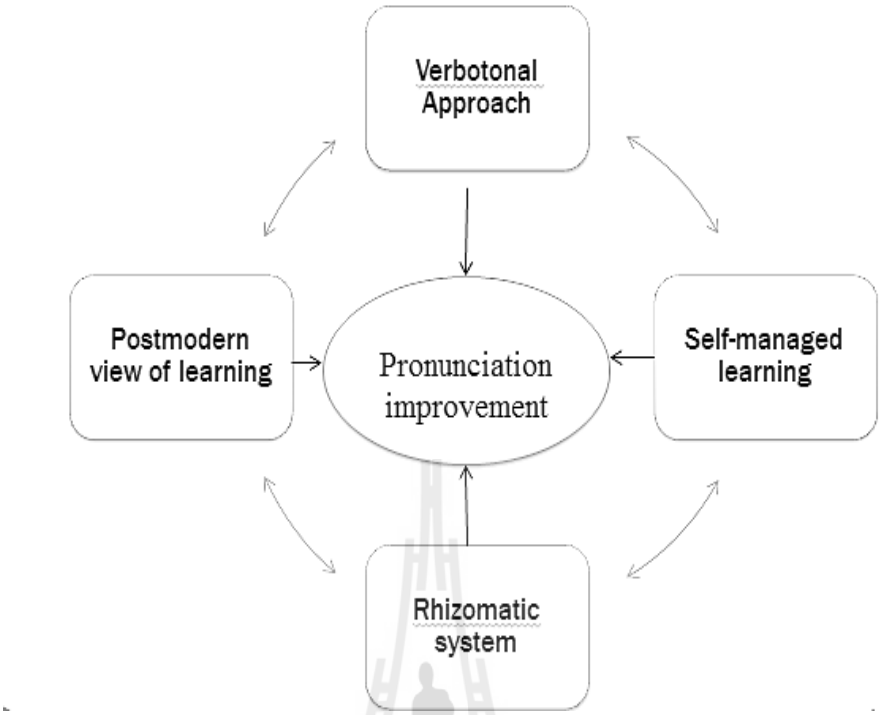


Figure 2.2 Theoretical framework

2.7 Summary

This chapter provided the theoretical framework of the present study with a review of the relevant literature on pronunciation instruction, learning theory, verbotonal system, autonomous learning, and current self-directed CALL learning environments. Main issues are discussed both in terms of theory and an adequate review of previous studies. These theories and studies are synthesized by the researcher through relating to the origin of the theory, rationale and the design of the present study.

CHAPTER 3
METHODS & MATERIALS

This chapter discusses the research methods and materials of the present study. It begins with the research design, which includes participants, instruments, pedagogic considerations, and the description of the experiment. Then, follow the materials and approaches used, including the verbotonal approach, teaching materials, students' diary, pretest and posttest, questionnaires and oral interviews. Finally, the procedures for data collection and analysis are explained.

3.1 Research Design

The research objectives and research questions determined the research design of the present study. Popular worldviews suggest the use of two philosophical positions in research, *positivism* and *constructivism*. These are quite different from each other in epistemological terms. However, it is our argument that methods are not necessarily tied to a particular worldview, and can be flexibly mixed at different moments of the research process (Teddlie & Tashakkori, 2009; Teddlie et al., 2008; Weathington et al., 2010). This takes us to the logical conclusion that a mixed methods approach may be used here. The necessity for this decision is laid out below. Tashakkori and Teddlie (1998, 2003) argued convincingly for the validity of

mixed method research. They emphasize its benefits in many diverse research settings. They argue that mixed methods are often more efficient in answering research questions than either the qualitative or the quantitative approaches alone because mixed methods allow cross-method comparisons and provide grounds for triangulating data. Thus the weaknesses of one method may be offset by the strengths of the other. Similarly, Creswell (2003) reviewed the literature on mixed research method and advocated triangulating the two approaches.

Therefore, a mixed methods research design was employed in this study as it would be able to provide more consistent data, resulting in more valid and reliable conclusions. In this study, the term "mixed methods" is taken to mean a combination of quantitative and qualitative methods used from identifying research questions, and hypotheses, sampling and selection of participants, through to the collection, analysis, and interpretation of the data.

The main purpose of the study was to enhance the pronunciation of Chinese EFL learners by applying a new approach based on the verbotonal system for pronunciation improvement embedded in a self-regulating Computer Assisted Language Learning (CALL) environment. As Wiersma and Jurs (2005) stated, an intervention designed to improve students' achievement should take on the form of an experimental treatment. This advice was followed here. Further, taking account of the features of a language learning experimental design, a quasi-experimental design was adopted since quasi-experimental studies are commonly conducted under

conditions where variables are difficult to control (Seliger & Shohamy, 1989). Students' pronunciation progress could be influenced by many factors in spite of the treatment in the experiment. As many researchers agree (e.g., Charles & Mertler, 2002; Thomas, 2003; Wiersma & Jurs, 2005), quasi-experimental research is a part of experimental research whose most important characteristic is to deal with the phenomenon of cause and effect. This is why a quasi-experimental design was employed as a main approach in the present study. Of the mixed methods used in this study, the first part was a quantitative experimental intervention and the second part was based on qualitative interviews and students' diaries.

The quantitative phase of the study examined statistical relationships between the intervention and students' pronunciation scores as well as descriptive statistics of the results from the written questionnaire. In this context, comparisons are possible between the experimental group and the control group because they are fairly clear-cut and the researcher can have some control over when to measure outcome variables (Punch, 2013).

In order to better understand the impact of the intervention, the other phase of the research was directed towards students' attitudes towards and perceptions of the intervention. Before students began their respective tasks and also at the end of the experiment, questionnaires were administered and semi-structured interviews were conducted to investigate students' attitudes and perceptions. The survey phase of the study aimed to generate results to corroborate results from the experimental phase. The data from students' diary and oral interview were analysed qualitatively.

Data triangulation and method triangulation were employed to crosscheck results for consistency and to offset any bias, so as to reduce the chances of reaching false conclusions (Hammersley, 2008), to reduce the uncertainty of data interpretation (Webb et al., 2000), and to enhance confidence and accuracy in the overall conclusions drawn from the study (Brewer & Hunter, 1989; Morse, 1991; Spicer, 2004). Thus, the five research questions were subjected to both quantitative and qualitative analysis so as to yield a maximum amount of information. The data sources analysed included students' scores on the pronunciation tests, students' written diary, written questionnaires and semi-structured oral interviews. To be specific, when identifying problems, qualitative thematic analysis was used to obtain teachers' and students' perceptions of the pronunciation problems they encountered either in the teaching or the learning process. Students' scores in the pretest and posttest, as well as data from questionnaires administered both before and after the experiment were analysed quantitatively using SPSS 16.0 (IBM, 2008). Students' diaries and oral interviews on their perceptions of the intervention were analysed qualitatively.

3.1.1 Participants

The term "population" is identified as "the entire set of relevant units of analysis, or data" (Frankfort-Nachmias & Nachmias, 1996). The population within a single piece of research is supposed to share the same features, or at least be of a similar nature. It is not normally practical for a researcher to study the entire population in a piece of research since the population is normally too large to handle

(L. Cohen et al., 2000). Therefore, a sample is commonly selected to represent the population. However, the entire population for this study consisted of only 96 English major undergraduates. This was a manageable number and selecting a sample was not necessary. The entire population participated in the study though there was one withdrawal from the programme (whose partial results were removed from the data set).

Xingyi Normal University, the participating university, had recently been accredited by the Ministry of Education (MOE) to offer bachelor's degree programmes. The participants in the present study were first-year English undergraduate majors in the School of Foreign Languages. They were majoring in English Education and most intended to become middle school or primary school English teachers. Aged 17-21, they were enrolled in a class of 48, the standard class size in this university being 40-50. All had studied English for a minimum of 6 years. These 96 students came from two intact classes. They were randomly assigned to an experimental group and a control group. Students from Class 201301 constituted the control group. Students from Class 201302 constituted the experimental group. Participants in the experimental group were 48 first-year English majors. Participants in the control group were another 48 first-year English majors. Their English proficiency levels were similar. It is necessary to note that 95 students out of 96 actually took part in the tests because one student in the control group transferred to another university and was excluded from the experiment.

Control Group

The control group, Class 201301, took the class in a traditional way as previous classes had. They used the approved textbook for pronunciation learning. They were free to practice their pronunciation in the method normally used by the teacher. The teacher was present in the classroom. Basically, students repeated and practiced what the teacher presented. They also participated in classroom activities revolving around articulatory description of sounds. After class, they spent many hours practicing either in groups or on their own. However, no CALL-VT techniques for pronunciation training were permitted in the control group.

Experimental Group

The experiment could be described as appropriate encouragement for self-managed learning on the basis of a verbotonal approach in a CALL environment. The starting point of the experimental design was based in part on Lian's (1980) instructional method for French intonation and Zhang's Somatically Enhanced Approach (SEA) for teaching Mandarin tones and prosody (Zhang, 2006). Both of these approaches are heavily based on the principles of the verbotonal approach of corrective phonetics. Students played and practiced the filtered sentences on a computer and engaged in further perception and production activities (some in a CALL environment). These sentences came from the same teaching materials used by the control group. However, these sentences had been filtered through a low-pass filter (set at between 320 Hz and 400 Hz depending on the fundamental frequency of

the speaker's voice). The filtering removed vowel and consonant sounds and liberated the intonation and rhythm patterns effectively. Thus, the "tune" of English became quite salient. Individual language sounds (e. g., words) were no longer recognizable but the melody was left behind. This procedure reduced the processing load on the student and allowed for greater focus on intonation. Students could better feel the prosody through amplification of the parts of the sentences which were left. To improve students' autonomy, these materials were embedded in a CALL system enabling students to engage in self-regulated learning. The teacher's role was no longer central and the students were encouraged and helped to learn in their preferred ways. Hence, it was hoped that these activities might result in accelerated, more effective learning and it was expected that students' learner autonomy would be developed accordingly.

In order to comply with ethical requirements, a written consent form was filled out by each participant to ensure that all understood the purpose of the experiment and they participated in the experiment voluntarily and were aware of all possible challenges and risks. In addition, prior to the start of the project, ethical concerns were cleared by the academic committee of the university.

3.1.2 Variables

As reviewed earlier in Chapter 2, notions of learning theory together with the relevant intervention, students' perceptions of pronunciation learning, and students' learning environments were factors which might influence pronunciation learning.

Thus, according to the objectives and research questions of the present study, the independent variables were (a), the intervention applied to pronunciation learning and (b), the students' learning environments. The dependent variables were students' scores on the various pronunciation tests and their perceptions of pronunciation learning built into the intervention process.

3.2 Research Instruments

Collection of data entailed the use of different research instruments, including pronunciation pretest and posttest, student and teacher's semi-structured oral interviews, students' written questionnaires, an inventory of students' level of autonomy, and students' diaries.

3.2.1 Learning materials

This project was undertaken within the organizational structures of the Chinese Ministry of Education. As a consequence, the learning materials used needed to comply with the requirements set by the Ministry of education and consisted of two parts. The first consisted of textbooks prescribed in the approved syllabus. The material selected for use was the textbook *English Pronunciation & Intonation for Communication* edited by Professor Wang Guizhen (G. Z. Wang, 2005). This book was also set for the control group. The second consisted of any materials chosen by students themselves to achieve their purpose. These came from any resources available to students.

3.2.2 Low-pass filtered language training

The use of filtered language for improving pronunciation was introduced before the beginning of the experiment. The notion of low-pass filtering was explained to students and the way of performing the filtering was demonstrated. They needed to know about this so as to be aware of what they were listening to and of the potential benefits of using these materials. While recordings of filtered language were made available for their use, students were also free to use any other materials on their own. During the self-regulated learning time, students were free to choose their learning environment and the materials that they preferred. Study time and place were decided by themselves. However, a set of learning objectives (syllabus) acting as a guide for students to achieve were provided. They were expected to study for a total of 10 hours per week (including the 2 hours of formal class time) as were the students in the control group.

3.2.3 Student questionnaire

It is important to understand students' beliefs and opinions about learning English pronunciation. It was even more so in the case of this experiment as students' attitudes towards language learning are regarded as a major concern in foreign language education. A questionnaire designed to reveal students' attitudes was therefore administered at the end of the experiment. It investigated students' understandings and perceptions of the intervention in the present study. As most researchers know, questionnaires are a frequently used assessment tool concerning

pronunciation although it is claimed that this method is not reliable since questionnaires present the respondents' subjective opinions and judgments about a situation (Nowacka, 2012) rather than the reality of their pronunciation issues. Be that as it may, many research studies have been carried out on learners' attitudes towards English pronunciation learning via questionnaires or interviews (e.g., Cenoz & Lecumberri, 1999; C Dalton & Smit, 1997; Kang, 2010; T. H. Liu, 1999; Norris-Holt, 2002; Nowacka, 2012; Porter & Garvin, 1989; Tergujeff, 2013; Tokumoto & Shibata, 2011). We followed suit with care. The construction of questionnaires used in the present studies was based on a review of previous research studies in this field and the researcher's experience in teaching pronunciation with additional guidance from Dörnyei's (2003) and Presser et al.'s (2004) publications about the nature, merits and shortcomings of questionnaires.

The student questionnaire consisted of two parts: a) background information and b) students' perceptions and attitudes towards pronunciation learning using CALL-VT. The first part included age, gender, years of learning English, grade of study, time spent on learning pronunciation outside the class. Such information was necessary to provide a general description of the participants, and possibly added new variables for analysis.

The second part included 10 items concerning students' opinions of pronunciation learning via CALL-VT. These items were drawn from a review of previous study on questionnaires conducted on pronunciation perceptions, identified

problems in the pre-experimental stage, and the researcher's experience. The questionnaire used a 5-point Likert scale ranging from Strongly Disagree to Strongly Agree with values of 1 to 5 assigned to the descriptors respectively. Higher total scores indicated higher agreement on each particular statement.

The questionnaire was administered to the experimental group. To avoid misunderstandings and confusion, the questionnaire was written in both English and Chinese. Five experts specialized in phonetics were invited to validate the language and content in the questionnaire in order to check the validity of all items. The questionnaire was revised and improved on the basis of those experts' comments and suggestions. Furthermore, in order to determine the reliability of the questionnaire, Cronbach's Alpha Coeffient (α) was used to check the internal consistency of the questionnaire items by analyzing the data from the trial version ($\alpha = 0.92$).

3.2.4 Semi-structured interviews

A follow-up oral interview was conducted to acquire in-depth information of students' attitudes towards learning pronunciation in a self-regulated learning environment. Interviews are known as a useful tool for generating information in research dealing with personal experience and perspectives (Holstein & Gubrium, 2004). As a qualitative research instrument, an interview enables an understanding of "the meaning that the participants hold about the problem, not the meaning that the researchers bring to the research or writers express in the literature" (John W Creswell, 2009, p. 175). Three types of interviews have been identified: unstructured, semi-structured, and

structured. A semi-structured interview involves asking a list of structured questions and then, depending on the responses of the interviewees, probing more deeply with open questions to obtain additional information. The flexibility of the semi-structured interview enables the researcher to gather large amounts of standardized participants' data (Gall et al., 1999). The interview was conducted face-to-face since it is a "shortcut" for the researcher to find answers to research questions by interacting with the respondents directly (Robson, 1993). Therefore, for the purpose of collecting not only standardized but also in-depth data, the semi-structured interview technique was selected for both student interviews and teacher interviews.

3.2.5 Inventory of students' learning autonomy

An inventory was constructed to measure the development of students' level of autonomy. The use of an inventory was inspired by discussions on the nature of learner autonomy by researchers such as Benson (2001), Holec (1981), Dickinson (1987), Little (1991), and Tassinari (2012). As Tassinari (2012) asserted, an autonomous learning process should entail an evaluation of the learner's competencies for autonomy. Since pronunciation learning in the present study was supposed to happen in a self-regulated learning environment, an evaluation of the students' competencies for autonomy provided an interesting opportunity to look at autonomy development. Tassinari (2012) employed a dynamic model of learner autonomy as a tool to support the self-assessment and evaluation of learning competencies.

The present study adopted Tassinari's model of inventories of learner autonomy to evaluate students' learner autonomy (Tassinari, 2012). The dynamic model and the descriptors Tassinari developed have been validated in workshops and discussions with experts from the CRAPEL, Université Nancy 2 and the Language Centre of the Freie Universität Berlin. A questionnaire consisting of 10 items was administered to the students in the experimental group to investigate their development of learner autonomy.

3.2.6 Student diary

Students both in the control group and the experimental group were asked to keep a diary to record their activities every time they studied pronunciation, including the length of time, place, materials (content), effectiveness, feeling, and resources. The diaries provided information on what, where, with whom, and how they studied pronunciation. The format of the diary was designed according to the above criteria. In addition to the required information to be recorded, students were free to provide comments on anything they felt was relevant to their pronunciation study.

These diaries were used only for this study. The researcher read and categorized the content using a qualitative methodology.

3.2.7 Pretest & posttest

Description of Pronunciation Test

general introduction to the pronunciation test in the School of Foreign Languages is necessary before the pretest and posttest are described in this part.

This test aims at testing students' mastery of English pronunciation. It is designed to cover all the main features of pronunciation. This includes phonemes (consonants and vowels) as well as suprasegmental features (intonation and stress). The test paper consists of five parts: phoneme reading (20%), word reading (20%), sentence reading (20%), text reading (20%), and a free talk (15%). Part I (phoneme reading) includes 302 voiced and unvoiced sounds covering all consonants and vowels in English. Part II (word reading) includes 100 words from English textbooks ranging from Grade 3 to Grade 8 in China. Part III consists of 10 sentences covering various intonations and rhythm structures. Part IV is passage reading. The passage length is about 150 words. The materials of the passage are also from the English texts authorized by the Ministry of Education of China. Part V is a free talk. Five topics are available and selected randomly. The student is required to give a 3-minute talk on the topic assigned to them. There is a bank of test papers in the School from which five test papers are chosen for each year's test. They are identified as Test Paper A to Test Paper E. The students will use one of these tests randomly as his/her test paper. Two trained raters listen and score each performance individually. The final score is the mean of those given by the two raters if their scores are not very different (5-point maximum difference). A third rater will test the student again if the scores from two raters are significantly different.

This test has been administered in the School of Foreign Languages for more than ten years. Students are encouraged to practice their pronunciation to pass the

test. A separate certificate is awarded if students reach the set standards. All English majors take part in this test since the certificate is proof of their level of English pronunciation.

The control group, Class 201301, took the class conducted in the traditional manner as previously taught. They used the textbook set for pronunciation learning in Xingyi Normal University for Nationalities. They were free to practice their pronunciation in the approach taught by the teacher. However, no CALL-VT materials or training were permitted in the control group nor would the control group have access to the same facilities for private study.

Test paper for the pretest and the posttest

The test papers used in the pretest and posttest phase of this study consisted of four parts which were constructed on the basis of the above test but improved according to the research purposes: phoneme reading (20%), reading words (25%), reading passage (25%), and oral interview (30%). The pretest and posttest papers were constructed by the researcher on the basis of the test papers used for the pronunciation test in the School of Foreign Languages and a review of the oral interview part of the International English Language Testing System (IELTS).

The validity and reliability of the tests were checked. First, 5 experts were invited to rate each item of both the pretest and posttest paper. The experts rated the relevance of each item for the purpose of the test and the appropriateness of the content areas, and checked the evaluation form by using Item-Objective Congruence

Index (IOC) as a validation method for the relevance of the content and the objective of the questionnaire. The evaluation form used a 3-point scale (1 = relevant, 0 = uncertain, -1 = irrelevant). The results of the IOC analysis (see Appendix J and Appendix I) showed that both pretest and posttest papers were valid to be adopted as a trial out paper. Then, these two papers were sent to 30 students in Xingyi Normal University for Nationalities to check their reliability.

The teacher tested the students both in the experimental and the control groups before and after the experiment to check the overall improvement, if any, of their pronunciation. The tests were recorded and kept in a website created by the researcher for blind rating.

Raters: Chinese expert raters

In order to estimate the students' pronunciation in a comprehensive way, both Chinese expert raters and native speaker raters were invited to rate the students' recordings in both pretest and posttest. The four Chinese experts were professional pronunciation raters for many years. The average number of years of experience for each of the four raters was 7. However, to make the rating more reliable, an inter-rater reliability check was performed using Pearson's Correlation Coefficient. This calculation measures the strength and direction of the relationship between two variables. Table 3.1 below shows the results of the four Chinese experts' rating scores for the experimental group in the pretest using Pearson's Correlation Coefficient. The inter rater-reliability was deemed acceptable as all of the results were higher than 0.70.

Table 3.1 Inter rater reliability results (Pearson's Correlation Coefficient)

raters	rater1&4	rater1&3	rater1&2	rater2&4	rater2&3	rater3&4
PCCC result	0.782	0.955	0.790	0.794	0.860	0.827

Raters: Naïve Native Speaker Raters

In order to triangulate the data collection methods and also to look at the students' pronunciation from another perspective, four native speaker raters were invited to rate the students' performances on the pre and posttest. The reasons for inviting native raters are as follows. First, it is valuable to triangulate the Chinese experts' ratings with another group of people. Second, the native raters do not have any special training. Thus, they would perform the ratings according to their perceptions of pronunciation as ordinary native speakers of English and with a different point of view from that of an expert. It is beneficial to see results from another, less expert, perspective. As naïve raters, they are the representatives of ordinary speakers of English who are possible target interlocutors of EFL learners.

Rubrics were developed to guide both the four Chinese expert raters and the naïve raters. Expert raters were invited to rate all parts of the tests. The four native speakers were invited to rate the last two parts (reading-passage and oral interview) in terms of comprehensibility, pronunciation and fluency of the language the students produce before and after the intervention. The reason for inviting them to rate the last two parts was that both these two parts test the suprasegmental features of English

language and this was the goal of English pronunciation learning set in this present study.

Rating procedure

All recordings were renumbered and randomly ordered. They were kept confidential on a website which only raters could access. Raters clicked on the provided html links and, guided by the rubrics, rated each piece of the recordings. They were not able to identify whom the recording belonged to, as the whole rating procedure was blind.

3.3 Pedagogic procedures

In this section, pedagogic procedures are discussed in order to bring out the theory behind the CALL-VT system for pronunciation instruction and also, to provide an overall picture of the students' activities during the experiment. The starting point for the pedagogic consideration was the theory of verbotonalism. Pedagogic sequences consisted of two related parts: inside the classroom and outside the classroom.

Classroom activities

In the first phase (defeating students' "deafness" to the sounds of English), a sensitization session was conducted in order to lighten the load and also to raise students' awareness of the target language pronunciation characteristics. There were 7 steps in this phase:

Step 1: Students were asked to sit in their preferred position as relaxed as possible. In principle, they could even lie on their backs on the floor if they wished to (unfortunately, the experimental conditions did not provide a comfortable place to allow them to do so). With the classroom quiet and dark, they received 5 to 10 minutes' of mind-calming exercises. Baroque music was played to help them be more relaxed and therefore receptive to the language input (A.-P Lian, 1980; Lozanov, 2009). Students were even free to play any mind-calming music that they could access through their own smartphones. Mind-calming exercises included: yoga breathing, relaxation exercises and baroque music.

Step 2: This and subsequent steps focused on the perception and production of prosodic patterns of language (stress, rhythm and intonation) and not individual sounds as in traditional approaches to phonetics. Here, students repeatedly listened to natural language sequences digitally modified through a low-pass filter set at 320 Hz. Low-pass filtering has the effect of removing all vowel and consonant sounds (essentially the words) and leaving behind the prosody of language: stress, rhythm, and intonation (the melody of language) thus lightening the students' processing/cognitive load (no words and grammar as such to process, only beats and melody). Filtered sentences sounded as though they were being hummed rather than articulated and they were not intelligible in the usual way. Students listened to the filtered sentences at least ten times in succession. In the introductory lecture before the intervention, they were told that they did not have to understand the meaning of

the sentences but just to listen. Once the consonants and vowels have been removed, the elements left behind, intonation and rhythm can be perceived and integrated more effectively (Renard, 1975). While students could not understand the detailed meaning of the sentence content, intonation does carry meaning and they were encouraged to guess the meaning of the intonation patterns themselves, e. g., “Is this a yes/no question?”, “Is this a statement?” etc.

Step 3: While listening to the filtered sentences, students and the teacher hummed in unison to the melody and rhythm of the filtered sentences so as to practice intonation production (the fundamental frequency of the voice which is responsible for intonation - F0 - is produced primarily by the vocal cords: actually a form of humming. Humming is a way of practicing intonation-production without the burden of words). At various moments in the class, volunteers would spontaneously stand up and present their hummed versions of the studied patterns in order to demonstrate their understandings.

The idea behind the first three steps is to focus on the melody of the sentence without interference from consonant and vowel sounds, words and grammar, thereby reducing the processing load on the brain and the articulatory organs. At the same time use of low-frequency patterns preferentially activates the right brain where melodic signals are processed (Hesling et al., 2005) thus enabling better perception of patterns.

Step 4: The teacher and students clapped their hands to the rhythm and beat of the English language that they listened to. Students clapped and even danced. They walked hand in hand or on their own to "express their feel of the language" or to synchronize to the rhythm together (thus developing a joint awareness of the rhythm and communicating it to each other). In this step, on the one hand, students could hum and clap out the rhythm in their personal perceived ways. On the other hand, the teacher could also use this moment to present the correct rhythm to the students as a non-prescriptive model. This model was presented as a suggestion of the teacher's personal preferences and perceptions rather than as a prescription to be imposed. In other words, students were not required to model the teacher's gestures but created their own representations of appropriate movement to accompany speech production. As a result, students were able to experience more explicitly their understanding of the rhythm of the sentence at a physical level and to compare and contrast their personal understandings of the rhythm and melody of English against those of other students as well as the teacher's.

Step 5: Students and teacher "walked" the rhythm of the language presented with feet coming down on every stressed syllable. They used gestures to help express their perceptions of the rhythm and intonation since body movement and gesture were proposed as aids to intonation learning.

Step 6: The original unfiltered sentences were played. Students were required to mouth the words to the sounds of the filtered patterns but not actually utter

any sounds. Mouthing the words is an intermediate step towards articulation of the full sentences and gives students an opportunity to practice the articulation of the sounds without placing them on a self-generated intonation background (which adds another layer of difficulty). Again this step is designed to reduce the load on students.

Step 7: Students were asked actually to fully utter the words which they added to the "language tune" that they had been learning. Original sentences were played continuously. Then, students repeated the sentences in chorus. The teacher checked and corrected students as necessary.

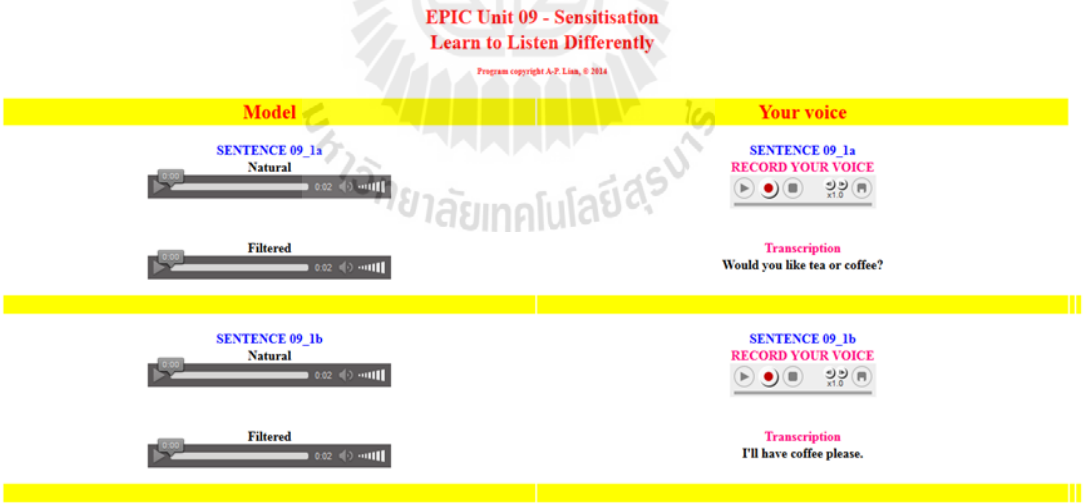
It should be noted that sequencing of the above steps was not linear and steps were not planned to occur in a fixed order. After listening to the filtered sentences approximately 10 times, the students were able to listen to the normal sentences and to make comparisons between the filtered and unfiltered versions. They were free to choose to listen to specific filtered or unfiltered sentences of their choice as many times as they felt necessary. They could also record their voices and play them back so as to compare their production with the models that they had been listening to. They could even filter their own voices if they wished to do so.

Out-of-classroom activities

The above listed classroom activities made up one part of the experiment. The other part of the experiment consisted of self-regulated pronunciation reinforcement exercises performed outside the classroom. Students were able to use

a computer room set up to provide access to filtered sentences and other resources (like authentic models of native speakers) for pronunciation learning. They could listen to and practice what they had studied in class and they could engage in other activities of their choice to improve their pronunciation. For example, they could make recordings of their voices and compare them with the correct intonation patterns or hum or gesture as they had been doing systematically in class.

A simple online computer assisted system was developed to help students to be self-managing. They could listen to filtered and unfiltered models and could practice and enhance their pronunciation of intonation patterns. At the same time they could essentially generate their own lessons by navigating through the entire set of course materials in a simple way. The graphic below shows the student interface.



The teacher

The researcher was not involved in the teaching. An experienced teacher who had taught the same course for many years took care of both the experimental

and control groups to avoid several other variables as much as possible (such as teaching methodology, teacher's personality and popularity etc.) from confounding the results. However, it was necessary for the teacher to understand the experiment design and the theory behind it to conduct the teaching task well. Therefore, the researcher talked to the teacher face to face about the experiment and provided all materials needed in teaching. The researcher also modeled a mini class for the teacher. That is, the teacher was trained to teach the experimental class.

It is also important to point out that while the teacher was not hostile to the CALL-VT system, and taught the CALL-VT class faithfully, she was skeptical of the nature of CALL-VT and was personally committed to the traditional teaching approach in which she was expert.

3.4 Data collection procedures

3.4.1 General procedures

This research was conducted in a normal English learning setting for a Chinese university, where two intact groups of students enrolled in the Phonetics Course participated in the study for an 18-week period. The focus of the study was to determine the effectiveness of the CALL-VT system, a filtered language training approach based on verbotonal theory of pronunciation learning. As discussed earlier, two groups of students enrolled in the Phonetics Course participated in the quasi-experiment during regular class time in the 18-week period. The study was

conducted from March 2014 to July 2014 - the second semester of the participants' course of study. However, due to the problems with the university server and other equipment, the experiment lasted only 14 weeks. Further, of these 14 weeks, two were make-up classes. In other words, the actual experimental time was reduced to 12 weeks.

As mentioned in Chapter 1, according to the English curriculum for English majors, the Phonetics Course aims to enhance the first-year English Majors' pronunciation in accuracy, fluency and effective communication. Students are expected to master basic knowledge of English phonetics, all consonant and vowel sounds, and natural intonation and rhythms in reading (reading aloud) and speaking skills. "English Pronunciation Intonation for Communication" (G. Z. Wang, 2005) was the text used as the formal teaching material. Five principles underpinned the textbook: 1) mastery of the correct pronunciation of 44 phonemes is necessary for good pronunciation; 2) word stress and sentence stress are emphasized as the basis for correct pronunciation of phonemes; 3) making links between words in a sentence can make a sentence comprehensible; 4) intonation and rhythm in sentences are important in communication; and 5) communicating effectively is the most important learning outcomes of this course. The goal is achieved through pronunciation development after a large amount of practice.

This textbook consists of 15 units which are intended to be studied over two semesters. Each of the units covers five parts: introduction, listening and speaking

exercises, main content, and pronunciation in communication. Answer keys to exercises, mouth diagrams of the English consonants, key points of texts, and a list of further reading are provided together with a CD of the book and mirrors to observe articulation.

3.4.2 Specific procedures

The specific procedures employed in this research were as follows. First, in order to determine if there were significant differences between the groups before the intervention, the two groups of participants were pretested with the pronunciation test developed by the researcher. There were two pronunciation tests of about the same difficulty level. The researcher randomly used one as a pretest and the other as a posttest.

Subsequently, the two intact groups of students were randomly assigned as control group and experimental group.

Next, the researcher applied the CALL-VT system to the experimental group. The control group did not receive any of the treatment given to the experimental group. Instead, treatment in the control group consisted of instruction typically conducted in EFL classrooms and the random learning activities normally engaged in by students in the course.

At the end of the 14-week period, the two groups of students were retested using the remaining pronunciation test (one test having been used in the pretest).

The test scores were reported in detail. The overall scores, scores on tests for individual phonemes, intonation, and flow of the text reading were recorded separately so as to get as much information as possible for further identification of student problems and also, to identify pronunciation improvements after the CALL-VT intervention.

The purpose of using different pronunciation tests for pretest and posttest was to compare the subjects' scores on the two tests and to observe their development after the intervention. Different pronunciation tests were used in the pretest and posttest to avoid the danger of the subjects' posttest scores being influenced by their pretest score. The data obtained from the pretest and posttest were submitted for quantitative analysis.

3.5 Data analysis

This section presents the methods of data analysis employed in the present study. Data obtained from the 14-week experiment on pronunciation pretest and posttest together with data from the written questionnaire were presented in terms of quantitative analysis, while data obtained from students' diaries and semi-structured oral interviews were analysed qualitatively.

3.5.1 Quantitative data analysis

The quantitative data were drawn from the tests for 95 students, 48 student questionnaires on their perceptions of both pronunciation learning and the

intervention in this study. These data were broken down into different parts for analysis. After the quantitative data had been collected, basic descriptive statistics were performed using SPSS (Version 16) (IBM, 2008) to obtain a general overview of the data. Cronbach's alpha (α) was calculated to test its internal consistency reliability.

3.5.1.1 *t*-test

Paired samples *t*-test was utilized to compare the participants' mean scores on the pretest and posttest. The purpose was to see whether there were statistically significant differences in the mean scores between students' pretest and post-test scores, thus, to decide on the effects of pronunciation improvement in a self-regulated learning environment.

3.5.1.2 ANCOVA

An Analysis of Covariance (ANCOVA) using Statistical Package in Social Sciences (SPSS) (IBM, 2008) was used to test whether the treatment had had an effect on the outcome variables (e. g., posttest scores) after removing the variance for any preexisting differences between the control and experimental groups.

3.5.2 Qualitative data analysis

Qualitative research describes an event in its natural setting as Abusabha and Woelfel (2003) define it. It is a subjective way to look at things as they are an attempt to explain the studied behavior (Walsh, 2003). It was hoped that data from the oral interview would provide the researcher with an overview and in-depth

information about the students' opinions and reflections on the present study. The data were analysed thematically and described a "live" picture of the situation since qualitative researchers use anthropological and ethnographic methods to study the participants rather than designing an experiment which artificially controls variables.

3.6 Summary

This chapter described the methods and materials employed in the present study. The population used in the experiment and the pedagogical considerations were presented together with the research instruments and the variables. In addition, the rationale for and construction of the instruments were discussed. Finally, the procedures for data collection and analysis were described. In the next chapter the research findings and relevant discussions will be presented.

CHAPTER 4

RESULTS

The main purpose of this chapter is to present the findings of the current study in response to the five research questions posed in Chapter 1. This chapter is organized in two sections. The first section deals with the quantitative analysis of the participants' performance on the pretest and posttest using statistical methods. The second section reports the data elicited through the questionnaire, the students' diaries, and the semi-structured interview from both quantitative and qualitative perspectives.

4.1 Assessment of pronunciation

The quasi-experimental design of the study made it possible to find answers to the five research questions (see Section **1.4**). This section describes the students' performances on the pretest and posttest as assessed by the validated tests (the validity and reliability of the tests were checked accordingly. See **3.2**).

4.1.1 Pretest results

The pretest was administered at the beginning of the experiment. As mentioned in **3.2.7**, a predesigned pretest was used to estimate the students' pronunciation ability in both experimental and control groups. Necessary reliability

and validity checks were conducted to make the test a valid tool. All the tests were recorded and renumbered for rating purposes and especially, to maintain anonymity.

4.1.1.1 Pretest results: Chinese raters

Four Chinese experts were invited to rate the pretests conducted with both experimental and control groups. The findings of the pretest were used to set a baseline for comparison and to help interpret the findings, particularly if any improvements or differences were discerned at the end of the experiment.

As described in **3.2.7**, the pretest consisted of four parts: phonemes, word-reading, passage-reading, and finally, an oral interview. Findings would be presented in a sequence of total score, scores for phonemes, word-reading, passage-reading, and scores for oral interview.

Total Score

A descriptive analysis based on total scores was employed to provide an overview of the participants' performances. Table 4.1 below shows the mean of the total scores on the pretest together with the standard deviation. The significance level of all tests was set at 0.05.

Table 4.1 Descriptive statistics of total scores in the pretest

Group	Mean	Number	Std. Deviation
Experimental group	70.89	48	8.38
Control group	75.20	47	8.37

To compare the means of the two groups, a two-tailed *t-test* was employed. The mean score of the experimental group was significantly different from that of the control group (p = 0.000). Specifically, the control group performed significantly better than the experimental group. The control group was ahead of the experimental group.

A descriptive analysis of the individual parts was then employed to see where the differences were situated.

Part I: Phonemes

In the first part, students were asked to read 20 phonemes which cover most phonemes required in the official pronunciation course book. The maximum sub-total score was 20.

Table 4.2 below shows the results of the independent samples two-tailed *t*-test. It was found that the differences between the experimental group and the control group were significant (p = 0.000). Specifically, the experimental group's ability to pronounce individual phonemes was poorer than that of the control group.

Table 4.2 Descriptive statistics of phonemes in the pretest

Group	Mean	Number	Std. Deviation
Experimental group	15.94	48	2.12
Control group	17.44	47	1.71

Part II: Word-reading

In the second part, students were asked to read 50 words chosen from a test bank of the vocabulary required for Chinese EFL learners. The maximum sub-total score was 25. Table 4.3 below shows the results of the independent samples *t*-test. It was found that the difference between the experimental group and the control group were significant (p = 0.003). Specifically, the experimental group's ability to read words correctly was poorer than that of the control group.

Table 4.3 Descriptive statistics of word-reading in the pretest

Group	Mean	Number	Std. Deviation
Experimental group	18.64	48	2.44
Control group	20.30	47	2.79

Part III: Passage-reading

In the third part, students were asked to read a passage of approximately 150 words. The maximum subtotal score was 30. Table 4.4 below shows the results of the independent samples *t*-test. It was found that the difference between the experimental group and the control group was significant (p = 0.009). Specifically, the experimental group was poorer than that of the control group in terms of the ability of passage-reading.

Table 4.4 Descriptive statistics of passage-reading in the pretest

Group	Mean	Number	Std. Deviation
Experimental group	21.04	48	2.35
Control group	22.40	47	2.56

Part VI: Oral Interview

In the fourth part, students were asked to attend a face-to-face oral interview. The sub-total score was 25. Table 4.5 below shows the results of the independent samples *t*-test. It was found that the difference between the experimental group and the control group was not significant (p = 0.586). This result indicates that two groups were at the same level before the intervention in terms of speaking English in a natural setting. In other words, in the case of free speech, students performed at the same level, maybe indicating that the difference between the two groups was due to the control group's better ability to adjust to artificial pedagogic exercises not necessary in real life contexts. For example, they were better students in a formal sense but not better language performers.

Table 4.5 Descriptive statistics of oral interview in the pretest

Group	Mean	Number	Std. Deviation
Experimental group	15.33	48	2.90
Control group	15.02	47	2.74

4.1.1.2 Pretest results: native speaker raters

Similar to the presentation of findings in **4.1.1.1**, findings from the native raters are reported in a sequence of sub-total scores; score in passage-reading in terms of comprehensibility, pronunciation, and fluency; score in oral interview in terms of comprehensibility, pronunciation, and fluency.

Sub-total Score: passage-reading

A descriptive analysis based on sub-total scores was employed to provide an overview of the participants' performances in reading passages. Table 4.6 below shows the mean of the sub-total score on the pretest together with the standard deviation. To compare the means of the two groups, a two-tailed t-test was employed. A significant difference was found between the experimental group and the control group ($p = 0.027$). That is, the control group was ahead of the experimental group in passage-reading according to the general rating in terms of comprehensibility, pronunciation and fluency.

Table 4.6 Descriptive statistics of sub-total score in passage-reading

Group	Mean	Number	Std. Deviation
Experimental group	10.11	48	1.35
Control group	10.72	47	1.33

Sub-total Score: Oral Interview

To estimate the students' abilities in oral interview, a descriptive analysis based on the sub-total score was employed to provide a picture of the participants' performances. Table 4.7 below shows the mean of the sub-total score on the pretest together with the standard deviation. A two-tailed *t*-test was employed to compare the means of two groups. There was no significant difference between the experimental group and the control group ($p = 0.185$). This result indicates that two groups started at the same level before the intervention in terms of speaking English in a natural setting. This finding is consistent with that of the Chinese expert raters.

Table 4.7 Descriptive statistics of sub-total score in oral interview

Group	Mean	Number	Std. Deviation
Experimental group	9.58	48	1.32
Control group	9.95	47	1.32

To take a closer look at the pronunciation ability, individual scores were provided by native raters in terms of comprehensibility, pronunciation and fluency. Therefore, a comparison of the two groups was conducted by a descriptive analysis of native raters' evaluation on students' performances. The following tables (from Table 4.8 to Table 4.13) provide a descriptive analysis of the above features.

Passage-reading: comprehensibility

Table 4.8 below shows the mean of the individual scores on the pretest in terms of comprehensibility together with the standard deviation. There was no significant difference between the experimental group and the control group in terms of comprehensibility (p = 0.187). That is, these two groups started at the same level in the terms of comprehensibility when reading passages before the intervention.

Table 4.8 Descriptive statistics of comprehensibility in passage-reading

Group	Mean	Number	Std. Deviation
Experimental group	9.58	48	1.32
Control group	9.95	47	1.32

Passage-reading: pronunciation

Table 4.9 below shows the mean of the individual scores on the pretest in terms of pronunciation together with the standard deviation. A significant difference was found between the experimental group and the control group (p = 0.002). That is, the control group performed better than the experimental group in terms of pronunciation when reading passages.

Table 4.9 Descriptive statistics of pronunciation in passage-reading

Group	Mean	Number	Std. Deviation
Experimental group	9.51	48	1.50
Control group	10.47	47	1.39

Passage-reading: fluency

Table 4.10 below shows the mean of the individual scores on the pretest in terms of fluency together with the standard deviation. There was no significant difference between the experimental group and the control group in terms of fluency ($p = 0.127$). That is, these two groups started at the same level in terms of fluency when reading passages before the intervention.

Table 4.10 Descriptive statistics of fluency in passage-reading

Group	Mean	Number	Std. Deviation
Experimental group	10.25	48	1.55
Control group	10.72	47	1.41

Oral Interview: comprehensibility

Table 4.11 below shows the mean of the individual scores on the pretest in terms of comprehensibility in the oral interview. There was no significant difference between the experimental and control groups in terms of comprehensibility ($p = 0.407$). That is, these two groups started at the same level in terms of comprehensibility when they had the oral interview before the intervention.

Table 4.11 Descriptive statistics of comprehensibility in oral interview

Group	**Mean**	**Number**	**Std. Deviation**
Experimental group	10.09	48	1.65
Control group	10.36	47	1.40

Oral Interview: pronunciation

Table 4.12 below shows the mean of the individual scores on the pretest in terms of pronunciation in the oral interview. A significant difference was found between the experimental group and the control group in terms of comprehensibility ($p = 0.006$). That is, the control group was ahead of the experimental group in terms of pronunciation when they had the oral interview before the intervention.

Table 4.12 Descriptive statistics of pronunciation in oral interview

Group	Mean	Number	Std. Deviation
Experimental group	9.14	48	1.16
Control group	9.81	47	1.19

Oral Interview: fluency

Table 4.13 below shows the mean of the individual scores on the pretest in terms of fluency in the oral interview. There was no significant difference between the experimental group and the control group in terms of fluency ($p = 0.652$). That is, these two groups started at the same level on in terms of fluency when they had the oral interview before the intervention.

Table 4.13 Descriptive statistics of fluency in oral interview

Group	Mean	Number	Std. Deviation
Experimental group	9.54	48	1.59
Control group	9.68	47	1.52

4.1.2 Posttest results

The posttest served to measure the effects of the pedagogical intervention on the students' pronunciation ability. It was administered when the pedagogical intervention had been completed. The assessment procedures were the same as those employed for the pretest.

4.1.2.1 Posttest results: Chinese raters

Total Score

The participants' performances on the pretest and posttest were compared in order to check for any improvements in the students' pronunciation ability, thus to examine the effects of the pedagogical intervention. Descriptive statistics was used as a tool to provide an overall picture of the students' performance (see Table 4.14). It is interesting to note that the outcomes of the experimental group were more consistent since the standard deviation was smaller than that of the control group. In other words, the outcomes of the control group were more scattered and less reliable than those of the experimental group.

Table 4.14 Descriptive statistics of total scores in the pre- and posttest

Group	Tests	Mean	Number	Std. Deviation
Experimental group	Pretest	70.89	48	8.38
	Posttest	84.93	48	6.48
Control group	Pretest	75.20	47	8.38
	Posttest	80.94	47	9.45

With regard to the improvement of each group, paired-sample t-test were used to perform the comparison of the pretest and posttest, thus to verify the potential effects of the pedagogical intervention on the EFL learners. This statistical analysis was appropriate because it compared the means of two variables – the pretest and the posttest – for each group. Statistical analysis showed that there was a significant difference between the pretest and the posttest in the experimental group ($p = 0.000$), and for the control group too ($p = 0.000$). That is, comparing the pretest and the posttest, both experimental group and control group improved significantly. However, the experimental group improved much more than the control group as Table 4.14 above shows. In the experimental group, the mean changed from 70.89 to 84.93, an increase of 14.4 (20.3%). In the control group, the mean changed from 75.20 to 80.94, an increase of 5.74 (8.09%). Moreover, there was a significant difference between the pretest means of the experimental group and the control group ($p = 0.014$) and there was a significant difference between the posttest means of the experimental group and the control group ($p = 0.018$). Initially, the control group was ahead of the experimental group. After the treatment, the experimental group had made up the difference with the control group and had overtaken it by a large margin.

In order to reduce within-group error variance and to eliminate confounds caused by any unmeasured variables, an Analysis of Covariance (ANCOVA) was used to remove the bias of the variables. In the pretest, there were

significant differences between the control group and the experimental group (p = 0.014), which indicated that the control group performed better than the experimental group before the treatment. However, the control group fell behind the experimental group in the posttest (p = 0.018). Thus, an ANCOVA on SPSS was used to test whether the treatment had an effect on the outcome variables (e. g., posttest score) after removing the variance for the preexisting differences. Using this method, we checked the differences between post scores of the experimental group and the control group when treating the pretest scores of those two groups as constant.

Referring to the output of SPSS, the p-value of Levene's test of equality of error variances is 0.826 (see Table 4.15). This indicates that there is no significant difference between the variances of the posttest scores of the experimental group and the control group. Thus, an analysis of covariance (ANCOVA) could be used to test the effect of the treatment as seen in the posttest scores. The p-value of the corrected model was 0.000 (see Table 4.16). This indicates that there was a significant difference between the experimental group and the control group on the posttest when treating the pretest scores as constant. That is, the experimental group performed significantly better than the control group and it seems clear that the intervention was highly effective.

Table 4.15 Levene's test of equality of error variances[a]

F	df1	df2	Sig.
0.049	1	93	0.826

Table 4.16 Tests of between-subjects effects

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	4328.619[a]	2	2164.309	93.392	0.000
Intercept	822.528	1	822.528	35.493	0.000
pre	3951.401	1	3951.401	170.507	0.000
EG	1198.248	1	1198.248	51.706	0.000
Error	2132.049	92	23.174		
Total	660243.540	95			
Corrected Total	6460.668	94			

To take a further look at where the differences were found, statistical analysis was also conducted on each part of the test.

Part I: phonemes

In the phonemes part, which tests the segmental features of English pronunciation, there was a significant difference between the pretest and the posttest scores in the experimental group ($p = 0.000$), and in the control group too ($p = 0.002$). In other words, comparing the pretest and the posttest, both experimental group and control group improved significantly in terms of segmental features. However, the experimental group outperformed the control group despite starting at a lower level (see Table 4.17).

Table 4.17 Descriptive statistics from the pretest and posttest: phonemes

Group	Tests	Mean	Number	S.D	p-value
Experimental group	Pretest	15.93	48	2.12	0.000
	Posttest	18.35	48	1.27	
Control group	Pretest	17.44	47	1.71	0.002
	Posttest	18.12	47	1.74	

Referring to the output of SPSS, the p-value of Levene's test of equality of error variances was 0.924 (see Table 4.18). Thus, an ANCOVA was used to test the effect of the treatment in the posttest. The p-value of the corrected model was 0.000 (see Table 4.19), which indicated that there was a significant difference between the experimental group and the control group in the posttest when treating the pretest scores as constant. That is, the experimental group performed significantly better than the control group thus the intervention was highly effective in terms of reading phonemes.

Table 4.18 Levene' s test of equality of error variances[a]

F	df1	df2	Sig.
0.009	1	93	0.924

Table 4.19 Tests of between-subjects effects

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	76.088[a]	2	38.044	25.063	0.000
Intercept	135.261	1	135.261	89.107	0.000
prepart1	74.811	1	74.811	49.284	0.000
EGCG	17.793	1	17.793	11.722	0.001
Error	139.652	92	1.518		
Total	31820.188	95			
Corrected Total	215.739	94			

Part II: word-reading

In the reading word part, which also tests the segmental features of English pronunciation, there was a significant difference between the pretest and the posttest in the experimental group (p = 0.000), and in the control group too (p =

0.000). That is, comparing the pretest and the posttest, both experimental and control groups improved significantly in terms of reading words. However, the experimental group outperformed the control group although it started at a lower level (see in Table 4.20) in reading words.

Table 4.20 Descriptive statistics of the results from the pretest and posttest

Group	Tests	Mean	Number	S.D	p-value
Experimental group	Pretest	18.64	48	2.44	0.000
	Posttest	22.28	48	1.67	
Control group	Pretest	20.30	47	2.79	0.000
	Posttest	21.53	47	2.47	

Referring to the output of SPSS, the p-value of Levene's test of equality of error variances was 0.145 (see Table 4.21). Thus, an ANCOVA was used to test the effect of the treatment in the posttest. The p-value of the corrected model was 0.000 (see Table 4.22), which indicated that there was a significant difference between the experimental group and the control group in the posttest when treating the pretest scores as constant. That is, the experimental group performed significantly better than the control group thus the intervention was deemed to be highly effective in terms of reading words.

Table 4.21 Levene's test of equality of error variances[a]

F	df1	df2	Sig.
2.155	1	93	0.145

Table 4.22 Tests of between-subjects effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	259.806[a]	2	129.903	72.004	0.000
Intercept	158.820	1	158.820	88.032	0.000
prepart2	246.623	1	246.623	136.700	0.000
EGCG	67.792	1	67.792	37.576	0.000
Error	165.979	92	1.804		
Total	46041.310	95			
Corrected Total	425.785	94			

Part III: passage-reading

In the passage-reading part, which also tested the suprasegmental features of English pronunciation, there was a significant difference between the pretest and posttest in the experimental group (p = 0.000), and in the control group too (p = 0.000). That is, comparing the pretest and the posttest, both experimental group and control group improved significantly in terms of reading passages. However, the experimental group outperformed the control group since they started at a lower level but overtook the control group in the posttest (see Table 4.23) in reading words.

Table 4.23 Descriptive statistics of the results from the pretest and posttest

Group	Tests	Mean	Number	S.D	p-value
Experimental group	Pretest	21.04	48	2.35	0.000
	Posttest	25.76	48	2.10	
Control group	Pretest	22.40	47	2.56	0.000
	Posttest	24.83	47	2.62	

Referring to the output of SPSS, the p-value of Levene's test of equality of error variances was 0.569 (see Table 4.24). Thus, an ANCOVA was used to test the effect of the treatment in the posttest. The p-value of the corrected model was 0.000 (see Table 4.25), which indicated that there was a significant difference between the experimental group and the control group in the posttest when treating the pretest scores as constant. In other words, the experimental group performed significantly better than the control group thus the intervention was highly effective in terms of passage-reading.

Table 4.24 Levene's test of equality of error variances[a]

F	df1	df2	Sig.
.327	1	93	0.569

Table 4.25 Tests of between-subjects effects

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	341.877[a]	2	170.939	77.690	0.000
Intercept	93.369	1	93.369	42.436	0.000
prepart3	320.936	1	320.936	145.864	0.000
EGCG	84.907	1	84.907	38.590	0.000
Error	202.423	92	2.200		
Total	61352.850	95			
Corrected Total	544.300	94			

Part IV: oral interview

In the interview part, which tests the natural production of spoken English, there was a significant difference between the pretest and the posttest in the experimental group (p = 0.000), and in the control group too (p = 0.001) (see table

4.26). That is, comparing the pretest and the posttest, both experimental group and control group improved significantly in terms of speaking English in a natural setting. However, the experimental group had more improvement than the control group. There was no significant difference between these two groups in the pretest (p = 0.586) but there was a significant difference in the posttest (p = 0.001).

Table 4.26 Descriptive statistics of the results from the pretest and posttest

Tests	Groups	Mean	Number	S.D	p-value
Pretest	Experimental group	15.33	48	2.90	0.586
	Control group	15.03	47	2.74	
Posttest	Experimental group	18.60	48	2.68	0.001
	Control group	16.50	47	3.44	

4.1.2.2 Posttest results：native speaker raters

Total score of passage-reading and oral interview

The sum of the scores of passage-reading and oral interview was presented and analysed because these two parts tested the suprasegmental features of the spoken English. Therefore, an overview of the subtotal scores of these two parts was worthy looking at. Descriptive statistics was used as a tool to provide an overall picture of the students' performance. As shown in Table 4.27 below, the average mean score of the 95 participants on the posttest was higher than that on the pretest.

Table 4.27 Descriptive statistics of the results from the pretest and posttest

Group	Tests	Mean	Number	Std. Deviation
Experimental group	Pretest	19.69	48	2.56
	Posttest	24.38	48	3.81
Control group	Pretest	20.67	47	2.39
	Posttest	22.26	47	3.12

With regard to the improvement of each group in passages-reading and oral interview, paired-sample t-test were used to perform a comparison of the pretest and posttest, so as to verify the potential effects of the pedagogical intervention on the EFL learners. The statistical analysis showed that there was a significant difference between the pretest and the posttest in the experimental group ($p = 0.000$), and in the control group too ($p = 0.000$). That is, comparing the pretest and the posttest, both experimental group and control group improved significantly. However, the experimental group improved much more than the control group as Table 4.28 above shows. In the experimental group, the mean changed from 19.69 to 24.38, an increase of 4.69 (23.8%). In the control group, the mean changed from 20.67 to 22.26, in increase of 1.59 (7.6%). There was a big difference between the range of the improvement in the experimental and control groups.

More importantly, as shown in table 4.28 below, there was no significant difference between the pretest means of the experimental group and the control group ($p = 0.0574$) whereas there was a significant difference between the posttest means of the experimental group and the control group ($p = 0.004$). This

indicated that the experimental group improved significantly more than the control group in the posttest.

Table 4.28 Descriptive statistics of the results from the pretest and posttest

Tests	Groups	Mean	Number	S.D	p-value
Pretest	Experimental group	19.69	48	2.56	0.057
	Control group	20.67	47	2.38	
Posttest	Experimental group	24.38	48	2.68	0.004
	Control group	22.26	47	3.44	

Further analysis for the results from the native raters was conducted for each individual rating criterion, comprehensibility, pronunciation, and fluency.

Part III: passage-reading (experimental group)

Generally, in reading passages, significant differences were found when comparing the pretest and the posttest of the experimental group for every rating criterion. Table 4.29 below shows the detailed mean, standard deviation, and p-value of each *t*-test check. In terms of comprehensibility, the mean changed from 10.55 to 13.30, an increase of 2.75 (26%). The *t*-test p-value was 0.000. In terms of pronunciation, the mean changed from 9.51 to 11.51, an increase of 2 (21%). The *t*-test p-value was 0.000. And the mean for fluency changed from 10.26 to 12.42, an increase of 2.16 (21%). The *t*-test p-value was 0.000. These findings demonstrated a great improvement made by the experimental group.

Table 4.29 Descriptive statistics of passage-reading for the experimental group

Rating criterion	Tests	Mean	Number	S.D	p-value
Comprehensibility	pretest	10.55	48	1.72	0.000
	posttest	13.30	48	2.09	
Pronunciation	pretest	9.51	48	1.50	0.000
	posttest	11.51	48	1.80	
Fluency	pretest	10.26	48	1.55	0.000
	posttest	12.42	48	1.94	

Part IV: oral interview (experimental group)

Also, in the oral interview, there were significant differences between the pretest and the posttest as shown in Table 4.30 below. In terms of comprehensibility, the mean changed from 10.09 to 12.41, an increase of 2.32 (22%). The mean of pronunciation changed from 9.14 to 11.41, an increase of 2.27 (24.8%). In terms of fluency, the mean changed from 9.54 to 12.06, an increase of 2.52 (26.4%). All these changes were considerable and the p-value was 0.000, which indicated a significant difference statistically.

Table 4.30 Descriptive statistics of oral interview for the experimental group

Rating criterion	Tests	Mean	Number	S.D	p-value
Comprehensibility	pretest	10.09	48	1.65	0.000
	posttest	12.41	48	2.45	
Pronunciation	pretest	9.14	48	1.16	0.000
	posttest	11.42	48	2.02	
Fluency	pretest	9.54	48	1.59	0.000
	posttest	12.06	48	2.22	

Part III: passage-reading (control group)

It is not surprising to find that students in the control group also improved in pronunciation learning. Detailed information of findings is presented in Table 4.31 below. In terms of comprehensibility, the mean changed from 10.98 to 11.99 (p = 0.000), and the mean for pronunciation changed from 10.47 to 11.11 (p = 0.008). The mean for fluency also increased from 10.72 to 11.47 (p = 0.001).

Table 4.31 Descriptive statistics of passage-reading for the control group

Rating criterion	Tests	Mean	Number	S.D	p-value
Comprehensibility	pretest	10.98	47	1.39	0.000
	posttest	11.99	48	1.86	
Pronunciation	pretest	10.47	47	1.39	0.008
	posttest	11.11	47	1.39	
Fluency	pretest	10.72	47	1.41	0.001
	posttest	11.47	47	1.51	

Part IV: oral interview (control group)

Also, in the interview part, all aspects, including comprehensibility, pronunciation and fluency, were found improved significantly. As shown in Table 4.32 below, the mean for comprehensibility changed from 10.36 to 10.97, and the mean for pronunciation changed from 9.81 to 10.54. With regard to fluency, the mean changed from 9.68 to 10.70 with a p-value of 0.000. That is, the traditional method for pronunciation learning also produced results.

Table 4.32 Descriptive statistics of oral interview for the control group

Rating criterion	Tests	Mean	Number	S.D	p-value
Comprehensibility	pretest	10.36	47	1.40	0.030
	posttest	10.97	47	1.99	
Pronunciation	pretest	9.81	47	1.18	0.002
	posttest	10.54	47	1.52	
Fluency	pretest	9.68	47	1.52	0.000
	posttest	10.70	47	1.84	

However, when compared to the improvement of the experimental group, the control group showed less improvement. In order to verify this statement, a comparison of the pretest and the posttest was made within both experimental and control groups.

Passage-reading: comprehensibility

As reported in **4.1.1.2**, there was no significant difference between the experimental group and the control group in the pretest of the reading passage in terms of comprehensibility ($p = 0.187$). Nevertheless, in the posttest, a significant difference was found in the passage-reading in terms of comprehensibility of the language ($p = 0.002$). Table 4.33 below shows the mean of two groups in the posttest of reading passages.

Table 4.33 Descriptive statistics of comprehensibility in passage-reading

Group	Mean	Number	Std. Deviation
Experimental group	13.30	48	2.09
Control group	11.99	47	1.86

Passage-reading: pronunciation

A comparison of mean scores for passage-reading in the posttest showed no statistically significant difference between the experimental group and the control group in the posttest (p = 0.231). Table 4.34 below shows no statistical difference but the experimental group was still ahead of the control group.

Table 4.34 Descriptive statistics of pronunciation in passage-reading

Group	Mean	Number	Std. Deviation
Experimental group	11.51	48	1.81
Control group	11.11	47	1.39

There was a significant difference between the control group and the experimental group (p = 0.002) in the pretest (see 4.1.1.2). Specifically, the control group (mean = 10.47) was ahead of the experimental group (mean = 9.51) in the pretest but fell behind the experimental group in the posttest. Thus, an ANCOVA using SPSS was performed to test whether the treatment had an effect on the outcome variables after removing the variance for the preexisting differences.

Levene's test of equality of error variances was performed and returned a value of 0.093 (see Table 4.35). This indicated that there was no significant difference between the variances of the posttest scores of the experimental group and the control group. Thus, an analysis of covariance (ANCOVA) could be used to test the effect of the treatment on the posttest. The p-value of the corrected model was 0.001 (see Table 4.36), which indicated that there was a significant difference between

the experimental group and the control group in the posttest when treating the pretest scores as constant. In other words, the experimental group performed significantly better than the control group and the intervention was effective. The experimental group outperformed the control group, indicating that the CALL-VT treatment was significantly more effective than the traditional approach to pronunciation training.

Table 4.35 Levene's test of equality of error variances[a]

F	df1	df2	Sig.
2.873	1	93	0.093

Table 4.36 Tests of between-subjects effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	32.918[a]	2	16.459	7.119	.001
Intercept	106.113	1	106.113	45.897	.000
preRP	29.143	1	29.143	12.605	.001
EGCG	12.635	1	12.635	5.465	.022
Error	212.703	92	2.312		
Total	12404.438	95			
Corrected Total	245.621	94			

Passage-reading: fluency

In terms of fluency in passage-reading, as described in 4.1.1.2, there was no significant difference between the experimental and control groups in the pretest ($p = 0.127$). However, in the posttest, a significant difference was found between the two groups ($p = 0.009$) as shown in Table 4.37 below. Specifically, the

improvement of the experimental group was significantly greater than that of the control group after the intervention.

Table 4.37 Descriptive statistics of fluency in passage-reading

Group	Mean	Number	Std. Deviation
Experimental group	12.42	48	1.94
Control group	11.47	47	1.51

Oral interview: comprehensibility

In relation to the oral interview, there was no significant difference between the experimental and control groups in the pretest ($p = 0.407$) in terms of comprehensibility. However, in the posttest, a significant difference was found between the two groups ($p = 0.002$) as shown in Table 4.38 below. Specifically, the improvement of the experimental group outperformed the control group after the intervention.

Table 4.38 Descriptive statistics of comprehensibility in oral interview

Group	Mean	Number	Std. Deviation
Experimental group	12.42	48	2.45
Control group	10.29	47	1.99

Oral interview: pronunciation

In relation to the oral interview, the control group (mean = 9.81) performed better than the experimental group (mean = 9.14) in the pretest. It was found that there was a significant difference between these two groups in the posttest ($p = 0.017$) in the posttest. Not only did the experimental group catch up the

differences with the control group, but it also overtook it. Table 4.39 shows the means of both groups together with the standard deviation. That is, the experimental group started behind the control group but caught up and overtook it.

Table 4.39 Descriptive statistics of pronunciation in oral interview

Group	Mean	Number	Std. Deviation
Experimental group	11.42	48	2.02
Control group	10.54	47	1.52

Oral interview: fluency

When considering fluency in the oral interview, there was no significant difference between the experimental and control group in the pretest ($p = 0.652$) whereas the experimental group performed significantly better in the posttest with a p-value of 0.002. Table 4.40 shows the detailed mean and standard deviation by the descriptive analysis from SPSS.

Table 4.40 Descriptive statistics of fluency in oral interview

Group	Mean	Number	Std. Deviation
Experimental group	12.06	48	2.22
Control group	10.70	47	1.84

As summary of all the above findings, it can be said, that whether starting at the same level or not, the experimental group's improvement was significantly higher than that of the control group after the intervention for all aspects that were being tested.

4.2 Students' opinions of the CALL-VT system

This section seeks to answer research question 4 and deals with the students' opinions of the CALL-VT system, the problems encountered by both the teacher and students, students' perceptions of the intervention, and the perceived effects of the intervention on pronunciation. Data obtained from the written questionnaire, and the semi-structured interview, were submitted for either quantitative or qualitative analysis.

4.2.1 Data from the written questionnaire

The written questionnaire was administered to the 48 students in the experimental group after completion the posttest. All 48 questionnaires distributed were returned. Whenever each respondent submitted a questionnaire, the researcher checked carefully that no blank or incomplete sheet had been submitted. All questionnaires were then analysed quantitatively.

In the written questionnaire, 5-point Likert-scale questions ranging from "strongly disagree" to "strongly agree" were utilized in order to make clear distinctions between students who agreed with the statement and those who did not. The students' responses to the questionnaire were coded and keyed into the SPSS programme 16.0 for statistical analysis. The five-point items were coded as follows:

Strongly disagree = 1

Disagree = 2

Not sure = 3

Agree = 4

Strongly agree = 5

In scoring the students' responses, one point was allocated to *Strongly Disagree*, two for *Disagree*, three for *Not Sure*, four for *Agree*, and five for *Strongly Agree*. That is, each number represented a statement of their opinion on each item in the questionnaire. It is noteworthy that the students' scores on the questionnaire did not represent their pronunciation performance but only their opinions of the CALL-VT system.

The quantitative data elicited through the questionnaire revealed the students' opinions of the CALL-VT system from one side (see Table 4.41). Most of them (mean = 4.33) thought that CALL-VT was helpful (mean = 4.33) and effective (mean = 4.15) in their pronunciation learning. Many of them also thought that this approach was interesting (mean = 3.85) and they were happy (mean = 3.83) when learning pronunciation with it. Many also believed that their pronunciation could be improved by applying CALL-VT in learning (mean = 3.90). Compared with the traditional approach, most students preferred the CALL-VT approach. As mentioned before, the CALL-VT system can be applied in the classroom, the computer lab, and anywhere students would like to use it. In terms of where

students chose to study, many students tended to use it in a computer lab (mean = 3.90). Many of the students preferred to apply this approach on their own (mean = 3.88) in the place where they liked being and they believed that their learner autonomy would be improved. This result will be verified thorough the use of another questionnaire on learner autonomy.

Table 4.41 Mean of students' opinions of CALL-VT

	N	Mean	S.D
1. CALL-VT is helpful in pronunciation learning.	48	4.33	0.63
2. CALL-VT is effective in pronunciation learning.	48	4.15	0.65
3. CALL-VT is interesting in pronunciation learning.	48	3.85	0.80
4. I am happy with CALL-VT.	48	3.83	0.81
5. I believe that my pronunciation is improved by using CALL-VT.	48	3.90	0.78
6. I prefer CALL-VT to a traditional approach in pronunciation learning.	48	4.15	0.83
7. I would like to learn pronunciation via CALL-VT in a computer lab.	48	3.90	0.99
8. I would like to learn pronunciation via CALL-VT in a classroom.	48	3.44	1.03
9. I would like to learn pronunciation via CALL-VT on my own.	48	3.88	0.87
10. My learner autonomy is improved via CALL-VT.	48	3.69	0.75
Valid N (listwise)	48		

4.2.2 Data from the semi-structured interview

A semi-structured interview was conducted to investigate the students' opinions of the CALL-VT system in pronunciation learning. The researcher purposively selected 10 students from the experimental group to be the interviewees.

The predesigned and validated interview questionnaire consisted of 10 questions (see Appendix M) concerning the effectiveness, the interests, and the popularity of the CALL-VT system. Also, students were required to describe their

opinions of and feelings towards the intervention. Two questions about learner autonomy were included in the interview questionnaire. Moreover, interviewees were free to express their ideas of any aspect of pronunciation learning. They were also encouraged to give their comments and suggestions for consideration when upgrading the CALL-VT system.

All interviewees regarded CALL-VT as a helpful approach to pronunciation teaching and learning. Of the 10 interviewees, 7 said that CALL-VT was "very helpful" to their pronunciation, whereas 3 of them thought the approach "was helpful to some extent". One even thought that the CALL-VT was much more effective than what she had expected before the intervention: "*I found myself improved amazingly after I learned pronunciation with the CALL-VT system.*" They were asked to give more detailed information about why they thought this approach was helpful and in what ways. As one interviewee said, *"The CALL-VT is helpful in many ways I think. First, it helps me to 'feel' the English language in a different way. I can see the pattern of a sentence better when vowels and consonants are gone. Second, I am more sensitive to the intonation and rhythm when I close my eyes and calm down while listening to the filtered sentences. Third, body-movements and gestures help me to express my understanding of the intonation and rhythm."* Surprisingly, 6 of them said this intervention had a positive impact not only on their pronunciation but also on their ability to listen in English. As one interviewee said, *"whenever I listened to the listening materials, I recalled the pattern we learned in*

the pronunciation class. This helped me grasp the gist of what was playing". Another interviewee said that when she read a passage, she were always aware of the rhythm. She would "dance" with the "waves" the sentences conveyed to her: *"I suddenly felt like dancing to the rhythm of the passage I was reading in the Integrative English text-book. I guess this reaction came from what I had experienced in the pronunciation class."*

All the interviewees thought the CALL-VT system was interesting. They had a good time both in and outside the class. One of them expressed that *"There were a lot of laughers when our friends 'danced' along with the intonation and rhythm. Thomas was like a bear when he walked out the stressed syllables".* However, five of them said that they got confused in the first class. They didn't make any sense of the filtered version of the recordings: *"I couldn't help guessing the meaning of the filtered sentences we listened to but it didn't make any sense to me at all".* After the teacher explained to them, they came to realize what was happening. As one of them said, *"I felt more comfortable when the teacher told us not to guess the meaning of the sentences but to 'feel' the intonation and rhythm of them".* Then, in the class, they found it interesting to move, clap, hum, and dance to the rhythm of the sentences. They both practiced on their own and observed others' performances. After class, they claimed to practice pronunciation with the CALL-VT system either by themselves or with their friends, using materials from different resources.

All interviewees agreed that their pronunciation improved after learning pronunciation with the CALL-VT system. However, one interviewee said, *"I did find my pronunciation improved as the test says and as my friends told me, but I didn't know whether it was because of this approach or because of my high motivation to practice".*

Students liked the CALL-VT system for many reasons. Most of them were curious about this new method. *"It's amazing when I hear my voice filtered by the programme. I can compare the rhythm with the model."* More than one interviewee expressed similar feelings. Byron said that he liked this approach because he could avoid keeping watching and moving his own speech organ movement as he did in the traditional class: *"I felt more relaxed and delighted when I hummed and clapped the rhythm I felt. I hated observing myself in the mirror when I mouthed the word."*

To describe their feelings towards the CALL-VT system in pronunciation, 6 of the interviewees used the word "interesting", and 3 of them used "helpful or beneficial." One of them thought that she believed this approach was effective but that it took time to make it happen. Interviewees were asked to comment and give suggestions in regard to the CALL-VT system. They expressed their wishes to have more software to facilitate their pronunciation learning. Some problems with the experiment were revealed from the data in this semi-structured interview. Two of them commented that it would be better if there were a bigger and quieter classroom.

Thus, they could move more freely. They also wished for a longer length of time for the experiment. As one interviewee said, *"It's a pity that we had only a semester to experience the CALL-VT pronunciation learning system. I wish we could learn pronunciation in this approach later."* Three of them thought that more instruction of theory or explanations would be beneficial since they could learn better if they understood the theories behind this system.

Compared with the traditional approach, all the interviewees preferred the CALL-VT system for learning pronunciation both in and outside the classroom. Most of them said that the traditional approach made them very tired in class because they were required to bring a mirror into the classroom and look at their tongues whenever they were trying to produce a single sound. This made them tired and annoyed them sometimes. *"I felt exhausted every time when I finished a pronunciation class. It seemed I had no energy left for the rest of the week,"* one interviewee complained. In the CALL-VT system, students can clap, hum, and dance with the rhythm. Students favored this. They didn't have to care much about the articulatory organs and their movements which kept them "busy" managing all the elements of articulation they had been taught. Sometimes, they would be overloaded and forget one or more of the elements that they were meant to control. They just listened, imitated, and expressed whatever they felt in their own way. It is worthy to note that most students felt shy when they were encouraged to dance and use gestures to express the rhythm. They enjoyed humming and clapping in class.

Most of them just “hum in their heart”. Shirley said, *“I was less nervous in a CALL-VT class. I just did what I wanted to express. It’s good for me not to care about how I pronounce the word. I believe I did better in this approach.”* After class, they would step on each stressed syllabus and even dance with their friends sometimes. These activities made them enjoy pronunciation learning.

Half the students preferred to learn on their own whereas the other half tended to learn in groups because their friends could help to correct them. They wanted to learn on their own because they felt shy (as mentioned earlier) when they faced their friends. They could behave freely, use gestures and step on stressed syllable in their own way.

All of the interviewees claimed that the CALL-VT system could improve their learner autonomy. A further question of why this would be so was asked. They said that they could apply this approach when they practiced alone.

4.3 Teacher’s opinions of the CALL-VT system

A pre-designed interview questionnaire was administered to the teacher who took care of the experiment to investigate the teacher’s opinions of pronunciation teaching within a CALL-VT system. This teacher had been specially trained in teaching pronunciation. She had been teaching this course for about 5 years. She was involved in the CALL-VT system because she would teach both the experimental group and the control group to reduce the teacher’s impact in teaching.

In her opinion, the CALL-VT system was very helpful and interesting to her pronunciation teaching in several ways. First, the computer, not the teacher, modeled the pronunciation. This freed her from the exhaustion of modelling in class. Second, the students were interested in humming and clapping when listening to the filtered version of the sentences. They had a lot of fun in the "wow" stage when they heard the normal sentence afterwards. Third, the atmosphere was active in class, and this would encourage the teacher's participation to a great extent. Fourth, the teacher found it easy to teach intonation with the help of the CALL-VT system.

She believed that CALL-VT would improve students' pronunciation. However, she was still very surprised by the extent of the improvement of the students in the experimental group. As she said, *"I couldn't believe they (the students in the experimental group) could improve so much. I thought they were weaker than the control group at the beginning. And I found them less cooperative in class."* The teacher did warn that the students in the experimental group were less popular among teachers.

Generally, she liked CALL-VT and she would like to apply this approach in pronunciation. Compared with the traditional approach, she preferred to use CALL-VT because of its flexibility. The teacher would like to give students freedom to learn in groups or on their own. She said that she seemed to lose "control" of the class but it was great to do so since the students' learner autonomy was developed. It is worth mentioning that students learn and practice pronunciation

with “chorus” drills in a traditional class. The teacher was like a conductor in a performance. In a CALL-VT class, students would learn and practice pronunciation in their own ways and at their own pace ideally. For example, they could listen to the filtered version of sentences for many times if it was necessary or if they liked. They could also make as many comparisons as many times as they wished.

4.4 Assessment of students’ learner autonomy

In order to estimate the development of students’ learner autonomy, a written questionnaire was developed on the basis of the inventory of students’ learning autonomy. The questionnaire consisted of 10 items taken from Tassinari’s (2012) dynamic model of learner autonomy. Before the experiment started, the questionnaire was administered. Then, after the 14-week-intervention, the questionnaire was administered again to discover the extent of learner autonomy development. After collecting the answers from the questionnaire, descriptive statistics was calculated as the first step for getting a general picture of the participants’ state of autonomous learning.

As described in the SPSS results (see Table 4.42), the general state of students’ autonomy was not high. Not many students could evaluate their pronunciation ability on their own (mean = 2.21). They evaluated their pronunciation ability mainly together with a language advisor (mean = 3.56) or with a test (mean = 3.50). However, after learning pronunciation with a CALL-VT, many of them could evaluate

their pronunciation ability on their own (mean = 3.25). In terms of setting goals for pronunciation learning, students mainly relied on a learning advisor (mean = 3.17) or a test (mean = 3.17). After 14 weeks, many of them claimed to set pronunciation goals on their own (mean = 3.31). The most promising change was found in the time and place of learning pronunciation. They did not plan their time well (mean = 2.75) or place (mean = 2.75) of pronunciation learning before the intervention. However, after the intervention, most of them claimed to plan their pronunciation learning on their own in terms of time (mean = 3.37) and place (mean = 4.19).

Table 4.42 Descriptive statistics of students' learner autonomy

Items	N	Before		After	
		Mean	S. D	Mean	S. D
1. I can evaluate my pronunciation ability on my own	48	2.21	0.97	3.25	1.10
2. I can evaluate my pronunciation ability together with a learning advisor	48	3.56	0.90	3.79	0.90
3. I can evaluate my pronunciation ability with a test	48	3.50	1.03	3.58	1.07
4. I can set myself goals on pronunciation learning on my own	48	2.85	1.20	3.31	1.11
5. I can set myself goals of pronunciation learning with a learning advisor	48	3.17	1.32	2.87	1.28
6. I can set myself goals of pronunciation learning with a test	48	3.17	1.19	3.35	1.26
7. I can plan a time of pronunciation learning for my learning on my own	48	2.75	1.19	3.37	1.25
8 I can plan a time of pronunciation learning with a learning advisor	48	3.19	1.10	2.85	1.24
9. I can plan a place of pronunciation learning for my learning on my own	48	2.75	0.96	4.19	1.10
10. I can plan a place for my pronunciation learning with a learning advisor	48	3.29	1.17	2.92	1.38
Valid N (listwise)	48				

4.5 Student diaries

Students both in the experimental group and control group were required to keep study diaries. A sample of the diary was presented to students as a guide. As mentioned in **3.3.6**, the information in students' diaries included the time length, place, materials (content), the effectiveness and their feelings of any kinds (both formal and informal ones) of pronunciation learning. Data from the student diaries were analysed thematically.

4.5.1 Student diaries in the experimental group

All 48 notebooks were submitted at the end of the experiment. Information from the diaries were grouped, coded, and categorized to be reported as results.

According to the diaries, students' reported the length of time spent learning pronunciation, both in class and outside class, was about 9 hours per week. This time length is approximately equal to the expected number of hours (10 hours per week). Their study time varied from early morning to late night. Most of them tended to practice pronunciation in the afternoon. More detailed information in terms of study time revealed that that pronunciation learning activities were carried out a greater number of time periods than previously: many of them made use of the few minutes available at dusk and before bed, something that they had never done before. One student mentioned his experience of tapping while waiting in line in the canteen. More interestingly, not a few said that they hummed the intonation patterns in the shower. As one student wrote in her diary, *"I like singing songs when I have a*

shower. Sometimes I also practice English. . One day, I hummed the sentence pattern we learned in pronunciation class and I found it was very funny. . I am not afraid of that someone will hear me practicing English during a shower. I feel good to do that."

Their study place varied from the dormitory, English Corner, the Stone Garden (a garden in the campus), near the lotus pond, to library, classroom, and language lab. Compared with their study place before the intervention, they were found to study in more places than before, many in places previously unfrequented, such as by the fish pond, in the sports ground, and in the garden. One participant even practiced her pronunciation in the gym between work-outs.

The content of their pronunciation exercises was found mainly in the textbooks, listening book, and comprehensive English book. A few reported downloading some materials from the Internet. Data revealed that students were able to obtain more materials of a broader variety on the basis of their preferences and availability. For example, several student diaries indicate the use of additional materials, either text or audio, from different sources such as the Internet and other courses rather than their official textbook. As one student says in her diary, *"The homework for today's speaking class was to make up a conversation with friends. Suddenly an idea came to me. I filtered the conversation I made up and listened to the rhythm of it. It was very interesting. "*

In terms of participants, more than half (about 60 %) of students reported a preference for learning pronunciation in groups but they also liked to practice pronunciation on their own. As one student reported in the feedback section, *"I like to learning pronunciation with my friends in class as it is interesting to move with the rhythms of the sentences. We can clap together, walk hand in hand, and humming. We have a lot of fun then. However, I would like to practice on my own because I want to do it in my own good time. I don't want to have an appointment with my friends."* Another student said she wanted to practice by herself because she would feel embarrassed if other students were there.

In terms of feelings in learning pronunciation, most students used the words "interesting", "enjoyable", "new", "different", "funny", and "happy" to describe their feelings while learning this course via the CALL-VT system. As one student wrote in the diary, *"Time flies in the pronunciation class. We moved, danced, and also laughed a lot in class"*. Many students recorded that they practiced the reading exercise from the comprehensive English class in a verbotonal approach. And they found it was interesting and beneficial for their English study. As one student says, *"I used the verbotonal approach when I read passages in the textbook for comprehensive English. I found that I could read better and understand better in this way. It is interesting to use the new approach in learning English."* Some of them reported that they found it was useful and helpful for their listening learning as well. They could catch the "chunks" of the listening materials easier than before: *"I*

can identify the collocations and patterns in the materials I listened to after I had verbotonal training in the pronunciation class. I guess I find easier to tell where a 'group' of words start and end."

4.5.2 Student diaries in the control group

There were 47 notebooks collected in the control group. Information was categorized, coded, and grouped to be analysed.

Students' reported length of time spent learning pronunciation, both in class and outside class, was about 9.2 hours per week, which was nearly the time length reported by the students in the experimental group. That is, both the experimental group and the control group spent a similar number of hours studying pronunciation. Their study time varied from early morning to late night. Most of them tended to practice their pronunciation in the afternoon.

Their study places included the dormitory, corners on the campus, the library, and the classroom. This was similar to what was reported in student diaries from the experimental group although students in the experimental group seemed to have more variety in terms of study place. The big difference was that they didn't study in the computer lab as the experimental group did.

The content of their pronunciation exercises were limited in their pronunciation book and the training exercises for the pronunciation test of the School of Foreign Languages.

In terms of participants, many (about 70 %) of students reported a preference for learning pronunciation in groups since they would like the group members to help them identify the problems. For example, one student kept her preference for pronunciation learning as: *"My pronunciation is poor. I would like to practice my pronunciation with my classmates because they can correct me when I make a mistake."* Some of them (about 20%) preferred a self-managed learning environment. One student said in her diary: *"I would like to study on my own because I didn't want to lose face in front of my friends."* A few of them (about 10%) could learn either in groups or on their own: *"I don't care about practicing pronunciation in groups or on my own. It all depends."*

Feelings of the students in the control group could be summarised as difficult in general. Nearly all of them felt it was "difficult" to learn pronunciation. They claimed to make great efforts in this course but they were still poor in pronunciation. A passage in a student diary says: *"Pronunciation was the most difficult course I took. I spent most of my spare time practicing but still, I made the least improvement among the courses I took in this term."* Many of them use the words "stressed", "nervous", "burden", and "exhausted" to describe their feelings while learning this course. Kate wrote in her diary, *"I felt disappointed with myself after a half hour's practicing. I didn't know what's wrong with me. I just couldn't pronounce the word like the teacher did. I used a mirror to observe my tongue movement and tried to monitor my pronunciation but I failed to produce a good one. I hate myself. I am not the person who can learn English well."*

It is worthy of mention that the students in the control group used mirrors in the class to look at their tongues and other parts to articulate the sounds. They would compare the positions of their articulatory organs with those of the teacher and tried to imitate what the teacher was modelling. In the articulatory approach, they felt stressed since they couldn't know exactly "where their tongue was" as some students wrote in their diaries. For example, one male student said in his diary: *"I feel embarrassed when I look at myself in the mirror. It doesn't help me to pronounce better. I even cannot move my tongue properly when I am trying to put it in the right position as the teacher models."*

4.6 Answers to research questions

4.6.1 Answer to research question 1: Is the CALL-VT system effective in improving pronunciation learning? If yes, in what ways?

Students' pronunciation ability in the experimental group improved significantly after being subjected to the CALL-VT system. This finding indicated that the CALL-VT system was effective in improving Chinese EFL learners' pronunciation learning. Their production ability improved greatly in terms of comprehensibility, pronunciation, and fluency.

4.6.2 Answer to research question 2: Is there a significant difference in pronunciation improvement between the experimental and the control groups? If so, what is the nature of these differences?

There was a significant difference in pronunciation improvement between the experimental and control groups although both groups improved significantly in their pronunciation ability. In short, students in the experimental group learned pronunciation more effectively than those in the control group. They achieved a higher level of performance in all aspects of their learning despite the fact that the control group was ahead of the start of the experiment.

4.6.3 Answer to research question 3: What are the students' opinions of the CALL-VT system for pronunciation learning?

Students had positive attitudes towards the CALL-VT system for pronunciation learning. They thought that the approach was interesting and effective in improving their pronunciation ability. They enjoyed learning pronunciation via the CALL-VT system. And they believed that this system was beneficial not only in pronunciation course but also in other courses like comprehensive English and listening.

4.6.4 Answer to research question 4: What are the teacher's opinions of the CALL-VT system for pronunciation learning?

The teacher who taught both classed was expert at teaching pronunciation through the traditional approach and she believed in the articulatory approach rather than the CALL-VT system. However, she changed her mind after the experiment. Specifically, she held positive opinions to the CALL-VT system and she regarded this approach an interesting, effective and helpful although it was not easy to conduct in the first stage.

4.6.5 Answer to research question 5: Is student autonomy developed through the CALL-VT system for pronunciation learning? If so, in what ways and to what extent?

Students' learner autonomy was developed through the CALL-VT system to some extent. Students could plan and manage their study time and places better and more freely after the intervention. Data from the students' diaries revealed that they had more freedom to choose their learning materials also.

4.7 Summary

This chapter described the results of the present study. The findings of the statistical analysis revealed that participants in both the experimental group and the control group improved in pronunciation ability over the course of the study. Both Chinese expert raters and native speaker naïve raters were invited to rate the recordings and thus provided valid findings. The results were promising and showed that the experimental group involved in the CALL-VT system was found to perform significantly better than the control group in comprehensibility, pronunciation, fluency, and phoneme production. Further, the findings indicated that both students and the teacher have a positive attitude towards the treatment. Besides, students in the experimental group developed their learner autonomy to an extent after being subjected to CALL-VT approach. The next chapter will present a discussion of the research findings of this study.

CHAPTER 5

DISCUSSION

This chapter discusses the findings reported in Chapter 4. The discussion is organized on the basis of the research questions presented in Chapter 1. First, the results of the pronunciation pretest and posttest are discussed. Second, the students' and the teacher's opinions of the CALL-VT system are explored. Third, the development of learner autonomy is discussed. Finally, a proposed model for pronunciation learning is discussed.

5.1 Discussion of ratings by Chinese experts

This part discusses the findings reported in Chapter 4 about students' overall performances, phonemes, word-reading, passage-reading, and oral interview in terms of comprehensibility, pronunciation, and fluency. Findings from other data sources such as questionnaires and interviews will also be discussed.

5.1.1 Students' overall performances

In terms of the overall assessment, a significant difference was found between the experimental group and the control group in the pretest according to the Chinese expert raters. As reported in 4.1.1.1, the control group (mean = 75.20)

outperformed the experimental group (mean = 70.89). This finding was not surprising as the control and experimental groups came from two intact classes assigned randomly, and the sample could not be manipulated so as to be balanced. This finding was also consistent with the teacher's impression that the control group was stronger than the experimental group in a traditional pronunciation instruction setting.

As reported in Chapter 4, both control and experimental groups improved significantly in the posttest. The control group, whose mean increased from 75.20 to 80.94, demonstrated a significant difference between the pretest and the posttest ($p = 0.000$), with an increase of overall score 5.74 (8.09%). There was also a significant difference between the pretest and posttest scores of the experimental group. The mean changed from 70.89 to 84.93, an increase of 14.04 (20.3%), which is approximately 2.4 times larger than that of the control group. Furthermore, in the pretest the standard deviations (SD) of the experimental and control group were both 8.38. However, in the posttest, the SD for the experimental group was 6.48 whereas the SD for the control group was 9.45. This indicates that the performances for students in the experimental group were more consistent and more predictable than those of the control group (which was less reliable). This means that, as a result of treatment with the CALL-VT system, students were able to perform in a more homogeneous manner, indicating that their learning was also more homogeneous and stable in its content and quality. This enabled students to enjoy a more stable, predictable, and successful learning experience.

Moreover, output from analysis of covariance calculations (ANCOVA) indicates that there was a significant difference between the experimental group and the control group in the posttest when treating the pretest scores as constant. To clarify, the control group was ahead of the experimental group to begin with. However, after the treatment, the experimental group had not only made up the difference with the control group but had also overtaken it by a large margin. These findings indicate that the CALL-VT system is both successful as a pronunciation learning approach and offers a better alternative to the learning of pronunciation than the traditional approach used so far in the university and many other places.

5.1.2 Pronunciation of phonemes

The first part of the pretest and posttest involved reading aloud a list of words written in phonetic script and designed to test the pronunciation of phonemes embedded in words. This part covered nearly all 44 phonemes of English. Statistical results show that both the experimental and control group improved significantly in the posttest. The improvement of the control group in terms of phonemes was expected because they had had plenty of practice in studying individual phonemes when learning pronunciation through the traditional approach which was part of their programme. Students spent many hours practicing the pronunciation of phonemes both in and outside class as reported in their diaries. It is noteworthy and surprising to see that the experimental group's pronunciation of individual phonemes also improved significantly despite the fact that they did not

have any specific training at all in the pronunciation of phonemes within the CALL-VT system. This surprising (and counter-intuitive) result indicates that students in the experimental group improved their pronunciation of phonemes without any work on the study of each individual phoneme. Significantly, too, in the pretest, the pronunciation of phonemes by the experimental group (mean = 15.94) was poorer than that of the control group (mean = 17.44) as reported in 4.1.1.1. However, the experimental group (mean = 18.35) caught up the differences with the control group and outperformed it (mean = 18.12) in the posttest. These findings demonstrate that the holistic approach to pronunciation learning proposed by the CALL-VT system was effective not only in prosody but with individual phonemes by automatically adjusting the pronunciation of individual phonemes without explicit intervention. An analysis of the mechanism of this phenomenon will follow the general discussion of results. In the meantime, it should be made clear that this outcome is entirely non-trivial and may well have a significant impact on pronunciation-learning systems everywhere.

5.1.3 Word-reading

The second part of the pretest and posttest involved word-reading, which could also be regarded as testing segmental features. It covered 50 words required by the syllabus. Statistical results show that both the experimental and control groups improved significantly. The improvement of the control group in terms of word-reading was not surprising because they had had plenty of practice in

pronouncing prescribed individual words in the traditional approach. Students spent many hours practicing reading vocabulary as reported in their diaries. The experimental group also improved significantly even though they did not have special training on the pronunciation of individual word within the CALL-VT system. This promising result indicates that students in the experimental group improved their pronunciation in word-reading without studying any specific individual vocabulary items. Practice in the CALL-VT approach focused entirely on the sentence level rather than on individual words.

Again, it is worth mentioning that the control group (mean = 20.30) was ahead of the experimental group (mean = 18.64) in the pretest. However, once again, in the posttest, the experimental group caught up the initial difference with the control group and overtook it. This demonstrates the power of the CALL-VT system. As reported in 4.1.2, the great improvement in the experimental group in this part indicates the students' progress in pronunciation of individual words, an essential component of communication from a traditional point of view. Output from the ANCOVA calculations show a significant difference between the experimental and control groups in the posttest when treating the pretest scores as constant. In other words, the experimental group improved more in recognizing syllables, word stress, and unusual correspondences between spelling and sound. They were also better able to produce correct grapheme to sound correspondence for all 40+ phonemes of English, including, where applicable, alternative sounds since most phonemes were covered in this part.

Students learned the pronunciation of individual words inductively by developing a “feel” for how they were pronounced in English.

This finding indicates that the CALL-VT system was more effective than the traditional approach in improving word-reading which relies heavily on segmental features. Thus the CALL-VT approach seems more effective than the traditional approach for correcting pronunciation in short language segments without specific training in phoneme-production.

5.1.4 Passage-reading

The third part of the pretest and posttest was passage-reading, in which the fluency, intonation, and rhythm of English were tested in longer contexts. Students were required to read aloud a passage about 150 words. The recordings were rated by the Chinese raters in terms of comprehensibility, pronunciation (nativeness), and fluency. In this part, both the experimental and control groups demonstrated a significant improvement in the posttest. It is noticeable that in the pretest, the experimental group (mean = 21.04) was behind the control group (mean = 22.40). However, in the posttest, the experimental group (mean = 25.76) caught up with and outperformed the control group (mean = 24.83).

The improvement of the control group in terms of passage-reading reveals that the traditional approach was effective in helping students enhance their comprehensibility, pronunciation, and fluency. However, compared to the traditional approach, the CALL-VT system offered a better alternative because of the higher

performance scores of the experimental group. Further, the CALL-VT system was helpful in identifying chunks in reading by enabling the students to develop a sense of the length, order, content and complexity of chunks (Sangarun, 2014). As Valsecchi et al. asserted, native speakers tend to recognize and work with chunks or lexical bundles while foreign language learners do not (Valsecchi et al., 2008). Foreign language learners need to become familiar with the structure and collocations of chunks (thought groups) so as to recognize them and learn to process text efficiently and also develop their spoken or written language. The CALL-VT system makes the students sensitive to the chunks and, as a result, they are better able to read and to produce spoken language based on chunks. The CALL-VT system appears to act on two levels: reading and oral production.

The better performance in terms of comprehensibility, pronunciation and fluency indicates that the students in the experimental group also had better grammar and lexical ability than the students in the control group as a consequence of the above-discussed chunk sensitivity. More supportive evidence was found in the naïve native raters' results to be discussed in the next section.

5.1.5 Oral interview

The interview was a face-to-face oral interview where the students could not prepare any materials in advance. This part tested students' ability to speak English spontaneously in natural settings. In the pretest, there was no significant difference between the control and experimental groups ($p = 0.586$). This finding is

particularly interesting because the control group had significantly outperformed the experimental group in all other aspects of the pretest. With regard to this finding, we surmise that the control group was more accustomed to an exercise-based mode of pronunciation learning involving preparation together with the learning or memorization of set materials, but was not accustomed to normal, face-to-face, language interaction where language had to be retrieved quickly and injected appropriately into fluent speech. It also indicates that the control group was probably adept at conforming to educational practices imposed on them rather than improvising. This finding confirms the existence of problems identified in the preliminary phase of the present study, namely, that students performed well in class exercises but were unable to apply their skills in natural language contexts (B. He et al., 2013).

In the posttest, the experimental group significantly outperformed the control group ($p = 0.001$). This finding is especially promising because the ideal target of pronunciation learning, as proposed in many studies and reviewed in Chapter 2, is to produce comprehensible and fluent English in natural settings. Thus, the CALL-VT system considerably enhanced the student ability to function effectively in natural language contexts. This is a clear and important advantage in normal communication.

Furthermore, as stated by Weir (1990), it was easier to make comparisons across performances in the oral interview because the candidates were asked the same questions. Also, it was useful for eliciting the candidate's ability to use particular

grammatical forms. Thus, the better performance of the experimental group in term of oral interview indicated that they were better not only in terms of pronunciation ability but also in terms of grammatical knowledge.

To sum up, there is strong evidence that the CALL-VT system, as indicated by Chinese expert raters, is significantly more effective than its traditional counterpart. The fact that CALL-VT can be deemed responsible for the improvement of this aspect of students' performances makes it a valuable programme and, perhaps surprisingly for a "pronunciation" programmeme, automatically acts on grammaticality as well as on speed of language retrieval and the production of a stream of naturally-flowing, rapidly sequences languages.

5.2 Discussion of rating results: native speaker

5.2.1 Overall performances

Native speakers rated the reading passages and the oral interview in terms of comprehensibility, pronunciation (nativeness), and fluency and gave individual scores on each of them. In terms of overall pronunciation, it was necessary to consider passage-reading and oral interview scores together since both parts evaluated the suprasegmental features of spoken English.

From the results reported in Chapter 4, in the pretest, in the case of assessments by native speakers, there was no significant difference between the experimental group and the control group in terms of the combined passage-reading

and oral interview. Specifically, these two groups were at the same level in terms of suprasegmental features of pronunciation. Then, in the posttest, the experimental group outperformed the control group significantly. Thus, the CALL-VT system is demonstrably effective in improving students' pronunciation ability in terms of prosody, an aspect of speech production that many researchers believe to be fundamental in improving pronunciation (e.g., Celce-Murcia et al., 2010; T. M. Derwing & Munro, 1997; Hahn, 2004; Molino, 2000).

5.2.2 Comprehensibility: Passage-reading & oral interview

The experimental group was demonstrated to outperform the control group in terms of comprehensibility in passage-reading and oral interview. This finding indicates that the experimental group performed better not only in terms of pronunciation but also in terms of other aspects of English learning. As reported in a study conducted by Trofimovich and Isaacs (2012), accent is uniquely related to aspects of phonology, while comprehensibility is chiefly linked to grammatical accuracy and lexical richness. Therefore, the CALL-VT treatment acted not only on accent but also on grammatical accuracy and lexical richness. This is a remarkable and surprising finding in that it demonstrates an influence on language learning which extends beyond the realm of pronunciation per se, and unexpectedly impacts on several other apparently unrelated aspects of language. Using Trofimovich and Isaacs's (2012) definition, it is arguable that the improvement in comprehensibility could, perhaps, be explained by other aspects of the students' English course.

However, if that had been the case, one would have expected both experimental and control groups to perform equally well. This did not happen. In fact, the experimental group significantly outperformed the control group indicating direct impact of the CALL-VT system on this important feature of speech. A discussion of this phenomenon will be provided in the "reflections" section of this research. To summarize, comprehensibility according to Trofimovich and Isaacs (2012) implies improved accuracy and the experimental group outperformed the control group in comprehensibility.

5.2.3 Pronunciation: passage-reading & oral interview

The experimental group started at a lower level than the control group but overtook it in the posttest in terms of pronunciation. Passage-reading tested the ability to read with correct intonation at the sentence level for materials not requiring any rehearsable linguistic creativity. In this test, students do not need to create language but simply to reproduce someone else's formulation. This is meant to enable load-reduction, and give control by the teacher of what students will be tested on. It also removes the need by students to engage in any form of in-depth linguistic creativity. This finding indicated that the experimental group had better ability in reading sentences.

The better pronunciation means that the experimental group improved in terms of phonemes, word-stress, sentence stress, rhythm, and intonation. Both segmental and suprasegmental features improved after using the CALL-VT system. Therefore,

no matter whether the target of pronunciation instruction focuses on the articulation of consonants and vowels as in the past, or whether it emphasizes stress and intonation as in recent years, the CALL-VT system can be of benefit to students in both cases.

5.2.4 Fluency: passage-reading & oral interview

The control group was ahead of the experimental group in the pretest in terms of fluency but it fell behind the experimental group in the posttest. In other words, the experimental group significantly outperformed the control group in terms of fluency in both passage-reading and oral interview.

As Goodwin (2001) asserted, with the advent of Communicative Language teaching, the focus of pronunciation instruction shifted to fluency rather than accuracy, encouraged a greater emphasis on suprasegmentals. Better fluency in reading aloud or speaking means better performances in producing spoken language in an intelligible and effortless way. The findings in terms of fluency in passage-reading indicate that the experimental group could identify letter-sound correspondences more accurately and quickly than the control group. Students could identify the spelling patterns and also the intonation patterns necessary to decode efficiently. These findings also demonstrated that the experimental group could apply more resources to the difficult task of blending together isolated phonemes to make words phrases and sentences. They could apply knowledge of the alphabetic code more effectively to identify words and to connect text fluently.

As reviewed in Chapter 2, fluency is not an end in itself but a critical gateway to comprehensibility. It bridges word decoding and comprehension. The Oxford English dictionary defines "fluency" as a) speaking or writing in an articulate and natural manner. b) (of a language) used easily and accurately. Good fluency in speaking means that one speaks without hesitating too much (very few ums and uhs; this is actually closer to the core meaning of the word "fluent" as in "flowing"), making very few mistakes (maybe two or three errors per minute of consistent talking), being very easy to understand for others (still with an accent, but not a strong one) and understanding the majority of what is said in normal and casual contexts. Therefore, the findings that the experimental group outperformed the control group meant that the experimental group produced more intelligible English with fewer hesitations and fewer mistakes than the control group. They could produce relatively correct, rich, well-articulated language at a reasonably fast rate.

Better fluency also means better flow. When people produce spoken language, they represent their thoughts using semiotic systems (including language) to communicate with others. Better fluency (flow) in oral interview indicates that the experimental group could produce fast, more effective encoding into spoken language. They retrieved all the necessary pieces of language, putting them together and uttering the message more rapidly and effectively than those in the control group and produced more effective, fast and correct spoken language. In other words, they could represent what they had in their minds better and more effectively than the control group. It also

means that they could produce more comprehensible, richer and syntactically correct language according to Trofimovich and Isaacs's definition (2012).

In summary, the CALL-VT system in contradiction with Goodwin (2001), who sacrificed accuracy to fluency in the name of communicative language teaching, helps students to produce language which is both accurate and fluent.

5.2.5 Correlation between comprehensibility and fluency: native raters

Since comprehensibility and fluency are related in the judgment of pronunciation according to some scholars (e.g., Tracey M Derwing & Munro, 2005; Trofimovich & Isaacs, 2012), it seems appropriate to examine the correlation between those two constructs. One would expect a high correlation between comprehensibility and fluency. Essentially, if someone were highly comprehensible one would expect them to be highly fluent and vice-versa: low fluency would imply low comprehensibility.

It may be possible to be comprehensible without necessarily being fluent, for example, one can produce intelligible language without if flowing smoothly or naturally. Yet, arguably, if language does not flow smoothly or naturally, it may not be comprehensible. But, if one is comprehensible, then according to Trofimovich and Isaacs (2012), one is more grammatically and lexically accurate and rich. Intuitively, comprehensibility and fluency seem to go together as discussed in the previous paragraph. We therefore decided to investigate this hypothesis by examining the correlation between comprehensibility and fluency.

The Pearson correlation coefficient, which measures the strength and direction of the relationship between two variables, was employed to check the correlation between comprehensibility and fluency in ratings by native speakers. A high level of correlation would be set at approximately 0.70 as for other correlation tests. Table 5.1 below shows the correlation between comprehensibility and fluency for both groups in the pretest and posttest.

Table 5.1 Correlation between comprehensibility and fluency in ratings (overview)

Groups	Pretest		Posttest	
	Passage-reading	**Oral interview**	**Passage-reading**	**Oral interview**
CG	0.89	0.89	0.82	0.89
EG	0.47	0.70	0.74	0.93

Pretest: the experimental group: passage-reading

The value of R is 0.47. Although technically a positive correlation, the relationship between the comprehensibility and fluency is weak (the nearer the value is to zero, the weaker the relationship). In other words, the performances of the experimental group were scattered in the pretest. The correlation between comprehensibility and fluency was not strong. This means that comprehensibility and fluency were dislocated and were not properly connected to produce natural sounding language and comprehensibility and fluency were not acting in harmony as one would expect in good speech. In other words, the desirable connection between comprehensibility and fluency had essentially collapsed.

Further, this result can account, in part, for the pronunciation teacher's initial assessment that the control group was stronger than the experimental group and is also reflected in the Chinese expert raters' judgments that the control group was stronger than the experimental group. It may also explain the teacher's surprise when the experimental group outperformed the control group in the posttest.

Pretest: the control group: passage-reading

The value of R is 0.89. This is a strong positive correlation, which means that high comprehensibility scores went with high fluency scores (and vice versa). In other words, the performances of the control group were consistent and comprehensibility and fluency acted together. This means that the control group could produce language where comprehensibility and fluency were strongly correlated and where grammatical accuracy, lexical richness and fluency were in balance. This finding is in line with rating results by both Chinese and native speakers, in which the control group already had a tight connection between comprehensibility and fluency at a desirable level in passage-reading in the pretest.

Posttest: the experimental group: passage-reading

The value of R is 0.74. This is a moderate positive correlation, which means there is now a tendency for high comprehensibility scores go with high fluency scores (and vice versa). This result indicates that the experimental group had learned to perform well in passage-reading with natural-sounding language in which comprehensibility and fluency had good correlation. Moreover, compared to the

value of R in the pretest, it also increases greatly. This mean that the experimental group improved significantly in connecting comprehensibility and fluency and this result confirms the ratings by both Chinese experts and native speakers. In other words, their ability to connect the comprehensibility and fluency was developed to a desirable level.

Posttest: the control group: passage-reading

The value of R is 0.82. This is a strong positive correlation, which means that high comprehensibility scores go with high fluency scores (and vice versa) in the posttest of the control group in terms of passage-reading. This finding indicates that the control group's performance was stable and consistent. However, there was a non-significantly drop from 0.89 to 0.82 in the value of R. This means that the control group's ability to connect comprehensibility and fluency did not increase, nor was it reduced.

Pretest: the experimental group: oral Interview

The value of R is 0.70. This is a moderate positive correlation, which means there is a tendency for high comprehensibility scores go with high fluency scores (and vice versa). This result indicates that, in the pretest of oral interview, the experimental group's performance was not ideally consistent. This finding is also in line with the ratings from the Chinese experts and native speakers, in which the experimental group and the control group were not significantly different. However, the control group was still stronger than the experimental group in the pretest.

Pretest: the control group: oral Interview

The value of R is 0.89. This is a strong positive correlation, which means that high comprehensibility scores go with high fluency scores (and vice versa). This result demonstrates that, in the pretest, the scores of the control group in terms of comprehensibility and fluency were consistent and even. On the one hand, this finding indicated the validity of the ratings. On the other hand, it indicates the stable performance of the control group in the pretest.

Posttest: the experimental group: oral Interview

The value of R is 0.93. This is a strong positive correlation, which means that high comprehensibility scores go with high fluency scores (and vice versa) in the posttest of the oral interview. That is, the experimental group's performance was extremely consistent and stable. Further, compared to the value of R in the pretest (0.70), there is a great increase which indicates that their ability to connect the comprehensibility and fluency was developed significantly.

Posttest: the control group: oral Interview

The value of R is 0.89. This is a strong positive correlation, which means that high comprehensibility scores go with high fluency scores (and vice versa) in the posttest of the control group. This result indicates that the performance of the control group was consistent in terms of oral interview. However, compared to the value of R in the pretest (0.89), which was exactly the same, indicating the control group's ability to connect comprehensibility and fluency had not changed.

Summary

From the above reports on the correlation check in term of comprehensibility and fluency, a summary of the correlations could be made as follows.

First, the correlations between comprehensibility and fluency in the control group were always strong whether in the pretest or posttest in both passage-reading and oral interview. This indicates that the control group's performance was stably harmonious and at a high/ desirable level in both the pretest and posttest. After their course, there is no significant increase in their ability to connect comprehensibility and fluency in both passage-reading and oral interview. It belies Goodwin's (2001) statement that it is necessary to sacrifice accuracy for fluency (or vice-versa). The two can be developed hand-in-hand.

Second, in the pretest, the correlations between comprehensibility and fluency in the experimental group were weak in both passage-reading and oral interview. This indicates that the performance of the experimental group was scattered and less connected. However, the correlations between comprehensibility and fluency in the experimental group became strong in both passage-reading and oral interview now indicating a strong connection between accuracy and fluency which was not present previously. And their ability to connect comprehensibility and fluency in free conversation not only caught up but overtook that of the control group. This means that their ability to connect the comprehensibility and fluency was developed significantly. We can surmise that the CALL-VT system was responsible for that change.

5.2.6 Pronunciation in passage-reading and oral interview

As reported in Chapter 4, naïve native raters provided individual scores in each rating criterion. In order to see in which context students pronounce better, a comparison between the pronunciation in passage-reading and oral interview was made within each group in both pretest and posttest.

Pretest: experimental group

In the pretest, in terms of pronunciation, the experimental group performed significantly better in the passage-reading than in the oral interview ($p = 0.001$) as shown in Table 5.2 below. In other words, the experimental group had slightly better pronunciation in prepared materials than in a free context (oral interview) where they could not prepare in advance.

Table 5.2 Mean of pronunciation in passage-reading & oral interview

Contexts	Mean	Number	Std. Deviation
Passage-reading	9.51	48	1.50
Oral interview	9.13	48	1.16

Pretest: control group

Also in the pretest, in terms of pronunciation, the control group performed significantly better in the passage-reading than in oral interview ($p = 0.000$) as shown in Table 5.3 below. In other words, the control group had slightly better pronunciation in prepared materials than in a free context (oral interview) where they could not prepare in advance.

Table 5.3 Mean of pronunciation in passage-reading & oral interview

Contexts	Mean	Number	Std. Deviation
Passage-reading	10.47	47	1.39
Oral interview	9.81	47	1.86

Posttest: experimental group

In the posttest, in terms of pronunciation, the experimental group caught up the initial difference in the pretest and performed at the same level in the passage-reading and oral interview ($p = 0.636$) as shown in Table 5.4 below. In other words, the experimental group pronounced just as well in a free, unprepared, context (oral interview) as in passage-reading where they could prepare for a few minutes in advance. Also, it was possible for students to read the passage in the testing bank when they prepared for the exam. The fact is that both experimental and control groups could not know the interview questions in advance since these questions were not included in the testing bank that they had been given. Thus, the preparation will no longer influence the test scores. This result showed that the experimental group improved significantly in the oral interview in terms of pronunciation.

Table 5.4 Mean of pronunciation in passage-reading & oral interview

Contexts	Mean	Number	Std. Deviation
Passage-reading	11.51	48	1.81
Oral interview	11.43	48	2.02

Posttest: control group

In terms of pronunciation, the control group continued to perform significantly better in the passage-reading than in oral interview (p = 0.000) as shown in Table 5.5 below. In other words, the control group pronounced better in prepared materials than in a free context (oral interview) where they could not prepare in advance. This finding is in line with the findings of their performances by ratings from Chinese expert raters. As discussed before in 5.1.4, the control group was more accustomed to conforming to educational practices imposed on them rather than improvising (e. g. oral interview).

Table 5.5 Mean of pronunciation in passage-reading & oral interview

Contexts	Mean	Number	Std. Deviation
Passage-reading	11.11	47	1.39
Oral interview	10.54	47	1.52

From the discussion above, it seems that the CALL-VT system helped the experimental group pronounce better in English speaking in a free context. In other words, the CALL-VT system gave learners a set of ready-made skills not requiring preparation, much like those of a native speaker. It is interesting to find that, in terms of pronunciation, the experimental group improved both in passage-reading and oral interview. Further, it is more promising and exciting to find that the experimental group pronounced equally well in both passage-reading and oral interviews despite the fact that they were poorer in the oral interview in the pretest.

No such change was found in the control group. In terms of pronunciation, the control group was better at passage-reading than in the oral interview in both pretest and posttest. After the treatment by a traditional approach to pronunciation, the control group remained better at pronouncing in the passage-reading than the oral interview. In other words, in view of the inequality of performance between prepared and unprepared language productions, it is arguable that the traditional approach was effective in improving students' pronunciation in prepared contexts, but less effective in enabling students to produce well-pronounced language in natural settings: essentially the desired outcome of any pronunciation programme. This result also confirms findings from the Chinese experts' rating who concur with this outcome.

5.3 Discussion of comparison ratings: Chinese experts and native speakers

Both similarities and differences were found in the ratings from Chinese experts and native speakers. The comparison between Chinese experts and native speakers focused on the last two parts of the test since the native speakers rated only the last two parts: reading passages and oral interview.

Generally, both Chinese experts and naïve native raters displayed the same overall direction in their ratings of the recordings. Both groups agreed on whether improvements had occurred or not and on the magnitude of that improvement. This

indicated that, even though one group consisted of experts in phonetics and the other not, they had similar concepts of good and bad pronunciation resulting in consistent ratings, and demonstrating a) shared understanding of what constituted good pronunciation despite a disparity of backgrounds and b) that naïve native speakers are actually capable of making good judgments equivalent to those of trained experts. This shows that both groups share the same set of values and that academic raters share the "real-world" values of naïve listeners and that they are able to support students accordingly. Also, it proved the validity and reliability of the ratings from another perspective.

Similarities were as follows:

a) both Chinese experts and native raters thought that the control group was better than the experimental group in terms of passage-reading. They also found that the experimental group overtook the control group in the posttest.

b) both Chinese experts and native raters thought that the control group and the experimental group started at the same level in terms of oral interview. They also found that the experimental group outperformed the control group in the posttest.

c) in terms of each rating criterion, comprehensibility, pronunciation, or fluency, both Chinese and native raters of English thought that the experimental group overtook the control group although the experimental group started at a lower level.

d) in terms of subtotal scores which combined passage-reading and oral interview, both Chinese experts and native raters thought that the control group and

experimental groups were at the same level in the pretest. However, in the posttest, both Chinese experts and native raters thought that the experimental group outperformed the control group significantly.

Table 5.6 Sub-total score of passage-reading and oral interview: Chinese experts

Tests	Groups	Mean	Number	S.D	p-value
Pretest	EG	36.31	48	4.87	0.279
	CG	37.38	47	4.65	
Posttest	EG	44.32	48	4.40	0.005
	CG	41.28	47	5.77	

Table 5.7 Sub-total score of passage-reading and oral interview: native raters

Tests	Groups	Mean	Number	S.D	p-value
Pretest	EG	19.69	48	2.56	0.057
	CG	20.67	47	2.38	
Posttest	EG	24.38	48	2.68	0.004
	CG	22.26	47	3.44	

The only difference found was that the means from the Chinese experts were higher than those of the native raters' although they rated in the same direction in terms of both passage-reading and oral interview. This demonstrated that although both Chinese experts and native raters shared similar notions regarding the concept of pronunciation the naïve native raters were actually stricter in scoring.

Taken together, the above findings clearly indicate the success of the CALL-VT system in improving not only the mechanical, articulatory, aspects of Chinese EFL learner's pronunciation but also comprehensibility and fluency, aspects

of speech production not directly correlated with articulation per se but, as discussed below in the "reflections" section, based on the development of syntactic and lexical abilities (comprehensibility) together with retrieval and production functions of language resulting in a stream of language(fluency).

This confirms that good articulation of phonemes is in fact not sufficient to enable persons to function socially even though it is often. This has been the focus of traditional pronunciation-teaching classes. Rather than producing sequence of correctly pronounced phonemes, what is needed is acceptable sounds embedded in syntactically and lexically acceptable language delivered as a flow at an acceptable rate. Therefore, the CALL-VT system does much more than get people to pronounce well.

5.4 Students' and teacher's perceptions of the CALL-VT system

5.4.1 Students' perceptions

Generally, students displayed a positive attitude towards the CALL-VT system though some of them thought it was difficult to apply in the early phase of the treatment. In general, the CALL-VT approach was perceived as very enjoyable and interesting. This reinforces Green' research finding which indicates a direct correlation between the enjoyment of an activity and its effectiveness in learning (Green, 1993). This finding also partially explains the great improvement of the experimental group in all aspects of pronunciation.

Thus, we can surmise that, with a positive attitude and because of the fun it involves, students were more willing to engage in pronunciation learning with the CALL-VT system. This is, in turn, contributed to pronunciation learning.

5.4.2 The teacher's perceptions

The teacher in charge of the intervention received training before beginning the experiment. In a face-to-face oral interview, she described her opinions of the CALL-VT approach as "enjoyable, helpful, and doable". Her liking of this approach encouraged the research from the teacher's perspective. Also, the popularity of the CALL-VT system promised the availability of this approach in pronunciation learning. Most noteworthy, though, the teacher was committed to CALL-VT's "competitor", the articulatory approach. The fact that she changed her mind is critically important.

5.5 Development of learner autonomy

The study indicates a positive impact on the development of students' autonomy when learning pronunciation through the CALL-VT system. Further discussions follow.

First, students' motivation for learning pronunciation was high. They were interested in learning pronunciation and in practicing pronunciation as indicated by an opinion survey of the CALL-VT system. This considerably enhanced their level of autonomy.

Second, students were free to use any materials which they chose according to their interests and their availability. This made their autonomous learning more possible. They could listen to and compare filtered sentences (produced by *Audacity*) with unfiltered sentences. Thus they were able to reinforce their "feel" for English with their understanding of the intonation and rhythms.

Third, students had considerable flexibility in choosing the time and place of pronunciation learning. This is a sign of autonomous learning. Further, students were able to engage in more self-regulated learning than before since they found more time periods and places to do so. In a sense, their personal preferences were better realized hence the rhizomatic learning.

The role of learning advisers in learner' autonomous learning has been highlighted in many studies (e.g., Mozzon-McPherson, 2007). The contributions of language learning advisers to the development of learner autonomy cannot be overlooked and some studies have indicated that learner autonomy was dependent on teacher autonomy (Little, 1995). A higher mean (mean = 3.79) of the dependence on a learning advisor when evaluating their pronunciation ability demonstrated that students' autonomy was developed to some extent. However, when it came to planning of study time (mean = 2.85) and place (mean = 2.87), students tended to plan them alone, without a learning advisor.

The above analysis suggests that the CALL-VT system was beneficial to the development of students' autonomous learning as well as the improvement of pronunciation ability.

There were only four male students among the 48 participants in the present study. Thus there was little possibility of conducting a statistical analysis on the answers from the questionnaire. For further research on learner autonomy, similarities and differences in gender in autonomous learning activities might be an interesting topic as Nunan (1992) argues that it is not absolute that all learners will develop autonomy in the same way and to the same degree. Varol and Yilmaz (2010) have conducted research on similarities and differences between female and male learners in terms of autonomous language learning activities, showing more similarities between male and female learners.

5.6 Brief reflections on the CALL-VT system

In this section we discuss aspects of this research which seem of relevance but which may extend beyond the scope of the research questions as there were a number of interesting aspects emerging from the research. Thus the following reflections on the CALL-VT system in relation to its theories and principles will provide a deeper and more complete picture of the present study.

5.6.1 Well-foundedness of verbotonal theory

Both Chinese experts and naïve native speakers assessed the experimental group as outperforming the control group in terms of all aspects tested: phonemes, word-reading, passage-reading and oral interview. This finding is a clear indication that the CALL-VT system is more effective than the traditional approach (an

articulatory approach) to pronunciation development currently used at Xingyi Normal University for Nationalities and many other universities around the world. It is clearly better practically and consequently, it may be deduced that it is also better theoretically. While it is part of the dogma of verbotonalism to assert that the study of prosody will simultaneously adjust the pronunciation of many if not all individual phonemes, this seems to be the first data-driven study to confirm the hypothesis. It is certainly the only study to do so in the context of Chinese EFL learners. These findings confirm the well-foundedness of the theoretical premises of verbotonalism.

5.6.2 Improvement in phoneme production: correct prosody stimulates correct pronunciation of individual sounds

It is particularly worth noting that the experimental group outperformed the control group in phoneme production ***without*** undergoing any training on individual phonemes. The control group, on the other hand, did undergo specific training in phoneme production. While this finding is part of the rhetoric of verbotonalism, it remains an exciting finding. It means that specific training in individual phoneme production is not required or is greatly reduced if prosodic training is implemented first. The only phoneme training is targeted only at those phonemes that remain incorrect after completion of prosody training. It seems that, as predicted by standard verbotonal theory, a holistic prosodic and gestural approach, prosody training automatically adjusts the pronunciation of individual sounds simply by developing sensitivity to the global tension system of the language (Guberina, 1971; A.-P Lian,

1980). Therefore, by internalizing the prosodic features of language, the phonemes were internalized as well. In other words, the CALL-VT system acted on parts (individual sounds) by acting on the whole (intonation and rhythm). Supporting evidence may be found in studies conducted by other researchers (e.g., Christiane Dalton & Seidlhofer, 1994; Guberina & Asp, 1981; Leppänen et al., 2010). This finding could have a very significant impact on the teaching of pronunciation resulting in a substantial reduction of work and effort by both teachers and students who will not need to study the detailed articulation of sounds but will acquire them directly from the prosody (i.e., without teacher intervention).

A possible explanation for this phenomenon follows: In the classroom, students listened to and practiced (repeated) sentence level utterances while focusing on rhythm and intonation. By removing the consonants and vowels through low-pass filtering, students' attention to individual sounds was bypassed automatically since there were only intonation and rhythm left. The prosody of utterances/sentences necessarily depends on individual sounds, each of which contains a contribution to the prosody of the sentence as a whole. This is especially the case for vowels and their F_0 formants.

When we speak naturally, we never produce prosody in isolation. Prosody is always linked to the production of individual phonemes or language sounds. In fact prosody is produced by the fundamental frequency (F_0) of individual sounds which then connects to the F_0 of the following sounds in the phonetic chain. This, in essence, is the phenomenon of co-articulation.

In keeping with the notion of VT theory, each sound has an optimal range of prosody phenomena (Guberina, 1972; Mildner & Bakran, 2001) beyond which any perception and, therefore, production of the sound would be distorted. These prosodic events, even though they are produced by what we call individual sounds, contribute to the characteristic melodies, rhythms and stress patterns of the language. In this perspective, individual sounds do not exist as discrete, isolated entities but are part of a greater dynamic of sound, an envelope, that they actually help to create.

To put it another way, individual sounds contribute to the overall variations in melody and tensions of the language. Correctly pronounced sounds will produce correct tensions. Incorrectly pronounced sounds will produce incorrect tensions.

In the CALL-VT system, the focus is on filtered intonation patterns of correct (never incorrect) language. These patterns necessarily carry the correct tensions of the language and therefore reflect the correct pronunciation (and tensions) of individual sounds. The filtering process separates these out from the actual articulation of sounds, makes them salient and places the focus on the correct melody and tensions rather than on the mechanics of articulation. These correct tensions are then learned and integrated into the students' perceptual systems at a holistic level. They are learned independently of any sound associated with them and integrated independently, although clearly associated with the language being learned.

In the CALL-VT approach, listening and exercising with the filtered patterns occurs first as part of the sentisation process. Then the words are added back through a gradual process.

When the words are added back to the filtered language, integrated, prosodic patterns which are now well rehearsed, predictable and expected by the students within the lesson context are actually imposed (or forced) onto the individual sounds, thus helping to (a) adjust students' (self-)perceptions of each sound in its dynamic context and (b) providing strong clues as to the value of F_0 for each sound in that particular context which, together, (c) will help the student to produce correct sounds. Articulation is considered by verbotonalism to be relatively unproblematic once good perception has occurred. This is heavily assisted by the fact that, at least in the beginning stages of the exercise, the prosody is provided as a background to sound/word production in much the same way as, in a song, music provides the background against which the words as sung. This is a very much a top-down approach which emphasizes the realities of speech and its dynamic nature. It is in marked contrast with the bottom-up approach characteristic of traditional articulation-based models of pronunciation learning which uproot sounds from their natural settings and place them (at least in the beginning) into unnatural perceptual and articulatory contexts requiring the performance of articulatory gymnastics rather than highlighting perception. Thus, the focus on prosodic elements of language "automatically" enables learning of the correct pronunciation of individual sounds and explains the intuitively unexpected outcome of phonetic improvement with no explicit phonetic training.

5.6.3 Problems of the traditional approach identified in the preliminary study

A major problem with the traditional approach at Xingyi Normal University for Nationalities is the fact that students take the teacher as their only model (B. He et al., 2014) although sometimes supported with pre-fabricated recordings. According to data from student diaries, students taught by the traditional approach (the control group) were required to memorize numerous rules and were required to perform many memorization exercises. As a result, when they had the time to practice this knowledge at leisure, they were able to perform well (e.g., as in the reading-aloud components of the pretest and the posttest). Unfortunately, they could not perform well in a natural English speaking context where there was pressure to mobilize knowledge quickly as they could not easily retrieve what they had in mind, encode it quickly and present it fluently and correctly. This did not happen in the case of the experimental group who did much better in the face-to-face high pressure contexts of the oral interview test. This finding confirms the existence of problems in pronunciation learning under a traditional approach and suggests the need to reduce memorization as much as possible. And also, this problem motivates the present study to focus on students' performances in speaking English in natural settings.

5.6.4 Comprehensibility and fluency

The findings in terms of comprehensibility and fluency in both passage-reading and oral interview show that the experimental group performed significantly better than the control group in the production of comprehensible

language in both mechanical/rehearsed settings and natural settings. The control group lagged behind. This indicates that the experimental group was better at rapidly retrieving grammatical and lexical items (comprehensibility (Trofimovich & Isaacs, 2012)) and injecting them quickly into the speech flow (fluency (Trofimovich & Isaacs, 2012)). This is a surprising finding. How can a pronunciation improvement programme result in better syntax, vocabulary and a higher rate of retrieval and delivery of language items?

5.6.5 Correlation between comprehensibility and fluency

Logically, one would expect comprehensibility and fluency to have a close relationship. This can be demonstrated by correlation calculations. As accuracy and richness improve so should fluency. This was certainly the case with the control (initially better-performing) group on the pronunciation tests. In the initially less well-performing experimental group there seemed to be a dislocation between comprehensibility and fluency. The CALL-VT system clearly helped to restore the expected connections and, again, the experimental group overtook the control group and produced a better result: greater grammaticality and lexical richness with speed of retrieval and delivery. As one ability grows through training, so does the other. In the kind of development provided by CALL-VT, accuracy and fluency go hand in hand with one another resulting in high levels of spoken proficiency in natural spoken settings. This result also confirms the ability of CALL-VT to connect comprehensibility and fluency.

5.6.6 Load lightening process in CALL-VT

The concept of load lightening was introduced earlier in this research. In the specific context of the CALL-VT classroom activities, load-lightening can be considered to be a 2-stage process: Stage 1, when students listened to the filtered sentences (the melody), their right brain was preferentially activated: melody is treated by the right hemisphere. The intonation patterns keep acting on the right brain by repeated exposure to the filtered recordings. It is arguable that students could have better understanding of the structure of intonation patterns since they become salient and exercises connect patterns to general meanings.

In stage 2, after the intensive sensitisation of stage 1, the words were added back and the normal left brain activities were activated as discussed before and left and right hemispheres resumed coordinated processing.

5.6.7 Reflections on raters

Both Chinese expert raters and naïve native speaker raters scored student performances in similar ways, thus showing that they shared essentially the same standards for judging students. The consistent ratings indicate that (a) Chinese expert raters are in fact able to make real-world judgments on pronunciation. They do not live in an isolated academic world. (b) Naïve native raters can in fact provide reliable assessments of pronunciation-related phenomena in the same way that experts do although they are ordinary speakers of English with no special expertise in pronunciation. Interestingly, the native raters seemed to be stricter than Chinese experts in their judgments since they tended to give lower scores than the expert raters.

In short, it seems that both Chinese expert and non-expert native speaker raters can make good judgments about the pronunciation of speakers. Being an expert gives no special advantage and being a naïve rater creates no special disadvantage.

5.6.8 Body movements and gestures

Body movements were involved in the process of sensitization and reinforcement in pronunciation learning in this research. While body movement is seen as an integral part of the VT system (Guberina, 1972), our findings confirm the findings by Condon and Ogston that self-synchrony exists and, more importantly, can be used for the purpose of improving pronunciation (Condon & Ogston, 1971). As reviewed earlier, Condon and Ogston asserted that speech and body were in synchrony when people produced or even, listened to speech.

The findings also show that the phenomenon of self-synchrony can be used to support the learning of pronunciation. Instead of the body following the language (as is normal), the language can be made to follow the body in order to generate the expected and desirable self-synchrony between body and language.

5.7 Summary

This chapter discussed some of the important findings arising from the present study in reference to the research studies and theories relevant to these findings. Chapter 6, the final chapter, will discuss the limitations of the study, establish the pedagogical implications and suggest some directions for further research.

CHAPTER 6
CONCLUSIONS, IMPLICATIONS AND RECOMMENDATIONS

This chapter concludes the dissertation. It is organized into four sections. Section one summarizes the present study, including the major findings and procedures. Section two considers the implications of the study for EFL pronunciation learning and its possible applications to instruction. Section three describes the strengths and limitations of the study. Finally, section four proposes recommendations for further research in EFL pronunciation learning.

6.1 Summary of the study

The present study was conducted to examine the effects of the CALL-VT system on Chinese EFL learners in relation to pronunciation learning and to depict their perceptions of this approach. Together with the check of effectiveness of this system, students' and teacher's perceptions of the system and the development of student autonomy were also investigated. The study employed a mixed methods design. A quantitative framework was used to assess the students' pronunciation performances at the beginning and at the end of the pedagogical intervention, as well as students' perceptions and their level of autonomy. A qualitative framework was

used to explore the students' and teacher's perceptions of the CALL-VT system and to analyse information from student diaries. As an aide-mémoire for the reader, the research questions are re-stated below in this study.

- Is the CALL-VT system effective for pronunciation learning? If yes, in what ways?
- Is there a significant difference in pronunciation improvement between the experimental and the control groups? If so, what is the nature of these differences?
- What are the students' opinions of the CALL-VT system?
- What are the teacher's opinions of the CALL-VT system?
- Is student autonomy developed through use of the CALL-VT system? If so, in what ways and to what extent?

In order to examine these questions, a quasi-experimental design consisting of pretest-treatment-posttest was used. The duration of the treatment was 14 weeks (28 hours) distributed through an 18-week semester, together with approximately 9 hours per week of self-regulated private learning. The tasks undertaken by students involved the use of low-pass filtered sentences to enhance perceptions in both in and out of classroom activities. The measuring instruments used in this study were a pronunciation test, students' written questionnaire, face-to face interview, questionnaire on learner autonomy, and student diaries. Four Chinese expert raters and four naïve native speakers of English were invited to assess students'

performances in pre and posttest. The 96 participants of this study belonged to two intact classes enrolled in the pronunciation course at Xingyi Normal University for Nationalities, China, in the second semester of the 2014 academic year. ANCOVA and *t*-test were performed, using SPSS 16.0, to analyse the test scores, and answers from written questionnaires were analysed by a one-way ANOVA.

Major findings of the study were summarized as follows:

First, the CALL-VT system was effective in improving Chinese EFL learners' pronunciation ability in terms of both suprasegmental (prosody) and segmental (phonemes) features without training in segmental features. Students' pronunciation was also significantly improved in terms of comprehensibility, pronunciation, and fluency. The experimental group overtook and outperformed the control group on all aspects of the study as assessed by both expert and naïve raters even though the control group was significantly ahead of the experimental group at the beginning of the experiment.

Second, the CALL-VT system offers a better alternative to pronunciation instruction for Chinese EFL learners compared to the traditional approach since students in the experimental group overtook and outperformed the control group in all aspects of pronunciation even though the control group was significantly ahead of the experimental group at the beginning of the experiment.

Third, the experimental group significantly outperformed the control group in comprehensibility and fluency, indicating that their performance was significantly

better in both syntactic and lexical areas (Trofimovich & Isaacs, 2012) when faced with the difficult task of producing natural, face-to-face, language.

Fourth, students regarded the CALL-VT system as an enjoyable environment where they could learn well and happily. They also believed that the system was beneficial to their English learning in general including listening. They claimed to experience a greater sense of peace and harmony.

Fifth, the teacher although sceptical at first and committed to the traditional articulation-based approach also expressed positive opinions about this system. She preferred the CALL-VT system over the articulatory approach not only because of its effectiveness in pronunciation learning but also because it relieved the teacher from the exhaustion of repeated modelling.

Sixth, students' learner autonomy was found to be increased. They could choose the time, place and materials for study according to their availability and personal preferences and could construct their own lessons or even generate their own filtered materials (including recordings of themselves) if they wished.

Finally, data from the student diaries revealed that members of the experimental group learned pronunciation more happily and in a more enjoyable way than those in the control group.

In summary, the CALL-VT system was effective and enjoyable to use and led to the enhancement of Chinese EFL learners' pronunciation learning while also revealing some exciting additional and unexpected findings. It also appeared to be

beneficial to their English learning in general. Furthermore, learner autonomy was developed through the system though much improvement is possible. These findings also suggest the well-foundedness of verbotonal theory as well as rhizomatic learning theory.

6.2 Pedagogic implications

The research findings point to some significant implications for English as a foreign language programmes for Chinese university students.

1. Teachers and other stakeholders should rethink the teaching of pronunciation to Chinese EFL learners.

Results in this study are sufficiently compelling to suggest the possibility of re-evaluating current practices in the teaching of English pronunciation to Chinese EFL learners. A change in content and approach may be desirable. In turn, this will have implications for EFL teachers and other stakeholders such as the government and other educational policy makers. Policy-makers and educators should consider at least three things. The first is to conduct a replication of the current study on a larger scale in order to test the validity and generalizability of the present findings. The second is to enrich teaching approaches and materials on the basis of the results of this study since students expressed disappointment with the current approach, especially the pronunciation book which was full of rules to be memorized and, consequently, created overload. A more effective alternative now

seems available. The third is to consider supporting the development of English pronunciation, listening, grammar and lexical abilities on the same principles as those of the CALL-VT system (since it seems that students' grammatical and lexical abilities were also automatically improved during the process of pronunciation learning).

2. The CALL-VT system may be applicable to English pronunciation teaching for native speakers of languages other than Chinese (Mandarin)

The CALL-VT system was effective with native speaker of Chinese. Chinese is a tonal language. This may have facilitated or enhanced the process. It would be interesting to investigate the effects of the CALL-VT system with native speakers of other languages such as Thai or Vietnamese. While some successes have been reported with native speakers of non-tonal languages (e. g., French to English speakers, (A.-P Lian, 1980)), which focuses on intonation and rhythm in general, might be also effective to EFL learners whose mother tongue was a tonal language like Chinese and Thai. Also, there is a need to conduct comparative studies based on the CALL-VT system between tonal and non-tonal alphabetical languages or non-tonal but character-based languages like Japanese.

3. Impact on theories of learning

This study concerns how learning happens and how learners optimally construct knowledge. The positive results achieved so far encourage us to rethink current learning theories. In particular, this study raises the issue of the right-brain

and left-brain relationship and the question of fixed-content curriculum and the possibility of students' self-construction of a curriculum based on need (as in the rhizomatic approach) rather than using a pre-organised, one size-fits-all model for both content and pedagogy. Critical to the success of this new approach will be the issue of awareness-raising and the discovery of optimal approaches for doing so (perhaps similar in spirit to the awareness-raising procedures used in the CALL-VT system).

4. Implications for teacher training

Results from the present study have clear implications for teacher training especially for a normal university like Xingyi Normal University for Nationalities. In terms of pronunciation instruction, pre-service and in service teachers should be aware that pronunciation can be taught with good results using the principles built into a system like CALL-VT and, furthermore, that it may be possible to extend these principles into other non-phonetic areas of language-learning.

5. Students' self-managed pronunciation learning

Students' development of learner autonomy has implications for self-regulated pronunciation learning. In a system like CALL-VT, unlike traditional models, students do not have to rely heavily on the teacher. As reported in Chapter 1, in a previous study, even when the government doubled the class time for pronunciation, students' pronunciation performances were still not satisfactory (B. He et al., 2014). Using a self-managed approach (e. g., CALL-VT) may, under the right circumstances,

be more profitable, flexible, and reliable system with students learning in their preferred, self-directed, autonomous way.

6.3 Strengths and limitations of the study

This study triangulated data collected from many sources including pretest, posttest, questionnaires, oral interviews and student diaries. Triangulation of multiple measures enabled the researcher to verify the research findings. Triangulating quantitative and qualitative data as well as methods contributes to a better understanding of the effects of the CALL-VT approach and provides an overall picture of how students and the teacher evaluated the whole system. Arguably, triangulation of the data collection procedures made the study more rigorous and the results more reliable.

Although this study yielded many promising and, in some cases, surprising insights and perspectives into the improvement of English pronunciation and other language-learning phenomena, some limitations should be addressed.

First, the extent of the data collection procedures gave rise to complaints by computer lab support staff about increased workload and this, in turn, reduced opportunities for students to exercise their autonomy. It is possible that results could have been better than they were.

Second, technically, the classroom could have been darkened more easily and been made quieter to create better relaxation conditions for the students.

Third, the participants were chosen on the basis of convenience and availability. The samples were not randomized and balanced and learners were participants in the study on the basis of their classroom enrollment. Consequently, there were not equal numbers of male (8) and female (87) students. Even though an equal number of males and females is not mandatory in studies of gender difference (e.g., Brantmeier, 2002, 2003, 2004; Young & Oxford, 1997), the unequal numbers of male and female students still represents a limitation. Besides, the experimental group and the control group were assigned randomly. Consequently, they performed differently in the pretest although they were assumed to be equal since they came from the same group of students who had been randomly assigned to different classes. The ANCOVA statistical procedure was used to compensate for this factor. Further, the fact that the experimental group outperformed the control group despite being behind the control group to begin with is significant in itself.

Fourth, the participants in this study were 96 first-year undergraduate English majors in Xingyi Normal University for Nationalities, China, who were advanced EFL learners. Students from other majors and other levels were not included in this study. Thus the findings of this study should be treated with caution in relation to generalizations.

6.4 Recommendations for further research

In light of the limitations discussed above and remaining issues regarding the CALL-VT approach, more research should be conducted to further explore the effects of this pronunciation learning system.

First, the study was a preliminary attempt to improve Chinese EFL learners' pronunciation. A large-scale replication study is clearly needed.

Second, this study set out to help students learn pronunciation in an autonomous learning environment. Due to limited access to the Internet in the laboratory, the research suggests that further empirical studies could be conducted so as to provide more opportunities for students' self-regulated, rhizomatic learning.

Third, the participants in the present study were first-year university students, who were already relatively advanced EFL learners. As a result, the interpretation and generalizability of these findings are limited. More empirical studies could be conducted at other levels. Future research may be targeted at different-sized groups of students or to other levels.

Fourth, because of government regulations, the predesigned learning materials for the experimental group were selected from a book that the control group was using although the experimental group had to be reported using other materials. Students were not totally free to choose what they wished to learn though they could supplement this core with other materials found on the Internet and elsewhere. Therefore, empirical studies focusing on the nature of materials for learners are recommended.

Fifth, when analyzing the students' recordings, the researcher found that they were a great source for identifying difficulties and problems of Chinese EFL students as the tests covered almost all phonemes and most suprasegmental features of English. Further analysis and categorizations of these problems in the context of a formal error analysis is strongly recommended. Results would certainly be informative. Error analyses of this kind as an aid to diagnostic evaluation would also be valuable.

Sixth, according to the verbotonal theory (Mildner & Bakran, 2001) every sound, word and sentence has its optimal frequency band. It would be interesting to investigate this claim and identify where those optimal bands are situated (for any specific language). Identifications of these bands will be extremely beneficial for correct pronunciation as they could then serve as the basis for "corrective" optimals (Renard, 1975).

Seventh, there is a need to develop learner analytics in order to track students' path through the system and give them additional support and assistance as and when they need it. While only probabilistic, the data collected here could help form the basis for a range of tools to assist students with.

A final suggestion for future research is that more variables should be considered. In addition to gender, age, and major, learners with different mother tongues or of different ethnicities could be investigated with potentially valuable results.

These remarks bring the dissertation to a conclusion. Some questions have been answered (at least tentatively), many others have been raised to be tackled in the future. Hopefully, the questions answered provide us with a principled starting point for improving the English pronunciation of Chinese EFL majors.

REFERENCES

REFERENCES

Abusabha, R., & Woelfel, M. L. (2003). Qualitative Vs Quantitative Methods: Two Opposites That Make a Perfect Match. *Journal of the American Dietetic Association, 103*(5), 566-569.

Andresen, J. T. (1990). Skinner and Chomsky Thirty Years Later. *Historiographia Linguistica, 17*(1-2), 145-165.

Archee, R. (2012). Reflections on Personal Learning Environments: Theory and Practice. *Procedia-Social and Behavioral Sciences, 55*, 419-428.

Asp, C. W., Guberina, P., & Pansini, M. (1981). *The Verbo-Tonal Method for Rehabilitating People with Communication Problems*. New York, NY: World Rehabilitation Fund.

Asp, C. W., Kline, M., & Koike, K. J. (2012). Verbotonal Worldwide. In R. Goldfarb (Ed.), *Translational Speech-Language Pathology and Audiology*. San Diego, CA: Plural Publishing

Bax, S. (2003). Call—Past, Present and Future. *System, 31*(1), 13-28.

Beatty, K. (2003). *Teaching and Researching Computer-Assisted Language Learning*. Edinburgh: Pearson Education Limited.

Benson, P. (2001). *Teaching and Researching Autonomy in Language Learning*. London: Longman.

Benson, P. (2007). Autonomy in Language Teaching and Learning. *Language teaching, 40*(1), 21-40.

Benson, P., & Voller, P. (1997). *Autonomy and Independence in Language Learning*. London: Longman.

Berthon, P., Pitt, L., & Watson, R. T. (2000). Postmodernism and the Web: Meta Themes and Discourse. *Technological Forecasting and Social Change, 65*(3), 265-279.

Blin, F. (2004). Call and the Development of Learner Autonomy: Towards an Activity-Theoretical Perspective. *ReCALL, 16*(02), 377-395.

Boje, D., Gephart, R. P., & Thatchenkery, T. J. (1995). *Postmodern Management and Organization Theory*. NY: Sage Publications.

Bourdieu, P. (2005). Habitus. *Habitus: A sense of place, 2*, 43-49.

Brantmeier, C. (2002). Second Language Reading Strategy Research at the Secondary and University Levels: Variations, Disparities, and Generalizability. *The Reading Matrix, 2*(3), 1-14.

Brantmeier, C. (2003). Does Gender Make a Difference? Passage Content and Comprehension in Second Language Reading. *Reading in a foreign language, 15*(1), 1-27.

Brantmeier, C. (2004). Gender, Violence-Oriented Passage Content and Second Language Reading Comprehension. *The Reading Matrix, 4*(2), 1-19.

Brazil, D., Coulthard, M., Johns, C., & Johns, C. (1980). *Discourse Intonation and Language Teaching*. London: Longman.

Breitkreutz, J., Derwing, T. M., & Rossiter, M. J. (2001). Pronunciation Teaching Practices in Canada. *TESL Canada Journal, 19*(1), 51-61.

Brewer, J., & Hunter, A. (1989). *Multimethod Research: A Synthesis of Styles*. Newbury Park, CA: Sage Publications, Inc.

Brown, A. (1992). Twenty Questions. In A. Brown (Ed.), *Approaches to Pronunciation Teaching*. London: Macmillan.

Brown, G., & Yule, G. (1983). *Teaching the Spoken Language*. Cambridge: Cambridge University Press.

Brown, H. D. (2006). *Principles of Language Learning and Teaching*. New York: Pearson Education.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. Cambridge: Cambridge University Press.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching Pronunciation Hardback with Audio Cds (2): A Course Book and Reference Guide*. Cambridge: Cambridge University Press.

Celce-Murcia, M., & Olshtain, E. (2000). *Discourse and Context in Language Teaching: A Guide for Language Teachers*. Combridge: Combridge University Press.

Cenoz, J., & Lecumberri, M. L. G. (1999). The Acquisition of English Pronunciation: Learners' Views. *International Journal of Applied Linguistics, 9*(1), 3-15.

Chambers, A., & Bax, S. (2006). Making Call Work: Towards Normalisation. *System, 34*(4), 465-479.

Charles, C. M., & Mertler, C. A. (2002). *Introduction to Educational Research*. Boston: Allyn and Bacon.

Chen, H. (2006). Intonation Phrasing in Chinese Efl Learners Read Speech. *Proceedings of the SP*.

Chen, H. (2013). Use of Praat in Visualization of English Pronunciaiton Teaching. *Journal of Lanzhou Institute of Education, 29*(9), 53-54.

Cohen, A. D. (1998). *Strategies in Learning and Using a Second Language*. London: Routledge.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research Methods in Education*. London: Routledge.

Condon, W. S., & Ogston, W. D. (1966). Sound Film Analysis of Normal and Pathological Behavior Patterns. *The Journal of Nervous and Mental Disease, 143*(4), 338-347.

Condon, W. S., & Ogston, W. D. (1971). Speech and Body Motion Synchrony of the Speaker-Hearer. *The perception of Language*, 150-184.

Cotterall, S. (1995a). Developing a Course Strategy for Learner Autonomy. *ELT journal, 49*(3), 219-227.

Cotterall, S. (2000). Promoting Learner Autonomy through the Curriculum: Principles for Designing Language Courses. *ELT journal*, *54*(2), 109-117.

Cova, B. (1996). What Postmodernism Means to Marketing Managers. *European Management Journal*, *14*(5), 494-499.

Creswell, J. W. (2003). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks: Sage Publications.

Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. London: Sage.

Dörnyei, Z. (2001). *Teaching and Researching Motivation*. Harlow: Pearson Education Limited.

Dörnyei, Z. (2003). *Questionnaires in Second Language Research: Construction, Administration, and Processing*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Dabbagh, N., & Kitsantas, A. (2012). Personal Learning Environments, Social Media, and Self-Regulated Learning: A Natural Formula for Connecting Formal and Informal Learning. *The Internet and higher education*, *15*(1), 3-8.

Dabbagh, N., & Reo, R. (2010). Impact of Web 2.0 on Higher Education. In D. W. Surry (Ed.), *Technology Integration in Higher Education: Social and Organizational Aspects* (pp. 174-187). Hershey, PA: IGI Global.

Dalton, C., & Seidlhofer, B. (1994). *Pronunciation*. Oxford: Oxford University Press.

Dalton, C., & Smit, U. (1997). On the Motivation of Advanced Pronunciation Learners. *Speak Out!, 21*, 5-10.

Dam, L., & Little, D. (1998). *Autonomy in Foreign Language Learning: From Classroom Practice to Generalizable Theory.* Paper presented at the JALT 24 th Annual International Conference on Language Teaching/Learning, Omiya Sonic City, Japan.

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic Skills and Speaking Fluency in a Second Language. *Applied Psycholinguistics, 34*(5), 893-916.

Deleuze, G., & Guattari, F. (1980). *Mille Plateaux*. Paris: Ed. de minuit.

Deleuze, G., & Guattari, F. (1987). Introduction: Rhizome. *A thousand plateaus: Capitalism and schizophrenia*, 3-25.

Derwing, T. M., & Munro, M. J. (1997). Accent, Intelligibility, and Comprehensibility: Evidence from Four L1s. *Studies in Second Language Acquisition, 19*, 1-16.

Derwing, T. M., & Munro, M. J. (2005). Second Language Accent and Pronunciation Teaching: A Research-Based Approach. *TESOL quarterly, 39*(3), 379-397.

Derwing, T. M., Thomson, R. I., & Munro, M. J. (2006). English Pronunciation and Fluency Development in Mandarin and Slavic Speakers. *System, 34*(2), 183-193.

Dickinson, L. (1987). *Self-Instruction in Language Learning*. Cambridge: Cambridge University Press

Dickinson, L. (1995). Autonomy and Motivation: A Literature Review. *System*, *23*(2), 165-174.

Doughty, C. J., & Long, M. H. (2003). *The Handbook of Second Language Acquisition.* . Malden, MA: Blackwell.

Dron, J. (2007). *Control and Constraint in E-Learning: Choosing When to Choose*. Hershey, PA: Idea Group.

Dulay, H. C., Burt, M. K., & Krashen, S. D. (1982). *Language Two*. Oxford: Oxford University Press.

Egel, I. P. (2009). Learner Autonomy in the Language Classroom: From Teacher Dependency to Learner Independency. *Procedia-Social and Behavioral Sciences, 1*(1), 2023-2026.

Fan, & Nang. (2005). A Survey on Phonetics Teaching for College Non-English Majors. *Journal of Xi'an International Studies University, 2005*(4), 43-48.

Fathman, A. K. (1976). Variables Affecting the Successful Learning of English as a Second Language. *TESOL quarterly*, 433-441.

Fidelman, C. (1993a). Dans La Peau Des Français, in Deutsche Haut Geschlüpft, *In Intercultures No. 23, Nov.-Déc*. Paris.

Fidelman, C. (1993b). Teaching Nonverbal Communication Skills. *Athelstan Newsletter on Technology and Language Learning, 5*(3), February, 5.

Fosnot, C. T. (2013). *Constructivism: Theory, Perspectives, and Practice*. NY: Teachers College Press.

Frankfort-Nachmias, C., & Nachmias, D. (1996). *Research Methods in the Social Sciences*. NY: Worth Publishers.

Fraser, H. (2000). *Coordinating Improvements in Pronunciation Teaching for Adult Learners of English as a Second Language*. Canberra: DETYA.

Gall, J. P., Gall, M. D., & Borg, W. R. (1999). *Applying Educational Research: A Practical Guide*. New York: Longman.

Gao, X. (2002). Negative Transfer in Pronunciation Learning and Teaching. *Foreign Language Teaching Abroad, 2002*(3), 46-49.

Garrigues, S. (1999). Overcoming Pronunciation Problems of English Teachers in Asia. Retrieved Jan. 24, 2014, from http://asianbridges.com/pac2/presentations/garrigues.html

Gass, S., Svetics, I., & Lemelin, S. (2003). Differential Effects of Attention. *Language Learning, 53*(3), 497-546.

Gassin, J. (1990). The Notion of Synchrony in Second Language Learning. *Learning, keeping and using languages*, 433-443.

Gilbert, J. B. (1987). Pronunciation and Listening Comprehension. *Current perspectives on pronunciation*, 29-40.

Gilbert, J. B. (2008). *Teaching Pronunciation: Using the Prosody Pyramid*. Cambridge: Cambridge University Press.

Gilbert , J. B. (2010). Pronunciation as Orphan: What Can Be Done. *Speak Out, 43*, 3-7.

Gimson, A. C., & Ramsaran, S. (1970). *An Introduction to the Pronunciation of English*. London: Edward Arnold.

Godwin-Jones, R. (2011). Emerging Technologies Autonomous Language Learning. *Language Learning & Technology, 15*(3), 4-11.

Goodwin, J. (2001). Teaching Pronunciation. In M. Celce-Murcia (Ed.), *Teaching English as a Second or Foreign Language*. Boston, MA: Heinle & Heinle.

Gorjian, B., Hayati, A., & Pourkhoni, P. (2013). Using Praat Software in Teaching Prosodic Features to Efl Learners. *Procedia-Social and Behavioral Sciences, 84*, 34-40.

Green, J. M. (1993). Student Attitudes toward Communicative and Non‐Communicative Activities: Do Enjoyment and Effectiveness Go Together? *The Modern Language Journal, 77*(1), 1-10.

Gremmo, M.-J., & Riley, P. (1995). Autonomy, Self-Direction and Self Access in Language Teaching and Learning: The History of an Idea. *System, 23*(2), 151-164.

Guattari, F., & Deleuze, G. (2000). *A Thousand Plateaus: Capitalism and Schizophrenia*. London: Athlone Press.

Guberina, P. (1971). *Studies on the Verbo-Tonal System*: University of Tennessee, Department of Audiology and Speech Pathology.

Guberina, P. (1972). *Case Studies in Restricted Bands of Frequencies in Auditory Rehabilitation of Deaf*. U. S. Office of Vocational Rehabilitation.

Guberina, P., & Asp, C. (1981). The Verbo-Tonal Method for Rehabilitating People with Communication Problems. *Monograph 13, World Rehabilitation*.

Guiora, A. Z., Beit-Hallahmi, B., Brannon, R. C., Dull, C. Y., & Scovel, T. (1972). The Effects of Experimentally Induced Changes in Ego States on Pronunciation Ability in a Second Language: An Exploratory Study. *Comprehensive psychiatry, 13*(5), 421-428.

Hahn, L. D. (2004). Primary Stress and Intelligibility: Research to Motivate the Teaching of Suprasegmentals. *TESOL quarterly, 38*(2), 201-223.

Hammersley, M. (2008). Troubles with Triangulation. In M. M. Bergman (Ed.), *Advances in Mixed Methods Research* (pp. 22-36). New Delhi: Sage.

Han, Y. (2013). On the Negative of Hexi Dialects in Gansu Province to English Pronunciation Acquisition. *Journal of Hexi Insititue, 29*(1), 112-115.

Hang, J. (2012). An Exploring Study on the Verbo-Tonal Method in Japanese Tones Teaching. *Journal of Changchun Education Institute, 28*(11), 124-125.

Hayta, R. A. F., & Yaprak, R. A. Z. (2013). Learner Autonomy and Computer Technology as a Facilitator of Autonomous Language Learning. *Journal of Education and Instructional Studies in the World, 3*(2), 57-65.

He, B., Ai, J., & Lian, A.-P. (2014). Reflections on Pronunciation Teaching for English Majors--a Case Study in Xingyi Normal University for Nationalities. *Journal of Xingyi Normal University for Nationalities*(2), 72-78.

He, B., Puakpong, N., & Lian, A. (2013). Factors Affecting the Normalization of Call in Chinese Senior High Schools. *Computer Assisted Language Learning*(ahead-of-print), 1-13.

He, S. F. (1987). Several Aspects on Comparison between English and Chinese in Pronunciation Teaching. *Foreign Languages, 1987*(6), 70-73.

Hedge, T. (2001). *Teaching and Learning in the Language Classroom*. Oxford: Oxford University Press

Hesling, I., Dilharreguy, B., Clément, S., Bordessoules, M., & Allard, M. (2005). Cerebral Mechanisms of Prosodic Sensory Integration Using Low - Frequency Bands of Connected Speech. *Human brain mapping, 26*(3), 157-169.

Ho, J., & Crookall, D. (1995). Breaking with Chinese Cultural Traditions: Learner Autonomy in English Language Teaching. *System, 23*(2), 235-243.

Holec, H. (1981). *Autonomy and Foreign Language Learning*. Oxford: Pergamon.

Holstein, J. A., & Gubrium, J. F. (2004). The Active Interview. In D. Silverman (Ed.), *Qualitative Research: Theory, Method and Practice* (2nd ed., pp. 140-161). Thousand Oaks, CA: Sage.

Hu, Y. H., & Uno, S. (2005). Effectiveness of a New Teaching Method Based on the Verbo-Tonal Method for Japanese Beginners' Learning of Chinese Voice Tones. *The Japanese Journal of Education psychology, 53*(4), 61-65.

Hutcheon, L. (2003). *A Poetics of Postmodernism: History, Theory, Fiction*. London: Routledge.

IBM. (2008). Statistical Package for the Social Sciences [Computer Program] (Version 16.0).

Johnson, M., & Liber, O. (2008). The Personal Learning Environment and the Human Condition: From Theory to Teaching Practice. *Interactive Learning Environments, 16*(1), 3-15.

Jones, R. H. (1997). Beyond "Listen and Repeat": Pronunciation Teaching Materials and Theories of Second Language Acquisition. *System, 25*(1), 103-112.

Joshi, K. R. (2011). Learner Perceptions and Teacher Beliefs About Learner Autonomy in Language Learning. *Journal of NELTA, 16*(1-2), 12-29.

Kang, O. (2010). *Esl Learners' Attitudes toward Pronunciation Instruction and Varieties of English.* Paper presented at the Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference. Ames, IA: Iowa State University.

Kelly, G. (1969). *25 Centuries of Language Teaching*. Rowley, MA: Newbury House.

Kelly, G. (2000). *How to Teach Pronunciation*. London: Bluestone Press.

Kenworthy, J. (1987). *Teaching English Pronunciation*. New York, NY: Longman

Kilgore, D. W. (2001). Critical and Postmodern Perspectives on Adult Learning. *New directions for adult and continuing education, 2001*(89), 53-62.

Klimova, B. F., & Semradova, I. (2012). The Teaching of Foreign Languages and Ict. *Procedia Technology, 1*, 89-93.

Krashen, S., Scarcella, R., & Long, M. (Eds.). (1982). *Child-Adult Differences in Second Language Acquisition*. Rowley, MA: Newbury House.

Larsen-Freeman, D., & Long, M. H. (1991). *An Introduction to Second Language Acquisition Research*. London: Longman.

Le, J. M., & Han, T. L. (2006). A Review of Foreign Phonetics Teaching and Current Situations--Comments on Jazz Chants as a Phonetics Teaching Method. *Foreign Language World, 2006*(1), 16-21.

Lennon, P. (1990). Investigating Fluency in Efl: A Quantitative Approach*. *Language Learning, 40*(3), 387-417.

Leppänen, P. H., Hämäläinen, J. A., Salminen, H. K., Eklund, K. M., Guttorm, T. K., Lohvansuu, K., . . . Lyytinen, H. (2010). Newborn Brain Event-Related Potentials Revealing Atypical Processing of Sound Frequency and the Subsequent Association with Later Literacy Skills in Children with Familial Dyslexia. *Cortex, 46*(10), 1362-1376.

Levis, J. M. (2005). Changing Contexts and Shifting Paradigms in Pronunciation Teaching. *TESOL quarterly, 39*(3), 369-377.

Levis, J. M., & Grant, L. (2003). Integrating Pronunciation into Esl/Efl Classrooms. *Tesol Journal, 12*(2), 13-19.

Levy, M. (1997). *Computer-Assisted Language Learning*. Now York: Oxford University Press.

Lian, A.-P. (1980). Intonation Patterns of French (Teacher's Book). Melbourne: River Seine Publications.

Lian, A.-P. (1987). Awareness, Autonomy and Achievement in Foreign Language Learning. *Revue de Phonétique Appliquée*, 82-84.

Lian, A.-P. (1996). *The Challenges of Technology.* Paper presented at the National Conference of the Australian Federation of Modern Language Teachers' Association, Perth, AFMLTA.

Lian, A.-P. (2004). Technology-Enhanced Language-Learning Environments: A Rhizomatic Approach. In J.-B. Son (Ed.), *Computer-Assisted Language Learning: Concepts, Contexts and Practices* (pp. 1-20). NY: iUniverse.

Lian, A.-P. (2011). Reflections on Language-Learning in the 21st Century: The Rhizome at Work. *Rangsit Journal of Arts and Sciences, 1*(1), 5-17.

Lian, A.-P. (2014a). On-Demand Generation of Individualised Language Learning Lessons. *Journal of Science, 9*(1), 25-38.

Lian, A.-P., & Gonzalez, J. A. (2008). *Technology-Supported Rhizomatic Language-Learning.* Paper presented at the 7th AsiaCALL Conference, Suranaree University of Technology, Nakhon Ratchasima, Thailand.

Lian, A.-P., & Lian, A. (1997). The Secret of the Shao-Lin Monk: Contribution to an Intellectual Framework for Language Learning. *On-CALL, May*, 2-19.

Lian, A. B. (2014b). “New Learning” and Call: A Diy Paradigm. *AsiaCALL Online Journal, 9*, A 14- A 26.

Little, D. (1990). Autonomy in Language Learning. In A. Swarbrick (Ed.), *Teaching Modern Languages* (pp. 81-87). London: Routledge.

Little, D. (1991). *Learner Autonomy: Definitions, Issues and Problems*. Dublin: Authentik.

Little, D. (1995). Learning as Dialogue: The Dependence of Learner Autonomy on Teacher Autonomy. *System, 23*(2), 175-181.

Littlewood, W. (1996). “Autonomy”: An Anatomy and a Framework. *System, 24*(4), 427-435.

Liu, H. X. (2012). An Empirical Study on Visible Software Enhancing Pronunciation Teaching. *Computer-assisted Foreign Language Education, 230*(6), 94-99.

Liu, T. H. (1999). Learners' Attitudes, Contact and Oral Esl Proficiency: A Study of Chinese Graduate Students at Us Universities. Retrieved Dec. 26, 2013, from http://docs.lib.purdue.edu/dissertations/AAI3017643/

Lodge, K. (2009). *Fundamental Concepts in Phonology: Sameness and Difference*. Edinburgh: Edinburgh University Press.

Lozanov, G. (2009). *Suggestopedia--Reservopedia: Theory and Practice of the Liberating-Stimulating Pedagogy on the Level of the Hidden Reserves of the Human Mind*. Sofia, Bulgaria: St. Kliment Ohridski University Press.

Luo, L. S., & Zhang, L. X. (2002). Reviews and Suggestions on English Phonetics Teaching. *Foreign languages and their teaching, 10*(1), 17-21.

Mak, B. (2011). An Exploration of Speaking-in-Class Anxiety with Chinese Esl Learners. *System, 39*(2), 202-214.

Martindale, T., & Dowdy, M. (2010). Personal Learning Environments. In G. Veletsianos (Ed.), *Emerging Technologies in Distance Education* (pp. 177-193). Edmonton: Athabasca University Press.

McLoughlin, C., & Lee, M. J. (2010). Personalised and Self Regulated Learning in the Web 2.0 Era: International Exemplars of Innovative Pedagogy Using Social Software. *Australasian Journal of Educational Technology, 26*(1), 28-43.

Mildner, V., & Bakran, J. (2001). Acoustic Correlates of Phonetic Correction. *Clinical linguistics & phonetics, 15*(1-2), 151-155.

Mitra, S., Tooley, J., Inamdar, P., & Dixon, P. (2003). Improving English Pronunciation: An Automated Instructional Approach. *Information Technologies & International Development, 1*(1), 75–84.

MOE. (2000). *Curriculum for English Majors*. Beijing: Foreign Language Teaching and Research Press.

Molino, J. (2000). Toward an Evolutionary Theory of Music and Language. In N. L. Wallin, B. Merker & S. Brown (Eds.), *The Origins of Music* (pp. 165-176). Combridge, MA: The MIT Press.

Morley, J. (1991). The Pronunciation Component in Teaching English to Speakers of Other Languages. *TESOL quarterly, 25*(3), 481-520.

Morse, J. M. (1991). Approaches to Qualitative-Quantitative Methodological Triangulation. *Nursing research, 40*(2), 120-123.

Mozzon-McPherson, M. (2007). Supporting Independent Learning Environments: An Analysis of Structures and Roles of Language Learning Advisers. *System, 35*(1), 66-92.

Mueller, J. L., Friederici, A. D., & Männel, C. (2012). Auditory Perception at the Root of Language Learning. *Proceedings of the National Academy of Sciences, 109*(39), 15953-15958.

Munro, M. J., & Derwing, T. M. (1995a). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning, 45*(1), 73-97.

Munro, M. J., & Derwing, T. M. (2006). The Functional Load Principle in Esl Pronunciation Instruction: An Exploratory Study. *System, 34*(4), 520-531.

Murray, L., & Barnes, A. (1998). Beyond the "Wow" Factor—Evaluating Multimedia Language Learning Software from a Pedagogical Viewpoint. *System, 26*(2), 249-259.

Norris-Holt, J. (2002). An Investigation of Japanese High School Students' Attitudes Towards the Study of English. *Second Language Learning and Teaching, 2*, 116-122.

Nowacka, M. (2012). Questionnaire-Based Pronunciation Studies: Italian, Spanish and Polish Students' Views on Their English Pronunciation. *Research in Language, 10*(1), 43-61.

Nunan, D. (1992). *Collaborative Language Learning and Teaching*. Cambridge: Cambridge University Press.

Nunan, D. (1999). *Second Language Teaching & Learning*. Boston, MA: Heinle & Heinle.

Nunan, D. (Ed.). (2003). *Practical English Language Teaching*. New York: McGraw-Hill ESL/ELT.

O'malley, J. M., & Chamot, A. U. (1990). *Learning Strategies in Second Language Acquisition*. Cambridge: Cambridge University Press.

Oyama, S. (1976). A Sensitive Period for the Acquisition of a Nonnative Phonological System. *Journal of Psycholinguistic Research, 5*(3), 261-283.

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

Piaget, J. (2013). *The Construction of Reality in the Child*. London: Routledge.

Pineda, M. V. (2014). Open Teaching and Personal Learning Networks (Plns) as Avenues of Enhanced Participation and Reflection. *Rangsit Journal of Arts and Sciences, 3*(2), 99-112.

Porter, D., & Garvin, S. (1989). Attitudes to Pronunciation in Efl. *Speak Out!, 5*, 8-15.

Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (2004). Methods for Testing and Evaluating Survey Questions. *Public opinion quarterly, 68*(1), 109-130.

Punch, K. F. (2013). *Introduction to Social Research: Quantitative and Qualitative Approaches*. London: Sage.

Reed, M., & Michaud, C. (2005). *Sound Concepts: An Integrated Pronunciation Course*. NY: McGraw-Hill.

Renard, R. (1975). *Introduction to the Verbo-Tonal Method of Phonetic Correction (B. Morris, Trans.)* Paris: Didier.

Richards, J., & Schmidt, R. (Eds.). (2002) Longman dictionary of applied linguistics and language teaching. Harlow, UK: Longman.

Richards, J. C., & Schmidt, R. (Eds.). Longman Dictionary of Language Teaching and Applied Linguistics. London: Pearson Education.

Roach, P. (2010). *English Phonetics and Phonology: A Practical Course* (4th ed.). Bonn: Ernst Klett Sprachen.

Robson, C. (1993). *Real World Research: A Resource for Social Scientists and Practitioners-Researchers*. Oxford: Blackwell Publishers.

Rossiter, M. J. (2009). Perceptions of L2 Fluency by Native and Non-Native Speakers of English. *Canadian Modern Language Review, 65*(3), 395-412.

Rubin, N. (2010). Creating a User-Centric Learning Environment with Campus Pack Personal Learning Spaces. *PLS Webinar, Learning Objects Community*.

Sangarun, P. (2014). Developing Reading Comprehension Skills in Another Language: A Computer-Based 21st Century Prototype. *Beyond Words, 2*(2), 121-153.

Scarcella, R. C., & Oxford, R. L. (1994). Second Language Pronunciation: State of the Art in Instruction. *System, 22*(2), 221-230.

Schmidt, R. W. (1990). The Role of Consciousness in Second Language Learning. *Applied linguistics, 11*(2), 129-158.

Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency*. London: Routledge.

Seliger, H. W. (1975). Maturational Constraints in the Acquisition of Second Language Accent. *Language Sciences, 36*, 20-22.

Seliger, H. W., & Shohamy, E. (1989). *Second Language Research Methods*. Oxford: Oxford University Press.

Selinker, L., & Rutherford, W. E. (2013). *Rediscovering Interlanguage*. London: Routledge.

Skehan, P. (2003). Task-Based Instruction. *Language teaching, 36*(1), 1-14.

Spicer, N. (2004). Combinating Qualitative and Quantitative Methods. In C. Seale (Ed.), *Researching Society and Culture* (2nd ed., pp. 293-303). Thousand Oaks, CA: Sage.

St. Louis, R. (2006). Helping Students Become Autonomous Learners: Can Technology Help? *The Journal of teaching English through technology, 6*(3).

Su, D. F., & Zhuang, Z. X. (1996). *Modern Foreign Language Teaching: Theories, Practice, and Approaches*. Shanghai: Shanghai Foreign Language Education Press.

Sui, H. L. (1997). Innovation on Multimedia and English Pronunciation. *Journal of Heilongjiang University, 1997*(2), 65-67.

Sun, B., & Ye, C. C. (2006). Reassessment of Computer-Assisted Language Learning. *Computer Assisted Foreign Language Education, 2006*(1), 98-110.

Tarone, E. (1978). The Phonology of Interlanguage. In J. C. Richards (Ed.), *Understanding Second & Foreign Language Learning: Issues & Approaches*. Rowley, MA: Newbury House.

Tashakkori, A., & Teddlie, C. (1998). *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. Thousand Oaks, CA: Sage.

Tashakkori, A., & Teddlie, C. (2003). *Handbook of Mixed Methods in the Social and Behavioural Science*. Thousand Oaks, CA: Sage.

Tassinari, M. G. (2012). Evaluating Learner Autonomy: A Dynamic Model with Descriptors. *Studies in Self-Access Learning Journal, 3*(1), 24-40.

Tavakoli, P., & Skehan, P. (2005). Strategic Planning, Task Structure, and Performance Testing. In R.Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 239-273). Amsterdam: John Benjamins.

Teddlie, C., & Tashakkori, A. (Eds.). (2009). *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences* (3rd ed.). New Delhi: Sage.

Teddlie, C., Tashakkori, A., & Johnson, B. (2008). Emergent Techniques in the Gathering and Analysis of Mixed Methods Data. In S. N. Hesse-Biber & P. Leavy (Eds.), *Handbook of Emergent Methods* (pp. 389-413). London: Guildford Press.

Tergujeff, E. (2013). Learner Perspective on English Pronunciation Teaching in an Efl Context. *Research in Language, 11*(1), 81-95.

Thomas, R. (2003). *Blending Qualitative & Quantitative Research Approach to Language Learning*. Englewood Cliffs, HJ: Prentice Hall.

Tincoff, R., & Jusczyk, P. W. (1999). Some Beginnings of Word Comprehension in 6-Month-Olds. *Psychological Science, 10*(2), 172-175.

Tokumoto, M., & Shibata, M. (2011). Asian Varieties of English: Attitudes Towards Pronunciation. *World Englishes, 30*(3), 392-408.

Trofimovich, P., & Isaacs, T. (2012). Disentangling Accent from Comprehensibility. *Bilingualism: Language and Cognition, 15*(04), 905-916.

Trubetzkoy, N. S. (1939). *Principle of Phonology*. Los Angeles: University of California Press.

Valsecchi, M., Künstler, V., White, B. J., Saage, S., Mukherjee, J., & Gegenfurtner, K. R. (2008). Advantage in Reading Lexical Bundles Is Reduced in Non-Native Speakers. *Journal of Eye Movement Research, 6*(5), 2.

Valtonen, T., Hacklin, S., Dillon, P., Vesisenaho, M., Kukkonen, J., & Hietanen, A. (2012). Perspectives on Personal Learning Environments Held by Vocational Students. *Computers & Education, 58*(2), 732-739.

Varol, B., & Yilmaz, S. (2010). Similarities and Differences between Female and Male Learners: Inside and Outside Class Autonomous Language Learning Activities. *Procedia-Social and Behavioral Sciences, 3*, 237-244.

Walsh, K. (2003). Qualitative Research: Advancing the Science and Practice of Hospitality. *Cornell Hotel and Restaurant Administration Quarterly, 44*(2), 66-74.

Wang, G. Z. (2005). *English Pronunciation Intonation for Communication*. Beijing: Higher Education Press.

Wang, Y., & Feng, J. C. (2010). An Investigation on Phonetics Teaching in College English and the Use of International Phonetic Alphabet. *Media in Foreign Language Instruction, 37*, 158-160.

Wardhaugh, R. (1970). The Contrastive Analysis Hypothesis. *TESOL quarterly, 4*(2), 123-130.

Warren, P., Elgort, I., & Crabbe, D. (2009). Comprehensibility and Prosody Ratings for Pronunciation Software Development. *Language Learning & Technology, 13*(3), 87-102.

Warschauer, M., & Healey, D. (1998). Computers and Language Learning: An Overview. *Language teaching, 31*(2), 57-71.

Warschauer, M., Shetzer, H., & Meloni, C. (2000). *Internet for English Teaching*. Alexandria, VA: TESOL Publications.

Weathington, B. L., Cunningham, C. J., & Pittenger, D. J. (2010). *Research Methods for the Behavioral and Social Sciences*. Hoboken, NJ: John Wiley & Sons.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (2000). *Unobtrusive Measures*. Thousand Oaks, CA: SAGE Publications.

Weir, C. J. (1990). *Communicative Language Testing*. NY: Prentice Hall.

Wen, Q. F., & Zhuang, Y. L. (2005). A Study on Improving Advanced English Learners' Self-Correction Ability in Oral English. *Foreign Language World, 106*(2), 33-51.

Wiersma, W., & Jurs, S. G. (2005). *Educational Measurement and Testing*. Englewood Cliffs, NJ: Allyn & Bacon.

Wu, H. (2011). Construction of Business English Curriculum in Higher Vocational College Based on Learners' Needs Analysis. *Journal of Guangdong Technical College of Water Resources and Electric Engineering, 3*(9), 76-80.

Wylie, L. W. (1977). *Beaux Gestes: A Guide to French Body Talk*. Cambridge, MA: The Undergraduate Press, E.P. Dutton.

Xiao, J. H. (2007). Meta-Learning Environment: On the Normalization of Call. *Academic Forum, 21*(5), 27-28.

Xie, G. H., & Liu, Z. B. (2004). Structural Features of Chinese and English and Their Learning Habits. *Journal of Qiandongnan Teachers' College, 3*(1), 32-34.

Xu, N. Z. (2004). English Suprasegmental Features and Their Pragmatic Functions. *Overseas Foreign Language Teaching, 2004*(2), 42-45.

Yan, Q. L. (2008). A Student-Centered Pronunciation Teaching Model. *China Education Innovation Herald, 26*, 122-123.

Yates, L., Zielinski, B., & Pryor, E. (2011). The Assessment of Pronunciation and the New Ielts Pronunciation Scale. *IELTS Research Reports, 12*, 1-44.

Young, D. J., & Oxford, R. (1997). A Gender-Related Analysis of Strategies Used to Process Written Input in the Native Language and a Foreign Language. *Applied language learning, 8*(1), 43-73.

Zhang, F. (2006). *The Teaching of Mandarin Prosody: A Somatically-Enhanced Approach for Second Language Learners.* (Unpublished PhD), University of Canberra, Australia.

Zhao, W. H. (2002). New Attempts to English Phonetic Teaching. *Media in Foreign Language Instruction, 84*(2), 3-5.

Zheng, Y. Y. (2010). *Vocabulary Knowledge Development of Chinese Learners of English in China: A Longitudinal Multiple-Case Study of Eight University Students.* (Ph.D Dissertation in Education), The University of Hong Kong.

APPENDICES

APPENDIX A

A Semi-Structured Interview Questionnaire on Pronunciation Learning

(for Students)

Dear students,

This oral interview is to investigate your opinions of pronunciation learning. There is no right or wrong answer. Please feel free to say what you think about pronunciation learning. The interview content will be kept confidential. Thank you for your cooperation.

Part I Personal Information

Major	❑ English		❑ Non-English	
Gender	❑ Male		❑ Female	
Age (years)	❑ 16-18	❑ 19-21		❑ 22 & above
Level	❑ Year One	❑ Year Two	❑ Year Three	❑ Year Four

Part II Interview Questions

1. Have you ever taken the course "Phonetics" which focuses on pronunciation?

❑ Yes ❑ No

If the answer is "No", please specify how you have learned English pronunciation.

If the answer is “Yes”, please continue with the following questions.

2. What is your opinion of this course? Please describe your feelings. (e. g., helpful, useful, beneficial, easy, enjoyable, difficult, boring, interesting, etc.)

3. Did you have any problems when you were learning English pronunciation? If yes, what problems did you have?

4. Do you think that your pronunciation improved after taking this course? If yes, to what extent did your pronunciation improve?

5. Have you ever used any technology (CDs, software, websites, etc.) to assist your study to supplement the teacher’s teaching? If yes, describe the technology.

6. What (Who) was your “model” when you learned to pronounce?

7. Did you like your teacher’s teaching? What aspects of your pronunciation did you expect him or her to help improve?

8. What aids did you need to supplement your teachers’ classroom teaching?

9. How much time did you spend on learning and practicing pronunciation?

10. Did you continue to study pronunciation after you finished that course? If yes, how?

11. Do you still have problems in pronunciation after that course? If yes, what problems do you still have?

12. Did you rely on the teacher's teaching or did you learn pronunciation by yourself? (Or both, please specify.)

13. Can you speak English in a natural way when you talk to people outside the class? (Or, do you think that what you learn in class can be used in real life? If yes, how? If not, why not?)

14. Have you sat for the Pronunciation Test in the school of Foreign Languages? If yes, what was your score?

15. Do you have anything else to say in terms of pronunciation learning?

APPENDIX B

A Semi-Structured Interview Questionnaire on Pronunciation Learning (for Students) (Chinese Version)

关于语音学习的调查问卷

（学生部分）

亲爱的同学，你好！

本访谈是为了全面了解英语专业学生在学习语音方面的的情况、遇到的问题、以及你们的期望等。此访谈内容只作为研究目的使用，没有正确或者错误的答案之分，只是为了真实反应你们语音学习的情况。我们将对访谈的内容严格保密。谢谢你的配合！

第一部分　个人信息

专业　❑ 英语专业　❑ 非英语专业

性别　❑ 男　❑ 女

年龄　❑ 16-18　❑ 19-21　❑ 22 及以上

年级　❑ 大一　❑ 大二　❑ 大三　❑ 大四

第二部分　访谈问题

1. 你学过《语音》课吗?

❑ 学过　❑ 没学过

如果答案是“没学过”,请说明你是怎么学习英语语音的。

如果答案是“学过”,请继续以下的问题。

2. 的感受如你对语音课何？请具体描述 (比如：对提高英语发音很有帮助、有益、有用, 学习很轻松, 或者很艰难，等等)。

3. 你在语音学习个过程中遇到过什么困难吗? 如果有, 遇到哪些困难呢?（比如，元音、辅音的发音问题，语音语调问题，连读，节奏问题等）

4. 你认为经过语音课的学习，你的英语发音是否有提高? 如果有提高，提高到什么程度（比如，学到基础发音，有很大提高，掌握了语音的基本知识，全面学习了英语语音知识等）。

5. 在学习过程中，老师的教学是否使用了现代技术手段（比如 CD,语音学习软件，网站，电脑等)？如果有，请具体说明。

6. 学习语音时，你的模仿对象是什么？（比如，上课的老师，录音等）

7. 你喜欢老师的教学吗？如果可能，你希望老师在什么方面改进？

8. 除了老师的课堂教学以外，你认为你还需要什么样的帮助（比如 CD,语音学习软件，网站，电脑等）来提高你的英语发音？

9. 你花了多少时间在学习语音上？（比如，每周大致多少小时？）

10. 修完语音课以后，你还继续学习英语发音吗? 如果继续，是怎么学习的，采取什么方式学习?

11. 经过语音课的学习以后，你认为你的英语发音还存在哪些问题?

12. 你主要是依赖老师的教学来进行学习呢，还是自己学习？或者是两者结合？请具体说明。

13. 你认为你能把语音课上学到的发音自然地用到真实场景交谈中吗？如果能，请说明是如何做到的；如果不能，请说明原因。

14. 你参加过外国语学院组织的“语音测试”吗？如果参加过，你的成绩如何？

15. 你还有其他关于语音学习的需要说明吗？你可以自由表达你想说的关于语音学习的任何问题。

APPENDIX C

A Semi-Structured Interview Questionnaire on Pronunciation Learning (for Teachers)

Dear teachers,

This interview is to investigate your opinions on pronunciation instruction. There is no right or wrong answer. Please feel free to say what you think about pronunciation instruction. The interview content will be kept confidential. Thank you for your cooperation.

Part I Personal Information

Gender	❑ Male	❑ Female	
Age (years)	❑ 20-25	❑ 26-31	❑ 32 & above
Teaching experience	❑ 1-5 years	❑ 6-11years	❑ 12 years above

Part II Interview Questions

1. Have you ever taught the course "Phonetics"?

❑ Yes ❑ No

If the answer is "Yes", please continue with the following questions.

2. What is your opinion of this course? Please describe your feelings. (e. g., helpful, useful, beneficial, easy, enjoyable, difficult, boring, interesting, etc.)

3. Did you encounter any problems when teaching English pronunciation? If yes, what problems did you meet?

4. Did you use any technology (CDs, software, websites, etc.) to assist your teaching? If yes, what was it?

5. What (Who) was the "model" when you taught pronunciation?

6. Did you encourage your students to participate the pronunciation test in the School of Foreign Languages? If yes, why?

7. How were your students motivated and what were the levels of their motivation when learning pronunciation?

8. How do you describe your classroom teaching? (e. g., student-centered, teacher-centered, active, interactive, dull, difficult)

9. How would you like to improve your teaching?

10. Do you have anything else to say in terms of pronunciation teaching?

APPENDIX D

A Semi-Structured Interview Questionnaire on Pronunciation Learning (for Teachers) (Chinese version)

关于语音教学的访谈问卷

(教师部分)

亲爱的老师，你好！

本访谈是为了全面了解你在英语专业语音课方面教学的情况、遇到的问题、以及你的期望等。此访谈内容只作为研究目的使用，没有正确或者错误的答案之分，只是为了真实反应你的语音教学的情况。我们将对访谈的内容严格保密。谢谢你的配合！

第一部分　个人信息

性别	❑ 男	❑ 女	
年龄	❑ 20-25	❑ 26-31	❑ 32 以上
教龄	❑ 1-5 年	❑ 6-11 年	❑ 12 年以上

第二部分　访谈问题

1. 你修过 “语音” 课吗?

❑ 修过　　❑ 没修过

如果回答是“修过”, 请继续下面的问题。

2. 你对这门课程的评价是什么？你描述你的感觉 (比如,对学生有帮助,有用，有益，简单，困难，枯燥，有趣等等）。

3. 你在教授英语语音的时候遇到过困难吗？如果遇到过，是什么问题？

4. 在教学过程中，你使用过诸如 CDs、语音软件和网页之类的技术手段来辅助你的教学吗？如果使用过，具体是什么技术手段？

5. 什么（或者谁）是你教授语音的模仿对象？

6. 你鼓励你的学生参加外国语学院组织的语音测试吗？如果是，为什么要鼓励？

7. 你对学生在学习语音过程中的动机来源和水平？

8. 你怎样描述你的课堂教学？（比如，以学生为中心，以教师为中心，课题气氛活跃，师生互动多，枯燥，困难等等。）

9. 你打算怎样改进你的教学?

10. 对于语音教学你有什么其他需要说明的吗?

APPENDIX E

IOC Analysis for the semi-structured interview with students

Item	Experts					Result of analysis
	1	2	3	4	5	
Q1	+1	+1	+1	+1	+1	√
Q2	+1	+1	0	+1	+1	√
Q3	0	+1	+1	+1	+1	√
Q4	+1	+1	+1	+1	+1	√
Q5	+1	+1	+1	+1	0	√
Q6	+1	+1	+1	+1	+1	√
Q7	-1	+1	+1	+1	0	√
Q8	+1	0	+1	+1	+1	√
Q9	+1	+1	-1	+1	+1	√
Q10	0	0	+1	+1	+1	√
Q11	+1	+1	+1	+1	+1	√
Q12	+1	+1	+1	+1	+1	√
Q13	+1	0	+1	+1	+1	√
Q14	+1	+1	+1	+1	+1	√
Q15	+1	+1	+1	+1	+1	√
Total	11	12	12	15	13	

Notes: 1. +1 = *the item is congruent with the objective*

2. -1 = *the item is not congruent with the objective*

3. 0 = *uncertain about this item*

Result of IOC:

(IOC = ∑R / N)

Item number: 15

R = 11 + 12 + 12 + 15 + 13 = 63 (Scores given by experts)

N = 5 (Number of experts)

IOC = 63/5 = 12.6

Percentage: 12.6/15 x 100% = 84%

The table above shows that the analysis result of IOC is 12.6, and the percentage is 84% which is higher than 80%. Therefore, the items are suitable for adoption in an interview questionnaire.

APPENDIX F

IOC Analysis for the semi-structured interview to teachers

Item	Experts					Result of analysis
	1	2	3	4	5	
Q1	+1	+1	+1	+1	+1	√
Q2	+1	+1	0	+1	+1	√
Q3	+1	+1	+1	+1	+1	√
Q4	+1	+1	+1	+1	+1	√
Q5	+1	+1	+1	+1	0	√
Q6	+1	+1	+1	+1	+1	√
Q7	-1	+1	+1	+1	0	√
Q8	+1	0	+1	+1	+1	√
Q9	+1	+1	--1	+1	+1	√
Q10	0	+1	+1	+1	+1	√
Total	7	9	7	10	8	√

Notes: 1. + 1 = *the item is congruent with the objective*

2. -1 = *the item is not congruent with the objective*

3. 0 = *uncertain about this item*

Result of IOC:

(IOC = ∑R/ N)

Item number: 10

R = 7 + 9 + 7 + 10 + 8 = 41 (Scores from experts)

N = 5 (Number of experts)

IOC = 41/5 = 8.2

Percentage: 8.2/10 x 100% = 82%

The above table shows that the analysis result of IOC is 8.2, and the percentage is 82% which is higher than 80%. Therefore, the items are suitable for adoption in an interview questionnaire.

APPENDIX G

Pretest Paper

Part I: Phonemes reading（20%). Direction: There are 20 phonemes in this part, please read them out. Every phoneme should be read only once.

1. [best]	2. [hæd]	3. [sæd]	4. [ʃə:t]	5. [kænt]
6. [stiə]	7. [sip]	8. [mil]	9. [ʃip]	10. [θiŋ]
11. [rait]	12. [si: m]	13. [neim]	14. [trʌŋk]	15. [drai]
16. [krɔp]	17. [gri:n]	18. [greip]	19. [preiz]	20. [praud]

Part II: Words reading (25%). Direction: There are 50 words in this part. Please read them out. Either British or American pronunciation is acceptable.

1. yesterday	2. daughter	3. drive	4. baby	5. worker
6. pair	7. building	8. friends	9. rain	10. encourage
11. visitor	12. win	13. less	14. accident	15. twentieth
16. spend	17. remember	18. minute	19. greeting	20 national
21. furniture	22. hear	23. bank	24. football	25. traffic
26. cough	27. question	28 strawberry	29. vacation	30. continue
31. birth	32. gesture	33. express	34. choice	35. century
36. pool	37. human	38. through	39. pale	40. goose
41. watermelon	42. herself	43. jeep	44. only	45. kill
46. nobody	47. rather	48. X-ray	49. ugly	50. zero

Part III: Passage reading (30%). Direction: Read the following passage.

Last October, the farmers needed help with the apple harvest. There were many apples on the trees. But there were not enough people to pick them. So Jim and his friends went to help the farmers. It was hard work, but they really enjoyed it. The children picked many apples. Sometimes the apples were quite high, and they needed to climb up the trees with ladders. The children carried the apples to the trucks.

It was a cold spring morning in the city of London in England. The weather was very cold, and many people were ill. So there were many people in the doctor's waiting room. At the head of the queue was an old woman. The woman was a visitor. She lived in the country. She was in the city to visit her daughter.

Part IV: Interview (25%). Direction: Answer the following interview questions. Use 1-3 sentences to answer each question.

Question 1. Tell me a little bit about yourself.

Question 2. Do you work well under pressure?

Question 3. What does success mean to you?

Question 4. How do you handle change?

Question 5. If you want to borrow your friend's bike, what would you say?

APPENDIX H

Posttest Paper

Part I: Phonemes (20%). Direction: There are 20 phonemes in this part, please read them out. Every phoneme should be read only once.

1. [ben]	2. [bæd]	3. [ʃə:t]	4. [hænd]	5. [fiə]
6. [su:p]	7. [tʃil]	8. [ʃəun]	9. [θimbl]	10. [lait]
11. [frai]	12. [trip]	13. [trʌŋk]	14. [dres]	15. [dri:m]
16. [praud]	17. [print]	18. [kreiv]	19. [krɔp]	20. [gri:n]

Part II: Words reading (25%). Direction: There are 50 words in this part. Please read them out. Either British or American pronunciation is acceptable.

1. choose	2. dress	3. film	4. cabbage	5. farmer
6. leave	7. throw	8. snow	9. sweater	10. rainy
11. everywhere	12. form	13. towards	14. information	15. object
16. decision	17. wife	18. farther	19. pleasure	20. arrive
21. group	22. weekday	23. noisy	24. eye	25. know
26. problem	27. judge	28. walkman	29. yourself	30. giraffe
31. drink	32. ants	33. idea	34. kitchen	35. quarter
36. hour	37. bright	38. homework	39. room	40. August
41. shine	42. X-ray	43. cruel	44. underground	45. middle
46. result	47. lovely	48. which	49. volleyball	50. zebra

Part III: Passage reading (30%). Direction: Read the following passage.

It is a fine Sunday morning. There are many people in the park. Many of them

are young and some are old. Some Young Pioneers are playing games over there. Two boys are playing with their yo-yos. A girl is flying a kite. Two children are mending a toy boat. There is a big lake in the park. The water is clear. There are some boats on the lake. Near the lake a young man is running. There is a house near the lake. Beside the house two men are working. A woman is watering the flowers. The flowers look very nice. A girl is drawing. Three other girls are taking photos. Look at that big tree. Under it are some men. Three of them are playing cards. One of them is looking at two cats. The cats are running up the tree.

Part IV: Interview (25%). Direction: Answer the following interview questions. Use 1-3 sentences to answer each question.

Question 1. Tell me something about yourself.

Question 2. Do you manage your time well?

Question 3. What do you want to be doing five years from now?

Question 4. How do you make important decisions?

Question 5. If you want to ask for leave in class, what would you say to the teacher?

APPENDIX I

IOC Analysis for pretest paper

Part I Phonemes

Item	Content	Experts					Result of analysis
		1	2	3	4	5	
1	[best]	+1	+1	+1	+1	0	√
2	[hæd]	+1	+1	+1	+1	+1	√
3	[sæd]	+1	+1	0	+1	+1	√
4	[ʃəːt]	+1	+1	+1	+1	+1	√
5	[kænt]	0	+1	+1	+1	+1	√
6	[stiə]	+1	+1	+1	+1	+1	√
7	[sip]	+1	+1	+1	+1	+1	√
8	[mil]	+1	+1	+1	+1	+1	√
9	[ʃip]	+1	+1	+1	+1	+1	√
10	[θiŋ]	0	+1	+1	+1	+1	√
11	[rait]	+1	+1	+1	+1	+1	√
12	[siːm]	+1	+1	+1	+1	+1	√
13	[neim]	+1	+1	+1	+1	+1	√
14	[trʌŋk]	+1	+1	+1	+1	+1	√
15	[drai]	+1	0	+1	+1	+1	√
16	[krɔp]	+1	+1	+1	+1	+1	√
17	[griːn]	+1	+1	+1	+1	0	√
18	[greip]	+1	+1	+1	+1	+1	√
19	[preiz]	+1	+1	+1	+1	+1	√
20	[praud]	+1	+1	+1	+1	+1	√
Total		18	19	19	20	18	

Notes: 1. +1 = *the item is congruent with the objective*

2. -1 = *the item is not congruent with the objective*

3. 0 = *uncertain about this item*

Result of IOC:

(IOC = ∑R/ N)

Item number: 20

R = 18 + 19 + 19 + 20 + 18 = 94 (Scores from experts)

N = 5 (Numbers of experts)

IOC = 94/5 = 18.8

Percentage: 18.8/20 x 100% = 94%

The table above shows that the analysis result of IOC is 18.8, and the percentage is 94% which is higher than 80%. Therefore, the items are suitable for adoption in an interview questionnaire.

Par II Words

Item	Content	Experts					Result of analysis
		1	2	3	4	5	
1	yesterday	+1	+1	+1	+1	+1	√
2	daughter	+1	+1	+1	+1	+1	√
3	drive	+1	+1	+1	+1	+1	√
4	baby	+1	+1	+1	+1	+1	√
5	worker	+1	+1	+1	+1	+1	√
6	pair	+1	+1	+1	+1	+1	√
7	building	+1	+1	+1	+1	+1	√
8	friends	+1	+1	+1	+1	+1	√
9	rain	+1	+1	+1	+1	+1	√
10	encourage	+1	+1	+1	+1	+1	√
11	visitor	+1	+1	+1	+1	+1	√
12	win	+1	+1	+1	+1	+1	√
13	less	+1	+1	+1	+1	+1	√
14	accident	+1	+1	+1	+1	+1	√
15	twentieth	+1	+1	+1	+1	+1	√
16	spend	+1	+1	+1	+1	+1	√
17	remember	+1	+1	+1	+1	+1	√
18	minute	+1	+1	+1	+1	+1	√
19	greeting	+1	+1	+1	+1	+1	√
20	national	+1	+1	+1	+1	+1	√
21	furniture	+1	+1	+1	+1	+1	√
22	hear	+1	+1	+1	+1	+1	√
23	bank	+1	+1	+1	+1	+1	√
24	football	+1	+1	+1	+1	+1	√
25	traffic	+1	+1	+1	+1	+1	√
26	cough	+1	+1	+1	+1	+1	√
27	question	+1	+1	+1	+1	+1	√
28	strawberry	+1	+1	+1	+1	+1	√
29	vacation	+1	+1	+1	+1	+1	√
30	continue	+1	+1	+1	+1	+1	√
31	birth	+1	+1	+1	+1	+1	√
32	gesture	+1	+1	+1	+1	+1	√
33	express	+1	+1	+1	+1	+1	√
34	choice	+1	+1	+1	+1	+1	√
35	century	+1	+1	+1	+1	+1	√

36	pool	+1	+1	+1	+1	+1	√
37	human	+1	+1	+1	+1	+1	√
38	through	+1	+1	+1	+1	+1	√
39	pale	+1	+1	+1	+1	+1	√
40	goose	+1	+1	+1	+1	+1	√
41	watermelon	+1	+1	+1	+1	+1	√
42	herself	+1	+1	+1	+1	+1	√
43	jeep	+1	+1	+1	+1	+1	√
44	only	+1	+1	+1	+1	+1	√
45	Kill	+1	+1	+1	+1	+1	√
46	nobody	+1	+1	+1	+1	+1	√
47	rather	+1	+1	+1	+1	+1	√
48	X-ray	+1	+1	+1	+1	+1	√
49	ugly	+1	+1	+1	+1	+1	√
50	zero	+1	+1	+1	+1	+1	√
Total		50	50	48	47	47	

Notes: 1. + 1 = *the item is congruent with the objective*

2. -1 = *the item is not congruent with the objective*

3. 0 = *uncertain about this item*

Result of IOC:

(IOC = ∑R/ N)

Item number: 50

R = 50 + 50 + 48 + 47 + 47 = 242 (Scores given by experts)

N = 5 (Number of experts)

IOC = 242/5 = 48.4

Percentage: 48.4/50 x 100% = 96.8%

The table above shows that the analysis result of IOC is 48.4, and the percentage is 96.8% which is higher than 80%. Therefore, the items are suitable for adoption in an interview questionnaire.

APPENDIX J

IOC Analysis for pretest paper

Part I Phonemes

Item	Content	Experts					Result of analysis
		1	2	3	4	5	
1	[ben]	+1	+1	+1	+1	+1	√
2	[bæd]	0	+1	+1	+1	+1	√
3	[ʃə:t]	+1	+1	+1	+1	+1	√
4	[hænd]	+1	+1	+1	0	+1	√
5	[fiə]	+1	+1	+1	+1	+1	√
6	[su:p]	+1	+1	+1	+1	+1	√
7	[tʃil]	+1	+1	+1	+1	0	√
8	[ʃəun]	+1	+1	+1	+1	+1	√
9	[`θimbl]	+1	+1	+1	+1	+1	√
10	[lait]	+1	+1	+1	+1	+1	√
11	[frai]	+1	+1	0	+1	+1	√
12	[trip]	+1	+1	+1	+1	+1	√
13	[trʌŋk]	+1	+1	+1	+1	+1	√
14	[dres]	+1	+1	+1	0	+1	√
15	[dri:m]	+1	+1	+1	+1	+1	√

16	[praud]	+1	+1	+1	+1	+1	√
17	[print]	+1	+1	+1	+1	+1	√
18	[kreiv]	+1	+1	+1	+1	+1	√
19	[krɔp]	0	+1	+1	+1	+1	√
20	[gri:n]	0	+1	+1	+1	+1	√
Total		17	20	19	18	19	

Notes: 1. +1 = *the item is congruent with the objective*

2. -1 = *the item is not congruent with the objective*

3. 0 = *uncertain about this item*

Result of IOC:

(IOC = ∑R/ N)

Item number: 20

R = 17 + 20 + 19 + 18 + 19 = 94 (Scores given by experts)

N = 5 (Number of experts)

IOC = 93/5 = 18.6

Percentage: 18.6/20x100% = 93%

The table above shows that the analysis result of IOC is 18. 6, and the percentage is 93% which is higher than 80%. Therefore, the items are suitable for adoption in an interview questionnaire.

Par II Words

Item	content	Experts					Result of analysis
		1	2	3	4	5	
1	choose	+1	+1	+1	0	+1	√
2	dress	+1	+1	+1	+1	+1	√
3	film	+1	+1	+1	+1	+1	√
4	cabbage	+1	+1	+1	+1	+1	√
5	farmer	+1	+1	+1	+1	+1	√
6	leave	+1	+1	+1	+1	+1	√
7	throw	+1	+1	+1	+1	+1	√
8	snow	+1	+1	+1	+1	+1	√
9	sweater	+1	+1	+1	+1	+1	√
10	rainy	+1	+1	+1	+1	+1	√
11	everywhere	+1	+1	+1	+1	+1	√
12	form	+1	+1	+1	+1	+1	√
13	towards	+1	+1	+1	+1	+1	√
14	information	+1	+1	+1	+1	+1	√
15	object	+1	+1	+1	+1	+1	√
16	decision	+1	+1	+1	+1	+1	√
17	wife	+1	+1	+1	+1	+1	√
18	field	+1	+1	+1	0	+1	√
19	passage	+1	+1	+1	+1	+1	√
20	arrive	+1	+1	+1	+1	+1	√

21	group	+1	+1	+1	+1	0	√
22	weekday	+1	+1	+1	+1	+1	√
23	noisy	+1	+1	+1	+1	+1	√
24	eye	+1	+1	+1	+1	+1	√
25	know	+1	+1	+1	+1	+1	√
26	problem	+1	+1	+1	+1	+1	√
27	jeep	+1	+1	0	+1	+1	√
28	walkman	+1	+1	+1	+1	+1	√
29	yourself	+1	+1	+1	+1	+1	√
30	giraffe	+1	+1	+1	+1	+1	√
31	drink	+1	+1	+1	+1	+1	√
32	juice	+1	+1	+1	+1	+1	√
33	idea	+1	+1	+1	+1	+1	√
34	kitchen	+1	+1	+1	+1	+1	√
35	quarter	+1	+1	+1	0	+1	√
36	hour	+1	+1	+1	+1	+1	√
37	bright	+1	+1	+1	+1	+1	√
38	homework	+1	+1	0	+1	+1	√
39	room	+1	+1	+1	+1	+1	√
40	August	+1	+1	+1	+1	+1	√
41	shine	+1	+1	+1	0	+1	√
42	X-ray	+1	+1	+1	+1	+1	√
43	cruel	+1	+1	+1	+1	+1	√

44	underground	+1	+1	+1	+1	0	√
45	middle	+1	+1	+1	+1	+1	√
46	result	+1	+1	+1	+1	+1	√
47	lovely	+1	+1	0	+1	+1	√
48	which	0	+1	+1	+1	+1	√
49	volleyball	+1	+1	+1	+1	+1	√
50	zebra	+1	+1	+1	+1	+1	√
Total		49	50	47	46	48	

Notes: 1. +1 = *the item is congruent with the objective*

2. -1 = *the item is not congruent with the objective*

3. 0 = *uncertain about this item*

Result of IOC:

(IOC = ∑R/ N)

Item number: 50

R = 49 + 50 + 47 + 46 + 48 = 240 (Scores from experts)

N = 5 (Numbers of experts)

IOC = 240/5 = 48

Percentage: 48/50x100% = 96%

The table above shows that the analysis result of IOC is 48, and the percentage is 96% which is higher than 80%. Therefore, the items are suitable for adoption in an interview questionnaire.

APPENDIX K

A Questionnaire on Students' Perceptions of Pronunciation Learning Via CALL-VT

Part I Personal Information

Gender	❑ Male		❑ Female	
Age	❑ Below 16-18	❑ 16-19	❑ 20 & up	
Level	❑ Year One	❑ Year Two	❑ Year Three	❑ Year Four
Nationality	❑ Han	❑ Buyi	❑ Miao	❑ other
Years of EFL study	❑ 0-5years	❑ 6-11years	❑ 12 years & up	
Time spent per week	❑ 10 hours	❑ 20 hours	❑ 30 hours	

Part II Students' Perceptions of pronunciation learning

Directions: This questionnaire is designed to gather information about your opinions on learning English pronunciation via CALL-VT. CALL-VT refers to learning pronunciation by applying a verbotonal approach in a self-regulated Computer Assisted Language Learning environment. Please read each statement carefully and tick (√) the response which represents your opinions. The numbers 5 to 1 stand for the following responses: 5 = strongly agree 4 = agree 3 = undecided 2 = disagree 1 = strongly disagree

No.	Statements of students' perceptions	5	4	3	2	1
1	CALL-VT is helpful in pronunciation learning	5	4	3	2	1
2	CALL-VT is effective in pronunciation learning	5	4	3	2	1
3	CALL-VT is interesting in pronunciation learning	5	4	3	2	1
4	I am happy with CALL-VT	5	4	3	2	1
5	I believe that my pronunciation is improved by using CALL-VT	5	4	3	2	1
6	I prefer CALL-VT to the traditional approach in pronunciation learning	5	4	3	2	1
7	I would like to learn pronunciation via CALL-VT in a computer lab	5	4	3	2	1
8	I would like to learn pronunciation via CALL-VT in a classroom	5	4	3	2	1
9	I would like to learn pronunciation via CALL-VT on my own	5	4	3	2	1
10	My learner autonomy is improved via CALL-VT	5	4	3	2	1

APPENDIX L

A Questionnaire on Students' Perceptions of Pronunciation Learning Via CALL-VT

(Chinese Version)

学生关于用 CALL-VT 方式学习语音的观点调查问卷

第 I 部分 个人信息

性别	❑ 男		❑ 女	
年龄	❑ 16—18 岁以下	❑ 16—19 岁		❑ 20 岁以上
年级	❑ 大一	❑ 大二	❑ 大三	❑ 大四
民族	❑ 汉族	❑ 布依族	❑ 苗族	❑ 其他
学习时间	❑ 0-5 年	❑ 6-11 年		❑ 12 年以上
每周学习时间	❑ 10 小时	❑ 20 小时		❑ 30 小时

第 II 部分 学生对于利用 CALL-VT 学习语音的观点

说明：本问卷旨在收集你对于通过 CALL-VT 系统学习英语语音的观点。CALL-VT 指的是应用视觉听觉法（verbo-tonal approach ）在计算机辅助语言环境下进行英语语音学习。请认真阅读问卷的题项并在代表你观点的题项后面打勾（√）数字 5 到 1 分别代表以下答案： 5 = 非常同意； 4 = 同意； 3 = 不确定； 2 = 不同意； 1 = 非常不同意

题项	学生观点描述	5	4	3	2	1
1	CALL-VT 有助于语音学习	5	4	3	2	1
2	CALL-VT 对语音学习有效	5	4	3	2	1
3	CALL-VT 很有趣	5	4	3	2	1
4	我乐意利用 CALL-VT 学习	5	4	3	2	1
5	我相信我能利用 CALL-VT 提高语音水平	5	4	3	2	1
6	比较传统的语音学习方式， 我更喜欢 CALL-VT	5	4	3	2	1
7	我更愿意在语言实验室利用 CALL-VT 学习	5	4	3	2	1
8	我更愿意在语言教室利用 CALL-VT 学习	5	4	3	2	1
9	我更愿意单独利用 CALL-VT 学习	5	4	3	2	1
10	通过 CALL-VT 学习系统，我的自主学习能力提高了	5	4	3	2	1

APPENDIX M

A Semi-structured Interview on Students' Perceptions of Pronunciation Learning Via CALL-VT

Part I Personal Information

Gender	❑ Male		❑ Female	
Age	❑ Below 16-18	❑ 16-19	❑ 20 and up	
Grade	❑ Year One	❑ Year Two	❑ Year Three	❑ Year Four
Nationality	❑ Han	❑ Buyi	❑ Miao	❑ other
Years of EFL study	❑ 0-5years	❑ 6-11years	❑ 11 years & up	

Part II Interview Questions

Q1. Is CALL-VT helpful to your pronunciation learning? If yes, in what ways? If not, why not?

Q2. Is CALL-VT interesting? If yes, in what ways?

Q3. Do you believe that CALL-VT will improve your pronunciation? Why?

Q4. Do you like CALL-VT? Why?

Q5. Please describe your opinions of CALL -VT.

Q6. What are your feelings when learning pronunciation with the CALL -VT approach?

Q7. Comparing the traditional approach to CALL -VT, which one do you prefer?

Why?

Q8. Do you like learning pronunciation in groups or on your own? Why?

Q9. Do you think CALL-VT can help you to develop learner autonomy? Why?

Q10. What else would you like to say about pronunciation learning?

APPENDIX N

A Semi-structured Interview on Students' Perceptions of Pronunciation Learning

（Chinese Version）

第 I 部分 个人信息

性别	❑ 男		❑ 女	
年龄	❑ 16-18 岁以下	❑ 16-19 岁		❑ 20 岁以上
年级	❑ 大一	❑ 大二	❑ 大三	❑ 大四
民族	❑ 汉族	❑ 布依族	❑ 苗族	❑ 其他
学习时间	❑ 0-5 年	❑ 6-11 年		❑ 12 年以上
每周学习时间	❑ 10 小时	❑ 20 小时		❑ 30 小时

第 II 部分 访谈问题

问题 1. CALL-VT 对你的语音学习有帮助吗？如果有，在哪些方面有帮助？如果没有，为什么？

问题 2. CALL-VT 有趣吗？如果有趣，怎么有趣？

问题 3. 你认为 CALL-VT 能提高你的语音水平吗？为什么？

问题 4. 你喜欢 CALL-VT 吗？为什么？

问题 5. 请你描述你对 CALL-VT 的观点

问题 6. 你用 CALL-VT 这种方法学习语音的感受是什么？

问题 7. 比较传统的语音学习方法而言，你是否更喜欢 CALL-VT？为什么？

问题 8.　你喜欢单独一个人还是与其他人一起学习语音？为什么？

问题 9.　你认为 CALL-VT 有助于提高你的自主学习能力吗？为什么？

问题 10.　对于 CALL-VT 系统下的语音学习，你还什么其他补充看法？

APPENDIX O

IOC Analysis for the Questionnaire on Students' Perceptions of Pronunciation Learning

Item	Experts					Result of analysis
	1	2	3	4	5	
Q1	+1	+1	+1	+1	+1	√
Q2	+1	+1	+1	+1	+1	√
Q3	+1	+1	+1	+1	+1	√
Q4	+1	+1	+1	+1	+1	√
Q5	+1	+1	+1	+1	0	√
Q6	+1	+1	0	+1	+1	√
Q7	-1	+1	+1	+1	+1	√
Q8	0	+1	+1	+1	+1	√
Q9	+1	+1	--1	+1	+1	√
Q10	+1	0	+1	+1	+1	√
Total	7	9	7	10	9	√

Notes: 1. +1 = *the item is congruent with the objective*

2. -1 = *the item is not congruent with the objective*

3. 0 = *uncertain about this item*

Result of IOC:

(IOC = ∑R/ N)

Item number: 10

R = 7 + 9 + 7 + 10 + 9 = 42(Scores given by experts)

N = 5 (Number of experts)

IOC = 42/5 = 8.4

Percentage: 8.4/10 x 100% = 84%

The above table shows that the analysis result of IOC is 8.4, and the percentage is 84% which is higher than 80%. Therefore, the items are suitable for adoption as a trial questionnaire.

APPENDIX P

IOC Analysis for the semi-structured interview of students

Item	Experts					Result of analysis
	1	2	3	4	5	
Q1	+1	0	+1	+1	+1	√
Q2	+1	+1	0	+1	+1	√
Q3	+1	+1	+1	+1	+1	√
Q4	+1	+1	+1	+1	+1	√
Q5	+1	+1	+1	+1	0	√
Q6	+1	+1	+1	+1	+1	√
Q7	--1	+1	+1	+1	0	√
Q8	+1	0	+1	+1	+1	√
Q9	+1	+1	-1	+1	+1	√
Q10	+1	+1	+1	+1	+1	√
Total	8	8	7	10	8	√

Notes: 1. +1 = *the item is congruent with the objective*

2. -1 = *the item is not congruent with the objective*

3. 0 = *uncertain about this item*

Result of IOC:

(IOC = ∑R/ N)

Item number: 10

R = 8 + 8 + 7+ 10 + 8 = 41 (Scores from experts)

N = 5 (Numbers of experts)

IOC = 41/5 = 8.2

Percentage: 8.2/10 x 100% = 82%

The above table shows that the analysis result of IOC is 8.2, and the percentage is 82% which is higher than 80%. Therefore, the items are suitable for adoption in an interview questionnaire.

APPENDIX Q

A Semi-structured Interview on Teachers' Perceptions of CALL-VT

Q1. Is CALL-VT helpful to your pronunciation teaching? If yes, in what ways? If no, why not?

Q2. Is CALL-VT interesting? If yes, in what ways?

Q3. Do you believe that CALL-VT will improve students' pronunciation? Why?

Q4. Do you like CALL-VT? Why?

Q5. Please describe your opinions of CALL-VT.

Q6. What was your feeling when teaching pronunciation in the CALL-VT approach?

Q7. Comparing the traditional approach with CALL-VT, which one do you prefer? Why?

Q8. Would you like your students to learn pronunciation in group or on their own? Why?

Q9. Do you think CALL-VT can help you develop students' learner autonomy? Why?

Q10. What else would you like to say about pronunciation teaching?

APPENDIX R

A Semi-structured Interview on Teachers' Perceptions of CALL-VT

(Chinese version)

Q1. 你认为 CALL-VT 语音学习系统 对语音教学有帮助吗? 如果有，在什么方面有帮助? 如果没有，为什么？

Q2. 你认为 CALL-VT 语音学习系统有趣吗？如果有，请说出理由。

Q3. 你相信 CALL-VT 会提高学生的语音能力吗? 为什么?

Q4. 你喜欢使用 CALL –VT 进行语音教学吗? 为什么?

Q5. 请描述你对 CALL –VT 的评价。

Q6. 你对 CALL –VT 的感受怎样?

Q7. 比较 CALL –VT 与传统的语音教学模式，你更喜欢那一种教学方法? 为什么？

Q8. 你喜欢学生以小组的形式还是以个人的形式学习语音？为什么？

Q9. 你认为 CALL –VT 有助于学生的自主学习能力提高吗？请说出理由。

Q10. 对于语音教学，你还有什么补充说明吗？

APPENDIX S

IOC Analysis for the semi-structured interview of teachers

Item	Experts					Result of analysis
	1	2	3	4	5	
Q1	0	+1	+1	+1	+1	√
Q2	+1	+1	0	+1	+1	√
Q3	+1	+1	+1	+1	+1	√
Q4	+1	+1	+1	+1	+1	√
Q5	+1	+1	+1	+1	0	√
Q6	+1	+1	+1	+1	+1	√
Q7	-1	+1	+1	+1	+1	√
Q8	+1	0	+1	+1	+1	√
Q9	+1	+1	-1	+1	+1	√
Q10	+1	+1	+1	+1	0	√
Total	7	9	7	10	8	√

Notes: 1. +1 = *the item is congruent with the objective*

2. -1 = *the item is not congruent with the objective*

3. 0 = *uncertain about this item*

Result of IOC:

(IOC = ∑R/ N)

Item number: 10

R = 7 + 9 + 7+ 10 + 8 = 41(Scores from experts)

N = 5 (Numbers of experts)

IOC = 41/5 = 8.2

Percentage: 8.2/10 x 100% = 82%

The above table shows that the analysis result of IOC is 8.2, and the percentage is 82% which is higher than 80%. Therefore, the items are suitable for adoption in an interview questionnaire.

APPENDIX T

Autonomous Pronunciation Learning Questionnaire

Directions: This questionnaire is designed to gather information about your state of autonomous pronunciation learning. Please read each statement carefully and tick (√) the response which represents your opinions. The numbers 1 to 5 stand for the following responses: 1 = strongly disagree 2 = disagree 3 = undecided 4 = agree 5 = strongly agree

Part I Personal Information

Gender	❑ Male		❑ Female	
Age	❑ Below 16	❑ 16-19	❑ 20 up	
Level	❑ Year One	❑ Year Two	❑ Year Three	❑ Year Four
Nationality	❑ Han	❑ Buyi	❑ Miao	❑ other
Years of EFL study	❑ 0-5years	❑ 6-11years	❑ 11 years up	
Time spent per week	❑ 10 hours	❑ 20 hours	❑ 30 hours	

Part II Students' state of pronunciation learning

Item	1	2	3	4	5
1. I can evaluate my pronunciation ability on my own					
2. I can evaluate my pronunciation ability together with a learning advisor					
3. I can evaluate my pronunciation ability with a test					
4. I can identify my needs on pronunciation learning on my own.					
5. I can identify my needs on pronunciation learning with a learning advisor					
6. I can identify my needs with a test					
7. I can set myself goals on pronunciation learning on my own					
8. I can set myself goals of pronunciation learning with a learning advisor					
9. I can set myself goals of pronunciation learning with a test					
10. I can plan a time for pronunciation learning on my own					
12. I can plan a time for pronunciation learning with a learning advisor					
13. I can plan a place for pronunciation learning on my own					
14. I can plan a place for my pronunciation learning with an advisor					

APPENDIX U

Autonomous Pronunciation Learning Questionnaire

(Chinese Version)

英语语音自主学习问卷

说明: 本问卷旨在调查你关于语音自主学习的状况。请仔细阅读并在代表你情况的相应题项后面打勾。数字 1 到 5 分别代表 1 = 非常不符合 2 = 不符合 3 = 不确定 4 = 符合 5 = 非常符合

第一部分 个人信息

性别	❑ 男		❑ 女	
年龄	❑ 16 以下	❑ 16-19		❑ 20 以上
年级	❑ 大一	❑ 大二	❑ 大三	❑ 大四
民族	❑ 汉族	❑ 布依族	❑ 苗族	❑ 其他
英语学习时间	❑ 0-5 年	❑ 6-11 年		❑ 12 年以上
每周学习时间	❑ 10 小时	❑ 20 小时		❑ 30 小时

第二部分 问卷题项

题项	1	2	3	4	5
1. 我能自己评价我的语音水平	1	2	3	4	5
2. 我通过老师评价我的语音水平	1	2	3	4	5
3. 我通过测试来评价我的语音水平	1	2	3	4	5
4. 我能自己确定语音学习的目标	1	2	3	4	5
5.老师和我一起确定语音学习的目标	1	2	3	4	5
6. 我通过测试来确定语音学习的目标	1	2	3	4	5
7. 我能自己计划语音学习的时间	1	2	3	4	5
8. 老师和我共同计划语音学习的时间	1	2	3	4	5
9. 我能自己选择语音学习的地点	1	2	3	4	5
10. 老师和我一起选择语音学习的地点	1	2	3	4	5

APPENDIX V

Rubrics for Pronunciation

Dear Colleague,

Thank you for agreeing to help with this research project. I would like you to rate students' recordings as outlined below.

There are two parts to be rated: (a) a read-aloud passage and (b) a face-to-face interview. Please rate only the interviewees. The following consists of rubrics for rating the recordings in terms of comprehensibility, pronunciation (nativeness), and fluency.

You are required to give two scores. The first is a band rating between 1 to 5 as specified below, for each of the recordings.

The second is a score within the band range specified below. The maximum score for the read-aloud passage is 30. The maximum score for the interview is 25.

In arriving at your rating and score, please take the time to listen thoroughly to the speakers. They sometimes start slowly or hesitantly. Please try to avoid making hasty judgments and try to make judgments on samples that seem typical of their normal speech e.g., once students are in full flight.

Part I: Rubrics for comprehensibility.

Rating	Score (passage-reading)	Score (interview)
1 = incomprehensible	0-5	0-5
2 = slightly comprehensible	6-11	6-10
3 = significantly comprehensible	12-17	11-15
4 = mostly comprehensible	18-23	16-20
5 = fully comprehensible	24-30	21-25

Example:

You rate a student for comprehensibility at band 4 (mostly comprehensible), and give him/her a score of 20 (this means that in your opinion, the student is in the middle of the band).

Part II Rubric for pronunciation (nativeness)

Rating	Score (passage-reading)	Score (interview)
1 = intrusive	0-5	0-5
2 = fairly good	6-11	6-10
3 = good	12-17	11-15
4 = very good	18-23	16-20
5 = advanced	24-30	21-25

Part III Rubrics for fluency.

Rating	Score (passage-reading)	Score (interview)
1 = intrusive	0-5	0-5
2 = Intermediate	6-11	6-10
3 = Good	12-17	11-15
4 = Advanced	18-23	16-20
5 = Native-like	24-30	21-25

Example:

Student number	Passage-reading						Interview					
	C		P		F		C		P		F	
1200	5	27	4	20	3	16	3	13	4	18	3	15

C = Comprehensibility

P = Pronunciation (nativeness)

F = Fluency

CURRICULUM VITAE

Bi He was born on January 8, 1978 in Xingyi, China. She graduated from Guizhou University for Nationalities in 2000 with a Bachelor of Arts degree in English Language and Literature. In 2010, she earned a Master of Arts in English Language Studies in Suranaree University of Technology, Thailand.

Upon graduation from Guizhou University for Nationalities in 2000, Bi He started her teaching career at Xingyi Normal University for Nationalities. She is currently an associate professor at the School of Foreign Languages, Xingyi University for Nationalities, China.

Since 2011, she has been enrolled in the Ph.D. program of English Language Studies at the School of Foreign Languages, Institute of Social Technology, Suranaree University of Technology, Thailand. Her research interests include verbotonalism, autonomy, language-learning methodology and Computer Assisted Language Learning.