

การสร้างแบบฝึกโดยใช้เทคโนโลยีการรู้จำเสียงพูดในการพัฒนาทักษะการออกเสียงของนักเรียนนักศึกษา ปริญญาตรีที่เรียนภาษาอังกฤษเป็นภาษาต่างประเทศในจังหวัดเชียงใหม่

จตุยานุณี เพิ่มทันจิตต์¹

มหาวิทยาลัยราชภัฏเชียงใหม่

บทความวิจัย

รับต้นฉบับ: 17 มกราคม 2566

รับตีพิมพ์: 6 มิถุนายน 2566

การวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อสร้างแบบฝึกโดยใช้เทคโนโลยีการรู้จำเสียงพูดในการพัฒนาทักษะการออกเสียงพยัญชนะควบกล้ำภาษาอังกฤษและศึกษาค่าประสิทธิภาพของแบบฝึก 2) เพื่อเปรียบเทียบผลสัมฤทธิ์ทางการเรียนทางการออกเสียงพยัญชนะควบกล้ำภาษาอังกฤษหลังเรียนของกลุ่มทดลองที่เรียนโดยใช้แบบฝึกโดยใช้เทคโนโลยีการรู้จำเสียงพูดและกลุ่มควบคุมที่เรียนโดยใช้วิธีปกติ และ 3) เพื่อศึกษาความพึงพอใจของนักศึกษาที่มีต่อการใช้แบบฝึกโดยใช้เทคโนโลยีการรู้จำเสียงพูด กลุ่มตัวอย่างแบ่งออกเป็น 2 กลุ่ม ได้แก่ นักศึกษาชั้นปีที่ 2 สาขาวิชาครุศาสตร์ภาษาอังกฤษ จำนวน 29 คน ที่เรียนโดยใช้แบบฝึกโดยใช้เทคโนโลยีการรู้จำเสียงพูดและกลุ่มควบคุมได้แก่ นักศึกษาชั้นปีที่ 2 สาขาวิชาครุศาสตร์ภาษาอังกฤษ จำนวน 27 คนที่เรียนโดยใช้การสอนปกติในภาคการศึกษาที่ 1 ปีการศึกษา 2565 มหาวิทยาลัยราชภัฏเชียงใหม่ ส่วนเครื่องมือในการวิจัยประกอบด้วยแบบฝึกโดยใช้เทคโนโลยีการรู้จำเสียงพูด แบบทดสอบวัดผลสัมฤทธิ์ทักษะการออกเสียงภาษาอังกฤษ และแบบสอบถามความพึงพอใจ ส่วนค่าสถิติในการวิเคราะห์ข้อมูล ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐานและการทดสอบค่าที

ผลการวิจัยพบว่าข้อมูลจากการเปรียบเทียบต่างระหว่างโครงสร้างพยางค์ พบพยัญชนะควบกล้ำต้นคำ 2 เสียง จำนวน 23 พยัญชนะควบกล้ำ และพยัญชนะควบกล้ำต้นคำ 3 เสียง จำนวน 7 พยัญชนะควบกล้ำ ส่วนพยัญชนะท้ายคำ 2 เสียง พบจำนวน 43 พยัญชนะควบกล้ำ พยัญชนะท้ายคำ 3 เสียงจำนวน 29 พยัญชนะควบกล้ำ และ พยัญชนะท้ายคำ 4 เสียงจำนวน 6 พยัญชนะควบกล้ำ และค่าประสิทธิภาพของแบบฝึกโดยใช้เทคโนโลยีการรู้จำเสียงพูด เท่ากับ 79.40/74.71 นอกจากนั้นผลสัมฤทธิ์ทางการอ่านออกเสียงพยัญชนะควบกล้ำหลังเรียนของนักศึกษาที่เรียนโดยการใช้แบบฝึกโดยใช้เทคโนโลยีการรู้จำเสียงพูดสูงกว่านักศึกษาที่เรียนโดยใช้วิธีการเรียนแบบปกติอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และนักศึกษามีความพึงพอใจต่อการใช้แบบ

¹ การติดต่อและการร้องขอบทความนี้ กรุณาส่งถึง จตุยานุณี เพิ่มทันจิตต์ มหาวิทยาลัยราชภัฏเชียงใหม่ e-mail : jutiyaneep@hotmail.com

ฝึกโดยใช้เทคโนโลยีการรู้จำเสียงพูดในภาพรวมในระดับมากที่สุดคิด (ค่าเฉลี่ยเท่ากับ 4.52 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.50)

คำสำคัญ: แบบฝึกทักษะการออกเสียง เทคโนโลยีการรู้จำเสียงพูด สหวิทยาภาษาอังกฤษ

THE CONSTRUCTION OF A SPEECH RECOGNITION TECHNOLOGY-BASED PRACTICE
MODULE TO IMPROVE THE PRONUNCIATION SKILL OF UNDERGRADUATE STUDENTS
STUDYING ENGLISH AS A FOREIGN LANGUAGE IN CHIANG MAI PROVINCE

Jutiyanee Permthanjit¹

Chiang Mai Rajabhat University Thailand

Research Article

Received: 17 January 2023

Accepted: 6 June 2023

The objectives of this research were 1) to construct a speech recognition technology-based practice module to improve pronunciation skill of English consonant clusters and find out the efficiency of the practice module 2) to compare the learning achievement on pronunciation skill of English consonant clusters between the experimental and controlled groups 3) to explore the satisfaction of the students with the implementation of the practice module. The sample group consisted of 29 year 2 English major students in Faculty of Education who studied with the supplement of practice module and the controlled group of 27 year 2 English major students in Faculty of Education who studied via the traditional methods in first semester of 2022 academic year at Chiang Mai Rajabhat University. The research instruments consisted of the practice module, the achievement test on English consonant cluster pronunciation, and a satisfaction questionnaire. The data were statistically analyzed by mean and standard deviation, and t-test.

The research results revealed the contrastive analysis of the syllable structures of English and Standard Thai, the wordlist contains 23 two-initial consonant cluster words, 7 three-initial consonant cluster words, 43 two-final consonant cluster words, 29 three-final consonant cluster words, and 6 four-final consonant cluster words. The efficiency of the practice module was 79.40/74.71. The comparison of the achievement test between

¹ Correspondence concerning this article and requests for reprints should be addressed to Jutiyanee Permthanjit Chiang Mai Rajabhat University E-mail : jutiyanee_p@hotmail.com

the experimental group and the controlled group revealed that the achievement of the experimental group was significantly higher than that of the controlled group at the .01 level. The overall satisfaction with the use of the practice module was at the highest level ($M = 4.52, S = 0.05$).

Keywords: Pronunciation Practice Module, Speech Recognition Technology, English
Phonology

Introduction

Pronunciation skill is fundamental to learning and teaching English. Unclear pronunciation can lead to unintelligibility or communication breakdown (Gilakjani & Ahmadi, 2011). Thus, clear pronunciation is crucial to language use efficiency for humans to communicate. The attention to pronunciation errors is important for explanations why we pronounce incorrectly. Ur (2000) summarizes three causes of mispronunciation among Thai students studying English. Firstly, some English sounds do not exist in Thai, making learners unfamiliar with those sounds, so they replace them with Thai equivalents. Secondly, some English sounds do exist in Thai, but are not distinct phonemes. This does not make learners realize that they can make a difference in meaning. Finally, sometimes learners are able to pronounce correctly but do not know how to stress words and use sentence intonation. Instead, they use features in their native language and subsequently cause miscommunication. Moreover, syllable structures of Thai and English are not completely the same. In Thai, two initial consonants and only one final consonant are allowed, while as many as three initial consonants and four final consonants are possible in English. This is another major hurdle for Thai students to be unable to pronounce English consonant clusters correctly (Pokaisawan, 2008).

The 4.0 Thai education emphasizes the learner-centered approach with an integration of technology and innovations to cultivate problem-solving, analytical and creative processes in learners, so that they can become inquisitive and autonomous as well as apply technology to improve their potentials efficiently. Speech recognition technology is a speech-to-text technology that creates texts from dictation by understanding the meanings of data series extracted from speech signals and converting them into texts. Nowadays, the technology has been improved for better precision and shorter response duration. It is also convenient and compatible with various types of equipment. As a consequence, it can be well applied to teaching and learning a second or foreign language (Hincks, 2003; Livescu, 2005; Liakin, Cardoso & Liakina, 2015; McCrocklin, 2015; and Evers & Chen, 2020). Another way for second language learners to improve pronunciation is the use of a skill practice module. Previous studies revealed that skill practice modules could actually help solve the problem of reading out loud as illustrated by higher posttest scores of pronunciation abilities of learners (Mahaphainun, 2004; Wongphanit, 2005; Sriwilai, 2007; Thanyakijanukit, 2010; and Etiratana, 2010). For pronunciation problem, the speech recognition technology is suitable

for solving the problem because learners are able to listen to the pronunciation or speech of English native speakers. After pronouncing a word, a learner gets an instant feedback to inform him/her whether his/her pronunciation is correct or not. Furthermore, learners are able to practice outside of classroom, which would concretely improve their pronunciation skills. Additionally, it is an instructional medium suitable for the current learning and teaching situation where COVID-19 has been spreading and a physical distancing learning approach is encouraged. This would be advantageous for learners to practice English pronunciation with instant feedbacks via computers. It is an effective means to reduce a chance to transmitting or contacting the virus.

Chiang Mai province has several tertiary educational institutions with a goal of local development. Thus it is found that a large number of students are from diverse ethnic, cultural and linguistic backgrounds (Arunsirot, 2021). From 15 years of personal teaching experience, a major cause of incorrect pronunciation of students is due to the fact that the phonological systems of their mother tongues and English are different. Arunsirot (2017) further pointed out that English is regarded as the third language for ethnic minority students besides Standard Thai and their respective mother tongues. Another factor affecting the efficiency of students' English pronunciation is a large classroom, depriving students of sufficient time to practice in class. A lack of English native teachers is another important cause for ineffectiveness of pronunciation learning since imitating sounds from native speakers is crucial for correct pronunciation. However, in the English teaching and learning situation at the universities, most instructors are Thais, thus somewhat depriving the students of good role models in English pronunciation. Moreover, having no instructional media that meet learners' needs is another important hurdle in improving their pronunciation skills.

In this research, the speech recognition technology is used to design the practice module to improve the pronunciation of English consonant clusters of undergraduate students learning English as a foreign language in an attempt to correct and solve their English pronunciation problem. It also provides instructors an efficient, technology-based instructional medium for teaching and learning English pronunciation. This is because the practice module based on speech recognition technology focuses on a learning approach that is totally different from what usually goes on in a regular classroom. The emphasis is on correcting and improving pronunciation errors by students themselves, which promotes

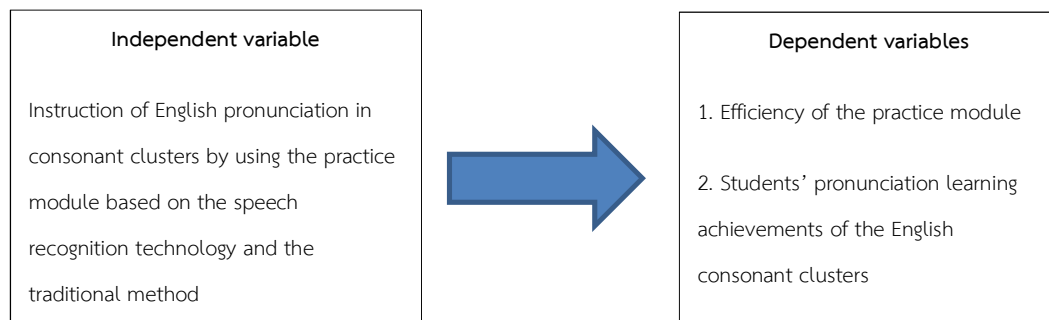
learner autonomy and improves their learning capabilities (Kochevar, cited in Numee, 1996). This speech recognition technology-based practice module would ultimately result in students having a better English language use, communicating more efficiently, and using this fundamental knowledge to improve other skills.

Objectives of the research

1. To construct a practice module based on the speech recognition technology to improve the pronunciation of English consonant clusters of undergraduate students in Chiang Mai province and find out the efficiency of the practice module,
2. To compare learning achievements in pronunciation of English consonant clusters between the experimental group after learning via the practice module and the controlled group learning via the traditional method, and
3. To explore the satisfaction with the implementation of the speech recognition technology-based practice module of the experimental group

Conceptual framework

The conceptual framework of this research is illustrated as follows.



Literature review

Problems of teaching English pronunciation in Thailand

Besides helping students to pronounce correctly, teaching English pronunciation enables students to greatly improve their speaking skills (Harmer, 2001). Nawayut (2013) stated that a large number of English learners are unable to pronounce correctly or native-like. There are several root causes as follows. Firstly, there is little or no opportunity to use

English on a daily basis. Most Thai students do not have many opportunities to listen to or use English on a continuous basis as the language is neither a home or community language, and neither is it an official language in the country. Therefore, Thai people cannot be exposed to good English models in different situations. This is different from those who are in an environment where native speakers are the majority or those who grow up in an English speaking society. In these two cases, learners gradually and naturally absorb the language. When they start using the language, they are able to utter with the near-native accent. Secondly, age is an essential factor for foreign or near-native accent. Generally, children learn a foreign language better than adults do. This is because adults' language mechanisms in their brains have limited capacities to learn a new language. Thirdly, aptitude plays a crucial role in learning a new language. Some learners can hear and imitate sounds better than others. Aptitude is also related to age. People with older ages have lower capabilities to learn a new language. Fourthly, the differences in phonological systems and syllable structures between Thai and English are a major hurdle to acquire English efficiently. Each language has its own phonological, morphological, syllabic, syntactic and semantic structures. The more different these structures are, the more difficult they become for learners to acquire. The next factor is knowledge of English teachers. They cannot be good role models on pronunciation for learners due probably to being underqualified for teaching English, lack of skills for teaching pronunciation, or being deprived of additional training. Another factor is class size and instructional media. Classes of 40 or more students do not facilitate learners to practice pronunciation. In addition, schools may not have adequate learning facilities, such as, computer programs for pronunciation practice, sound laboratories, or self-access centers. These facilities can provide learners an opportunity to practice pronunciation. Another factor is a test-driven curriculum. This type of curriculum overlooks correct pronunciation but focuses on reading and writing that most examinations require. Thus, both teachers and learners put more emphasis on more urgent issues. The next factor is misunderstanding about communicative learning. Many understand that as long as learners are able to communicate, pronunciation is not so important. The problem is that pronunciation is a major component of the ability to communicate verbally. When pronunciation is incorrect, communication is subsequently ineffective. Another factor is the concept of being able to sound native-like. The concept that a good or correct pronunciation must be near-native or native-like is to possess the British or American accent. This is far beyond the expectation of most teachers and discourages learners. Finally, it is

the attitude of some Thai learners. These learners do not wish to speak like native speakers because they think that other people may not like it or perceive that it does not sound natural. This is another cause of pronunciation failure. For those learning English at older ages, they are often worried about their pronunciation styles, making it even more difficult to acquire and master the skill. It has thus become another important problem for correct English pronunciation.

The importance of speech recognition for learning languages

Levis & Suvorov (2013) defined ‘automatic speech recognition (ASR)’ as an independent, machine-based process of decoding and transcribing oral speech and it plays a vital role in English language learning especially in improving pronunciation. ASR is a leading technology that allows students to interact with data-processing programs through vocalization (Sidgi & Shaari, 2017). At the same time, it can provide immediate feedback that can help students to correct mispronunciations in their learning as soon as they make a mistake. Furthermore, it allows the students to listen to and practice their pronunciation in an enjoyable environment. It also gives each individual student corrections, which is difficult to achieve in class with a huge class size (Sidgi & Shaari, 2017). With distinct benefits of ASR mentioned above, Eskenazi (1999) used it for training students to improve their accents in a foreign language. Additionally, ASR can reflect the pronunciation level of English learners and is a good and effective way to practice pronunciation (Tu & Chen, 2011). Neri et al. (2003) stated that it offers extra learning time and material as well as the possibility to practice in a stress-free environment. It can improve pronunciation of targeted problematic speech sounds and the errors being addressed in the training. Also, Kim (2006) mentioned that ASR software can be exploited as a valuable tool for teaching EFL students’ pronunciation due to shortages of native English speaking teachers.

Research methodology

Population and sample group

The population was undergraduate students who were studying English as a foreign language in Chiang Mai province. The purposive sampling method was applied to select the sample group who were the second-year English education majors at Chiang Mai Rajabhat University. They were divided into two groups: the experimental and the controlled groups.

The experimental group comprised 29 students who studied by using the speech recognition technology-based practice module and the controlled group was 27 students who studied by the traditional method. They were enrolling in the ENG 2706 English Phonetics and Phonology course in the first semester of the 2022 academic year.

Research instruments

The research instruments are as follows:

1. The speech recognition technology-based practice module. The steps for constructing the module are as follows.

1.1 Theories and studies about the construction of skill practice modules and basic operational principles as well as development guidelines of the speech recognition system were investigated.

1.2 A contrastive analysis was performed to find out about problematic English consonant clusters for Thai students. Studies about syllable structures of Standard Thai and the English phonological system were examined.

1.3 English words with consonant clusters found to be problematic to Thai students were selected. There are 240 words divided into four lessons. They include problematic words with two initial consonant clusters, problematic words with three initial consonant clusters, problematic words with two final consonant clusters, and problematic words with three and four final consonant clusters.

1.4 The draft of the practice module was outlined, based on the practice module construction steps posited by Sanohsamniang (1999) and Mascharas et al. (2007). For each lesson, the students were required to read the instructions, take the pre-test, study the content, do the pronunciation practice, examine the summary, and take the posttest respectively.

1.5 The design of the module was intended for the Android Operation System on smartphones. The Adobe Photoshop or Adobe Illustrator was used to design the media screen and the Adobe Captivate and Android Studio were used to design and construct the practice module.

1.6 The storyboard was formulated that described the components, icons and pictures, and colors of the module, for instance, font sizes and shapes, flows and connections, or responses to users. After that, they were assessed by three experts in media techniques and production for accuracy and propriety in order to perform further improvement.

1.7 The module was constructed by using the speech recognition technology with the following details.

1.7.1 The Adobe Illustrator and Adobe Photoshop programs were used to create graphic works and decorate the still pictures prepared for the construction. The results were that the graphics and pictures were attractive and meaningful according to the contents. They were then uploaded onto the Adobe Captivate program in order to use them as the components of the module.

1.7.2 The Flash Program was utilized to create moving pictures.

1.7.3 The Adobe Captivate program was used to create the computer-assisted lessons.

1.7.4 The Android Studio program was employed to develop the module on the Android Operation System. After the completion of the development, the finished work was imported or installed on the touchscreen of the instrument in the *.apk* file. The test was performed to find out whether the module could work accurately and properly.

The construction steps are illustrated in Figure 1.

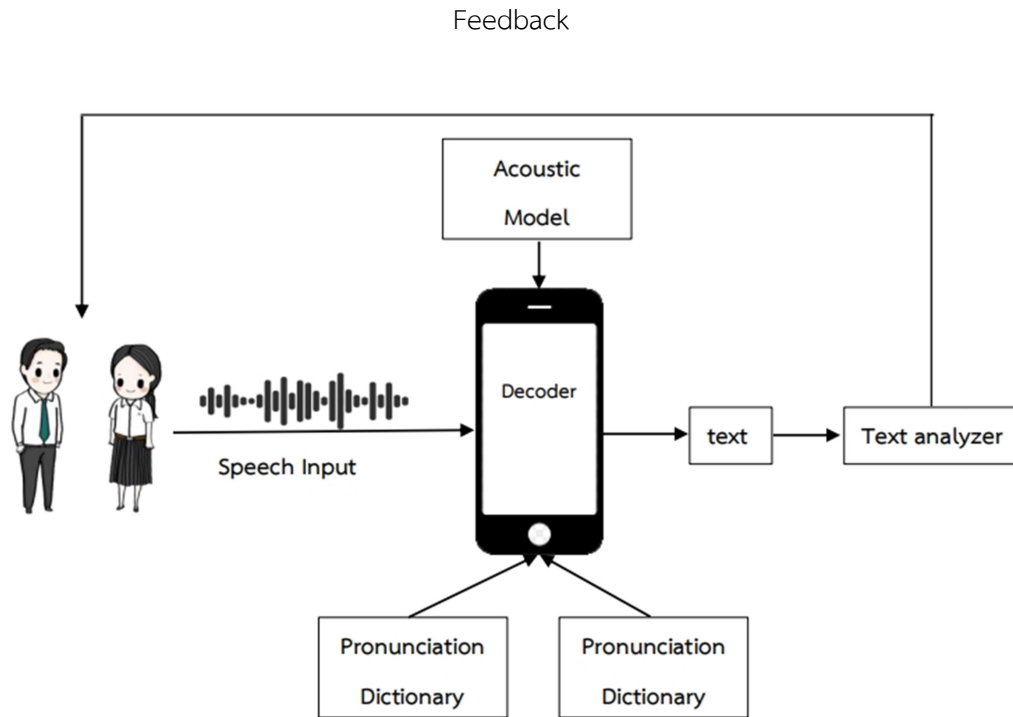


Figure 1 The construction steps of the practice module using the speech recognition technology

1.8 The module was then presented to five technical experts and five experts on contents to assess its quality. Improvement was performed according to the experts' suggestions before it was implemented with the experimental group. From the assessment of the contents, presentation, colors, pictures, sounds, fonts, and instructions by using the five-point rating scale of Likert (1967), it was revealed that the overall propriety of the module was at a high level ($M=4.36$, $S = 0.61$). The assessment on the contents, comprising the index of item-objective congruence (IOC), content division, story flow, pictures and language, tests, and lesson management, revealed that the overall content propriety was at a high level ($M=4.37$, $S= 0.50$).

2. The lesson plans in this study described a teaching plan of English pronunciation course focusing on consonant cluster words for second-year English pre-service teachers. The time required to complete the lessons was one month (four weeks). It consisted of four chapters: problematic words with two initial consonant clusters, problematic words with three initial consonant clusters, problematic words with two final consonant clusters, and problematic

words with three and four final consonant clusters. Each lesson comprised three steps: opening, self-study with the speech recognition technology-based practice module, and closing. The evaluation result of the lesson plans based on the index of item-objective congruence (IOC) indicated that all items in the lesson plans received the IOC scores higher than 0.92, indicating that they were all acceptable.

2. The achievement test was constructed to assess the English consonant cluster pronunciation skills of the experimental group who studied with the supplement of the practice module and the controlled group who studied with the traditional method. The steps of constructing the achievement test are as follows.

2.1 Thirty words with both initial and final consonant clusters regarded as problematic to Thai students were selected.

2.2 The passing criterion was set at 70%, which was based on the criteria established by the Department of Academic Affairs, Ministry of Education (2011).

2.3 The test was then assessed for its accuracy and propriety by three experienced English instructors. The index of item-objective congruence score was 0.86, which indicating that they were accepted.

Data collection and analysis

This study was conducted with two classes during the four weeks of the English phonetics and phonology course. The speech recognition technology-based practice module was implemented in the experimental group while the control group was given a traditional method. The oral exams took place in the exam periods before and after the instruction. It took approximately ten minutes with a total of 30 marks. In the oral examination, each student had to pronounce the consonant cluster words. If he/she could pronounce each word correctly, he/she would get one point per word. If not, he/she would get no point. Then, the students' achievements in consonant cluster pronunciation were compared between the experimental and controlled groups by using the t-test, standard deviation, mean, and percentage.

3. The satisfaction assessment of the students toward the practice module

3.1 A documentary research on the construction of a questionnaire was conducted in order to create the satisfaction questionnaire for this study. The questionnaire was divided into three parts: general information of the respondents, information about their satisfaction with the speech recognition technology-based practice module, and additional suggestions. It is a five-point rating scale: *highest, high, moderate, low, and lowest*.

3.2 The questionnaire was assessed for its content validity by three experts and the IOC was 0.97, indicating that the items in the questionnaire were congruent with the research objectives and could be used.

Analysis results

1. The contrastive analysis results revealed the structures of English consonant clusters which do not exist in Thai, as shown in Table 1.

Table 1 English consonant clusters which do not exist in Standard Thai

Initial consonant clusters		Final consonant clusters		
2 consonants	3 consonants	2 consonants	3 consonants	4 consonants
/pj-/	/spl-/	/-pt/	/-pts/	/-ksts/
/tw-/	/spr-/	/-kt/	/-tTs/	/-ksts/
/tj-/	/str-/	/-ks/	/-kts/	/-mpts/
/kj-/	/stj-/	/-st/	/-kst/	/-lfts/
/bl-/	/skl-/	/-sk/	/-kst/	/-ndts/
/br-/	/skr-/	/-lk/	/-fts/	/-rst/
/dr-/	/skw-/	/-lt/	/-fts/	
/dw-/		/-ld/	/-sps/	
/gl-/		/-lm/	/-skt/	
/gr-/		/-lp/	/-sks/	
/gw-/		/-lf/	/-mpt/	
/fl-/		/-ls/	/-mps/	
/fr-/		/-lv/	/-mft/	
/tr-/		/-lt/	/-ndz/	

/Σr-/		/-mp/	/-ndt/	
/sp-/		/-mf/	/-nst/	
/st-/		/-nt/	/-nkt/	
/sk-/		/-nd/	/-lpt/	
/sl-/		/-nz/	/-lps/	
/st-/		/-ft/	/-lts/	
/sn-/		/-nk/	/-lks/	
/sw-/		/-ps/	/-lkt/	
/sf-/		/-bz/	/-ldz/	
		/-ts/	/-lft/	
		/-nz/	/-lft/	
		/-dz/	/-lfs/	
		/-ks/	/-lvz/	
		/-gz/	/-lmd/	
		/-lz/	/-lmz/	
		/-md/		
		/-pt/		
		/-bd/		
		/-kt/		
		/-gd/		
		/-st/		
		/-vd/		
		/-zd/		
		/-pt/		
		/-nt/		
		/-dt/		
		/-nt/		
		/-ft/		
		/-tt/		

2. The results of the efficiency of the speech recognition technology-based practice module

The following table presents the efficiency of the practice module that was used for the experimental group in this study.

Table 2 The results of the efficiency of the speech recognition technology-based practice module (N = 29)

The formative scores				Total	Percentage	Post-test score	percentage
Chapter 1 (40 pts)	Chapter 2 (40 pts)	Chapter 3 (40 pts)	Chapter 4 (40 pts)	120 points			
923	924	931	906	3698	2302.50	650	2166.67
31.83	31.86	32.10	31.24	127.03	79.40	22.41	74.71

From the above table, the results revealed that the total posttest scores of the four lessons of the experimental group were 3,684 from the full scores of 4,640 or 79.40%. The total posttest scores were 650 from the full scores of 870 or 74.71%. Therefore it was higher than the predetermined criteria of 70/70.

3. Comparison of the pronunciation achievements between the experimental group and the controlled group

The table below illustrates the mean, standard deviation and the t-test of the scores on consonant cluster pronunciation skills of the experimental group who used the speech recognition technology-based practice module and the controlled group who studied via the traditional method.

Table 3 Comparison of the posttest results between the experimental group and the controlled group

Posttest scores	N	M	S	t	Sig.
Controlled group	27	17.15	.243	-13.645	.000
Experimental group	29	22.41	.296		

It was revealed from the above table that the posttest mean scores of the experimental group were 22.41 and those of the controlled group were 17.15. When the scores of the two groups are compared, it is found that those of the experimental group were significantly higher than those of the controlled group at the .01 level.

4. Results from exploring the satisfaction of the experimental group with the use of the practice module

After having implemented the practice module to the 29 students in the experimental group, their satisfaction with the use of the module was assessed through the questionnaire. The results of the assessment are shown in Table 3.

Table 4 Results of the satisfaction with the practice module of the experimental group

Description	Quality level		
	M	S	Level
1. The use of the practice module is useful in improving my pronunciation.	4.52	0.51	highest
2. My pronunciation has improved after using the practice module.	4.59	0.50	highest
3. I have adequate and relevant knowledge to use the practice module to improve my pronunciation.	4.59	0.50	highest
4. Instructions in this subject enable me to acquire the correct principles of English pronunciation.	4.72	0.46	highest
5. I am able to remember and acquire more English consonant clusters after using the practice module to improve my pronunciation.	4.45	0.51	high
6. I feel more confident in my English pronunciation after using the practice module to improve my pronunciation.	4.55	0.51	highest
7. I feel that using the practice module in class	4.41	0.50	high

motivates me to participate in classroom activities more.			
8. I am able to apply the instructional processes using the speech recognition technology to other subjects.	4.34	0.48	high
Mean	4.52	0.50	highest
Total	36.17	3.963	

From the above table, it is concluded that the overall satisfaction with the implementation of the speech recognition technology-based practice module of the experimental group was at the highest level with the mean of 4.52 and the standard deviation of 0.50

Conclusion and discussion

The efficiency of the practice module based on the speech recognition technology to improve the pronunciation of English consonant clusters is 79.40/74.71, which is higher than the predetermined criteria. The learning achievements after practicing via the module of the experimental group were higher than those of the controlled group. Their satisfaction with the implementation of the practice module was at the highest level with the mean of 4.52 and the standard deviation of 0.50. The findings indicate that the module can be used to improve the pronunciation of English consonant clusters efficiently. This is because the construction of the module is based on good and practical module construction principles and concepts, enabling the design to have meaningful and easy-to-understand icons and pictures that motivate the students to learn. This is in line with Krutprong (2021) and Yodinphrom (2006), stating that the construction of a learning module must be in correspondence with the objectives, ages and interests of learners. Pictures and illustrations should be colorful and attractive to make them interesting and eye-friendly to learners.

Additionally, the construction is based on phonetic principles with a contrastive analysis being performed on the syllable structures and consonant clusters of English and Standard Thai. The result of the analysis yielded four lessons, comprising two initial consonant clusters, three initial consonant clusters, two final consonant clusters, and three to four final consonant clusters respectively. For those with pronunciation difficulties, they

were provided explanations about correct pronunciation processes showing points of articulation for each sound in an attempt to correctly practice according to the phonetic principles. Most practice activities focus on the sights, hearing, and practices of the students. This is consistent with Thuaycharoen (1990), who applied linguistic knowledge about pronunciation by explaining points and manners of articulation of consonants, vowels, and tones for better and clearer pronunciation skills of learners.

Moreover, integrating and applying the computer technology into the practice module plays an essential role in bringing speech sounds and moving pictures to improve pronunciation skills. Hearing English speech sounds from native speakers, seeing moving pictures and illustrations indicating points of articulation, and an instant feedback of pronunciation assessment had enabled the students to correctly imitate the sounds and reduce their shyness during pronunciation. The module also promotes the students to learn by themselves anytime and anywhere, facilitating them to practice without limits, not only in class but also in any situation, resulting in their pronunciation skills to become more effective and efficient. This is in line with Spaai & Herms (1993); Lambacher (1996); Eskenazi (1999); and Wennerstrom (2000), noting that the integration between speech sounds and moving pictures plays a significant role in acquiring pronunciation skills of learners, whether it is segmental or suprasegmental features. Furthermore, the speech recognition technology enables learners to receive instant feedbacks, promote actual practices, create new knowledge and systematize their learning habits, follow up their pronunciation progress, and raise their awareness on their own pronunciation problems which is regarded as an important problem-solving step due to actual practice experiences. McCrocklin (2016) stated that the speech recognition technology promote learner autonomy. In addition, Liakin, Cardoso & Liakina (2015) reiterated that the use of technology could significantly improve pronunciation skills of learners.

From these distinct features of the speech recognition technology-based practice module, the pronunciation achievements of the experimental group after practicing via the module were significantly different at the .01 level. It is thus concluded that the module is an efficient tool to facilitate the students to practice the pronunciation of English consonant clusters by themselves as often as they wished. It is deemed a very useful instrument in the teaching and learning context where the corona virus has been spreading. Learners are able to study the lessons independently, motivating them to concretely improve their

pronunciation skills on the English consonant clusters as well as reducing the face-to-face interactions between learners and instructors in a new normal learning environment. Furthermore, it is advantageous for teachers to change their role from knowledge providers to learning facilitators, ensuring the improvement of the pronunciation skills and encouraging learners to achieve their learning goals. It is also a constructive learner-centered approach that could bring about a sustainable learning practice in the new social and environmental context.

Recommendations

1. Recommendations for implementing the research results

1) The results could be used as a foundation for investigating linguistic systems at the lexical, syntactic and discourse levels, especially differences of syllable structures between English and other ethnic languages.

2) The results could be utilized to construct practice modules in other foreign languages with the incorporation of the speech recognition technology.

2. Recommendations for further studies

1) There should be experimental and comparative studies in other ethnic languages regarding their language systems and phonemes that are different from those of English in order to tackle English pronunciation problems of ethnic students.

2) There should be an investigation on foreign language practice modules with the incorporation to the speech recognition technology with contents being suitable for learners' levels and with a longer period of experimental time in order for learners to have ample time to practice, retain, and be skillful in an intended language skill.

References

- Arunsirot, S. (2021). The use of game application to enhance learning English morphology system of English as a foreign language learner. *FEU Academic Review*, 15(1), 47-63.
- _____. (2017). Developing practical guide in Standard Thai pronunciation of Lower Secondary School Kachin ethnic students at Ban Maena School, Tambon Muangna, Amphur Chiang Dao, Chiang Mai province. *FEU Academic Review*, 11(1), 89-98.

- Damnoensawat, S. (1993). *How to teach Thai with fun*. Bangkok: Chulalongkorn University Press.
- Department of Academic Affairs, Ministry of Education. (2011). *Educational quality assurance system in educational institutions: Operation framework and guidelines*. Bangkok: Ministry of Education.
- Eskenazi, M. (1999). Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype. *Language Learning & Technology*, 2(2), 62-76.
- Etirattana, S. (2010). *The development of a cluster reading practice module for Mathayom Suksa 3 students* [Unpublished M. Ed. Thesis]. Ubon Ratchathani Rajabhat University.
- Evers, K., & Chen, S. (2020). Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. *Computer Assisted Language Learning*, 35(8), 1869-1889.
- Gilakjani, A. & Ahmadi, A. (2011). A study of factors affecting EFL learners' English listening comprehension and the strategies for improvement. *Journal of Language Teaching and Research*, 2(5), 977-988.
- Harmer, J. (2001). *The practice of English language teaching* (3rd ed.). Harlow: Longman.
- Hincks, R. (2001). Using speech recognition to evaluate skills in spoken English. Papers from Fonetik 2001: 58-61. Lund University Department of Linguistics.
- Kim, I.-S. (2006). Automatic Speech Recognition: Reliability and Pedagogical Implications for Teaching Pronunciation. *Educational Technology & Society*, 9 (1), 322-334.
- Krutprong, A. (2021). Development of final consonant writing skills using learning management via multimedia computer lessons for Mathayom Suksa 2 Karen students at Ratchathani Wittayalai School, Sukhothai. *Journal of Faculty of Education Pibulsongkram Rajabhat University*, 8(2), 216-227.
- Lambacher, S. (1996). A CALL tool for improving second language acquisition of English consonants by Japanese learners. *Computer Assisted Language Learning*, 12(2), 137-156.
- Levis, J.; Suvorov, R. *Automatic speech recognition*. In: CHAPELLE, C. (Ed.). The encyclopedia of applied linguistics. [S.l.: s.n.], 2013.
- Liakin, D., Cardoso, W. & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *Computer Assisted Language Instruction Consortium*, 32(1). 1-25.

- Likert, R. (1967). *The method of constructing an attitude scale: Reading in attitude theory and measurement*. New York: Wiley & Son.
- Livescu, K. (2005). *Feature-based pronunciation modeling for automatic speech Recognition* [Unpublished doctoral dissertation]. Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science.
- Mahapaibul, N. (2004). *The development of pronunciation and writing abilities on words with the -n final consonant of Prathom Suksa 1 Karen students in Mae Sot district, Tak province by using the pronunciation and writing practice module* [Unpublished M. Ed. Thesis]. Nakhonratchasima Rajabhat University.
- Masjarat et al. (2007). *An educational innovation: Exercises and supplementary practices*. Bangkok: Century.
- McCrocklin, S. (2015). Automatic speech recognition: Making it work for your pronunciation class. *Proceedings of the 6th Pronunciation in Second Language Learning & Teaching*, 126-133, CA: Santa Barbara.
- _____. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System*, 57, 25-42.
- Nawayut, S. (2013). *Using phonetic symbols to teach English pronunciation*. Chiang Mai: Rajamangala University of Technology Lanna.
- Neri, A., Cucchiari, C. & Strik, W. (2003) *Automatic speech recognition for second language learning: how and why it actually works*, in *Proc. ICPhS*, 1157–1160.
- Numee, N. (1996). *The construction of a practice module of words with difficult spelling for a remedial session of Mathayom Suksa 2 students at Chuang Pran Withaya School in Ratchaburi province* [Unpublished M. Ed. Thesis]. Kasetsart University.
- Pokaisawan, C. (2008). *Effect of Computer-Assisted Language Learning materials for practicing final consonant clusters in English for enhancing the students' listening skills and pronunciation abilities*. Ratchaburi: Muban Chombueng Rajabhat University.
- Sanosamnieng, S. (1999). *The use of a supplementary practice module to solve mathematic problems for Mathayom Suksa 1 students* [Unpublished M. Ed. Thesis]. Chiang Mai University.
- Sidgi, L. F. S., & Shaari, A. J. (2017). The usefulness of automatic speech recognition (ASR) Eyespeak software in improving Iraqi EFL students' pronunciation. *Advances in Language and Literary Studies*, 8(1), 1-6.

- Spaai, G. & Herms, D. (1993). A visual display for the teaching of intonation. *CALICO Journal*, 10(3), 19-30.
- Sriwilai, J. (2007). *The development of the r-l-w cluster practice module with illustrations for Prathom Suksa 3 students* [Unpublished M. Ed. Thesis]. Uttaradit Rajabhat University.
- Thanyakitjanukit, S. (2010). *The effects of a linguistics-based Thai reading practice module on reading abilities of Prathom Suksa 2 students* [Unpublished M. Ed. Thesis]. Nakhon Sawan Rajabhat University.
- Thuaycharoen, P. (1990). *Phonetics and practical phonetics*. Bangkok: Thammasart University Press.
- Tu, H. & Chen, Y. (2011). English learning system of oral phonation based on ASR and smart phone platform. *Computer Applications and Software*, 28 (9), 64-66/147.
- Ur, P. (2000). *A course in language teaching practice and theory*. Cambridge: Cambridge University Press.
- Wennerstrom, A. (2000). Intonational meaning starting from talk. In *Spoken English, TESOL and Applied Linguistics*, Palgrave, 72-98.
- Wongphanit, P. (2005). *The development of a remedial linguistics-based Thai reading practice module for Prathom Suksa 2 students* [Unpublished M. Ed. Thesis]. Nakhon Sawan Rajabhat University.
- Yodinphrom, S. (2006). *The use of a writing practice module to promote creative writing of students with physical impairment* [Unpublished M. Ed. Thesis]. Chiang Mai University.