教育部教學實踐研究計畫成果報告

Project Report for MOE Teaching Practice Research Program (Cover Page)

計畫編號/Project Number：PED1080358

學門專案分類/Division：教育

執行期間/Funding Period：2019/8/1 至 2021/1/31

(計畫名稱/Title of the Project)：探討語音科技對大一英文口語訓練的成效：線上單詞及句子覆誦練習對學生英語發音及流利度的影響

Exploring the Effects of Incorporating Speech Technology into a Freshman English Class：The Impacts of Online Shadowing Exercise on Learners' English Pronunciation and Fluency

(配合課程名稱/Course Name)：英文(附二小時英聽)一 二

計畫主持人(Principal Investigator)：高照明

執行機構及系所(Institution/Department/Program)：

國立臺灣大學外國語文學系

成果報告公開日期：2021/3/18

□立即公開 □延後公開(統一於 2023 年 3 月 31 日公開)

繳交報告日期(Report Submission Date)：2021/3/18

Motivations of the Study and Research Purposes

English oral proficiency is one of the most important language skills required of college students. Unfortunately, results of well-established English proficiency tests such as TOEFL and IELTS have shown that the average oral proficiency level of Taiwanese test-takers is quite unsatisfactory. It has become a consensus among college English instructors that more oral training is required for freshman English classes. However, due to the constraint of time and class size, traditional classroom-based teaching is unable to provide students with sufficient training for English pronunciation and oral fluency.

To bridge this gap, we have developed a web-based system for students taking Freshman English to practice pronunciation at word and sentence level. The system aims to improve learners' English pronunciation and fluency via shadowing without human intervention. Thirty subjects taking Freshman English from September 2019 to June 2020 at National Taiwan University participated in this study. Learners were requested to listen to the audio files produced by native speakers of English for words and sentences, imitate the pronunciation as closely as possible, and they repeat exactly what they had heard. When learners practiced, information about the words or sentences was shown on the webpage. We conducted experiments to investigate if learners were able to correct their pronunciation errors on their own simply by shadowing and if shadowing supplemented with automatic feedback enabled by Google speech-to-text API could further reduce pronunciation errors. Results from our experiments suggested that shadowing could indeed reduce learners' pronunciation errors. This was also supported by our survey. 78.3% of the respondents agreed that shadowing could improve pronunciation. However, the effects of automatic feedback enabled by Google speech-to-text API were not noticeable in our experiment. The subjects' responses to our survey question also supported this conclusion, as only 47.9% of the respondents agreed that automatic feedback was helpful to learning. In terms of fluency, only 45.7% of the respondents agreed that shadowing could improve their oral fluency. The data suggested while shadowing was effective to improving pronunciation, the effects on fluency and the effects of automatic feedback on pronunciation were unclear.

Literature Review

Oral fluency is the one of the ultimate goals of foreign language education. Byrne (1986, p. 9) points out that "the main goal in teaching the productive skill of speaking will be oral fluency. This can be defined as the ability to express oneself intelligibly … reasonably accurately and without too much hesitation." To attain this goal, Byrne (1986, p. 10) suggests that we have to "bring the students from the stage

of where they are mainly imitating a model of some kind, or responding to cues, to the point where they can use the language freely to express their own ideas." As oral fluency requires a lot of practice and training, Byrne proposes a realistic three-stage training, starting with imitating.

The term "imitating" is reminiscent of drills of the audio-lingual approach many decades ago, which emphasizes intensive exercises involving listening to sentences, memorizing, and repeating. The audio-lingual approach was criticized for lack of communicative functions and gave way to the communicative approach emphasizing practice involving speech functions such as asking directions, making requests, and other activities such as role play. Compared with the audio-lingual approach, the communicative approach places more emphasis on fluency than on grammatical accuracy. Hammerly (1991, p. 9) criticizes the communicative approach by pointing out that it ignores language structures and that "(its) advocates do not seem to care that students mispronounce sounds, use wrong stems or endings, or construct sentences following faulty rules." Summarizing the features of different approaches, Bygate (2010, p. 71) comments that "Audiolingual approaches aimed to develop speaking only in terms of pronunciation and fluent, accurate manipulation of grammar. Situational approaches introduced dialogue patterns into the range of features to be taught, and functional approaches added speech acts into the syllabus."

Repeating and imitating are required in communication. For example, repeating often occurs in a discourse by speakers with different intonations to express 'approval', 'emphasis', 'question', etc. One frequently used technique in oral training is shadowing, in which the learners repeat exactly what they have heard. The learners pay attention to the details in pronunciation such as vowels, consonants, stress, intonation, liaison, accent and imitate as closely as possible the way a native speaker speaks. Shadowing has long been used to train interpreters (cf. Kurz, 1992) and has been used to evaluate a learner's proficiency level under the name of "elicited imitation" since 1960s (cf. Murphey, 2001). While shadowing has long fallen out of favor by the ESL mainstream, its effectiveness on improving listening comprehension, oral fluency, and interpreting has been affirmed by several scholars.

For example, according to Murphey (1993), students interviewed in his study report that "conversational shadowing allowed them to assert some control over the process and content of conversations and to build better rapport through reflective listening." Murphey (1993) concludes that that "silent shadowing had a major impact on their learning, increasing attention and retention of material in short-term memory". Celce-Murcia et al. (1996) suggest a method of learning intonations by shadowing, mirroring, imitating native speakers' utterances. Yamada (2010) conducts experiments on the effect of shadowing on listening and concludes that shadowing is effective to

listening comprehension. Hsieh, Dong, and Wang (2013) confirm in their experiment that the technique of shadowing is useful to the learning of intonation, fluency, word pronunciation and overall pronunciation. The study by Foote and McDonough (2017) also indicates that shadowing exercise can improve comprehension and fluency.

As previous researches have revealed that shadowing is useful to fluency and pronunciation in a foreign language, we want to explore if a web-based system that supports online shadowing exercise is effective to reducing pronunciation errors in the ESL context. If shadowing is effective, it would suggest that learners can improve their pronunciation without any intervention or guidance from an instructor. Would an intelligent system capable of generating automatic feedback to learners be more effective to improving learners' pronunciation? In order to investigate these questions, we have designed a web-based shadowing system and its variant supplemented with automatic feedback,

The technology most relevant to a system with automatic feedback on learners' pronunciation is automatic speech recognition (ASR). Several papers discuss the applications of ASR in foreign language learning and assessment. For example, Coniam (1998) explores the potential use of ASR as an assessment tool in English in the ESL context. Dalby and Kewley-Port (1999) investigates the use of ASR technology in foreign language learning, including speakers of American English learning Spanish and speakers of Mandarin Chinese learning English. Their research suggests that ASR-derived feedback can improve pronunciation. Luo et al. (2010) explore the relationship of acoustic features such as phoneme intelligibility and prosodic fluency with the reference scores for learners' shadowing. Ginther et al. (2010) discuss the conceptual and empirical relationships between temporal measures of fluency and oral English proficiency. Franco et al. (2010) describe a speech recognition and pronunciation scoring toolkit for computer-aided language learning applications. Berstein et al. (2010) validate automated speaking tests. Chen (2011) develops a speaking website using Microsoft ASR toolkit. In Chen's study, students and in-service teachers using the system agree that ASR can offer different types of exercises which encourage students to produce more oral output. All these researches suggest that automatic speech recognition technology has become increasingly important and popular in foreign language learning.

An intelligent CALL system based on shadowing exercises and ASR is also congruent with Swan's Comprehensible Output Hypothesis. Swain (1985) proposes that language learning takes place not only because of comprehensible input as proposed by Krashen (1982) but also because of comprehensible output. Learners can improve their grammatical competence when they are pushed to produce and modify their output. The task of shadowing might have the potential advantage of making the

language output more comprehensible and appropriate. Swain (1995) presents the three functions of output in second language learning, namely, the noticing effect, the hypothesis/testing function, and the reflective function. In discussing the 'noticing' function, Swain and Lapkin (1995), p. 372) write that 'a learner will, on occasion become aware of (i.e. notice) a linguistic problem (brought to his/her attention either by external feedback (e.g. clarification requests) or internal feedback'. In other words, for language learners to acquire new language forms in L2, they need to encounter a linguistic problem first. Only when they notice a problem can they endeavor to overcome it. The noticing effect is obvious in the shadowing and interpreting tasks. Once learners cannot produce the oral output correctly or fail to pass the threshold required by our proposed system, they know their problems and need to try until they can overcome them.

Research Questions

In this study, we aim to investigate the effects of our proposed web-based shadowing system (including its AI-empowered version with automatic feedback), as well as learners' attitudes towards shadowing exercise of words and sentences. We hope we can find emprical evidence showing that our proposed web-based shadowing systems can indeed improve learners' pronunciation and oral fluency.

As our system supports the function of automatic feedback, we are interested in exploring the role of speech recognition and its effects on correction of pronunciation errors. It is widely known that most ASR systems have poorer recognition rates for certain consonants and vowels. Previous studies by Coniam (1998) and Chapelle and Chung (2010) conclude that speech recognition (ASR) systems are not reliable enough to be used in high-stake assessment. Chen (2011) also notes that it is challenging for users to speak a sentence that can be successfully recognized by an ASR system. We want to explore if the existing Google English speech recognition API, which makes automatic feedback possible in our system, is reliable enough to be used in learners' error detection tasks.

In summary, the research questions in this study are the following.

(1) Can the proposed web-based shadowing system improve learners' pronunciation of consonants, vowels, and stress?

(2) Can the proposed web-based shadowing system improve learners' fluency?

(3) Is automatic feedback provided by the proposed system helpful to learning pronunciation?

(4) What are learners' attitudes toward the web-based shadowing exercise of repeating a spoken sentence with script?

(5) What are learners' attitudes toward the web-based exercise of repeating a spoken word with script?
(6) What are learners' perceptions about the pronunciation of the words and sentences in the system?
(7) What are learners' perceptions about the performance of automatic feedback in our system?

Research Design and Methodology

A web-based shadowing system with automatic feedback was been developed for this research. The system was designed to empower users to learn speaking at their own pace. Exercises with different levels of difficulties were provided to ensure that users can learn the materials efficiently and effectively. Figure 1 is the user interface of our web-based shadowing system.

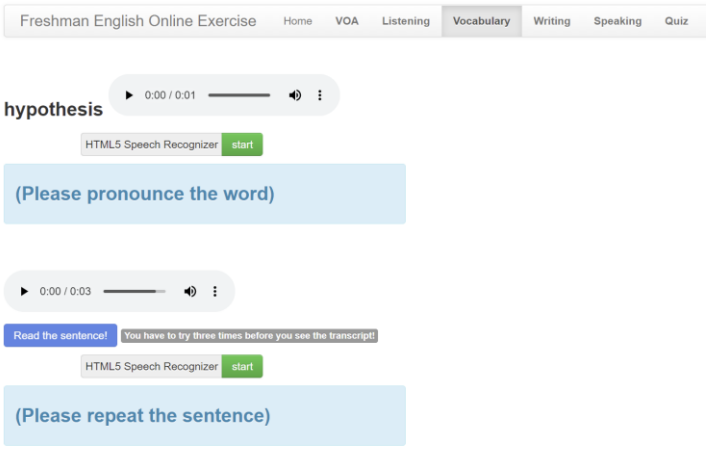

Figure 1. The user interface of our web-based shadowing system

Following Byrne (1986), we combined different exercises in different stages of oral training. The first stage involved the shadowing of a word. The second stage involved the shadowing of a sentence. The third stage contained conversational shadowing. In this study, we focused on the first two stages oral training. Learners started with the web-based shadowing exercise of words. After completing the web-based shadowing exercises of words, learners proceeded to the shadowing exercises of sentences. If they had difficulties understanding the meanings of the example sentences, they could read the English script, which is intended to reduce the cognitive load during the exercise and helped learners better focus on the identification and correction of pronunciation errors.

To keep track of the changes that occurred during and after the oral training, the subjects' pronunciations before and during the shadowing exercises were recorded.

When they did the web-based shadowing exercise, they needed to perceive the differences between the input sounds and their speech output and then imitate the input sounds as closely as possible. They were requested to repeat each word four times.

There are two types of shadowing exercise in our system, the traditional shadowing exercise with a web interface and the AI-empowered version with automatic feedback. The latter type of exercise employed speech recognition techniques to measure the similarity between the input word and learners' output speech. For instance, the pronunciation of 'odd' and 'at' were represented as /ɑːd/ and /æt/ in the Longman Dictionary of Contemporary, respectively, using International Phonetic Alphabets (IPA). With IPA, the similarity between the two words can be simplified as the similarity between two strings of phonetic symbols, which can be computed by algorithms such as the Minimal Edit Distance (MED). MED calculates the minimal steps of editing (i.e. insertion, deletion, and substitution) if the two strings are to be identical.

Google speech-to-text API and the Minimal Edit Distance algorithm were two key technologies that made instant feedback in our system possible. Google speech-to-text API converted users' speech input into words. If the similarity between a learner's speech (converted into words by Google speech-to-text API) and the word or sentence input by the instructor was below a given threshold, our system automatically generated feedback such as "Try again!" to request the learner to do the exercise again. Likewise, if the similarity is much higher or slightly higher than the threshold, the system generated feedback such as "Excellent" or "Not bad". In order to evaluate the performance of our system, we recorded learners' speech and the automatic feedback generated by the system.

We formulated two hypotheses. The first hypothesis was that shadowing could improve learners' pronunciation accuracy. The second hypothesis was that shadowing with automatic feedback enabled by the Google speech-to-text API could help learners correct their pronunciation errors.

To test these two hypotheses, we conducted two experiments. Thirty students taking Freshman English participated in this research. In the first experiment, a half of the class (15 students) were requested to pronounce 10 words, each for 5 times in sequence. In the first exercise, the subjects read the word or sentence without listening to any audio file. This was recorded as the original pronunciation before treatment. The subjects were asked to repeat the sounds they heard four times. The speech data served as the data after each treatment (in our case, shadowing exercise). The ten words included 'collapse', 'colleague', 'comment', 'commit', 'communicate', 'community', 'compatible', 'compensate', 'complex', and 'comprehensive'. These ten

words were chosen from the Academic Word List.

In the second experiment, the other half of the class (15 students) were requested to pronounce the same set of words, each for 5 times in sequence. In the first and second exercise, the subjects simply read the words without listening to audio files. In the third, fourth, and fifth trials, the subjects did the shadowing exercise by listening to the audio files using the automatic feedback function enabled by the Google speech-to-text API.

The 30 subjects' recorded speech data was transcribed by two annotators, who had passed the advanced level of General English Proficiency Test (GEPT) and had very good command of spoken English. The two annotators had received intensive training in English phonetics and phonetic transcriptions of IPA before they started to transcribe learners' speech and annotate errors. A program was written to check if there were any disagreements among the transcriptions annotated by the two annotators. All the disagreements in transcriptions and error tags were resolved via consensus reached in discussions.

Since the length of the ten words that subjects were required to pronounce was not the same, all the subjects' pronunciation errors were normalized over the number of phonemes in each word, giving us a value between 0 and 1, which served as an indication of the extent the subjects' pronunciations diverged from the "correct" one (0 being error-free and a complete match with the IPA of the word and 1 being completely different). The errors were obtained by use of edit distance, through which we compared the subject's pronunciation and the General American IPA listed in the Longman dictionary. The subjects' speech data were transcribed into IPA, and for every difference between the subjects' speech and the General American IPA, the algorithm classified them as either an insertion, a deletion, or a substitution, and counted them each as a one-step divergence.

Results and Discussions

Figure 2 showed that all the ten words included in the shadowing exercise of the first experiment exhibited roughly the same pattern. As the subjects did more shadowing exercise, normalized errors decreased and the subjects' pronunciation gradually approximated the correct pronunciations.

A few observations could be made based on Figure 2. While there was a general tendency that the subjects' errors decreased as they did more shadowing exercise, some words seemed easier to correct than others in pronunciation. For instance, the word 'comment' had a rather steep descent in its normalized errors after the first shadowing exercise, suggesting that the subjects were able to rectify some mistakes after they listened to the correct pronunciation of the word.

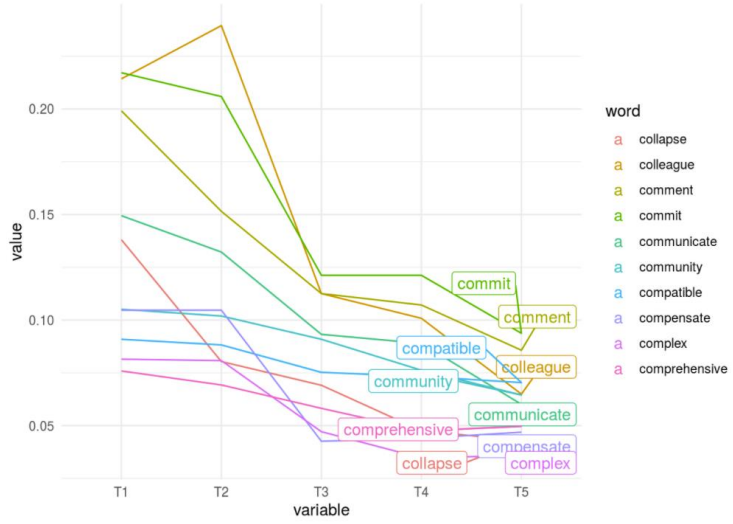

Figure 2. Improvement of the subjects' pronunciation of ten words during the shadowing exercise

Figure 3 revealed the results of the second experiment of shadowing, in which automatic feedback enabled by speech recognition was provided. Like the first experiment, we could see that the normalized errors decreased as the subjects did more shadowing exercise. However, the positive effects of automatic feedback enabled by Google speech-to-text API were unnoticeable. Comparisons between Figure 2 and Figure 3 suggested that the web-based shadowing system with automatic feedback did not reduce learners' pronunciation errors as we had expected. Unlike Figure 2, Figure 3 showed less consistent patterns. The occasional increase in errors from T1 to T2 might be attributed to the absence of audio files to imitate. The sudden rise of errors in T3, however, was due to different reasons. It might arise from mistakes in speech recognition and automatic feedback. This was probably the reason why T4, and to a lesser extent T3, showed tendencies for the students to relapse into prior errors, although such relapses were often self-corrected by the time of T5. These relapses appeared to be word-specific as well.

In short, contrary to our expectation, Figure 3 seemed to show that computer-generated feedback might confuse learners due to the program's inability to reliably identify pronunciation errors. The results of the first and second experiments thus indicated that shadowing without automatic feedback would appear more effective.

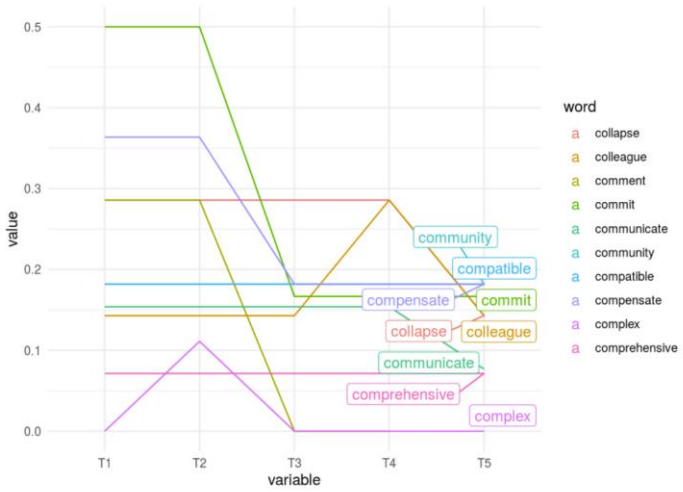

Figure 3. Improvement of the subjects' pronunciation of ten words during the shadowing exercise with automatic feedback

Our analyses were further borne out by data collected in the survey (cf. (1)). More than three fourths of the subjects' responses were positive about the effects of shadowing on improving pronunciation (cf. Question 1). However, regarding the subjects' perceived effects of shadowing on oral fluency (Question 2), only 45.7% of the respondents agreed and 39.1% were unsure. In terms of the subjects' perceived benefit of shadowing with automatic feedback, many subjects (41.3%) were unsure and only 47.9% of the respondents agreed (Question 3). Over 50% of the respondents liked shadowing at the word and sentence levels (Question 4 and 5). With respect to the subjects' attitudes toward the quality of the audio files in the system, over 85% of the subjects' responses were positive (Question 6 and 7). Responses to the two open-ended questions in the survey (cf. Appendix) indicated that while most subjects were positive about the effects of the web-based shadowing system, quite many of them were skeptical about the effects or benefit of automatic feedback enabled by Google speech-to-text API. The data collected in the two experiments and in the survey also support our hypothesis that the web-based shadowing system can help learners reduce pronunciation errors. However, as speech recognition technology was not reliable enough for nonnative speakers of English, automatic feedback enabled by Google's speech-to-text API was sometimes rather misleading and confusing to learners. Its effects on improving learners' pronunciation and fluency remained dubious.

The reason why we did not find effects of shadowing on fluency was presumably because very limited time was dedicated to the training of shadowing on the sentence level. Unlike the correction of pronunciation errors, oral fluency requires much longer time to develop. The statistics of the survey questions was in (1).

(1).

1. Repeating the pronunciation of an English vocabulary word after a native speaker of English can help me improve my pronunciation of consonants, vowels, and stress.
78.3% of the respondents agreed. 13% didn't agree. 8.7% were not sure.

2. The system can help me improve my oral fluency in English.
45.7% of the respondents agreed. 15.2% disagreed. 39.1% were unsure.

3. The automatic feedback provided by the system is helpful to learning pronunciation.
47.9% of the respondents agreed. 10.8% disagreed. 41.3% were unsure.

4. I like the online exercise of repeating a spoken sentence with script.
56.3% of the respondents agreed. 23.9% disagreed. 19.6% were unsure.

5. I like the online exercise of repeating the pronunciation of a vocabulary word with script. 60.8% of the respondents agreed. 15.3% disagreed. 23.9% were unsure.

6. The pronunciations of the words in the system are good.
87% of the respondents agreed. 2.1% disagreed. 10.9 % were unsure.

7. Overall, the performance of automatic feedback is acceptable.
86.9% of the respondents agreed. 13.1% disagreed. 10 % were unsure.

Conclusions and Suggestions for Future Research

We developed a web-based shadowing system with automatic feedback. The learning effects of shadowing on learners' pronunciation in the ESL context have been reaffirmed. However, due to the limited time allocated to the shadowing exercise of sentences in our experiment, we were unable to find noticeable improvement in learners' oral fluency. Moreover, the learning effects of automatic feedback empowered by the Google speech recognition engine in our proposed system have not been observed, which might arise from the lower accuracy rate of speech recognition engines for detecting mispronunciation. Our research suggests that a more appropriate way to incorporate speech technology into oral training is to understand the gist of a learner's input and respond to it naturally. A web-based conversation shadowing system that provides human-like interactions with learners is more likely to make oral training more engaging, interesting, and effective.

References

Bernstein, Jared, Van Moere, Alistair, and Cheng, Jian. (2010). Validating automated speaking tests. Language Testing, Vol. 27, No. 3, pp. 355-377.

Bygate, Martin. (2010). Speaking. In Kaplan, Robert. (ed). The Oxford Handbook of Applied Linguistics, p. 63. -73.

Celce-Murcia, Marianne, Donna Brinton and Janet M. Goodwin. (1996). Teaching pronunciation: A reference for teachers of English to speakers of other languages. Cambridge: Cambridge University Press.

Chapelle Carol, Chung, Yoo-Ree. (2010). The promise of NLP and speech processing

technologies in language assessment, Language Testing, Vol. 27, No. 3, pp. 301-315.

Chen, Hao-Jan. (2011). Developing and Evaluating an Oral Skills Training Website Supported by Automatic Speech Recognition Technology. RECALL, Vol. 23 , No 1, pp. 59-78.

Chung, Da-Un. (2010). The Effect of Shadowing on English Listening and Speaking Abilities of Korean Middle School Students. English Teaching. CALICO, Vol. 65, No. 3, p. 97.

Coniam, D. (1998). The Use of Speech Recognition Software as an English Language Oral Assessment Instrument: An Exploratory Study. CALICO, Vol. 15, No. 4, pp. 7.-23.

Cook, Guy. (2010). Translation in Language Teaching. Oxford: Oxford University Press. References.

Bernstein, Jared, Van Moere, Alistair, and Cheng, Jian. (2010). Validating automated speaking tests. Language Testing, Vol. 27, No. 3, pp. 355-377.

Bygate, Martin. (2010). Speaking. In Kaplan, Robert. (ed). The Oxford Handbook of Applied Linguistics, p. 63. -73.

Celce-Murcia, Marianne, Donna Brinton and Janet M. Goodwin. (1996). Teaching pronunciation: A reference for teachers of English to speakers of other languages. Cambridge: Cambridge University Press.

Chapelle Carol, Chung, Yoo-Ree. (2010). The promise of NLP and speech processing

technologies in language assessment, Language Testing, Vol. 27, No. 3, pp. 301-315.

Chen, Hao-Jan. (2011). Developing and Evaluating an Oral Skills Training Website Supported by Automatic Speech Recognition Technology. RECALL, Vol. 23 , No 1, pp. 59-78.

Chung, Da-Un. (2010). The Effect of Shadowing on English Listening and Speaking Abilities of Korean Middle School Students. English Teaching. CALICO, Vol. 65, No. 3, p. 97.

Coniam, D. (1998). The Use of Speech Recognition Software as an English Language Oral Assessment Instrument: An Exploratory Study. CALICO, Vol. 15, No. 4, pp. 7.-23.

Dalby, Johnathan and Kewley-Port, Diane. (1999). Explicit Pronunciation Training Using Automatic Speech Recognition Technology. CALICO, Vol. 16, No. 3, pp. 425.-445.

English Learning Programs at the VOA Chinese Website. http://www.voachinese.com/section/english-teaching/2506.html

Foote, J, McDonough, K (2017) Using shadowing with mobile technology to improve L2 pronunciation. Journal of Second Language Pronunciation 3(1): 33–56.

Franco, Horacio, Bratt, Harry, Rossier, Romain, Gadde, Venkata Rao, Shriberg, Elizabeth, Abrash, Victor, and Precoda, Kristin. (2010) EduSpeak®: A speech recognition and pronunciation scoring toolkit for computer-aided language learning applications Language Testing, pp. 401-418.

Ginther, April, Dimova, Slobodanka, and Yang, Rui. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring, Vol. 65, No. 3, pp. 379-399.

Hamada, Yo. (2012). “An Effective Way to Improve Listening through Shadowing.” The Language Teacher, Vol. 36, No. 1. pp. 3-10.

Hughes, Rebecca. (2002). Teaching and Researching Speaking. Routledge.

Hsieh, Kung-Ting, Dong, Da-Hui, and Wang, Li-Yi. (2013). A Preliminary Study of Applying Shadowing Technique to English Intonation Instruction. Taiwan Journal of Linguistics, Vol. 11 , No. 2, pp. 43-66.

Luo, Dean, Yamauchi, Yutaka and Minematsu, Nobuaki. (2010). Speech Analysis for Automatic Evaluation of Shadowing. In Proceedings of SLaTE.

Lynch, Tony. Listening: Sources, Skills, and Strategies. In Kaplan, Robert. (ed). The Oxford Handbook of Applied Linguistics, p. 74. -87.

Murphey, T. (2001). Exploring conversational shadowing. *Language Teaching Research*, *5*(2), 128-155.

Shiki, O., Mori., Y., Kadota, S., & Yoshida, S. (2010). Exploring differences between shadowing and repeating practices. *Annual Review of English Language Education in Japan, 21*, 81-90.

Dalby, Johnathan and Kewley-Port, Diane. (1999). Explicit Pronunciation Training Using Automatic Speech Recognition Technology. CALICO, Vol. 16, No. 3, pp. 425.-445.

English Learning Programs at the VOA Chinese Website. http://www.voachinese.com/section/english-teaching/2506.html

Franco, Horacio, Bratt, Harry, Rossier, Romain, Gadde, Venkata Rao, Shriberg, Elizabeth, Abrash, Victor, and Precoda, Kristin. (2010) EduSpeak® : A speech recognition and pronunciation scoring toolkit for computer-aided language learning applications Language Testing, pp. 401-418.

Ginther, April, Dimova, Slobodanka, and Yang, Rui. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring, Vol. 65, No. 3, pp. 379-399.

Hamada, Yo. (2012). "An Effective Way to Improve Listening through Shadowing." The Language Teacher, Vol. 36, No. 1. pp. 3-10.

Hughes, Rebecca. (2002). Teaching and Researching Speaking. Routledge.

Hsieh, Kung-Ting, Dong, Da-Hui, and Wang, Li-Yi. (2013). A Preliminary Study of Applying Shadowing Technique to English Intonation Instruction. Taiwan Journal of Linguistics, Vol. 11 , No. 2, pp. 43-66.

Luo, Dean, Yamauchi, Yutaka and Minematsu, Nobuaki. (2010). Speech Analysis for Automatic Evaluation of Shadowing. In Proceedings of SLaTE.

Lynch, Tony. Listening: Sources, Skills, and Strategies. In Kaplan, Robert. (ed). The Oxford Handbook of Applied Linguistics, p. 74. -87.

Murphey, T. (2001). Exploring conversational shadowing. *Language Teaching Research*, *5*(2), 128-155.

Shiki, O., Mori., Y., Kadota, S., & Yoshida, S. (2010). Exploring differences between shadowing and repeating practices. *Annual Review of English Language Education in Japan, 21, p.* 81-90.

Appendix

The Subjects' Responses to the Two Open-ended Survey Questions

1. What are the advantages of the speech recognition system used in this class for language learning?

(1) I pronounced some of the words incorrectly. Now I know the correct pronunciation.
(2) efficient

(3) Well, you don't need a teacher to make your pronunciation better so you can try many times and not feel shy because you are talking to a machine, not to a human
(4) It can help me know the problems of my pronunciation.
(5) Quick and painless ⇔ in ideal condition ≙ it works properly in a quiet environment
(6) It improves my listening skill.
(7) I think that the only advantage is that you have to repeat so many times the sentences to do it right, that you improve your pronunciation
(8) learn more effectively.
(9) Teachers don’t have to take a lot of time to teach students how to pronounce words.
(10) I can improve my speaking skills.
(11) convenient
(12) People with quite weak English could actually improve a lot and quickly.
(13) The system not only helps us get better with our pronunciation, but also we can figure out our mistakes when repeating the word.
(14) It can help us in producing the right pronunciations by the features that can repeating and evaluating our pronunciation.
(15) Improve my pronunciation of vocabulary words and speaking skills.
(16) help improving my pronunciation.
(17) We can test our oral skills by ourselves, then the teacher can have more time to prepare for the class.
(18) The advantages are: helps us improve our speaking and is quite effective as well.
(19) effectively provide feedback for speaking quality
(20) it help me improve my pronunciation
(21) help teachers to teach students to practice English pronunciation more effectively
(22) It seldom happens that it can't recognize our pronunciation.
(23) Convenient, flexible, quite complete
(24) It's good for improving the fluency and pronunciation of English.
(25) I think the advantage is that it allows me to improve both my listening and speaking skills. It is the best way to practice these two skills.
(26) This does improve my oral fluency. It's quite hard to pass the test, and I have to speak correctly to pass it.
(27) We can use technology while learning, making learning more fun.
improve my pronunciation.

(28) Give student instant feedback.
(29) The system is good. I like these types of exercises.
(30) I have learned different methods of learning English. The approach is more suitable for people who have gradated and now have a job.
(31) To force learners to repeat a sentence and imitate the sound.
(32) It helps me correct my mispronunciation.

2. What are the advantages of the speech recognition system used in this class for language learning?

(1) Everyone has a different accent. This is difficult for speech recognition.
(2) The way it grade my pronunciation is weird.
(3) It requires some appliance support.
(4) sometimes the system can't recognize the sound.
(5) not effective
I think it should record your voice and then let you listen it so that you can see your mistakes.
(6) The noise might affect the result a lot. Since the noise is loud, sometimes the human sound will be covered by the noise and the student has to speak much louder.
(7) Sometimes there are technical difficulties, for example the recognition system doesn't work on sounds like 'v' and 'f's.
(8) It was too noisy in the class to use the system.
(9) The system did not work smoothly.
(10) I just think it is not good, that's all... There are different kinds of English, and people always develop their own styles of English... Is it that important for all of us to mimic one person style of English?
(11) I think you need to improve the system. The recognition of the system needs improvement.
(12) The recognition system is sometimes not really smart. Sometimes the feedback from the system is not accurate.
(13) People with higher level of English proficiency may find this system less useful due to the bugs in the system.
(14) The speech recognition system still needs a lot of improvement on the sensitivity to some sounds.
(15) Sometimes the system recorded my words together with my classmates', which made the system less useful.
(16) It couldn't point out which part of our pronunciation was incorrect.

(17) The quality of microphone made it very difficult to reach the best level. Sometimes it doesn't recognize the words that we pronounce. I think it can be improved.
(18) Well.. it doesn't work that well realistically, especially when everyone uses it at the same time in the same area.
(19) When I don't get good grades on the test of repeating sentences, I don't know what I should practice more.
(20) Sometimes it can't recognize the voice.
(21) It lacks the ability to correctly judge our speech (word with short pronunciation) the results of judging is sort of ... weird?? It lacks real interactions.
(22) It does not recognize my voice even if I speak well the word or sentence
(23) Some words cannot be recognized by the computer.
(24) Sometimes, the system provides incorrect feedback.
(25) Names cannot be recognized easily.
(26) The system cannot improve fluency.
(27) It is hard to tell the differences of some sounds.

.