

A System for Computer Assisted Peer Assessment of EFL Pronunciation

Hengbin Yan and Jiexuan Lin

Faculty of English Language and Culture
Guangdong University of Foreign Studies
Guangzhou, Guangdong Province, China
{yhb, lin_jx28}@gdufs.edu.cn

Abstract

Pronunciation is an important yet often neglected subfield in Second Language Acquisition, both in pedagogy and research (Derwing 2012). One significant, under-researched area is the role peer assessment/review can play in shaping English-as-a-Foreign-Language (EFL) students' pronunciation proficiency. Whereas there have been many studies demonstrating the effectiveness and benefits of peer review on EFL writing/oral proficiency etc. (see Mendonça et al., 1994), few studies exist that test the efficacy of similar approaches as applied to pronunciation learning/training tasks.

To investigate the viability of computer assisted peer assessment of EFL pronunciation, we present in this study a prototypical web-based/mobile platform for peer review of EFL pronunciation with adaptively generated items for both training and testing purposes. We discuss some of the prominent features of the platform as well as the results from our preliminary studies involving more than 300 EFL students who used the platform for pronunciation training and peer review.

1 Introduction

Pronunciation is an important yet often neglected subfield in Second Language Acquisition, both in pedagogy and research (Derwing et al., 2012). One significant, under-researched area is the role that adaptive training and peer assessment/review can play in shaping English-as-a-Foreign-Language (EFL) students' pronunciation proficiency. Whereas

there have been many studies demonstrating the effectiveness and benefits of adaptive training and peer review on EFL writing/oral proficiency etc. (see Mendonça et al., 1994), few studies exist that test the efficacy of similar approaches as applied to pronunciation learning/training tasks.

Faced with the inherent difficulties and fossilization effects in pronunciation learning, for which traditional classroom instruction is unable to provide students with sufficient intervention/feedback to effect improvements, peer assessment, enhanced by adaptive training mechanisms, may prove a viable complement to teacher-centered instruction and Computer Aided Pronunciation Training (CAPT) whose accuracy and usefulness is currently still limited. Although some initial benefits of peer assessment in pronunciation has been demonstrated in recent studies (Luo, 2016), current review processes are limited in scope, depth and function and are exceedingly effort- and time- consuming, rendering it impractical for the majority of teaching curricula, and existing technologies for peer review in pronunciation prove inadequate for facilitating the review process.

To overcome some of the existing problems and further investigate the viability of computer assisted peer assessment of EFL pronunciation, we present in this study a prototypical web-based/mobile platform flexibly designed for a wide range of peer review tasks powered by adaptive item generation with functionalities ranging from online recording, audio submission, progress tracking, adaptive training, group assignment, peer annotation to automated visual feedback. We discuss some of the promi-

nent features of the platform as well as the results from our preliminary studies involving more than 300 EFL students who used the platform for pronunciation training and peer review.

2 Literature Review

2.1 Peer assessment

Peer assessment refers to the arrangement in which individuals evaluate the outcomes of learnings of peers of similar status. [Topping \(1998\)](#) argues that the immediacy, frequency and volume of peer assessment outweigh concerns over its quality sometimes not reaching the level of professional teaching staff. More importantly, by providing concrete examples for assessors to think about and discuss the quality of the peers as well as their own work, peer assessment can serve to consolidate and deepen the assessor's understanding of the learning task ([Van Lehn et al., 1995](#)).

2.1.1 PA in EFL

In EFL contexts, the formative effects of peer assessment have been recognized as a complement to summative teacher assessment ([Peng, 2010](#)). Despite concerns over the reliability and precision of peer assessment, it has been argued that the main goal and functions of peer assessment are not equal to teacher assessment, which is a form of summative assessment, while PA is a formative process with pedagogical benefits that can complement teacher assessment ([Peng, 2010](#)). It is noted that teacher assessment is not necessarily more accurate and less biased than peer assessment.

Although peer review is widely recognized for the constructive feedback given to the receiver of the feedback, as has been shown in various studies in which revisions resulting from peer feedback lead to improved performance in writing tasks, previous experiments have demonstrated the benefits of peer review for review givers, especially lower proficiency students, so much so that the benefits they receive are significantly greater than that of review receivers ([Lundstrom and Baker, 2009](#)).

In verbal communications, previous studies on peer assessment on oral presentation suggests that instruction on skill aspects may suffice for achieving a relatively high degree of inter-rater agreement.

Rater training, whether short or long, did not significantly improve inter-rater agreement. However, it is found to improve the frequency as well as relevancy of feedback comments ([Saito, 2008](#)).

2.1.2 Benefits

[Trofimovich et al. \(2016\)](#) finds that self-assessment of L2 speech by L2 learners is flawed (inaccurate) compared with assessment by native speaking listeners of the rated language, and that a Dunning Kruger effect, whereby lower proficiency self-raters tend to overestimate, while higher-proficiency ones underestimate, their own proficiency, is apparent. This effect is also present in other previous studies of self-assessment, and didn't easily fade away even after long periods of training and feedback ([Hacker et al., 2000](#)). The inaccuracies in self-assessment are found to be linked to factors such as listener rated measures of phonological accuracy, as well as temporal fluency.

One potential advantage of peer review is that it provides a venue to facilitate peer learning by providing samples that although not from native speakers, are intelligible and comprehensible and are relevant to their needs and can therefore serve as aspiration models ([Murphy, 2014](#)). In an increasingly globalized world where world Englishes are becoming more mainstream, the ability to recognize and understand common accent patterns is important for language learners.

2.2 Peer Assessment of EFL Pronunciation

Unlike peer assessment in other subfields of Second Language Acquisition, studies on peer assessment of EFL pronunciation, especially those utilizing computer technology are few and far between. [Luo \(2016\)](#) conducted a study in which students were asked to practice oral reading and then submit recordings to be reviewed by peer students through an online learning platform (Blackboard). The results of the study shows that students engaging peer review practice achieved greater gains than control groups in phoneme-, cluster- and sentence-level errors, although lower gains in word-level errors. She notes that after several weeks of training in giving peer feedback, student reviewers' reviewing ability substantially improved to the point of being able to independently point out problems

confirmed by teachers. In addition, two thirds of the students reported feeling that their feedback helped their classmates, while they didn't find feedback from classmates useful to them.

Compared with CAPT, Luo notes the peer review process could be easily applied to any arbitrary language in a cost-effective manner, and prove especially useful for languages whose resources are scarce. Although Luo's study on peer review in pronunciation has been demonstrated to be effective for EFL students, a few problems remain unsolved and questions unanswered.

While Luo's study makes use of an existing online learning system for the peer review process, a lot of the procedures are still largely manual and inefficient. For example, the students listen to, shadow read, practice, record their readings and then manually compare the recording with a standard model before uploading the recording to Blackboard for peer review. This is a time-consuming process that takes at least 5 days. Apart from the inefficiency, a lot of uncontrollable variables are left for the students to manage on their own, resulting in less optimal training. For example, it is not possible to track the way and number of times the students listen to and shadow read, and not every student has the capacity to make effective comparisons/perceive the subtle differences between their own reading and that of a standard model in order to correct and adjust their reading. Additionally, it's hard to monitor and track the progress of students as they practice and submit their assignments. The Blackboard platform is not specifically designed and therefore not particularly suited for pronunciation peer review. First, only one type of feedback - textual comment, is allowed. Other types of feedback, such as inline, categorized labels and speech feedback is not possible. Certain problematic areas are hard to pinpoint and demonstrate using texts, and recorded feedback can be a useful alternative. Luo admits that for students to upload aural comments to Blackboard would be too ineffective and therefore had to rely on text-only feedback. Second, assignment of groups is not flexible. Ideally, a peer review task should allow for anonymous, randomized group assignment with flexible, configurable conditions. One question is, can students of lower proficiency provide adequate feedback to higher proficiency students. Ide-

ally, teachers can create peer review tasks based on the results of a pretest, or the overall performance of ongoing tests. Current applications fall short in this area. Previous studies suggest that students were less confident in rating their peers' grammatical and pronunciation errors, especially if the learners under review have higher proficiency than the assessors (Lim, 2007). Finally, the interface is unintuitive and cluttered. Luo notes that students had problems recording and uploading their assignments, especially at the beginning of the semester, resulting in delayed or missed submissions, and in case of their peers missing the submission, affecting their assessment of the peers.

2.3 Computer Assisted Pronunciation Training

There have been a few longitudinal studies of the effects of CAPT on pronunciation training. One recent example is Hsu (2015), who reported some success of a longitudinal (16 months) study of a small group of non-English-major students (n=30) whose experimental group (n=18) received feedback from a commercial application (MyET) in fluency, stress, intonation and phoneme in imitation exercises.

However, systems relying entirely on CAPT suffer from many of the bottlenecks of current technologies and there have been reports investigating the drawbacks of such approaches.

Previous meta studies (Lee et al., 2015) comparing the effectiveness of automated feedback generated by computers, either in whole or in part, and exclusively human instructions show that automated training has comparatively smaller effect sizes. This might be attributed to the lack of adaptability, perceptual accuracy resulting in inappropriate feedback. It is still not certain how the effectiveness of CAPT can generalize/transfer to more general contexts outside the training (Thomson, 2011). Also, it is still unclear how ASR models used in previous research for non-native speech correlate with human judgments of intelligibility and it is possible that those models, not having been trained to adapt to the dialectal background of the speaker, fails to recognize speech perfectly intelligible by humans.

Fouz-González (2015) reviews state-of-the-art applications in use in recent years and concludes that while current CAPT technologies hold great promises, their usefulness in pedagogy is still lim-

ited and are not yet suited for autonomous practice by students. Instead of relying solely on CAPT, he suggests empowering students with self-monitoring skills. Such skills require building mental concepts and models that allow them to be able to evaluate their own performance, and to notice their own mistakes. Peer review, by developing the ability for self-monitoring, provides a viable complementary method for pronunciation training.

3 Platform Development

The overall architecture of the web-based platform is implemented by two major modules: first, an adaptive training module which consists of a perception submodule responsible for adaptively generating listening items to test and train students' ability in differentiating and identifying distinguishing features of English pronunciation and intonation and a production submodule for testing and training students' ability to orally produce similar sets of features effective for use in different pronunciation contexts; second, a peer review module for managing the peer review process in which sound recordings of student production are assigned based on configurable settings to different review groups who serve as anonymous, independent and therefore objective judges of the quality of the recordings.

3.1 Adaptive training module

Existing pronunciation training instruction suffers from a few problems including insufficient student motivation resulting from not seeing the benefits of teacher instruction, insufficient teacher instruction, especially one-on-one instruction, and inadequate time and resources (Foote et al., 2011), and traditional once-a-week classroom instruction is unlikely to be sufficient to produce the changes in the often fossilized pronunciation of EFL learners. Instead, substantial amounts of trackable practice and intervention in the form of exercises and feedback are necessary.

However, traditional forms of exercises where students repeat the same items over and over may not be the ideal way to practice since students are less likely to grasp the more general patterns underlying a particular category of perception or production skills, and the improvement in one particular

practice item does not mean transferable skills have been acquired but instead give students a false sense of achievement. For example, even a student has through practice mastered the pronunciation of contrastive pairs like *pitch* or *peach*, it is not guaranteed to be generalized to other words involving distinctions between */i/* and */il/*.

3.1.1 Automatic Generation

In the traditional paper-based approach, due to spatial constraints, it is impractical to enlist multiple variations of the same practice item, even when such practice is considered ideal. A system to generate different categories of problems with sufficient variations can be useful for this purpose.

For automatic item generation, we existing corpus and dictionary resources. Items used in pronunciation training are divided into two categories, segmental items concerning features of individual sounds, and suprasegmental items concerning aspects beyond the individual sound such as stress, intonation, linking and rhythm.

At the segmental level, items are generated through exhaustive dictionary-based searches using a number of predefined criteria. For example, minimal pair differentiation, a common perception exercise in pronunciation training, is automatically generated in the system. Pairs of phonemes known to be difficult to distinguish perceptually (e.g. */l/* vs */n/*, */ð/* vs */s/*) are collected and then used as criteria for searches in predefined dictionaries with IPA pronunciation to word mappings for words (Uncommon words are filtered out using a frequency-based word list) that differs by one of the phonemes but otherwise produces the same sound.

At the suprasegmental level, a small (still growing) sample of recordings of sentences and short texts by native speakers of English have been annotated by pronunciation instructors for suprasegmental features such as intonation, stress, linking and rhythm. The recordings are selected for exercises testing particular aspects of suprasegmentals.

With the generation mechanism in place, pronunciation instructors can define criteria for generating new items in accordance with the current teaching curriculum, double check the generated items, and roll out to students either as practice items or as quizzes. Results from the generated items are pro-

cessed by a real time tracking component responsible for statistical aggregation and analysis. The statistics will be fed back to the instructors and serve as guidance for future item generation.

3.1.2 Adaptive practice

One important consideration in the teaching of foreign languages is learner variability. Even within the same teaching institution, students differ greatly in proficiency, prior knowledge, interests, motivation, anxiety levels and learning styles. The traditional “one-size-fits-all” approach to language teaching in a classroom setting can lead to less optimal training effectiveness.

With the help of our adaptive system, teachers can easily assign tasks tailored to the needs of different students. In the item generation stage, each generated item is given at least one tag/category based on the task at hand. When the user completes and submits a new item, it is either automatically scored (if there is a well-defined key to the item) or assigned to at least one human for scoring. In either case, the score for that item is stored as part of the student’s performance statistics, and his/her accuracy rate of associated tags/categories to which the item belongs will be updated. The updated statistics, either Test-specific or longitudinal, can have an immediate effect on the item generation process which uses such information to generate targeted items for that user.

The adaptive component also provides real-time, ongoing feedback for students in summarized, comparative formats.

3.2 Peer Review Module

The peer review module is responsible for the generation of review items, filtering of reviewers based on predefined criteria, and dynamic group assignment. Review items can be generated either uniformly or based on individualized differences. Based on performance statistics from the adaptive training module, automatically generated recording items target individual students’ current weak points in a particular aspect of speech production. Items are recorded up to a number of times as defined by the teacher.

Filtering. The recorded items are then submitted to the system for peer review. Who can receive the items for reviewing is specified by the administrator. The system makes sure that the reviewers meet a

set of requirements such as differentiation scores of the target review items being higher than a certain threshold, or the overall pronunciation proficiency higher than that of the speaker.

Group assignment. Group assignment can be used in conjunction with/instead of filtering. Teachers/administrators can allow static or ad-hoc assignment of groups of varying sizes (pair, small group, class, grade etc.) according to the task at hand.

3.3 Web interface Design

As the interface needs to accommodate a wide range of controls, we design a pluggable architecture in which test/training items exist independently of each other while at the same time can be combined onto the same interface as needed. In this way, we create an interface that is highly uniform and easy to use.

Multimodal labeling interface. We built the interface on a browser (shown in Figure. 1) to make it immediately accessible to potentially geographically dispersed users and utilize the existing web architecture. Currently there are few existing tools suitable for peer assessment of EFL pronunciation. The ones that have been used in previous studies (e.g. the Blackboard system) have been found to be inflexible and difficult to use, lacking in the crucial functionalities necessary for efficiently performing the peer assessment task.

The interface we designed is a flexible browser-based interface for the peer review of arbitrary recordings. The interface is divided into several sections: 1) the play & contrast section where reviewers can choose to listen to recordings made earlier and to compare them with the standard model; 2) the scoring section where student gives an overall score to certain parts of the recording (stress, intonation); 3) the button next to the stars gives hints about the criteria; 4) the annotation section where students labeling problematic areas of the recording and make individualized comments. After students select a certain span of the text, a pop-up window that appears allow them to select pre-defined labels and write free-form comments.

Under the current architecture, the overall workflow of the system begins with perception and production items selected from the pool of automatically generated items and taken by students to measure their perceptual and productive proficiency at

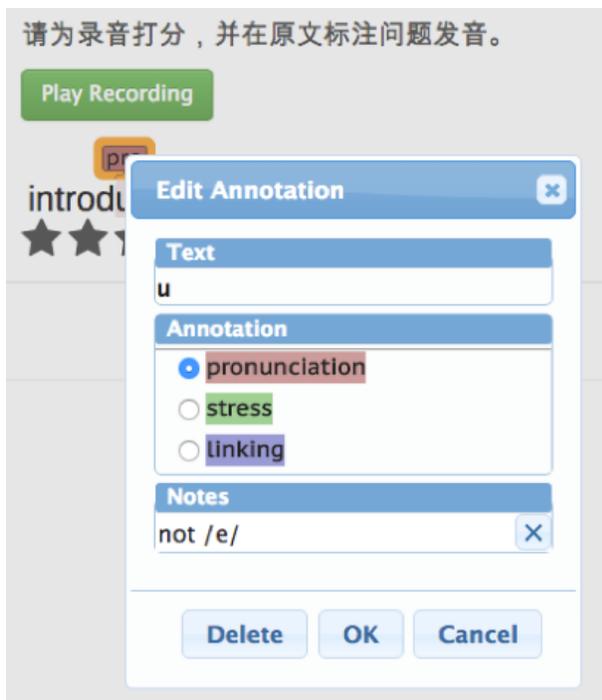


Figure 1: Web-based interface for peer assessment of pronunciation

a particular point in time. Perception items come with predefined keys and are automatically scored by the system. Production items, in the form of recordings made by users on the web-based interface, is given to peer users for peer review. A typical peer review workflow involves the student playing a recording made by another student, and (optionally) plays the standard reference model to make comparative analysis. The interface aims to allow reviewers to submit both quantifiable and qualitative data intuitively. Some of the basic input elements in the interface include: 1) scores: the overall rating that the reviewer assigns to a particular aspect of the recording; 2) tags: labeling of areas that the reviewer identifies as potentially problematic. The tags can be defined based on the needs of the task; 3) comments: free-form text targeted at the problem at hand. As the components on the interface can stand alone, each of them can be added to/removed from/made optional in the interface depending on the task. Such features of composability results in a compact, flexible, and easy-to-use interface. The results from the automatic scoring and peer review are then used to conduct statistical analysis which provides guidance

for the adaptive item generation process. Once the items are adaptively generated, they are displayed on the web-based interface to be taken by the users.

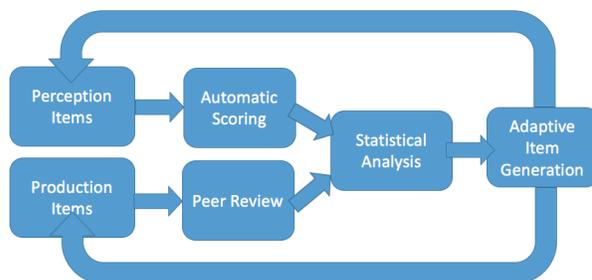


Figure 2: Overall workflow of the platform

4 Experiments

To test the efficacy of the platform as a viable means for peer review in pronunciation, we recruited 340 students from a university located in the southern China. The students first-year English majors with a least 6 years of experience in learning English. Data were collected of the students on their usage patterns.

4.1 Adaptive Practice

In our initial experiment, we instructed students to make imitative recordings as the homework for a couple of weeks. The students were allowed to practice on the platform and make as many recordings as they want to. They were informed that only their last recording will be scored. The system recorded the number of times they practiced online. By fitting an analysis of variance (anova) model (Table 1.), we find that the student's perceptual performance can be partially accounted for by the number of recordings they made as part of their exercises ($F=4.57$, $p=0.034$), which suggests that the more the students practice on the platform, the more likely their ability to differentiate different English sounds is going to improve.

Table 1: Analyses of Covariance for Perception Improvement

	Df	Mean Sq	F value	Pr(>F)
tries	1	5.73	4.40	0.0374
Residuals	169	220.27	1.30	

At the beginning of the study, students undertook a pretest on their proficiency in perceptual items, with a full score of 30. At the end of the study, a post-test using the same items were administered again to the same students. A paired t-test is then performed on the pretest and post-test, the results of which is shown in Table 2. The average score of the post-test for all students is 24.29, which significantly higher than the pretest score of 22.37 ($t=-15.50$, $p<0.000$), indicating significant improvement in students' perceptual performance after using the adaptive training system.

Table 2: Comparison of Pretest and post-test scores of perceptual items

	t
Test statistic	-15.50113
DF	312
p value	1.393535e-40
Alternative hypothesis	two.sided

Paired t-test: pretest and post-test

4.2 Peer Review

As part of their weekly exercises, students were required to complete a set of pronunciation and peer-review tasks. In the pronunciation tasks, students read either single words, contrastive pairs of words, a complete sentence, or a short text. Each of the items can be read a maximum number of two times. Then, each student's recording is submitted for peer review. Each of the recorded items is randomly dispatched to one of the other students taking the same course. Once a student completed and submitted the reviewing task, the feedback s/he gave will be visible to the recorder for reference.

4.3 User Survey

At the end of the month-long training, students were asked to volunteer to provide feedback on the experience of using the platform. 162 users responded by completing the online questionnaire hosted on the same platform. The items in the questionnaire are given in the Appendix.

Based on the survey results, we have been able to identify the overall perceptions of the students on

the system. In order to analyze the results of the survey more effectively, we performed a factor analysis after which 4 factors emerged.

The first factor concerns the capability of the platform to help improve students' abilities in oral and perceptual/productive pronunciation, and is thus named training improvement. Around half of the people believed that the system helped them to improve their proficiency in both the perception as well as production of English pronunciation. This figure indicates that although statistics shows that the majority of the students actually improved their test scores since the study, only half of them attributes part of their success to the use of the system. This definitely indicates room for improvement. As the system is still under active development, more in-depth analysis will be carried out to find out how to better cater for the needs of a wider range of students.

The second factor points to students' subjective perception of the platform as an enjoyable tool for their needs, and is summarized as 'subjective enjoyability'. Compared with the more traditional form of book-based exercises, the vast majority (83.02%) of the students surveyed prefer using the system to do exercises, while a considerably smaller percentage (but still more than half) of people prefer using it to doing pronunciation practice in a classroom setting. This reflects the fact that students are eager to adopt the use of the platform even the exercises they do prior to coming to college have been mostly paper-based. The teaching style in the author's university adopts a communicative approach to the teaching of language and therefore teachers pay close attention to interactive activities in the classroom. Multimedia is used for effective teaching, the level of teacher-student and student-student interaction. It is therefore surprising to see that more than half the students still prefer the Web-based approach to doing pronunciation practice.

The third factor reflects the system's capability to arouse the interests of and motivate students to engage tasks hosted on it, and is termed "Motivation". The majority of the students (75.31%) believe that the system provides them with enough motivation in pronunciation practice. Meanwhile, about half of the students think that the system stimulate their interests in learning English pronunciation. This

shows the advantage of online systems to monitor the progress of students' learning.

The fourth factor shows the perceived ease of use of the system by the users and is termed "ease of use". The system was perceived to be rather easy to use, as 94.44% of all surveyed users reported that they learned to use the system pretty quickly.

5 Conclusions

In this paper we introduce a Web-based platform for computer assisted peer review of EFL pronunciation. Modules for adaptive item generation and peer review are developed to overcome some of the bottlenecks in pronunciation training and instruction. We show that a Web-based platform can be effective in facilitating adaptive training and peer assessment in EFL pronunciation, with several obvious advantages: 1) The traditional lack of timely feedback and practice resources can be alleviated with the introduction of a centralized platform powered by Web-based technologies. 2) A user-friendly interface can motivate students to participate in more pronunciation practice than traditional paper-based and classroom exercises. 3) A web-based interface can be utilized to provide real-time or near-real-time feedback to students. 4) Computer-assisted peer review can serve as a complement to teacher/classroom instruction and automated speech technologies. 5) Web-based technologies can be utilized for more efficient peer review on EFL pronunciation. A web-based platform for peer review allows for random, anonymous grouping which is time-consuming and error-prone using traditional approaches. It allows real time tracking of progress and collection of user data beneficial for further enhancing pronunciation training and instruction. The successful deployment of the platform and positive feedback received from the users suggest that computer assisted peer review on pronunciation is not only feasible, but conducive to a wide range of tasks for providing timely and detailed feedback for students.

Despite these initial successes, this study also leaves a number of questions unanswered. Although it is shown in the study that computer assisted peer review complement teacher assessment, it remains unclear to what degree peer assessment following this approach can achieve the efficacy of teacher as-

essment. For example, with students serving as raters, are they able to reach a fair degree of inter-rater agreement? Are they able to provide the sort of quality feedback to their peers as can be expected from a teacher? What factors may influence the effectiveness of peer assessment? These questions serve as the starting points for our further research. Fortunately the current platform has provided us with the means to carry out such research in a way not possible previously, and this may serve as yet another benefit of the platform.

6 Appendix

Questionnaire for User Feedback on the System:

1. Easy of use of the interface

I learned how to use the system quickly

The operations of the system interface are highly efficient

2. Skill Improvement

The system improved my ability to distinguish English sounds

The system improved my ability to pronounce English sounds

The system improved my oral English abilities

3. Needs & Motivation

The functions provided by the system meet my needs

The exercises provided by the system motivate me to study pronunciation

The system arouses my interests in studying pronunciation

4. Item Design

The exercises on the system are challenging

The design of the exercises is reasonable

The number of exercises is appropriate

5. Preference

I like using the system to do exercises

I prefer to use the system for pronunciation practice to paper-based exercises

I prefer to use the system for pronunciation practice to doing pronunciation exercises in the classroom

I would continue to use it for my pronunciation study if it continues to be open for use

7 Acknowledgements

This study was supported by the Project for Innovating Young Scholars (Social Sciences) of the Department of Education of Guangdong Province, China (grant number: 2015WQNCX030)

References

- Tracey M Derwing, Lori G Diepenbroek, and Jennifer A Foote. How Well do General-Skills ESL Textbooks Address Pronunciation? *TESL Canada Journal*, 30(1):22–44, 2012. doi: 10.18806/tesl.v30i1.1124.
- Jennifer A Foote, Amy K Holtby, and Tracey M Derwing. Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal*, 29(1):1–22, 2011. ISSN 0826-435X. doi: 10.18806/tesl.v29i1.1086.
- Jonás Fouz-González. Trends and Directions in Computer-Assisted Pronunciation Training. In *Investigating English Pronunciation*, pages 314–342. Palgrave Macmillan UK, London, 2015. doi: 10.1057/9781137509437_14. URL link.springer.com/content/pdf/10.1057/9781137509437_14.pdfhttp://link.springer.com/10.1057/9781137509437_14.
- Douglas J. Hacker, Linda Bol, Dianne D. Horgan, and Ernest A. Rakow. Test prediction and performance in a classroom context. *Journal of Educational Psychology*, 92(1):160–170, 2000. doi: 10.1037/0022-0663.92.1.160. URL <https://doi.org/10.1037/0022-0663.92.1.160>.
- Liwei Hsu. An empirical examination of EFL learners' perceptual learning styles and acceptance of ASR-based computer-assisted pronunciation training. *Computer Assisted Language Learning*, 8221(May):1–20, 2015. ISSN 0958-8221. doi: 10.1080/09588221.2015.1069747. URL <http://www.tandfonline.com/doi/full/10.1080/09588221.2015.1069747>.
- Junkyu Lee, Juhyun Jang, and Luke Plonsky. The Effectiveness of Second Language Pronunciation Instruction: A Meta-Analysis. *Applied Linguistics*, 36(3):345–366, 2015. ISSN 1477450X. doi: 10.1093/applin/amu040.
- Hana Lim. A Study of Self- and Peer-Assessment of Learners' Oral Proficiency. *CamLing 2007: Proceedings of the Fifth University of Cambridge Postgraduate Conference in Language Research*, pages 169–176, 2007.
- Kristi Lundstrom and Wendy Baker. To give is better than to receive : The benefits of peer review to the reviewer 's own writing. 18:30–43, 2009. doi: 10.1016/j.jslw.2008.06.002.
- Beate Luo. Evaluating a computer-assisted pronunciation training (CAPT) technique for efficient classroom instruction. *Computer Assisted Language Learning*, 29(3):451–476, 2016. ISSN 09588221. doi: 10.1080/09588221.2014.963123. URL <http://10.0.4.56/09588221.2014.963123{%}5Cnhttp://jproxy.lib.ecu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true{%}&db=ehh{%}&AN=113740026{%}&site=ehost-live>.
- Cássia O. Mendonça, Karen E. Johnson, and Cassia O. Mendonca. Peer Review Negotiations: Revision Activities in ESL Writing Instruction. *TESOL Quarterly*, 28(4):745, 1994. doi: 10.2307/3587558. URL <https://doi.org/10.2307%2F3587558>.
- John M. Murphy. Intelligible comprehensible, non-native models in ESL/EFL pronunciation teaching. *System*, 42:258–269, feb 2014. doi: 10.1016/j.system.2013.12.007. URL <https://doi.org/10.1016%2Fj.system.2013.12.007>.
- Jui-ching Peng. Peer assessment in an EFL context: Attitudes and friendship bias. In Matthew T. Prior et Al, editor, *Selected*

- Proceedings of the 2008 Second Language Research Forum*, pages 89–107, 2010. URL <http://www.languagetestingasia.com/content/3/1/11>.
- H. Saito. EFL classroom peer assessment: Training effects on rating and commenting. *Language Testing*, 25(4):553–581, oct 2008. ISSN 0265-5322. doi: 10.1177/0265532208094276. URL <http://ltj.sagepub.com/cgi/doi/10.1177/0265532208094276>.
- Ron I. Thomson. Computer Assisted Pronunciation Training: Targeting Second Language Vowel Perception Improves Pronunciation. *CALICO Journal*, 28(3):744–765, 2011. ISSN 07427778. doi: 10.11139/cj.28.3.744-765. URL <http://www.equinoxpub.com/journals/index.php/CALICO/article/view/22985>.
- Keith Topping. Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3):249–276, 1998. ISSN 0034-6543. doi: 10.3102/00346543068003249. URL <http://rer.sagepub.com/content/68/3/249.short>.
- Pavel Trofimovich, Talia Isaacs, Sara Kennedy, Kazura Saito, and Dustin Crowther. Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition*, 19(01):122–140, jan 2016. ISSN 1366-7289. doi: 10.1017/S1366728914000832. URL http://www.journals.cambridge.org/abstract/_}S1366728914000832.
- Kurt Van Lehn, Michelene T Chi, William Baggett, and R Charles Murray. Towards a Theory of Learning During Tutoring. Technical report, DTIC Document, 1995.