ReCALL **23**(1): 59–78. 2011 © European Association for Computer Assisted Language Learning
doi:10.1017/S0958344010000285

*Developing and evaluating an oral skills training website supported by automatic speech recognition technology**

HOWARD HAO-JAN CHEN
Department of English, National Taiwan Normal University, He-ping East Road, Section 1, Taipei 10610, Taiwan125
(email: hjchen@ntnu.edu.tw)

Abstract

Oral communication ability has become increasingly important to many EFL students. Several commercial software programs based on automatic speech recognition (ASR) technologies are available but their prices are not affordable for many students. This paper will demonstrate how the Microsoft Speech Application Software Development Kit (SASDK), a free but powerful tool, can be used to develop an oral skills training website for EFL students. This ASR-based website offers six different types of online exercises which allow students to practise their oral skills and obtain immediate feedback on their performance. A group of 25 college students and a group of 35 pre-service English teachers were invited to use the website. Two surveys were conducted to investigate the students' and the pre-service teachers' perceptions of this site. The results indicated that most teachers and students enjoyed using this website, which they felt could help improve their English oral skills. They also pointed out that the main strength of the ASR-based learning system is that it offers several different types of exercises which can encourage learners to produce more output in a low-anxiety environment. The major limitations of the website are the insufficient feedback and the challenging standards one must meet in order to achieve a pass mark. These findings can be useful for teachers who are interested in using ASR in teaching and for CALL researchers who aim to develop better ASR-based systems for language learning.

Keywords: oral production, automatic speech recognition, Microsoft Speech Kit, Web-based

1 Introduction

1.1 The Importance of English Oral Ability

The importance of better English oral communication ability has been recognized at national and global levels. Many universities and institutes in English-speaking countries ask prospective students to demonstrate their oral proficiency by achieving

* This research project was funded by two research grants from the National Science Council, Taiwan (NSC 92-2411-H-003-039 and NSC93-2411-H-003-012).

a particular score on various language tests (IELTS and TOEFL). For those who hope to study abroad to advance their education, the pressure to enhance English oral proficiency is great.

The pressure to improve oral skills is not confined to the student population in countries where English is a foreign language. Researchers and professionals in a variety of fields need to be able to communicate in English to participate in activities such as business meetings and international conferences. It is thus clear that many individuals in EFL settings have a strong need to improve their oral abilities.

Educational institutes in EFL settings have begun to explore a variety of methods to help individuals achieve improved oral skills. For instance, some schools and universities have hired more native English speakers and provide students with more opportunities to interact with these teachers. Some institutions have reduced class sizes and expect that this will lead to more teacher-student interactions in the target language. Some schools and universities have also begun to explore the power of new speech recognition technologies.

1.2 Improving oral language skills with automatic speech recognition technologies

As computers have become increasingly powerful and affordable, speech recognition technologies have been incorporated into various language learning software programs (Bernstein, Najmi, & Ehsani, 1999; Egan & LaRocca, 2000; Ehsani & Knodt, 1998; Eskenazi, 1999a, 1999b; Harless, Zier, & Duncan, 1999; Hincks, 2003; Holland, Kaplan, & Sabol, 1999; LaRocca, Morgan, & Bellinger, 1999; Mostow & Aist, 1999; Rypa & Price, 1999; Wachowicz & Scott, 1999). Chen (2001) reviewed five commercial English learning programs (CNN Interactive English, Syracuse English Comprehensive Learning Series, TeLL Me More, TRACI Talk, and Encarta Interactive English Learning) which are based on automatic speech recognition technologies. These programs use different speech recognition engines and provide various language learning activities. Most of these English learning programs are sold in a CD-ROM or DVD format.

In addition to these commercial programs, several academic ASR-based language learning systems have been developed in various universities and institutes. Some of these ASR-based systems are briefly reviewed. The Subarashii program is a well-known computer-based interactive spoken language education (ISLE) system (Bernstein *et al.*, 1999). The ASR-based system allows Japanese language learners to engage in seemingly open dialogues. Authentic situations have been incorporated into Subarashii, and each situation includes realia and photos. The series of situations is designed for learners to practise partially constrained, goal-oriented spoken interactions.

Another project called Virtual Conversations was adapted from a speech-activated multimedia system at Interactive Drama Inc. (Harless *et al.*, 1999). This program allows users to engage in face-to-face dialogues with virtual characters who appear in full motion video on a CD-ROM. During the dialogues, users have to participate in simulated interviews and several 'real-life' situations. The main purpose of the program is to assist military linguists who are expected to exhibit full linguistic competence in Arabic.

The Voice Interactive Language Training System (VILTS) is a language training system developed at the Stanford Research Institute (Rypa & Price, 1999). The system was developed to improve the listening and speaking skills of French learners.

VILTS includes three types of activities: listening, speaking, and reading aloud. The goal of the speaking and reading-aloud activities is to achieve appropriate dialogue responses in certain conversational situations. Ten daily conversations were included in the prototype.

FLUENCY is a pronunciation training system for ESL learners created at Carnegie Mellon University (Eskenazi, 1999a). It uses the CMU SPHINX II recognizer to detect speakers' pronunciation errors. The recognizer adopts the "forced alignment mode" to detect speakers' pronunciation in terms of phone and prosody errors. The speakers' recognition scores are compared to the mean scores of native speakers' pronunciation, and then errors can be identified.

Several researchers (Bernstein *et al.*, 1999; Egan & LaRocca, 2000; Ehsani & Knodt, 1998; Eskenazi, 1999a, 1999b; Harless *et al.*, 1999; Holland *et al.*, 1999; LaRocca *et al.*, 1999; Mostow & Aist, 1999; Neri, Cucchiarini & Strik, 2001, 2003; Neri, Cucchiarini, Strik & Boves, 2002; Rypa & Price, 1999; Wachowicz & Scott, 1999) have pointed out that commercial and academic ASR-based software programs can provide the following benefits:

1. Students will have more opportunities to produce the target language and have extensive interaction with computers.
2. Students will be given individual attention; they do not need to compete with other classmates.
3. Students can learn to communicate under a less threatening environment and they often can get feedback quickly.
4. Students are provided with various types of direct or indirect feedback from computers.
5. Students have the opportunity to listen to English spoken by a range of native speakers.
6. Students can control the pace of their learning which might lead to increased confidence.

Although speech recognition software programs can be useful for pronunciation training and oral skills training, many students still do not have access to these products. One major problem for language learners is the high price of some commercial programs. For example, the price for TELL ME MORE is approximately US$250. In addition to the high retail price, another major problem associated with these PC-based programs is their limited accessibility. Many of them are generally installed in school language laboratories, which means that students can only access them during school hours. They often cannot have access to these useful ASR programs from their dormitory or from home.

Given that the Internet is the most effective and powerful platform for delivering learning resources to students, ASR-based websites can provide more convenient services to learners. In an extensive search using the Google search engine, it was found that very few free ASR-based programs or websites are available for second language learners. One exception is a recent project reported by Chiu, Liou, and Yeh (2007). A research team in Taiwan has developed a web-based conversation environment called CandleTalk. CandleTalk is equipped with an ASR engine that judges whether learners provide appropriate input. This system was developed to

help EFL learners receive explicit speech act training that leads to better oral competence. Six speech acts (greeting, parting, requesting, apologizing, complaining, complimenting) are presented as the foci of the materials. The results of an experimental study on CandleTalk showed that the application of ASR was helpful for college freshmen in the learning of the speech acts, particularly for less proficient students. Although the CandleTalk website offered some interesting online lessons with the help of ASR technologies, the learning content in this site was still very restricted and only focused on the teaching of a few speech acts.

In addition to the free CandleTalk site, there is another commercial web-based program available in Taiwan called My English Tutor (MyET). Developed by a company called L Labs, MyET is particularly strong in offering error diagnosis and feedback. MyET can analyze students' pronunciation, pitch, timing and emphasis, and even pinpoint individual sounds which are problematic. After students speak, they receive a score and specific feedback on how to improve their spoken output.

Chen (2006) investigated the impact of MyET on college EFL students. Forty students were assigned to two groups. The control group received classroom instruction on contrastive stress. The experimental group received the same classroom instruction, and was also asked to use the ASR program MyET after class for about six weeks. The post-test scores showed that the ASR program helped students in the experimental group improve their command of contrastive stress. Moreover, most students who used MyET held a positive attitude toward the use of the program.

Liao (2009) also investigated the impact of MyET on college EFL students. Forty-three college students were divided into three groups. The first group received instruction on contrastive stress in the classroom and used MyET, the second group used MyET only, and the third group had classroom instruction and printed handouts. After the treatment, it was found that ASR helped students improve their pronunciation. In addition, students showed a positive attitude toward MyET and indicated that they would be likely to recommend it to other learners.

Given that the Internet is the most effective platform for delivering learning content and resources to language learners and the participants in several aforementioned studies expressed positive feedback toward web-based ASR systems, it seems useful to develop ASR-based language learning systems on the Internet for students who cannot have access to commercial programs.

Although there are several different approaches that can be utilized in developing web-based speech applications, there are two major foci of these new speech technologies. One is VoiceXML and another is Speech Application Language Tags (SALT). Both are quite useful solutions for adding voice recognition to the Internet. The VoiceXML Forum was developed by AT&T, IBM, Lucent, and Motorola in 1999, in order to develop a standard markup language for specifying voice dialogues. It allows voice applications to be developed and deployed in the same way that HTML is used for visual applications.

SALT was developed as a competitor to VoiceXML and is supported by the SALT Forum. Founding members of the forum include Cisco Systems, Comverse, Philips Consumer Electronics, ScanSoft, Intel, and Microsoft. SALT is an XML-based markup language that is used in HTML and XHTML pages to add voice recognition capabilities to web applications. The Microsoft Speech Server product supports SALT.

1.3 Research questions

As previous research has shown, there are several advantages of using ASR programs. If a new web ASR-based training system can be developed with the help of these new technologies, then more language learners can benefit. This study aimed to develop a unique language learning site based on Microsoft's SALT language and further investigate EFL students' and pre-service teachers' perceptions of this innovative site. Two major research questions were proposed:

1. What do college-level EFL students perceive as the strengths and limitations of the oral skills training site supported by ASR technologies?
2. What do pre-service EFL teachers perceive as the strengths and limitations of the oral skills training site supported by ASR technologies?

In the following sections the subjects, the learning content and functionalities of the ASR-based site and the procedures of data collection are first introduced. Then the user survey results are presented. Pedagogical implications and possible directions for improvement of the website are also discussed.

2 Methodology

2.1 Subjects

To better understand students' and teachers' perceptions of the ASR-based web system, a group of college freshmen (low-intermediate level EFL learners) and a group of pre-service English teachers were invited to participate in this study. Teachers' perceptions were targeted because few existing studies have investigated language teachers' views about automatic speech recognition.

For the college student group, 25 Taiwanese college students who were taking a Freshman English course were invited to use the ASR website. These students were graduates of vocational high schools. The majority had difficulty with English pronunciation, and in a survey taken prior to the study, most described their English pronunciation as poor.

The pre-service teachers group was composed of 35 junior or senior English majors studying in a national university in Taiwan. At the time of the study, the students were taking a Computer-Assisted Language Learning course in which they had opportunities to experience various new computer technologies for language learning. They were instructed to use the ASR system and were asked to identify the strengths and weaknesses of the system in an evaluation report.

2.2 Instrument: the oral skills training site based on ASR technologies

The ASR-based oral skills training website was developed with funding from the National Science Council (NSC), Taiwan. The development team included a professor with a background in TESOL, a professor with a background in computer science, several student programmers and part-time research assistants.

The Microsoft SALT language was adopted to create a website to facilitate oral skills training. To help developers effectively utilize the SALT language, Microsoft

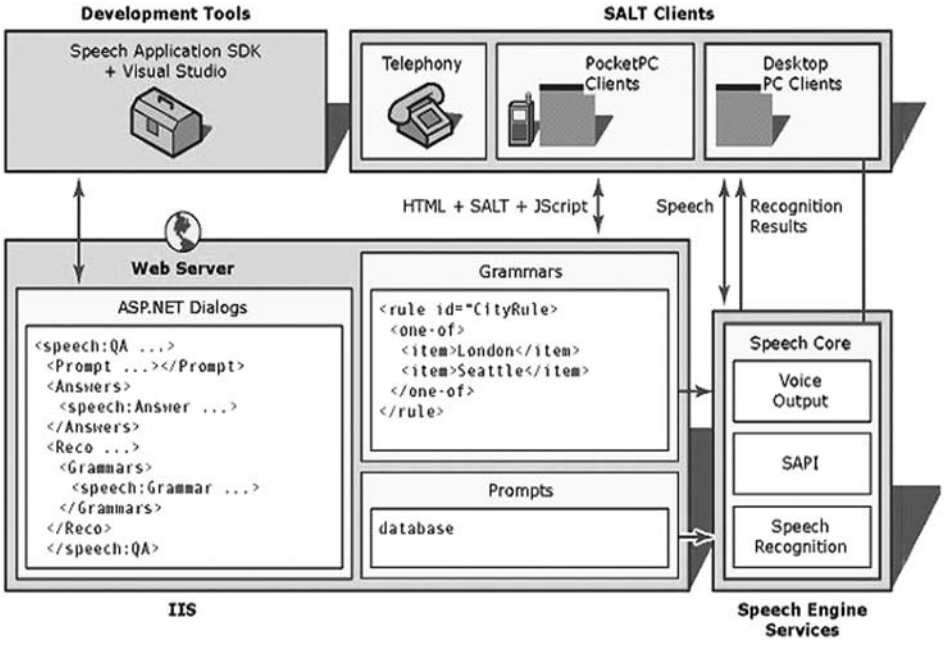

Fig. 1. ASP.NET Distributed Speech Application Scenarios

released a toolkit called Microsoft Speech Application Software Development Kit (SASDK). The toolkit can add speech interfaces to ASP.NET Web applications, and helps developers create, debug, and deploy these speech-enabled ASP.NET Web applications. Microsoft's SASDK and Speech Server can be used to develop both voice-only (telephony) and multimodal (desktop PC, notebook PC, pocket PC and other mobile devices) speech recognition/synthesis environments. This configuration is shown in Figure 1.

As shown in Figure 1, SASDK and Visual Studio ASP.NET can be used to edit web pages. Once these are put on the Web Server (IIS), the client can easily use a web browser to read a web page which contains speech tags. The web server processes the voice input from the client or sends results to the client. On desktop or notebook computers, users can download and install the speech add-in for Microsoft Internet Explorer. After users install the speech add-in software program, the computer will have the speech API for speech recognition and the TTS prompt engine for speech synthesis. When users employ Internet Explorer to read web pages, the speech add-in will automatically decode SALT grammar and perform speech recognition and speech synthesis. This process is shown in Figure 2.[1]

Based on the Microsoft speech technologies, a web-based automatic speech recognition system for oral skills was developed. A screenshot of what a user sees when accessing the site is shown in Figure 3. This ASR-based website can be designed to allow students at different proficiency levels to work on their pronunciation and oral

[1] Figures 1 and 2 are taken from Microsoft MSDN library documents.

https://doi.org/10.1017/S0958344010000285 Published online by Cambridge University Press

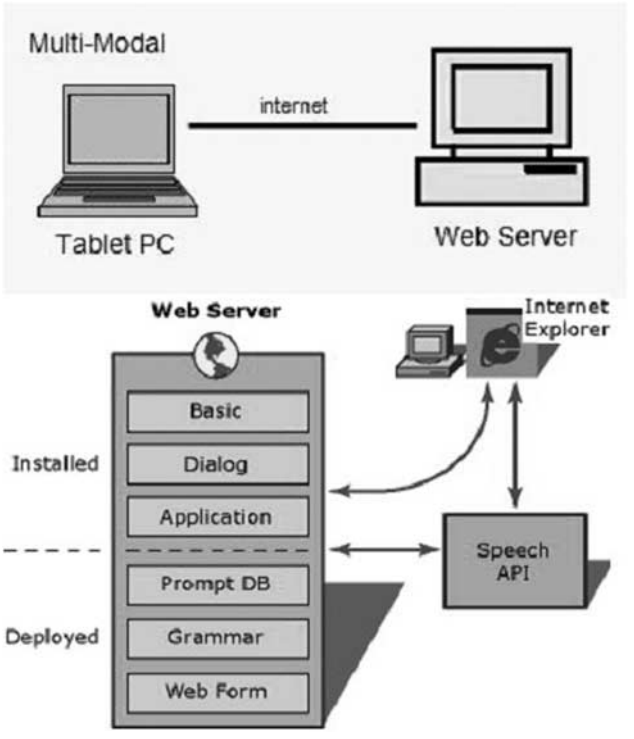

Fig. 2. The interactions of various components in Microsoft.Net speech

communication skills. When students log onto the website, they read the instructions and install the Microsoft speech add-in for Internet Explorer. After installing the speech add-in program, students can begin to use the various exercises available on the ASR-based oral skills training site.[2]

There are six different types of interactive exercises based on ASR, and the first three types operate in a similar fashion. The first type comprises the easiest exercises, and these are targeted at beginning students. A picture is shown on the screen, and students have to say the name of the object. The second type of exercise asks students to listen and repeat the sentence shown on the screen. This exercise is a popular test format used in various spoken English tests (e.g., TELL ME MORE). The website contains hundreds of commonly used sentences. All these sentences were recorded by native English speakers. Students can listen to the native speakers' pronunciation first and then begin to read the sentences aloud. The third type of exercise is one of the most commonly used exercises in various commercial ASR-based programs. This exercise asks students to listen to a sentence and then choose an appropriate response from four options (only one answer is correct).

In these first three types of interactive exercises, after students say the answer, the system will give immediate feedback. If the answer is acceptable, loud applause follows

[2] (http://140.122.83.197/speech or http://140.122.83.226:8080/speech/)

https://doi.org/10.1017/S0958344010000285 Published online by Cambridge University Press

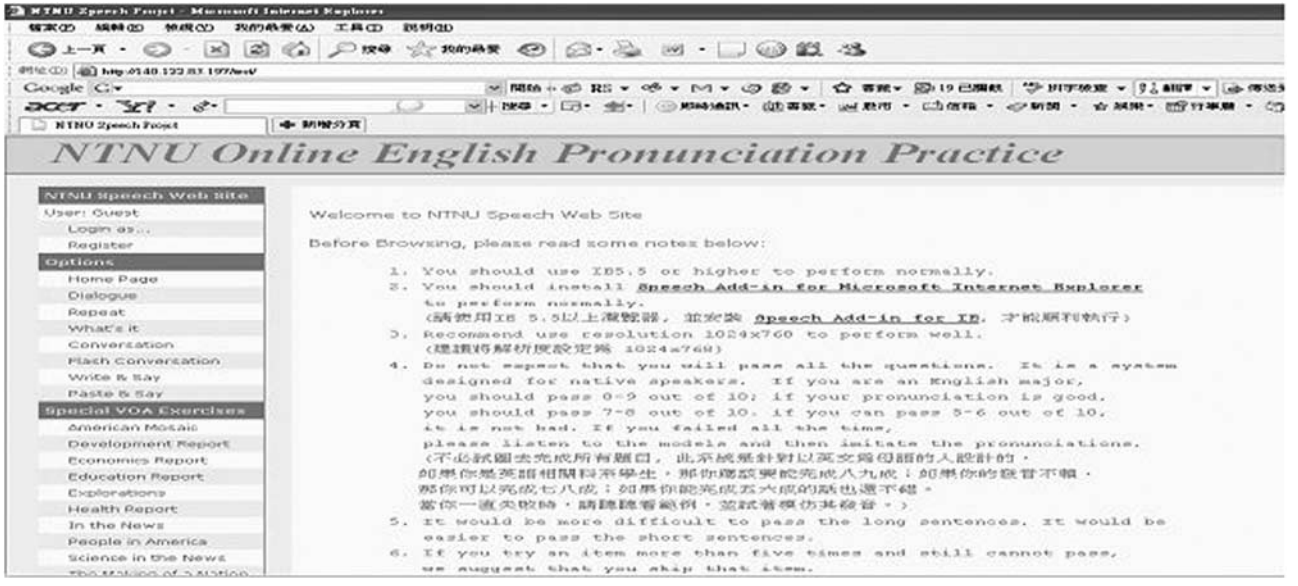

Fig. 3. The oral skills training site

and the system notes that the learner has passed this item. If the answer is not acceptable, then the system shows "No recognition". The learner can listen to the model and try again. Examples of the first three types of interactive exercises are shown in Figures 4–6.

The fourth type of exercise, "Role-Play", is more challenging. Students need to take on different roles in a conversation. In contrast to the previous exercises, students need to read all the sentences in a dialogue correctly before they can get a full score. This exercise is more suitable for intermediate level students. An example is shown in Figure 7.

The fifth type of exercise combines Flash animation and automatic speech recognition technologies. Students first view the conversation encoded in Flash, and then they are asked to play different roles and to practise the sentences. The animation can often make the interaction more interesting. An example is shown in Figure 8.

In the sixth type of exercise, 'Write and speak', students are allowed to create their own pronunciation or speaking exercises, thus setting a learning goal for themselves. This is a unique feature which many different commercial ASR software programs

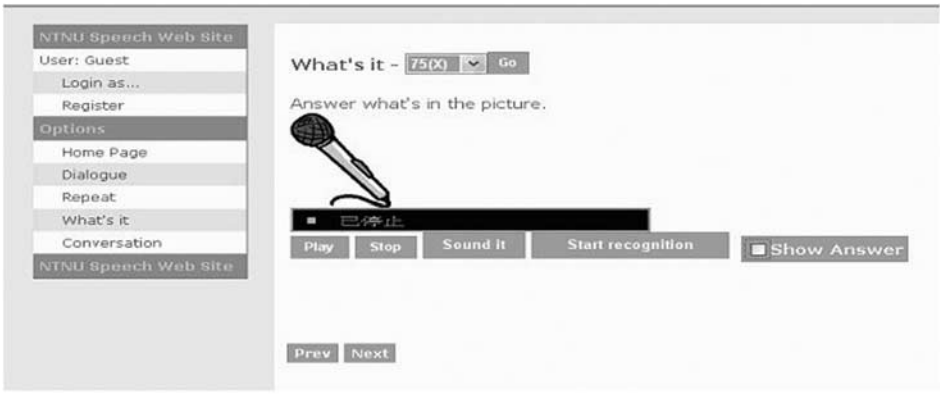

Fig. 4. Identify the objects

Fig. 5. Listen and repeat

https://doi.org/10.1017/S0958344010000285 Published online by Cambridge University Press

Fig. 6. Choose an appropriate response

Fig. 7. Role-Play

Fig. 8. Flash-based conversation practice

https://doi.org/10.1017/S0958344010000285 Published online by Cambridge University Press

Fig. 9. Write and Speak

do not have. Based on the Microsoft speech recognition technologies, a simple interface was created which allows students to input (copy and paste) any sentence they want to practise. After students input the sentence and click submit, the system automatically creates an exercise for them. Students can then say the sentences out loud, and the system evaluates their performance and gives them instant feedback. One example is shown in Figure 9. This type of on demand pronunciation exercise might better meet the needs of individual students.

In addition to the learning content and tools, a tracking device was also developed to help students monitor their own participation and progress, given that this system is mainly designed to encourage students to do these pronunciation exercises after classes. The learning record also offers language teachers an effective way of monitoring student performance.

2.3 The procedure for collecting user feedback

To obtain feedback from students and teachers, a group of 25 freshmen and a group of 35 pre-service English teachers were asked to use the website and then comment on their experience. The college students were asked to use the website over a ten-week period. Each week they were assigned to work on various exercises available on the website, and each session lasted for approximately two to three hours. At the end of the ten-week period, a questionnaire was distributed to solicit their feedback. The questionnaire consisted of nine questions in the format of a five-level Likert scale: (1. Strongly disagree; 2. Disagree; 3. Neither agree nor disagree; 4. Agree; 5. Strongly agree) and three open-ended questions asking students about the strengths and weaknesses of the website and possible suggestions for improvement. In addition, there were five questions about the students' background and their usage of the website.

The pre-service teachers were guided to use the various interactive exercises based on the ASR system for about two hours in a computer laboratory. After working with these different exercises, they were asked to submit an evaluation report in which they identified the strengths and weaknesses of the system.

https://doi.org/10.1017/S0958344010000285 Published online by Cambridge University Press

3 Survey results

3.1 Survey results from the college students

Based on the survey, most students felt that their pronunciation and speaking skills were not adequate. While using the ASR site, they chose to try three to five times before they gave up trying. When they found they were not sure about the pronunciation of a certain word, many students (65%) would choose to listen to the audio models provided by the system. Results from the students are summarized in Table 1.

The mean score for all the items was 3.89, with a standard deviation of 0.58. In general, the student participants stated that the ASR-based learning system was beneficial to their pronunciation/speaking and other English language skills and they would continue using the site in the future (item 7, 8, and 9). Among the various ASR-based exercises, the ''identify the objects'' and ''listen and repeat'' exercises ranked high (4.20; 4.04) in the survey. However, the results for item 1 and item 6 showed that students were not very satisfied with the navigation and the accuracy of the speech recognition engine (mean = 3.20 and 3.65 respectively).

Based on the responses to the three open-ended questions, the following positive feedback was provided by students.

1. This website is good for students who need to improve their pronunciation and listening.
2. The ''Identify the objects'' section is useful for reviewing vocabulary items.
3. The ''Listen and repeat'' section is quite useful.
4. The website is also useful for vocabulary learning and listening
5. The site is very interesting and easy to use.
6. For students who do not dare to speak, this is a good tool for them to practise conversation.
7. It is very convenient to have access to this website. Students can do the exercises from home or the dormitory.
8. The site has a wide range of options from very easy items to more difficult items.

Table 1 *Results of student responses to the oral skills training site*

Item No.	Items	Mean	SD
1.	The website was very easy to use and browse.	3.20	0.58
2.	This website provided different pronunciation exercises; these exercises are useful for improving pronunciation.	3.80	0.50
3.	I liked the ''Identify the objects'' exercise.	4.20	0.65
4.	I liked the ''Listen and repeat'' exercise.	4.04	0.68
5.	I liked the ''Choose an appropriate response'' exercise.	3.84	0.37
6.	The speech recognition mechanism was accurate.	3.64	0.57
7.	This system helped me improve my English pronunciation.	3.92	0.64
8.	This system helped me improve my overall English ability.	4.16	0.62
9.	In the future, if the system can have more learning materials, I will continue to use this website.	4.20	0.58

https://doi.org/10.1017/S0958344010000285 Published online by Cambridge University Press

In addition to the strengths, students also provided specific suggestions to further improve the content and design of this website. These are summarized below.

1. The content is rich and useful, but the user interface is not very attractive. The interface should be more attractive.
2. The standards set by the ASR engine are too high. No matter how hard students try, sometimes they cannot pass that item. It is very frustrating. Sometimes students have to skip an item.
3. Some TTS (synthetic) voices are not clear and natural enough. Human voices are clearer and better.
4. The site should add in more pictures and more videos to attract or motivate users.
5. The site would be better if it could be designed as an online game.

3.2 Survey results from the pre-service teachers

The pre-service teachers' opinions were collected via their evaluation reports. A total of 35 pre-service teachers submitted reports. Because all of these pre-service teachers have solid training in linguistics and TESOL methodology, their opinions are somewhat different from those of students. Their perceptions and suggestions are summarized in Table 2.

Table 2 *Strengths of the ASR-based oral skills training website identified by pre-service teachers*

Strengths	Summary of the comments
1. Creates a lower anxiety speaking environment.	*'ASR provides a low-anxiety environment in which students may decide their own pacing and level.'*
2. Various exercises for students are available. The exercises are arranged by difficulty level.	*'Different levels of difficulty and different kinds of activities are offered to enhance the effectiveness of the system. The various items in the system are arranged from easy (words) to difficult (dialogues).'*
3. Self-access learning for language learners.	*'It is a good tool for students especially those at the lower intermediate level. The self-access system might also be good for remedial teaching. This site allows learners to have more oral practice.'*
4. Allows students to practise pronunciation.	*'This kind of ASR system is beneficial to students who want to practise their pronunciation of individual sounds.'*
5. The Role play activity (animation) is very interesting.	*'The ''role play exercise'' can help students enhance their communicative competence. The ''Flash Conversation'' is great, because it makes the learning more active and relaxing.'*
6. Exercises included in 'Write and speak' are very useful.	*''Write and speak' is perhaps the most useful part. Students can write down everything they want to learn and learn the correct pronunciation from the system.'*
7. The tracking and log function is useful for both learners and teachers.	*'It is useful for teachers to see how the students performed. They can determine if their students answered the questions correctly, and how many questions students have completed. Teachers can use this system to encourage students to practise English at home.'*

https://doi.org/10.1017/S0958344010000285 Published online by Cambridge University Press

Table 3 *Weaknesses of the ASR-based oral skills training website and suggestions for improvement as identified by pre-service teachers*

Weaknesses	Summaries of teachers' comments
1. Limited corrective feedback cannot offer a clear guide for learners to improve their pronunciation.	*'If the system does not accept a student's pronunciation of a certain word or a sentence, the system only shows ''Please try again.'' Detailed feedback about students' problems is not provided. Students can only listen to the model one more time and then try again. They can only improve by imitation. There should be a better guide, so that problems like the position of tongue, the shape of mouth, intonation, and so on can be better explained to the students.'*
2. The ASR system is too strict for EFL students.	*'For longer sentences, some students might have difficulties passing these oral exercises. Therefore, the standard of the system should be lower if possible.'*
3. Human voices should replace synthesized voices.	*'For some exercises, a more authentic voice would be even better! The synthesized sound is a little bit artificial and is not natural enough.'*
4. Stress and intonation are not covered.	*'The practices only focus on pronunciation drills, but little on stress and intonation. It would be useful if the system could provide some intonation graphs to help students better imitate the model voice.'*
5. Teachers and students might not be able to solve some technical problems.	*'In some cases, teachers and students might not be able to solve the technical problems of using the system (like poor sound cards and microphones) or they may not have sufficient support or time to solve technical problems.'*
6. The fixed dialogues might not be very useful.	*'The dialogues in the system are not extensive enough for the students to get the whole picture of the English language, and the fixed dialogues may not be really useful in the real world.'*

The column on the left provides a general category of response while the column on the right includes more specific comments or suggestions.

In addition, the pre-service teachers helped to identify several weaknesses of the system based on their professional knowledge, and they also offered suggestions for improvement. A summary of these comments is shown in Table 3. The column on the left provides a general category of response while the column on the right includes more specific comments or suggestions.

3.3 Summary of student and teacher feedback

It is important to note that students and pre-service teachers indeed had different perspectives in examining the ASR-based website. For the strengths, students mostly focused on the richness of the content and the convenience in using the website. They found this site interesting and it also allowed them to practise speaking and other language skills. Pre-service teachers believed that the interaction with computers could create a lower anxiety speaking environment. This rich learning content also provided students with a self-access learning environment. Teachers also liked the

https://doi.org/10.1017/S0958344010000285 Published online by Cambridge University Press

Table 4 *Summary of the major suggestions*

Suggestions for Improvement	Students	Teachers
1. The content is rich, but the user interface is not very attractive. The interface should be more attractive.	✓	
2. The ASR system can only provide limited corrective feedback (acceptance vs. rejection).		✓
3. The ASR system is too strict for language learners and the standards are too high. It is very frustrating sometimes when the system does not accept what students say.	✓	✓
4. The TTS voices are not clear enough. Human voices are clearer. Authentic human voices should replace all the TTS voices.	✓	✓
5. Stress and intonation training are not adequately covered		✓
6. The site should add in more pictures and videos to attract and motivate users. The site would be better if it could be designed as an online game.	✓	
7. The practice of fixed dialogues might not be very useful. More diverse patterns should be added.		✓
8. Technical problems like poor sound card and microphone might appear.		✓

different activities and exercises; they in particular liked the 'Role play' and 'Write and speak' activities. The tracking and log function was also considered to be essential.

In addition to these strengths, both students and pre-service teachers made several useful suggestions about improving this site. Table 4 summarizes their major suggestions.

Based on the feedback from both students and teachers, the ASR-based oral skills training system can be further improved in different aspects. First, the system needs to provide learners with more detailed feedback. Second, the standards of the ASR engine might be too high, especially for lower level EFL learners. Third, some sections used synthesized voices and the quality of TTS voices can be improved. Fourth, the user interface can be made more attractive and game-like. Fifth, stress and intonation training should be adequately covered and the formats of the exercises should be more flexible.

4 Discussion

4.1 Students' and teachers' attitudes toward the ASR-based oral training site

Based on the user surveys, the ASR-based system can provide learners with more opportunities for producing target language output. Most students indicated that they liked the rich content (a variety of exercises) and the convenience of accessing the website. They felt that this site could be helpful for their speaking and listening skills and vocabulary knowledge. Moreover, this site offered a good learning environment for shy students who do not dare to speak in public. The majority of the pre-service teachers found that the ASR-based website could create a low anxiety

https://doi.org/10.1017/S0958344010000285 Published online by Cambridge University Press

environment for students to practise speaking, and that the site would also be useful for self-access learning. Teachers also liked the 'Role play' and 'Write and speak' activities which were unique to this website. In addition, the tracking and log function was also considered to be beneficial for teachers to track learner participation and progress. In summary, both students and teachers saw this website as a very useful addition to current programs that promote oral skill development for EFL students, and their overall reaction to the website was positive.

The findings of this study are similar to those of previous studies about users' positive attitudes toward other web-based ASR systems for language learning (Chen, 2006; Chiu *et al*., 2007; Liao, 2009). Students enjoyed using this site mainly because it allowed them to speak, produce more output, and get immediate feedback. This type of interaction and feedback is not found in many websites designed for language learning. Pre-service teachers liked this site because it offered a convenient web-based environment and various exercises for EFL students to practise speaking skills. It also creates a less stressful learning environment for EFL students who do not dare to speak in public.

Although students and teachers showed positive attitudes toward the ASR-based website, they also identified several weaknesses and limitations of this ASR-based website. These included insufficient feedback, the high threshold level of the ASR, poor TTS quality, lack of stress and intonation training, and the fixed patterns of the online exercises. Based on the comments and suggestions summarized in the previous section, it should be clear that developing a high-quality ASR-based language learning system is not easy. In the following sections, possible directions to improve these weaknesses are discussed in detail.

4.2 Possible directions to improve the ASR-based oral skills training site

Most pre-service teachers pointed out that the corrective feedback provided by the ASR-based system was too limited. The system either accepts or rejects students' oral input. Detailed feedback on students' pronunciation problems is not provided. This seems to be a major limitation of this system. However, many existing ASR software programs have similar limitations. O'Brien (2006) pointed out that it is very difficult to pinpoint the subtle problems of individual learners. Chun (2007) indicated that users were not satisfied with the feedback provided by TELL ME MORE, one of the most well-known commercial software programs based on ASR technologies. Chun suggested that a good ASR system should be able to provide more meaningful feedback to learners. Chiu *et al*. (2007) indicated that their web-based ASR program, CandleTalk, could not provide any feedback on pronunciation errors made by EFL learners.

The Microsoft speech recognition engine has great limitations in this respect because of its original design. The Microsoft speech system was designed for English native speakers and not specially designed for pedagogical purposes. Thus, the system cannot pinpoint individual students' pronunciation difficulties and problems.

The second problem noted by teachers and students is that the standard set by the Microsoft speech recognition engine is too high and demanding for some non-native English speakers. The standard set by the Microsoft speech recognition engine is based on the performance of many native speakers. It is thus challenging for non-native

speakers to reach the standard. In fact, how to determine a proper standard has been an important issue in building up any ASR system for second or foreign language learning. Chen (2001) and O'Brien (2006) pointed out that some commercial ASR systems (CNN Interactive, TELL ME MORE, and Microsoft Encarta) actually allowed learners to adjust the sensitivity settings (i.e., the standard of acceptance). There were advantages and disadvantages to this practice. If a student can adjust the settings, then he/she might feel less frustrated. However, it is also likely that students might set the threshold level too low and the system could thus not detect any deviations from the students' speech input.

In fact, there is an ongoing debate amongst researchers about the types of training data necessary to develop an adequate ASR engine for second language learners. Some would choose to use the spoken output of native speakers as the only target model, and some maintain that a "more tolerant" language model should be built upon intelligible second language learner output. For instance, several ASR researchers in Taiwan believe that computers should understand and accept Taiwanese English. In a large project called EAT (English Across Taiwan), several major universities and institutes collaborated to collect more than 100,000 sentences produced by 1,200 Taiwanese students (cf. Tang, 2005; Fan, 2006). It is expected that ASR systems which are developed with non-native spoken data will be more responsive to non-native English speakers. So far, little empirical research has compared user satisfaction with ASR systems which used different sets of "training data". Thus, more empirical evidence is needed to show which standards would be better.

The third problem was the poor quality of the TTS synthesized voices used in some exercises on this site. For instance, the synthesized voices were widely used in the activity 'Write and speak'. In this exercise, students were allowed to submit any sentence they wanted to practise. After listening to the TTS voice, the student would speak the sentence to the system. These default voices provided by the Microsoft Windows XP system were the free 8 kHz voices, and thus they were somewhat unnatural and machine-like. In fact, there is a simple solution to improve the TTS voice quality. Users can purchase the AT&T Natural Voices program and install it on their own computers. This allows the low quality 8 kHz voices to be easily replaced by the AT&T 16 kHz natural voices. The cost of the high quality human-like voices is about 35 US dollars. After installing the 16 kHz voices, users should have rather different opinions about the quality of the synthesized voices.

In addition to the aforementioned problems pointed out by both pre-service teachers and students, the students made some further suggestions. They expected that the interface would be more interesting and that the whole site would be designed like an online game. It is obvious that students like more entertaining learning content. It would be very complex to integrate ASR technologies with a 3D learning environment like Second Life. It would be more feasible to develop a chatbot with ASR capacities. Many existing online chatbots are text-based chatbots, and students cannot interact with the chatbot via their voice input.

Pre-service teachers have also pointed out some additional weaknesses of this ASR-based pronunciation site. They were concerned about the following two issues: lack of solid training on stress and intonation and the limitations of fixed dialogues. Although this website did not specifically target stress and intonation, the native

speakers' pronunciation models in many of the exercises might help students in this regard. O'Brien (2006) in her review article pointed out that most of the existing ASR tools do not handle prosodic elements adequately. Hardison (2005) suggested that speech visualization could be an important method to teach stress and intonation.

With respect to the problems about fixed dialogue patterns, it might be useful for beginning or lower level students to practise these fixed sentences before they produce their own. However, more proficient students should be encouraged to come up with their own sentences. One of the key limitations of many existing ASR systems is that they have difficulties in dealing with multiple sentences. If students produce too many sentences at a time, then the system cannot make a sensible judgment. This is certainly a limitation on a learner's creativity and productivity.

5 Conclusion

Oral communication skills are widely considered to be difficult for many EFL students. EFL teachers often have limited time available to work on pronunciation and speaking skills and EFL students also have few opportunities for using the target language for oral communication outside the classroom. Many existing websites offer conversation scripts and even downloadable audio/video files for teaching conversation. However, students still cannot orally interact with these websites. Very few existing sites supported by ASR technologies are freely available to language learners.

The ASR-based oral skills training website developed in this study can fill this important gap. According to the survey results, both students and teachers responded positively to this website. The EFL students indicated that the convenient ASR-based website encouraged them to produce more output in the target language, helped them improve speaking and listening skills, and enhanced vocabulary knowledge. Pre-service teachers also pointed out that the website provides students with rich learning content and offers students a non-threatening environment for practicing oral skills. The tracking function can also allow students to use the site as a self-access learning tool.

Although the ASR-based website has several useful features for oral training, there is still much room for improvement. First, the system cannot clearly pinpoint the errors of individual speakers. The students can only listen to the models carefully and try again and again. Based on the pre-service teachers' suggestions, a very important future research direction is to improve the feedback mechanism.

Second, to avoid the frustrations reported by college EFL students, perhaps researchers can provide a solution by making the speech recognition system less demanding (e.g. lowering the sensitivity of the speech recognition engine) for lower level learners. More empirical studies, however, are needed to compare the effects of ASR engines which have different evaluation standards on second or foreign language learners.

Third, pre-service teachers indicated clearly that the ASR system should include both segmental and suprasegmental feedback. Students need both types of feedback and more efforts should be made in this domain.

Fourth, feedback from both students and pre-service teachers indicated that the format of the ASR-based exercises is too rigid. The participants expected that students should be allowed to produce their own interlanguage output and the ASR-based system should be able to respond to these utterances. Although this free-conversation

type of ASR system is indeed very attractive, overcoming the difficulties involved in developing this type of interactive system will need more research.

ASR technologies at this stage are still not mature enough to support free conversations between second language learners and computers (O'Brien, 2006). It might take some time before we see significant breakthroughs in artificial intelligence and automatic speech recognition technologies. However, the ASR technologies available now such as the one described in this paper can still be carefully incorporated into various web-based language learning systems to encourage second and foreign language learners to produce more output in the target language and improve pronunciation and oral communication skills.

References

Bernstein, J., Najmi, A. and Ehsani, F. (1999) Subarashii: encounters in Japanese spoken language education. *CALICO Journal*, **16**(3): 361–384.

Chen, H.-J. H. (2001) *Evaluating five speech recognition programs for ESL learners*. Paper presented at the ITMELT 2001 Conference, Hong Kong. http://elc.polyu.edu.hk/conference/papers2001/chen.htm.

Chen, M.-W. (2006) The Impact of Automatic Speech Technology on Contrastive Stress among Adult EFL Learners. Unpublished master's thesis, Da-Yeh University.

Chiu, T., Liou, H. and Yeh, Y. (2007) A study of web-based oral activities enhanced by automatic speech recognition for EFL college learning. *Computer Assisted Language Learning*, **20**(3): 209–233.

Chun, D. M. (2007) Come ride the wave: But where is it taking us? *CALICO Journal*, **24**(2): 239–252.

Egan, K., and LaRocca, S. (2000) Speech recognition in language learning: A must. *Proceedings of InStill 2000*. Dundee: University of Abertay, 4–7.

Ehsani, F. and Knodt, E. (1998) Speech technology in computer-aided language learning: Strengths and limitations of a new CALL paradigm. *Language Learning & Technology*, **2**(1): 45–60.

Eskenazi, M. (1999a) Using automatic speech processing for foreign language pronunciation tutor: Some issues and a prototype. *Language learning and Technology*, **2**(2): 62–76.

Eskenazi, M. (1999b) Using a computer in foreign language pronunciation training: What advantages? *CALICO Journal*, **16**: 447–469.

Fan, T.-Y. (2006) A Design for a Personal Dictionary Inquiry System based on Taiwanese Accented English Speech Recognition. Unpublished master's thesis, National Cheng-Kung University.

Hardison, D. (2005) Contextualized computer-based L2 prosody training: Evaluating the effects of discourse context and video input. *CALICO Journal*, **22**: 175–190.

Harless, W. G., Zier, M. A. and Duncan, R. C. (1999) Virtual dialogues with native speakers: The evaluation of an interactive multimedia method. *CALICO Journal*, **16**: 313–337.

Hincks, R. (2003) Speech technologies for pronunciation feedback and evaluation. *ReCALL*, **15**(1): 3–20.

Holland, V. M., Kaplan, J. D. and Sabol, M. A. (1999) Preliminary tests of language learning in a speech-interactive graphics microworld. *CALICO Journal*, **16**: 339–359.

LaRocca, S. A., Morgan, J. J. and Bellinger, S. M. (1999) On the path to 2X learning: Exploring the possibilities of advanced speech recognition. *CALICO Journal*, **16**: 295–310.

Liao, C.-F. (2009) EFL Learners' Use of Contrastive Stress Supported with Automatic Speech Analysis System. Unpublished master's thesis, Da-Yeh University.

Mostow, J. and Aist, G. (1999) Giving help and praise in a reading tutor with imperfect listening–because automated speech recognition means never being able to say you're certain. *CALICO Journal*, **16**: 407–424.

Neri, A., Cucchiarini, C., and Strik, H. (2001) Effective feedback on L2 pronunciation in ASR-based CALL. *Proceedings of the workshop on Computer Assisted Language Learning, Artificial Intelligence in Education Conference*. San Antonio, Texas, 40–48.

Neri, A., Cucchiarini, C., Strik, H. and Boves, L. (2002) The pedagogy-technology interface in Computer Assisted Pronunciation Training. *Computer Assisted Language Learning*, **15**(5): 441–467.

Neri, A., Cucchiarini, C., and Strik, H. (2003) Automatic Speech Recognition for second language learning: How and why it actually works. *Proceedings of 15th International Congress of Phonetic Sciences*. Barcelona, Spain, 1157–1160.

O'Brien, M. G. (2006) Teaching pronunciation and intonation with computer technology. In: Ducate, L. and Arnold, N. (eds.), *Calling on CALL: From theory and research to new directions in foreign language teaching*. San Marcos,TX: CALICO, 127–148.

Rypa, M. E. and Price, P. (1999) VILTS: a tale of two technologies. *CALICO Journal*, **16**(3): 385–404.

Tang, Shih-Min. (2005) Error Pattern Analysis for Computer Assisted English Pronunciation Learning. Unpublished master's thesis, National Cheng-Kung University.

Wachowicz, K. A. and Scott, B. (1999) Software that listens: It's not a question of whether, it's a question of how. *CALICO Journal*, **16**: 253–276.