Computer Assisted Language Instruction Consortium

CALICO JOURNAL (ONLINE) ISSN 2056–9017

Article

Learners' Feedback Regarding ASR-based Dictation Practice for Pronunciation Learning

Shannon McCrocklin

Abstract

Although early ASR-based dictation programs were criticized for lack of accuracy and explicit feedback for L2 pronunciation practice, teachers and researchers have shown renewed interest. However, little is known about student reactions to ASR-based dictation practice. This qualitative study examines student perspectives, identifying advantages and challenges to working with dictation software and generating ideas for the ideal ASR dictation program. Advanced ESL participants (n=16) worked with Windows Speech Recognition *in a three-week hybrid pronunciation workshop. The study identifies many themes, including advantages such as ease of use, usefulness for pronunciation learning due to feedback provided, and heightened awareness of pronunciation issues, but also disadvantages, such as frustrating levels of recognition, particularly in the first attempt, doubts of the program's transcription abilities, and lack of convenience. Participants reported that convenience and greater support in pronunciation practice would be important for an ideal program.*

KEYWORDS: PRONUNCIATION; SECOND LANGUAGE ACQUISITION; SPEECH RECOGNITION; LEARNER AUTONOMY

Introduction

For the last 20 years, significant attention has been paid to Automatic Speech Recognition (ASR) in language learning. ASR is a technology that analyzes speech captured by a microphone and formulates an output, often written transcription (Levis & Suvorov, 2014). ASR is present in popular language

Affiliation

Southern Illinois University, USA.
email: shannon.mccrocklin@siu.edu

CALICO JOURNAL VOL 36.2 2019 119–137

https://doi.org/10.1558/cj.34738

learning software, such as *Rosetta Stone*, to give students practice speaking the language while providing feedback on pronunciation.

Pronunciation feedback is important because many second language (L2) learners miscategorize sounds in the L2 based on sound categories of their first language and struggle to notice errors in their speech (Blankenship, 1991; Flege, Munro, & Fox, 1993). In order for students to practice successfully on their own, they need feedback (Sheerin, 1997), and immediate feedback is thought to be particularly important for pronunciation (Saito & Lyster, 2012). Further, because practice with technology can be potentially endless, ASR can enable pronunciation feedback during extensive experimentation, which Schwienhorst (2007) argues is necessary for empowering students to become autonomous learners.

Much of the current research has focused on ASR-based Computer-Assisted Pronunciation Training (CAPT) programs. CAPT systems typically lead students through self-paced training with numerous speech samples and offer opportunities for production practice, often having students repeat words or respond to particular prompts (Neri, Mich, Gerosa, & Giuliani, 2008). CAPT researchers have made great strides in improving ASR recognition of non-native speech (Cucciarini & Strik, 2018; Tepperman, 2009) and the quality of feedback provided (Gao, Xie, Cao, & Zhang, 2015; Wang, Qian, & Meng, 2013). Today, for ASR-based CAPT programs that score pronunciation of known utterances, the scores are almost as reliable as human raters (Cincarek, Gruhn, Hacker, Nöth, & Nakamura, 2008; Neri, Cucchiarini, & Strik, 2003). Further, ASR-based CAPT programs have facilitated learning for diverse populations of learners (Hincks, 2003; Cucchiarini & Strik, 2018; Neri et al., 2008; Wang & Young, 2015). Based on not only the validity of ASR scores, but also students' positive reactions, Cordier (2009) calls for ASR-based CAPT to be made available to all students learning foreign languages.

Nevertheless, in order to score pronunciation accurately, dedicated CAPT programs must make use of predetermined practice items. Hincks (2015) describes two possibilities for arranging ASR practice: giving predetermined words/phrases to be read aloud, or providing questions that require certain responses. Using the expected response as a base for interpretation, the program provides the learner with feedback. Although a goal is to include more free, unguided practice opportunities (Hincks, 2015), in current programs the learner must follow the programmed choice of lessons. This limitation means that students have little opportunity to direct their learning and teachers may struggle to embed CAPT into their courses.

ASR-based dictation programs, on the other hand, do not provide analysis of speech; they were designed for native speakers and simply work by transcribing oral speech into a text form. The transcript includes each word that

the program can identify from the speech stream. Early studies criticized dictation programs for low rates of accurate recognition for L2 learners (Coniam, 1999; Derwing, Munro, & Carbonaro, 2000). Strik, Neri, and Cucchiarini (2008) further criticize the limited feedback provided by transcripts. They dismiss such programs as ineffectual for L2 learning because the output should not be considered useable feedback, concluding that learners deserve CAPT that employs ASR for specific and targeted pronunciation feedback.

Dictation programs do provide flexibility, however, as the learner can provide any utterance for the program to transcribe. Self-directed work facilitated by such flexibility may foster greater levels of student autonomy in pronunciation learning (McCrocklin, 2016). In the last few years, researchers have shown a renewed interest in dictation programs and have investigated learning outcomes following dictation practice (Liakin, Cardoso, & Liakina, 2014; McCrocklin, 2019). Liakin et al. (2014) used a pre-/post-test design in which participant productions of the French vowel /y/ were rated for accuracy by Francophone listeners and showed that, when comparing three groups (an ASR-dictation practice group, a non-ASR pronunciation training group, and a control group with no training), only the ASR-dictation group made statistically significant improvements in the French vowel /y/. As part of the larger study this paper reports on, McCrocklin (2019) also examined student improvement using listener ratings of accuracy for several targeted sounds in a pre-/post-test design. Participants in the workshop using ASR-dictation for half of their production practice improved as well as the fully face-to-face (F2F) instruction group, even slightly outperforming the F2F group on most segmentals. Mroz (2018) suggested work with ASR-dictation is useful because students feel that dictation practice replicates measures of human listener intelligibility by providing indications of meaning loss due to mispronunciation. Further, Wallace (2016) suggested that dictation practice may be particularly useful for noticing frequent errors.

Yet, because of the limited research in dictation programs, little is known about how students would react to such practice. In the two studies that have examined student perceptions, students reported perceiving educational value in dictation practice and finding the practice enjoyable (Liakin, Cardoso, & Liakina, 2017; Mroz, 2018). In the author's personal experience, however, many instructors have expressed hesitation to incorporate work with such programs because they worry low recognition levels will be frustrating to students. Levy (2015) calls for careful investigation of student perspectives through qualitative study that can help keep research and practice "aligned and connected" (p. 556). In response, the current study seeks to continue the line of research started by Liakin, Cardoso, and Liakina (2017) using interviews and focus groups to provide richer data and deeper analysis of

student perspectives regarding ASR-based dictation programs. Specifically, the research study seeks to answer the following questions:

1. What do second language learners perceive as the advantages and challenges of working with a dictation program, specifically *Windows Speech Recognition*, for practicing their L2 pronunciation?
2. What changes to the program do participants envision for creating an ideal resource for practicing L2 pronunciation?
3. To what degree do participants appreciate dictation practice for L2 pronunciation and how does their assessment affect plans for future use of the program?

Methods

Participants were introduced to an ASR-based dictation program as part of a three-week pronunciation workshop on several consonants and vowels of English. The workshop included hybrid units conducted half F2F and half using computers. For production practice during computer days, participants used *Windows Speech Recognition* (*WSR*), which is already installed on any PC running *Windows* and does not require access to the internet. If participants did not have a personal PC, they were able to use PC labs on campus or check out laptops for use at home.

Workshop

The study was conducted through a workshop that took place in two advanced ESL courses at a large Midwestern university in the United States. Given Derwing et al.'s (2000) finding that ASR dictation programs had high levels of errors for even advanced learners, high proficiency was considered ideal for this study to lessen the likelihood of overwhelming frustration. The researcher taught both class sections during the workshops using the same materials and lesson plans. The three-week pronunciation workshop covered one vowel and consonant contrast per week, focusing on sounds less common among the world's languages and more likely to be challenging for English learners to acquire: the vowel pairs /ɛ/ vs. /æ/, /ɔ/ vs. /ʌ/, and /i/ vs. /ɪ/ and the consonants /ɹ/, /θ/, /ð/, /ʒ/, and /dʒ/. Although the overall goal of the workshop was to improve intelligibility (Derwing et al.,[2000] show that phonemic accuracy is correlated with intelligibility for human listeners), sounds such as /θ-ð/ were included, despite their low functional load, as they allowed for participants across numerous language backgrounds to share in the same training. Ultimately, a future goal would be to use the advantages of technology to promote individualized instruction that focuses more heavily on higher functional load segments.

Each week, participants began with a F2F lesson followed by a computer workday, giving participants repeated opportunities to practice with *WSR* over the three-week period. Following suggestions from Celce-Murcia, Brinton, and Goodwin (2010) for teaching pronunciation as part of the communicative framework, participants were provided, in the F2F day, with minimal-pair listening practice, information about the manner and place of articulation, and spelling patterns for predicting sounds. Participants then engaged in controlled production (transformation drills, identification of keywords with targeted sounds) and guided production activities (asking partners about preferences, completing information gaps, cooperating in planning tasks). For the computer workday, participants were lead through a listening review utilizing instructor-recorded discrimination tasks followed by focused listening in a TED Talk. Then, participants practiced controlled (minimal pairs and scripted dialogues) and guided/free production (responding to discussion prompts) in *WSR*. Each week of the workshop, participants were provided with a guide sheet for work with *WSR* along with an optional work guide that provided additional ideas for practicing, such as finding a favorite poem to try reciting. At the end of each week, participants submitted a short recording (less than one minute) for instructor grading that focused on the targeted segmentals. Along with a grade for each recording assignment, participants received a printout of the words/phrases with any mispronounced targeted sounds highlighted. Participants submitted all assignments, including a Word document demonstrating *WSR* practice, through the course website.

At the beginning of the workshop, the researcher provided a basic set of instructions for opening and using *WSR*. The directions included a link for an English language pack for *WSR* as well as tips for working with *WSR*, which emphasized that participants did not have to practice endlessly to make the dictation perfect. Instead, one of the tips encouraged participants to try incorrectly transcribed words up to three times and then move on to the next item if recognition failed.

Participants

Sixteen participants were included in the present study. Of the 32 total participants enrolled in the two course sections, only 28 signed informed consent. Of those, only 16 participated to a satisfactory degree, attending/completing five of the six workdays and regularly submitting the *WSR* practice sheets, homework, and language logs. The participants were a mixture of undergraduate (n=12) and graduate (n=4) students. The majority (81%) of participants spoke Chinese as their L1. The group was evenly split between males and females. The average age of participants was 20.6 (range=18–26). The majority (81%) had spent over eight years learning English. Applicants to

the university are required to hold a minimum score of 71 on the TOEFL (IBT) or an overall score of 6.0 on the IELTS for admittance. Fifteen participants also provided a pre-workshop recording that was rated for accuracy on sounds targeted in the workshop; only one student, Chenglei (pseudonym), did not. Audio recordings were rated for segmental accuracy by both the researcher and an independent applied linguist. Although there was a wide range in segmental accuracy in the targeted sounds, the average accuracy was 71.09. This average was compared against the full set of 27 consent-providing students who submitted audio recordings, which was 71.97, suggesting the participants that completed the study were not self-selecting based on language proficiency (did not have a substantially lower or higher accuracy than those who consented but ultimately failed to adequately complete the study). Table 1 provides specific information about the 16 participants, each identified with pseudonyms.

Qualitative Data Collection

To learn about participants' reactions to *WSR* for practicing pronunciation, qualitative methods were employed. Participants provided feedback through answers to open-ended questions on weekly learning logs, individual interviews, and class focus-group discussions. Initially, participants' practice with *WSR*, submitted as *Word* documents in *Moodle*, were also collected. However, this data proved useless as several participants reported editing their documents, removing mistakes in the transcript. Therefore, the practice documents have not been analyzed as they were unreliable indicators of *WSR* practice.

Language Learning Logs

At the end of each week, participants reported their pronunciation work and reflected on their experiences. Participants reported time spent working on their pronunciation, types of activities used, and reactions to the work completed. Participants completed these logs through a quiz feature in *Moodle*.

Interviews

Interviews were considered useful in this study because they allow deeper lines of questioning than surveys (Johnson & Turner, 2003). Participants were asked to take part in an interview outside of class time after the completion of the workshop, and 12 participants completed the interview, which was conducted by a research assistant unfamiliar to participants and audio-recorded. The interviews were semi-structured with 13 primary questions, which elicited responses about participants' experiences in the workshop generally, participants' use of *WSR* for pronunciation work, and participants' reactions to *WSR* as a tool for pronunciation practice.

Table 1
Participant Information

Pseudonym	Gender	Age	L1	Level	Years Learning English	Time in U.S.	Pre-Intervention Targeted Sound Accuracy
Bohai	M	18	Chinese	Undergrad	> 8 yrs	< 6 mo	90.48
Chenglei	M	23	Chinese	Grad	5–6 yrs	< 6 mo	------
Daiyu	F	20	Chinese	Undergrad	> 8 yrs	< 6 mo	79.17
Eona	F	26	Hindi/ Marathi	Grad	> 8 yrs	6 mo–1 yr	84.38
Feng	M	23	Chinese	Grad	> 8 yrs	< 6 mo	69.35
Gan	M	19	Chinese	Undergrad	> 8 yrs	< 6 mo	66.67
Guowei	M	21	Chinese	Undergrad	> 8 yrs	< 6 mo	89.58
Hualing	F	18	Chinese	Undergrad	> 8 yrs	< 6 mo	64.58
Huojin	M	23	Chinese	Grad	> 8 yrs	< 6 mo	77.08
Jeong	F	22	Korean	Undergrad	> 8 yrs	< 6 mo	70.83
Liling	F	18	Chinese	Undergrad	7–8 yrs	< 6 mo	59.38
Liwei	M	20	Chinese	Undergrad	> 8 yrs	< 6 mo	64.58
Najwa	F	21	Malay	Undergrad	> 8 yrs	< 6 mo	73.96
Shoushan	M	20	Chinese	Undergrad	> 8 yrs	< 6 mo	66.72
Xiulan	F	20	Chinese	Undergrad	> 8 yrs	< 6 mo	57.58
Yuming	F	18	Chinese	Undergrad	7–8 yrs	< 6 mo	52.08

Focus Groups

All participants took part in focus groups that occurred during class time four weeks after the completion of the workshop. Focus groups allow participants to react off of one another and, perhaps, be emboldened when hearing similar stories (Madriz, 2000). Participants who had stopped using *WSR* could be hesitant to admit this behavior in a one-on-one interview. Similar to the interviews, research assistants conducted audio-recorded focus groups in order to encourage participants to respond freely. The discussions included eight questions about participants' use of *WSR*, including what participants perceived as the advantages and disadvantages *WSR* offered for pronunciation work, possible improvements envisioned for dictation programs, and whether participants had continued using *WSR* after the workshop ended, examining sources of motivation to continue or stop working with *WSR*.

Analysis

Each of the interviews and focus group discussions were first transcribed verbatim. The open-ended responses from the language logs were copied from *Moodle* and pulled into a data sheet, distinguishing responses by student and week. By far, the interviews and focus group discussions offered the most informative insights. However, all three sources were examined to determine student reactions. Using a general inductive approach, responses were labeled for emerging themes and ideas. The researcher then reevaluated the data in light of emerging themes, coding additional comments according to themes in subsequent data analysis. As recommended by Creswell and Plano Clark (2007) to enhance validity, the researcher used a peer review, asking a colleague to examine a subset of the data looking for themes that the main researcher may have missed. Based on the peer review, the researcher rechecked several emerging categories.

Results

Overall, 81.25% of participants found *WSR* useful for pronunciation practice. While participants gave reasons for their appreciation of *WSR*, they also identified several challenges to using dictation software for practice. Around 75% of participants reported at least one dissatisfaction with *WSR*. To address the first research question regarding what second language learners perceived as the advantages and challenges of working with a dictation program, the researcher examined the transcript data for themes. Numerous advantages and challenges emerged. Table 2 summarizes the identified themes with the numbers of participants reporting each.

Table 2
Summary of Themes: Advantages and Disadvantages to WSR Practice

Advantages	Challenges
• Easy to use (n=5) • Useful for pronunciation practice (n=13) • Provides feedback on pronunciation (n=9) • Heightened awareness of pronunciation weaknesses (n=8)	• Low levels of recognition frustrating (n=12) • First attempt is particularly difficult (n=8) • Participants doubt program (n=7) • Lack of convenience (n=7)

Advantages of Working with *WSR*

Participants used many positive adjectives to describe *WSR*; for example, participants thought the program was "helpful" (Chenglei, Daiyu, and Liling), "useful" (Hualing), "cool" (Daiyu), "interesting" (Liwei) and "fun" (Chenglei). Participants appreciated that the program was easy to use. The most positive review came from Daiyu, who stated, "The software is so good, and it's very useful … easy to use." Guowei elaborated slightly, stating, "It is easy to use. The settings are very clear." It is important to note that, although the program was regarded as easy to use in general, practicing pronunciation with *WSR* was reported as initially difficult or frustrating (see the following section for more detail).

The majority of participants (81%) thought the program was useful for pronunciation practice. Daiyu, who had very positive feelings towards *WSR*, said, "It can record my error directly and, when it recorded right, it inspired me." Similarly, Hualing stated, "I think it's useful, because it could help us to correct our pronunciation by ourselves … I didn't use it before, and now I think it's a useful software, so I feel a little excited."

Participants mentioned feedback as a primary benefit of the program. Eona pointed out, "It can give feedback. I say it, it can give me the information I said." The majority (56%) specifically mentioned feeling that they could use the transcript as feedback on their pronunciation accuracy. One student stated, "Yeah I think it's a really good software, because when I speak to it, if I didn't pronounce very well, it will make mistake, so I need to revise myself." Similarly, Chenglei described making use of the transcription, stating, "Cause it helps you; if you say the words wrong and the *WSR* reads it differently, that means you are saying it wrong, so you know what's the problem with your pronunciation." Similarly, Liwei stated, "It can give feedback. I say it, it can give me the information I said. It's interesting to learn English by." Xiulan appreciated not only that she felt that she got feedback on her pronunciation through *WSR*, but also that she could focus on her pronunciation in a way impossible in conversation. She explained, "When we talk with other people, we cannot focus on our pronunciation, because we need to finish the whole sentence and

make other people understand that … But if we use that program we can, you know, pay more attention to our pronunciation."

Because of the feedback, participants became more aware of issues in their pronunciation. Bohai stated, "Some words when I say, I think I say it correctly, but when I use the software it record what I said. I found, 'oh that's wrong.'" He continued by describing a particular experience in which *WSR* mis-recognized the word, "fill". The mistranscription led him to ask his roommate about the word and discover his pronunciation was closer to "few". While Bohai focused on particular words, some participants mentioned particular sounds they noticed through the training. Five participants discussed discovering issues with a vowel pairing, while one student reported noticing consistent errors with a consonant. However, most participants remarked more generally on discovering information about their pronunciation. Feng stated, "It lets you aware of your pronunciation. You will know your shortcomings, your weaknesses on each part. I think that is what the program gives to me." This awareness sometimes led participants to seek out more help. After Huojin's first attempt with the program was a struggle, he realized he might have pronunciation issues. He stated, "I [think] there is something wrong with my pronunciation. So after that I will find some website that teach you how to pronounce it."

Challenges of Working with *WSR*

While the majority of participants thought practice with *WSR* was useful, many still became frustrated with the perceived low recognition rates. Feng stated, "The Windows Speech Recognition made me crazy. It is hard for it to recognize my accent." Several participants, such as Bohai, Chenglei, and Guowei, echoed the idea that practice "made [them] crazy." Several participants pointed out that they had to try the words/phrases multiple times. Liwei stated, "[At] first, like, I mean I have to pronounce that word like several, it's not several times, like a thousand times!" Only one student, Hualing, reported limiting themselves to two attempts at a pronunciation, while several reported many more, such as Huojin who estimated at least 10 attempts in response to an error.

The perceived low recognition was a particular problem in the first practice attempt with the program. Bohai articulated this common sentiment in the interview, stating, "It's the hardest work I've ever done … I can't do well in assignment the first time, but I can get all the correct answers in the last time." Similar to Bohai, most participants found ways to make the work with *WSR* more successful. Shoushan described this change, "At the beginning, I feel uncomfortable with *WSR*. I tried a lot of times and I … at the beginning I feel it's a waste of time to practice it, but after it helped me correct my pronunciation on some words, I feel more comfortable".

Several participants reported that they began using additional practice strategies in their work with *WSR* in order to have more success. Chenglei, Hualing, and Feng reported that they began using e-dictionaries to look up the pronunciation of words before trying them in *WSR*. Daiyu began saying the words aloud before turning on speech recognition. Huojin mentioned that he began looking up pronunciation lessons online on the targeted sounds before working in *WSR*. Liwei actually described a multi-step process that began with an e-dictionary to hear the words, followed by covert rehearsal, then practice with *WSR*, and, if the program failed to recognize, he went back to the dictionary.

For some participants, however, *WSR* continued to be a source of frustration even in later attempts. Guowei stated, "Sometimes I pronounce, I read some words that people can very quickly understand, but this program is stupid." This frustration led some participants to doubt either themselves or the program. Chenglei described his frustration, stating, "I don't know about the performance other people have with *WSR*, but for me I just don't—I just can't. I just don't know why *WSR* can't figure out what I am saying. Every time I spoke a word, *WSR* always gave me the wrong word." Seven participants (44%) specifically mentioned concerns regarding the program's accuracy. These doubts led Bohai to test the program with native speakers, reporting, "So I asked my roommates, three American native speakers, [to record in *WSR*] and none of them got the correct answer." He lost faith in the program's abilities at that point. Curiously, Bohai and Guowei had the highest scores on their pronunciation diagnostic. This raises interesting questions, such as whether the program was particularly sensitive to their accentedness (both were Chinese speakers) or whether they were frustrated because they felt *WSR* did not match the intelligibility they had achieved with human listeners.

The final main challenge that emerged was that *WSR* lacked convenience. Three participants had their PCs set up for their native language and struggled to set up *WSR* in English. If they did not have access to a friend's computer, they ended up using campus computer labs, which was problematic because they were forced to make noise in a normally quiet lab in order to finish the assignment. Xiulan explained, "sometimes if I use the computer in the lab of the library of our college, it's not very convenient, because I will make the noisy sounds and have an effect on other people so I don't like to use it." Although participants were provided information on how to check out laptops from the university with the program installed, none of the participants reported having taken advantage of this service. Even for participants that were able to use *WSR* on their own private computers, *WSR* was contrasted with applications that could be used on mobile devices.

Suggested Improvements for the Dream Dictation Program

To address the second research question, which focused on changes to the program that participants envisioned for creating a more ideal resource for practicing L2 pronunciation, participants had the opportunity, during the focus groups, to describe modifications they would like to see to improve *WSR*. The main suggestion was improving the level of recognition so that the program was less frustrating to use. However, participants had additional suggestions to add features to make it more appealing and functional as a language learning software: make the program more convenient (n=8), provide language input (n=6), and add additional feedback mechanisms (n=3).

Participants wanted the program to be mobile so it could be used everywhere. Liling stated, "We can only use *WSR* on our computer … and, if we don't take our computers, we cannot use *WSR*. But we bring our cellphones everywhere so it is very useful." Participants often compared *WSR* to e-dictionaries, pointing out that e-dictionaries were much more convenient to use. Liwei explained, "I think e-dictionary is efficient. We can use our cellphones to search for words, and we can use it everywhere." In particular, this convenience was important when participants ran into communication issues in public. Hualing stated, "Usually I speak good, but the pronunciation is not good, so the shopper can't help me. So I usually use the dictionary to correct my pronunciation."

The appreciation of e-dictionaries, however, went beyond simply convenience. The dictionaries also supported ASR practice, providing crucial language input. Eona pointed out, "Sometimes if I face a new word, we will use a dictionary. In dictionary, it can speak—tell me how to pronounce this word." Similarly, Huojin stated, "No matter how many times I say the single word, it comes out another word. So I have to check the dictionaries and listen to the dictionary recording many times, so I can pronounce it right, so the machine can recognize my voice." In addition to valuing searchable recordings provided by e-dictionaries, Najwa suggested adding phonetic spellings, "I think the software would be better if it had like the way to pronounce it correctly or the way that we usually see in the dictionary, yeah the phonetics." Eona recommended a text-to-speech feature stating, "if I can type this as a whole sentence to the program, and the program can help me to pronounce and let me listen to it, I think it will be much more easy to pronounce correctly." Through these responses, it can be gleaned that participants appreciated being able to direct their work, but needed a way to get language input to enhance their practice.

Finally, although participants felt that they could extract feedback from the dictation provided, participants offered ideas regarding adding feedback mechanisms. Najwa indicated that she would also like tips on how to pronounce the sounds, "And also the sound—how to produce it." These tips may

require programs to provide articulatory descriptions and images of mouth positioning. Gan, who had also experimented with *Google Voice Search*, mentioned that it would be helpful if the program provided similar words that the student may have attempted. For example, if the program recognized the word "feet," the program could enable the student to see phonetically related words, such as "fit," to learn more about the difference. However, he was overwhelmed by the length of the list provided by *Google*. He recommended that upon initiating feedback (perhaps by clicking on a word that was transcribed improperly) the program list three words that sound similar. The student could click on the word they were attempting to hear it again or get a lesson on the differences between the sounds.

Participants' Final Conclusions and Plans for Future Use

To address the third research question, which focused on student's overall assessment of the program and possible plans for future use, participants were asked for not only feedback on the value of the program, but also if they had continued use after completion of the workshop or intended to continue use in the future. Overall, many participants saw value in *WSR* pronunciation practice and thought that it had helped them to improve. Feng explained:

> I think maybe it's a struggle at first, but it did improve my pronunciation … You know at start, when I'm talking to my American friends, they say "Oh, you have very nice speaking," but when I actually say something they always say, "Sorry, what did you say?" but recently that phenomena is less than before. It's a good start.

However, participants were less certain about continuing to use *WSR*. For example, Liwei ended up questioning whether the program was useful enough to continue, stating, "*WSR*, it's like kind-of like good … but I don't know that was really helpful for me, so I would rather practice on my own or listen to movies." He later revisited this debate stating, "But I think it's worth it. You know, if the machine can understand you speaking, it's much easier for people to understand you." Liwei ended by reporting that he was still considering using it in the future.

Several were considering working with another ASR program. Actually, three participants had already explored additional programs they preferred. While Feng appreciated Apple's *Siri*, Gan liked *Google Voice Search*. The student with the strongest dislike of *WSR*, Guowei, turned to *Dragon Dictation*. He explained, "Actually, I hate this program, I mean *WSR*. Sometimes I pronounce, I read some words that people can very quickly understand, but this program is stupid. I dunno the name—*Dragon Dictation*? I think this is very helpful." Only one student thought she would continue working with *WSR* specifically.

Half of the participants, however, reported in the focus group discussions no intention to continue working with an ASR program. Huojin stated, "I will continue to practice my pronunciation but maybe in different ways." Of those, six (75%) indicated that they were unlikely to continue because they were too busy. Jeong explained that she had not used *WSR* since completion of the workshop, stating, "I don't have enough time to practice my pronunciation."

Discussion

This study showed that despite finding ASR dictation practice challenging, the majority (81%) of participants saw dictation practice as useful for improving their pronunciation. Similar to Liakin et al. (2017), this study found that students deem dictation programs easy-to-use (even if the practice itself is challenging), but this study also highlighted the value of dictation practice for noticing errors, not only within individual words, but as patterns across words, which supports Wallace's (2016) argument. Participants began to see the program as a way to check their pronunciation, which suggests that it may be particularly useful as a form of formative assessment. Participants were also motivated to seek out additional resources and employ additional strategies to find ways to improve, checking their improvement through the dictation program.

These results, taken with Liakin et al. (2014) and McCrocklin (2019) which showed that students could improve their production of segmentals using ASR-based dictation equally well or better than from face-to-face instruction, show potential for ASR dictation practice to enable useful pronunciation practice. In particular, participants appreciated receiving feedback through the dictation transcript. This is a particularly noteworthy finding, considering that Strik et al. (2008) dismiss dictation programs as inappropriate for pronunciation learning partially because they do not provide what the authors consider appropriate feedback. On the contrary, participants felt that they were able to easily make use of the transcript as feedback to identify mispronounced words, and occasionally even specific sounds, which is important for enabling autonomous learning in pronunciation (Sheerin, 1997; Saito & Lyster, 2012).

However, given the frustrating levels of recognition, participants likely received inaccurate negative feedback through frequent transcription errors, a concern mentioned in early research by Derwing et al. (2000). Ideally, the transcript would show errors where a human might also fail to recognize speech (i.e. in places of lost intelligibility), but in Derwing et al. (2000) *Dragon Dictation* recognition rates were not correlated with human intelligibility scores. Concerns about accuracy were raised both by participants in this study and in Liakin et al. (2017). These concerns highlight the need to revisit accuracy rates for non-native speech in popular speech recognition programs. While

the transcription errors pushed participants to continue learning and attempting the utterance, inaccurate feedback could distract participants from real pronunciation issues affecting intelligibility and could demotivate them to continue.

Finally, in addition to highlighting the importance of convenience, this study found that, similar to Liakin et al. (2017), participants described that ideally they would like to be able access additional feedback mechanisms. Participants in this study suggested providing access to language input and adding feedback that can be activated when needed, such as lists of similar words and tips on how to make different sounds. Their recommendations in many ways point to possibilities for enacting the future envisioned by Hincks (2015) as CAPT begins to provide student-directed learning in which students' speech is unconstrained.

Implications for Pronunciation Teachers

Given that participants found *WSR* useful but frustrating, teachers should carefully consider the needs of their students in deciding whether to introduce ASR-dictation practice. While dictation may be particularly useful for advanced students seeking more self-directed learning, CAPT offers structured lessons and explicit feedback that is likely to be essential for beginners and children. If flexibility is a goal, ASR dictation may be a possible solution, but teachers should carefully think about which dictation programs to employ. While *WSR* was chosen because the researcher could guarantee access, teachers should explore a range of programs that may better meet their students' needs. Teachers could explore *Google Voice* with documents in *Drive*. Mroz (2018) has suggested that students did not report high levels of frustration with *Google*'s speech recognition and suggested it may be due to higher accuracy rates. Further, *Siri* for the iPhone and *Cortana*, available for Android, offer dictation capabilities with conversational interaction for mobile devices.

If implementing ASR-dictation practice, teachers should consider making guides to facilitate practice and be available should students need help. Given that the first attempt is the most frustrating, it might be useful to lead a class discussion after the first practice allowing students to voice their frustrations, but also providing tips for working with the program, such as strategies, like e-dictionary use, that may make practice more successful. Teachers could also consider creating practice guides that hyperlink to dictionary entries with pronunciation recordings or recommend text-to-speech applications to help students prepare longer stretches of speech. As students become more comfortable with dictation practice, guiding students in choosing their own materials could engage learner interest in the tool and support greater learner autonomy.

Implications for ASR- based CAPT Designers/Researchers

While there seems to be some potential in using dictation programs for pronunciation practice, more research is needed. In what ways has the accuracy of dictation programs changed since early research? How does the transcript compare to feedback students could obtain from native speakers? What is the relationship between student reactions to ASR work and their accentedness or intelligibility levels? Bohai and Guowei raise questions about which students are likely to get more frustrated by the work. Also, how do students work with the programs? Students mentioned developing strategies, but a detailed accounting of the actual practice would be useful for understanding behaviors of successful and unsuccessful users. The transcripts of work submitted in this study proved a significant limitation as students had edited them during practice. In the future, researchers could consider screen capture to avoid this problem.

As developers continue to work on CAPT programs that allow open, unguided speech, researchers will need to think carefully about ways of providing support without limiting student choice and autonomy. As students worked to describe the ideal dictation program, many pointed to features that already exist, such as mobile applications which already exist through *Dragon*. However, the program that many of these students envisioned with built in dictionary features and feedback while still allowing current levels of flexibility does not exist.

Participants indicated that they needed greater support from accessible language input while working with dictation programs. Building a program that allows students to not only look up words in the dictionary or use text-to-speech functions, but also to quickly attempt that same word and get feedback from ASR may provide a useful tool. Students also wanted to be able to access additional feedback, such as pronunciation tips, related words, or facial diagrams. Although Hincks (2015) mentions the importance of language input and feedback, students indicate that they also want the resource to be searchable and controllable.

Conclusion

This qualitative study examining student reactions to an ASR-based dictation program has highlighted potential usefulness of the program, but also identified several problematic issues in implementation that need to be addressed. For teachers, the results indicate participants appreciated the practice because they valued feedback that could help them improve. However, the frustration voiced by participants suggests that teachers should seek ways to better support the practice so it can be more successful, such as carefully considering ways to guide students in practice and help students make sense of feedback they receive through the transcript output. For researchers and designers,

participants appreciated the flexibility of dictation and felt that they could make use of the transcript for feedback. However, participants reported desiring greater convenience as well as more support, in particular more language input, as part of dictation practice.

About the Author

Shannon McCrocklin is an Assistant Professor in the Department of Linguistics at Southern Illinois University. Her research focuses on second language pronunciation learning and teaching, including computer-assisted pronunciation training.

References

Blankenship, B. (1991). Second language vowel perception. *Journal of the Acoustical Society of America*, *90*, 2252–2252. https://doi.org/10.1121/1.401514

Celce-Murcia, M., Brinton, D., & Goodwin, J. (2010). *Teaching pronunciation* (2nd ed.). Cambridge, England: Cambridge University Press.

Cincarek, T., Gruhn, R., Hacker, C., Nöth, E., & Nakamura, S. (2008). Automatic pronunciation scoring of words and sentences independent from the non-native's first language. *Computer Speech and Language, 23*, 65–88. https://doi.org/10.1016/j.csl.2008.03.001

Coniam, D. (1999). Voice recognition software accuracy with second language speakers of English. *System*, *27*, 49–64. https://doi.org/10.1016/S0346-251X(98)00049-9

Cordier, D. (2009). *Speech recognition software for language learning: Toward an evaluation of validity and student perceptions* (Doctoral dissertation). http://scholarcommons.usf.edu.

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage Publications.

Cucchiarini, C., & Strik, H. (2018). Automatic Speech Recognition. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (pp. 556–569). New York, NY: Routledge.

Derwing, T., Munro, M., & Carbonaro, M. (2000). Does popular speech recognition software work with ESL speech? *TESOL Quarterly, 34*(3), 592–603. https://doi.org/10.2307/3587748

Flege, J. E., Munro, M. J., & Fox, R. A. (1993). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America*, *95*, 3623–3641. https://doi.org/10.1121/1.409931

Gao, Y., Xie, Y., Cao, W., & Zhang, J. (2015). A study on robust detection of pronunciation erroneous tendency based on deep neural network. *Proceedings from INTERSPEECH 2015*, *Dresden, Germany, 693–696.* Available from https://www.isca-speech.org/archive/interspeech_2015.

Hincks, R. (2003). Speech technologies for pronunciation feedback and evaluation. *ReCALL. 15*(1), 3–20. https://doi.org/10.1017/S0958344003000211

Hincks, R. (2015). Technology and leaning pronunciation. In M. Reed & J. Levis (Eds), *The handbook of English pronunciation* (pp. 505–519). Malden, MA: John Wiley & Sons.

Johnson, B., & Turner, L. A. (2003). Data collection strategies in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 297–319). Thousand Oaks, CA: Sage Publications.

Levis, J., & Suvorov, R. (2014). Automated speech recognition. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. http://onlinelibrary.wiley.com/.

Levy, M. (2015). The role of qualitative approaches to research in CALL contexts: Closing in on the learner's experience. *CALICO Journal*, *32*(2), 554–568. https://doi.org/10.1558/cj.v32i3.26620

Liakin, D., Cardoso, W., & Liakina, N. (2014). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal, 32*(1), 1–25. https://doi.org/10.1558/cj.v32i1.25962

Liakin, D., Cardoso, W., & Liakina, N. (2017). Mobilizing instruction in a second-language context: Perceptions of two speech technologies. *Languages*, *2*(3), 1–21. https://doi.org/10.3390/languages2030011

Madriz, E. (2000). Focus groups in feminist research. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed.) (pp. 835–850). Thousand Oaks, CA: Sage Publications.

McCrocklin, S. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System, 57*, 25–42. https://doi.org/10.1016/j.system.2015.12.013

McCrocklin, S. (2019). ASR-based dictation practice for second language pronunciation improvement. *Journal of Second Language Pronunciation, 5*(1), 98–118.

Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals, 51*(3), 1–21. https://doi.org/10.1111/flan.12348

Neri, A., Cucchiarini, C., & Strik H. (2003). Automatic speech recognition for second language learning: How and why it actually works. *Proceedings from the 15th ICPhS*, Barcelona, Spain, 1157–1160.

Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning*, *21(5)*, 393–408. https://doi.org/10.1080/09588220802447651

Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning*, *62*(2), 595–633. https://doi.org/10.1111/j.1467-9922.2011.00639.x

Schwienhorst, K. (2008). *Learner autonomy and CALL environments*. New York, NY: Routledge.

Sheerin, S. (1997). An exploration of the relationship between self-access and independent learning. In P. Benson & P. Voller (Eds.), *Autonomy and independence in language learning* (pp. 54–65). London, England: Longman.

Strik, H., Neri, A., & Cucchiarini, C. (2008). Speech technology for language tutoring. *Proceedings of Language and Speech Technology (LangTech '08) Conference,* Rome, Italy, 73–76.

Tepperman, J. (2009). *Hierarchical methods in automatic pronunciation evaluation.* (Doctoral dissertation). Ann Arbor, MI: UMI Dissertation Services.

Wallace, L. (2016). Using Google Web Speech as a springboard for identifying personal pronunciation problems. *Proceedings of the 7th Annual Pronunciation in Second Language Learning and Teaching Conference*. Retrieved from https://apling.engl.iastate.edu/alt-content/uploads/2016/08/PSLLT7_July29_2016_B.pdf.

Wang, H., Qian, W., & Meng, H. (2013). Predicting gradation of L2 English mispronunciations using crowdsourced ratings and phonological rules. *Proceedings from Speech and Language Technology in Education 2013. Grenoble, France, 1–5.*

Wang, Y. H., & Young, S. S. C. (2015). Effectiveness of feedback for enhancing English pronunciation in an ASR-based CALL system. *Journal of Computer Assisted Learning, 31*, 493–504. https://doi.org/10.1111/jcal.12079