McCrocklin, S. (2015). Automatic speech recognition: Making it work for your pronunciation class. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 126-133). Ames, IA: Iowa State University.

AUTOMATIC SPEECH RECOGNITION: MAKING IT WORK FOR YOUR PRONUNCIATION CLASS

Shannon McCrocklin, University of Texas Rio Grande Valley

Automatic Speech Recognition (ASR), the technology behind language learning technology such as *Rosetta Stone* and *Burlington English*, is also available through many dictation programs freely accessible on the devices that students are likely to already own or have access to. Such ASR programs empower students to work on their pronunciation on their own, getting feedback based on the transcription provided by the program (McCrocklin, 2014). This paper introduces the potential benefits of using ASR in the classroom (such as fostering learner autonomy and supplementing course work), explores a few of the programs/technologies available (Siri, Google Voice Search, and Windows Speech Recognition), provides ideas for utilizing ASR as part of a pronunciation class (such as ideas for guiding student work), and addresses challenges that could potentially develop when using ASR programs (such as student frustration).

INTRODUCTION

Automatic Speech Recognition (ASR), the technology behind language learning technology such as *Rosetta Stone* and *Burlington English*, is also available through many dictation programs freely accessible on the devices that students are likely to already own or have access to. Such ASR programs empower students to work on their pronunciation on their own, getting feedback based on the transcription provided by the program. This paper introduces the potential benefits of using ASR in the classroom, explores a few of the programs/technologies available, provides ideas for utilizing ASR as part of a pronunciation class, and addresses challenges that could potentially develop when using ASR programs.

Background Information

Students learning a second language often recognize a need or desire to work on their pronunciation. Many students do not know how to work on their pronunciation outside of the language classroom, however. Unfortunately for these students, pronunciation is often treated as the "Cinderella" of language teaching (Kelly, 1969, p. 87), downgraded as a teaching goal and often pushed aside in favor of other skills (Isaacs, 2009; Lang, Wang, Shen, & Wang, 2012). As teachers, we want to help empower our students to practice and improve their pronunciation on their own, but few strategies and tools are readily available for students to use and we struggle to help our students become autonomous in their language learning.

One of the causes of student dependence on instructors in pronunciation learning is likely the classroom experience itself. Pronunciation instruction is often heavily led by the instructor. Many pronunciation classroom activities still rely on the teacher to model correct pronunciation and to monitor, evaluate, and give feedback on student production. Pronunciation teachers also often rely on drills or controlled production activities, giving students little room for free expression or communicative practice. These types of pronunciation classes seem unlikely to help students practice outside of class because students are not encouraged to develop skills or strategies for monitoring or evaluating their own pronunciation and are given very little room for free experimentation with their pronunciation. Of the ten main teaching techniques introduced by Celce-Murcia, Brinton, and Goodwin (2010) for teaching pronunciation as part of the Communicative Approach (p. 9-10), only one, "recording of learner's production" even mentions making use of the student in the evaluation. Admittedly, it is quite difficult for students to monitor their own production. In order to monitor and correct pronunciation accuracy, students must be able to hear when they make mistakes, which requires students to create aural discrimination categories appropriate to the L2, while research has indicated that sounds in an L2 are filtered through the phonological system of the first language (Beddor & Strange, 1982; Blankenship, 1991; Flege, Munro, & Fox, 1993).

Yet feedback is vital to the success of autonomous learning outside of the classroom (Sheerin, 1997). One technology that can help provide feedback is Automatic Speech Recognition (ASR), which allows students to experiment with the language in a safe, private setting. "[ASR] is an independent, machine-based process of decoding and transcribing oral speech. A typical ASR system receives acoustic input from the speaker through a microphone, analyzes it using some pattern, model or algorithm, and produces an output, usually in the form of a text" (Levis & Suvorov, 2014, p. 1). Many people have had experience with an ASR program, either through automated telephone lines, Siri on the iPhone, or speech dictation programs. ASR is also built into many language learning programs such as *Rosetta Stone* (2013), or *Burlington English* (2014). When used for pronunciation training, ASR is a tool that empowers students to practice at their own speed, getting feedback from the words recognized. Students that are trained to use ASR for their pronunciation practice have heightened self-efficacy, beliefs in their abilities to learn autonomously, and may engage in more autonomous learning behaviors (McCrocklin, 2014).

This paper explores the use of dictation technologies for pronunciation learning. While many students may enjoy learning from a dedicated language learning software, such as Rosetta Stone, such programs are often cost prohibitive and do not allow for flexibility in content. In order to make the pronunciation practice more accessible financially, this paper explores a few of the ASR technologies for pronunciation practice that are freely available (on various devices) for a wide range of target languages.

AUTOMATIC SPEECH RECOGNITION TECHNOLOGIES

Siri

Siri is described by Apple as an "intelligent personal assistant" available on the iPhone (Apple Support, 2014). It not only allows you to dictate messages, but also understands commands to complete actions. People can feel as if they are talking to Siri because the program can answer questions and will ask a question in return if the program does not understand a command or request. To begin using Siri on the iPhone, press and hold the home button (the physical button at the bottom of the screen). Once you hear two quick beeps, Siri is ready to listen to commands or questions. Siri can understand and speak the following languages: Cantonese, English, French (France, Canada), German, Italian, Japanese, Korean, Mandarin, Spanish (Mexico, Spain) (Apple

Support, 2014). Any language listed is available in any country in which you may be residing. To change languages for Siri, the user must enable a new language in the settings on the iPhone (Apple Support, 2014).

Benefit: Talking to Siri can feel like a conversation and therefore can feel more natural than dictating into other programs.

Drawback: Work with Siri may be more difficult to submit as homework, but students could dictate emails to the instructor through Siri.

Recognition Level: Due to its fairly high levels of recognition, Siri may work best for lower and intermediate learners.

Google Voice Search

Google's Voice Search is available on mobile devices as well as PCs and Macs in over 50 different languages (Chowdhry, 2014; Moon, 2014). Some of the languages available are: English, French, German, Italian, Japanese, Korean, Russian, Spanish, and Brazilian Portuguese (Google Support, 2014). One exciting new feature of the Google Voice Search application is the ability to update the settings to allow the program to detect and dictate in five different languages at a time, instead of changing the settings each time the user wanted to switch languages (Chowdhry, 2014). To utilize Google Voice Search without the application, simply go to www.google.com. To the right of the search bar, you will see an icon of a microphone. Click on the microphone icon and enable the microphone to get started with voice searches.

Benefit: Google voice search is particularly good with short utterances (one or two words) which could be useful for working with minimal pairs.

Drawback: As part of a class, work with Google Voice Search may be more difficult to submit as homework. Students would likely have to copy each of the search results individually into another document.

Recognition Level: Due to its fairly high levels of recognition, Google Voice Search may work best for lower and intermediate learners.

Windows Speech Recognition

Windows Speech Recognition (WSR) is an automatic speech recognition program that is already installed on PCs that use the Windows operating system (Microsoft, 2014). Once WSR is opened, it can be used to dictate into other programs, such as Microsoft Word. To get started with Windows Speech Recognition, click the start button and search for "windows speech recognition". Once completing a few start-up pages, the program will become available as a tab at the top of the screen (see Figure 1). Press the microphone icon to turn on the listening function. WSR is available in the following languages: Catalan, Chinese (China, Hong Kong, Taiwan), Danish, Dutch, English (Australian, Canadian, British, U.S., Indian), Finnish, French, German, Italian, Japanese, Korean, Norwegian, Polish, Portuguese (Brazil, Portugal), Russian, Spanish (Spain, Mexico), Swedish (Microsoft Developer Network, 2014). Upon purchase of a new computer, Windows (and therefore the speech recognition) is only able to operate in a single pre-installed language. To use another language, the user must download a language recognition pack for each language desired (to search for the language pack you are interested in, see http://www.microsoft.com/en-us/download/default.aspx).

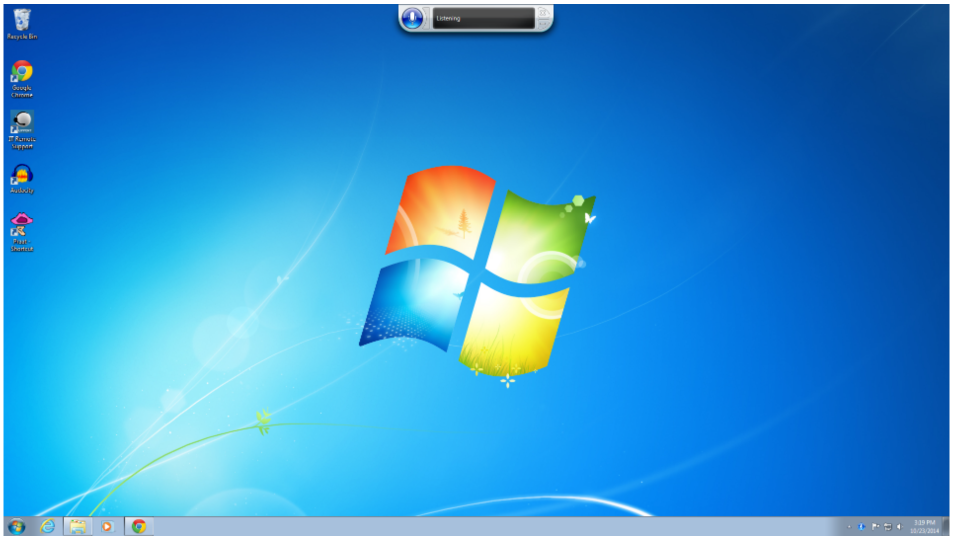

Figure 3. Image of PC computer screen with Windows Speech Recognition tab

Benefit: Windows Speech Recognition makes homework submission easier as students can dictate into a Word document and submit the file.

Drawback: Windows Speech Recognition relies on context to help guess words uttered and, as such, has a much harder time with single words or minimal pair work. Words will be best understood if embedded into sentences for grammatical context.

Recognition: Windows Speech Recognition is fairly sensitive to deviations in speech patterns and therefore offers lower rates of recognition. Intermediate learners are likely to find the program challenging. The program is best suited for intermediate-advanced learners, although with voice trainings with the program (right click on the tab at the top of the screen while the program is running, click “Configuration”, and click “Improve voice recognition”) students could make the program more accommodating.

INCORPORATING ASR IN THE CLASSROOM

While there is some potential to use ASR for suprasegmentals (perhaps looking at two syllable noun and verb pairs, such as “to exPORT” versus “an EXport” with stress shifts), ASR is primarily useful for segmentals because students receive feedback on sounds from the spellings in the dictation. ASR work will also be particularly useful as a supplement and follow-up to classroom work time. Students should be introduced to differences in sounds and be given practice in properly producing sounds before working with ASR so that when students run into difficulties with words or sounds (i.e. the program mis-transcribing speech) the students can use information and tips from the class sessions to continue practicing and improving their pronunciation.

Work with ASR is easily incorporated through homework assignments when using Siri or Windows Speech Recognition. Students can dictate into a Word document, submitting the file through a course management system, or email, emailing homework directly to the instructor. Work with ASR may also be potentially used during class time if the teacher has a computer lab available. One issue to be aware of, however, is that the microphones in computers and smartphones are often sensitive to ambient sounds and may struggle to function properly if background sounds, such as other students simultaneously working with ASR, are present. ASR is likely to work best in a lab with full or partial booths that will block some of the noise from travelling and high quality microphones, such as those designed for call center use.

When working with ASR, students understand relatively quickly and easily how the transcription provided from ASR can be used for feedback. The larger issue with working with ASR is that it can be frustrating to use, even for native and highly proficient speakers. In my own teaching experiences, students want to try the words and phrases until they get every single word transcribed perfectly, but sometimes this is simply not possible. It is important to help your students understand that the technology is not perfect and that they do not have to get every word and sentence transcribed perfectly. I tell my own students that the program is useful because it can help them identify areas to work on and provides feedback through the transcription. Then, I recommend to my students that they try saying a particular word up to three times, but if, after three tries, the program still has not recognized the word, they should move on. Some other tips that might be useful:

- If you are working with minimal pairs, ask students to focus only on the targeted sound. For example, if the target sound is /i/ in "beet" and they are able to get "be" they have accomplished the goal.
- Similarly, ask students to focus only on the targeted word(s) in a provided sentence.

Discussing the limitations of the program can help students set realistic goals for working with the program. Your role as a teacher using ASR in the classroom may also need to change as you help motivate students to push through the struggles and celebrate with them when they achieve successes.

Ideas for Practice Activities

To begin working with an ASR program, students simply need to have something to say. Teachers can provide guide sheets or materials for practice or students can find or develop their own materials. For teachers designing guide sheets, consider using ASR to follow-up on the sounds introduced in class. You can have lists of minimal pairs (beet-bit, green-grin, etc.) that students have to read through or sentences with words using the target sounds. It may be useful in these activities to highlight or bold the targeted words so that students can easily see how the ASR practice lines up with the practice done in class. The guide sheet could also ask questions that allow students to freely answer; for example, it could ask students to describe what they did over the weekend or to describe what they see in a picture. Moving from the minimal pairs which are a controlled activity to free responses will allow students to ease into practice with ASR while still focusing on their pronunciation accuracy on targeted sounds.

Students could also bring in their own practice with ASR. You could ask students to find a favorite poem or famous speech that they could read to the dictation program. Students could also prepare a presentation for your class (or another class) and practice the presentation with ASR. One of the great advantages of dictation ASR programs, is that any content can be brought in for practice and teachers and students are only limited by their creativity in direction and content of practice.

Challenges

In order for Automatic Speech Recognition practice to be successfully integrated into a course, teachers should be aware that there are challenges to using ASR. First, students may have trouble getting access to speech recognition technology. Students may not have access to a smartphone or they may have a Mac when you were hoping to use WSR on a PC. One way of approaching this is to be flexible, allowing some students to submit work through email with Siri and others to use WSR on PCs. Students could also be required to use one type of technology if the teacher ensures that all students can gain access to the same technology through school resources. For example, if the school has a PC computer lab or computers to check out, WSR is fairly easy to access for any student. Please note, however, that, if students must use a computer lab, they may feel uncomfortable speaking to the computer when there are other people working in the lab quietly on other projects. Finally, with a smartphone, PC, or Mac a student would be able to access Google Voice and may be a great way of making the technology available to all students (although Google Voice Search does not allow students to easily capture and demonstrate completed practice which may make homework submission more difficult).

Students are also likely to get frustrated with the technology for being too sensitive or not catching enough errors to provide helpful feedback. When using a more sensitive program such as WSR in my own classes, this frustration led students to doubt the program or even doubt themselves. Some students in my classes took their computers to their roommates or friends that were native speakers and asked them to dictate to the computer. When the computer failed to transcribe perfectly, my students became doubtful of the program's ability to help them practice their pronunciation. On the other hand, some of my students began to doubt themselves. Some indicated that they began to doubt their pronunciation abilities after using the program because the program was unable to transcribe their speech accurately. While it is useful to have a program show students where they are making pronunciation errors, the great amount of negative feedback can be overwhelming, particularly in the first practice with the program. I found my role as a teacher become much more about encouragement and acknowledgment that I was asking my students to complete a task which was frustrating. As students pushed through with repeated uses of the technology, however, they became more comfortable using the programs and saw the benefit of practicing with ASR, noting it helped them recognize their pronunciation problems and allowed them to practice repeatedly with items until they could get it right.

One final challenge is that students may cheat and there is not necessarily any way for teachers to know if students typed in their pronunciation work. While this is a possibility, it is fairly easy to tell when students type a dictation in because the submission of work is perfect, all words are identified perfectly and there are no errors, which is unlikely if students actually use an ASR program. In my own teaching experiences, students were more likely to type answers when they were confused about how to use the program or were frustrated by the low rate of recognition at the beginning of practice. Taking time in class to solve technology issues at the beginning of the semester and encouraging students to visit office hours if they continue to struggle with the program will be useful in helping students overcome issues related to the technology itself. Accepting and praising practice work that has transcription errors can also help students understand that the value of the practice lies more in doing the practice than in showing a perfect transcription at the end.

CONCLUSION

Automatic Speech Recognition can be a powerful tool for empowering pronunciation students to practice and get feedback on their pronunciation outside of class (McCrocklin, 2014). Although there may be challenges to re-envisioning and re-directing a dictation program for pronunciation purposes, the advantages of the program use make it worth trying. While this paper explored three main ASR technologies that cover a wide range of languages, there are likely many more reasonably priced or free options for the specific language of interest. Through exploring your options for integrating ASR into your classroom, you can provide students a tool that will allow them to work on their pronunciation outside of class, supplementing your in-class teaching of pronunciation and enabling students to work on their pronunciation autonomously.

ABOUT THE AUTHOR

Shannon McCrocklin is an assistant professor at the University of Texas-Rio Grande Valley. She completed her Ph.D. in Applied Linguistics and Technology at Iowa State University. She holds an M.A. in Teaching English as a Second Language from the University of Illinois at Urbana-Champaign where she developed an interest in pronunciation teaching and applied phonetics and phonology. Her research focuses on improving pronunciation training for students and CAPT (Computer-Assisted Pronunciation Teaching). She has presented at CALICO, NCTE, PSLLT, and AAAL. shannon.mccrocklin@utrgv.edu

REFERENCES

Apple Support. (2014). About Siri. Retrieved from http://support.apple.com/kb/ht4992.

Beddor, P.S. & Strange, W. (1982). Cross-language study of perception of the oral-nasal distinction. *Journal of the Acoustical Society of America*, *71*, 1551-1561.

Blankenship, B. (1991). Second language vowel perception. *Journal of the Acoustical Society of America*, *90*, 2252-2252.

Burlington English. (2014). Burlington English: Meeting the challenges of adult language acquisition. Retrieved from: http://www.burlingtonenglish.com/about.aspx.

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

Chowdhry, A. (2014). Google's Voice Search For Android Now Recognizes 5 Different Languages At Once. *Forbes*. Retrieved from http://www.forbes.com/sites/amitchowdhry/2014/08/22/googles-voice-search-for-android-now-recognizes-5-different-languages-at-once/.

Flege, J.E., Munro, M.J., & Fox, R.A. (1993). Auditory and categorical affects on cross-language vowel perception. *Journal of the Acoustical Society of America*, *95*, 3623-3641.

Google Support. (2014). Ok Google and voice search. Retrieved from

https://support.google.com/websearch/answer/2940021?hl=en

Isaacs, T. (2009). Integrating form and meaning in L2 pronunciation instruction. *TESL Canada Journal*, *27 (1),* 1-12.

Kelly, L.G. (1969). *25 centuries of language teaching: an inquiry into the science, art, and development of language teaching methodology, 500 BC-1969.* Rowley, Massachusetts: Newbury House Publishers.

Lang, Y., Wang, L., Shen, L., & Wang, Y. (2012). An integrated approach to the teaching and learning of zh*. Electronic Journal of Foreign Language Teaching 9(2*), 215-232.

Levis, J. & Suvorov, R. (2014). Automated speech recognition. In C. Chapelle (Ed.) *The encyclopidia of applied linguistics*. Retrieved from http://onlinelibrary.wiley.com/store/10.1002/9781405198431.wbeal0066/asset/wbeal0066.pdf?v=1&t=htq1z7hp&s=139a3d9f48261a7218270113d3833da39a187e74.

McCrocklin, S. (2014). *The potential of Automatic Speech Recognition for fostering pronunciation learners' autonomy* (Doctoral dissertation). Retrieved from http://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=4909&context=etd.

Microsoft. (2014). Microsoft accessibility: Products: Windows Speech Recognition. Retrieved from http://www.microsoft.com/enable/products/windows8/default.aspx.

Microsoft Developer Network. (2014) Language support. Retrieved from http://msdn.microsoft.com/en-us/library/hh378476(v=office.14).aspx.

Moon, M. (2014) Google Voice Search can now handle multiple languages with ease. *Engadget*. Retrieved from http://www.engadget.com/2014/08/15/google-voice-search-multi-language-default/.

Rosetta Stone. (2013). Cutting-edge technology: The software that started it all. Retrieved from http://www.rosettastone.com/features#link1.

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