SALUKISPEECH: INTEGRATING A NEW ASR TOOL INTO STUDENTS' ENGLISH PRONUNCIATION PRACTICE

Shannon McCrocklin, Southern Illinois University

Claire Fettig, Southern Illinois University

Simon Markus, Southern Illinois University

Recent research has shown that use of Automatic Speech Recognition (ASR)-based dictation for L2 pronunciation practice provides several benefits, including student improvement of segmental accuracy, noticing of probable errors at the word and sound level, and increased attention to segments with recurring errors. Further, the transcript can support greater learner autonomy by providing feedback and a sense of human intelligibility, which can be motivating for students. This paper introduces a new ASR-based tool that provides flexible, student-driven practice while also providing support for segmental learning. By providing communicative practice, the website offers learners the chance to practice English with guided, communicative learning activities while ASR provides implicit feedback in the form of a transcript. When students notice errors in the transcript, the system compares the intended word to the transcribed word to lead students to an appropriate segmental lesson. At the end of a practice session, students can email a summary report of their practice to their instructor. This paper introduces this new tool and discusses how it could be used for student pronunciation practice outside of class.

Cite as: McCrocklin, S., Fettig, C, & Markus, S. (2022). Salukispeech: integrating a new ASR tool into students' English pronunciation practice. In J. Levis & A. Guskaroska (eds.), *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference*, held June 2021 virtually at Brock University, St. Catharines, ON. https://doi.org/10.31274/psllt.13342

INTRODUCTION

Recent research has shown that Automatic Speech Recognition (ASR)-based dictation practice can support second language pronunciation learning. This paper introduces a flexible, student-driven tool, SalukiSpeech.com, which integrates ASR-dictation practice to support English L2 pronunciation practice.

Automatic Speech Recognition

Automatic Speech Recognition (ASR) is a "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). ASR is a key component for several technologies, including digital personal assistants, such as Apple's Siri, and language learning applications like Duolingo. ASR has quickly become popular in language learning applications as it can provide individualized feedback for learners on their pronunciation (for examples, see Duolingo, Rosetta Stone, and Blue

Canoe [Daniels & Taylor, 2021]). However, in many of these programs, the learners must follow a plan of study provided by the program, practicing only pre-determined words and phrases (Hincks, 2015) in order for the program to be able to assess the pronunciation and give feedback. The lack of communicative practice and choice to direct learning may limit learning achievement and learner autonomy.

However, there is another option for language learning. Studies have shown a range of benefits to practicing with ASR-based dictation programs. Dictation programs were not developed for language learning; instead, they are often offered as accessibility software or promoted as a way for native speakers to create texts more quickly and efficiently. Recent research has shown a range of benefits to practicing with dictation programs, including increased noticing of probable errors at the word and sound level (McCrocklin, 2019b) and increased attention to segments with recurring errors (McCrocklin, 2019c; Wallace, 2016). Students who practice with ASR-based dictation show improvement of segmental accuracy, both for vowels and consonants (Guskaroska, 2020; Liakin, Cardoso, & Liakina, 2014; McCrocklin, 2019a; Park, 2017), Further, the transcript can support greater learner autonomy by providing feedback (McCrocklin, 2016) and a sense of human intelligibility (Mroz, 2018), which can be motivating for students (Mroz, 2020).

However, researchers have raised concerns that the transcript may not be sufficient or usable as feedback (Strik et al., 2008). Although McCrocklin (2019c) showed that learners felt they were able to make sense of the transcript as feedback, learners in McCrocklin (2019b) recommended building in greater support for dictation practice activities. SalukiSpeech (available at https://www.salukispeech.com and launched by McCrocklin, Fettig, and Markus [2021]) is a new tool that has been created by the authors to meet this need by providing flexible dictation practice with greater learner support.

Introduction to SalukiSpeech

SalukiSpeech provides flexible, student-driven practice using ASR-dictation to support learners in identifying and addressing errors. Following recommendations from Celce-Murcia, Brinton, and Goodwin (2010), the website focuses on providing a range of pronunciation learning activities including explanations of English sounds, spelling patterns to support prediction of sounds, listening practice, controlled production, and guided communicative practice. It focuses primarily on the guided communicative language tasks (i.e., picture description tasks) with lessons that can be activated when learners notice mis-transcriptions of their speech from the ASR. When a learner notices a mis-transcription, the learner can click on the mis-transcription and supply the intended word. The website then compares the two words to identify one or more differences. The missing sounds from the intended word are then considered probable segmental errors, which are linked to lessons that lead students through explanations of the sound segment, listening exercises, and controlled production. This section provides detailed information about how the website functions.

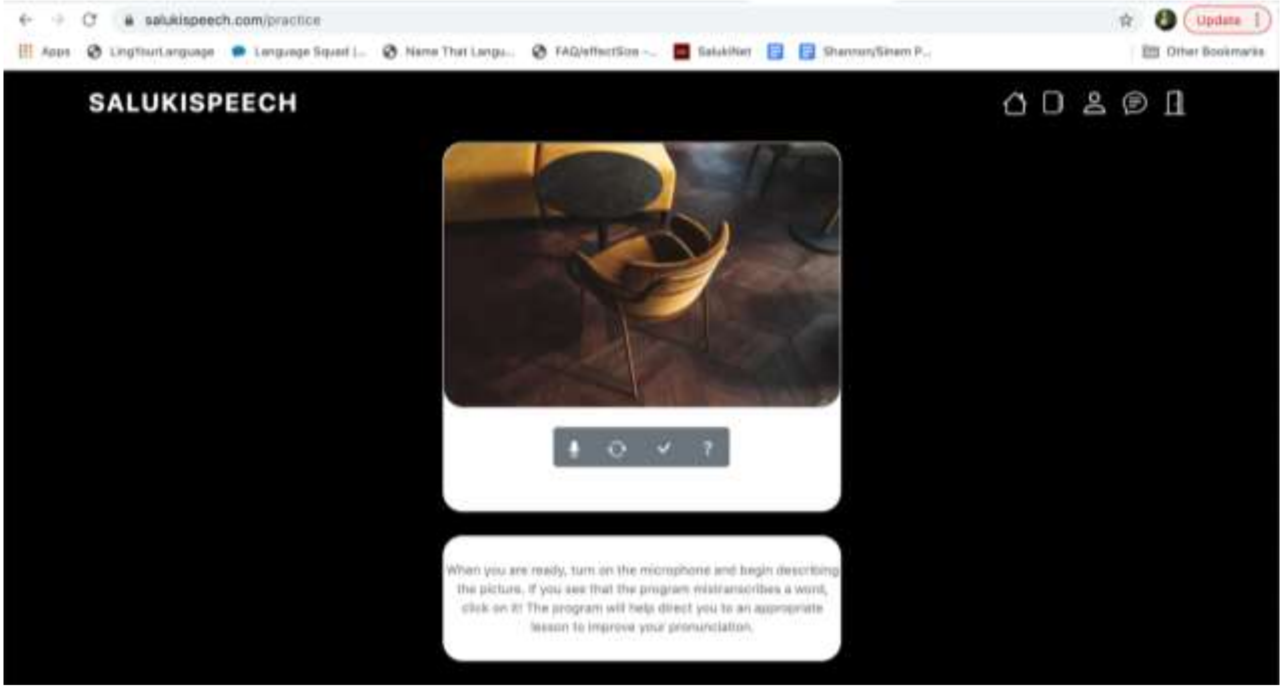
SalukiSpeech was developed as a Flask application. It relies on and draws from Flask/Python, Jinja, Pysle, Unsplash, Google API, and Postsgre SQL. Because it uses Google's speech recognition API, it must be used in Google Chrome to function properly.

To get started, learners must create an account within the site. In setting up accounts, learners are asked to provide basic demographic information (e.g., age, gender, native language) and are provided with a consent form notifying the student that their practice data will be collected and may be analyzed for research purposes or website improvements. Although the website collects user emails and usernames, the researcher view of the website does not link any identifiable information to the exportable data in order to maintain user anonymity.

Students begin practice by opening a practice page, identified by a speech bubble icon in the top right corner (see Figure 1, the second icon from the right). At the practice page, also shown in Figure 1, students are provided a picture to describe. Pictures are drawn randomly from Unsplash.com, a repository of free images online that features a wide variety of content and differing compositions. If students don't like the image provided, aren't sure what the image depicts, or lack the vocabulary to describe a given picture, they can get a new image by pressing the refresh icon below the picture (shown in Figure 1, below the practice picture, second icon from the left) as many times as desired.

Figure 1

Screen capture of SalukiSpeech featuring practice page with image from Unsplash.com

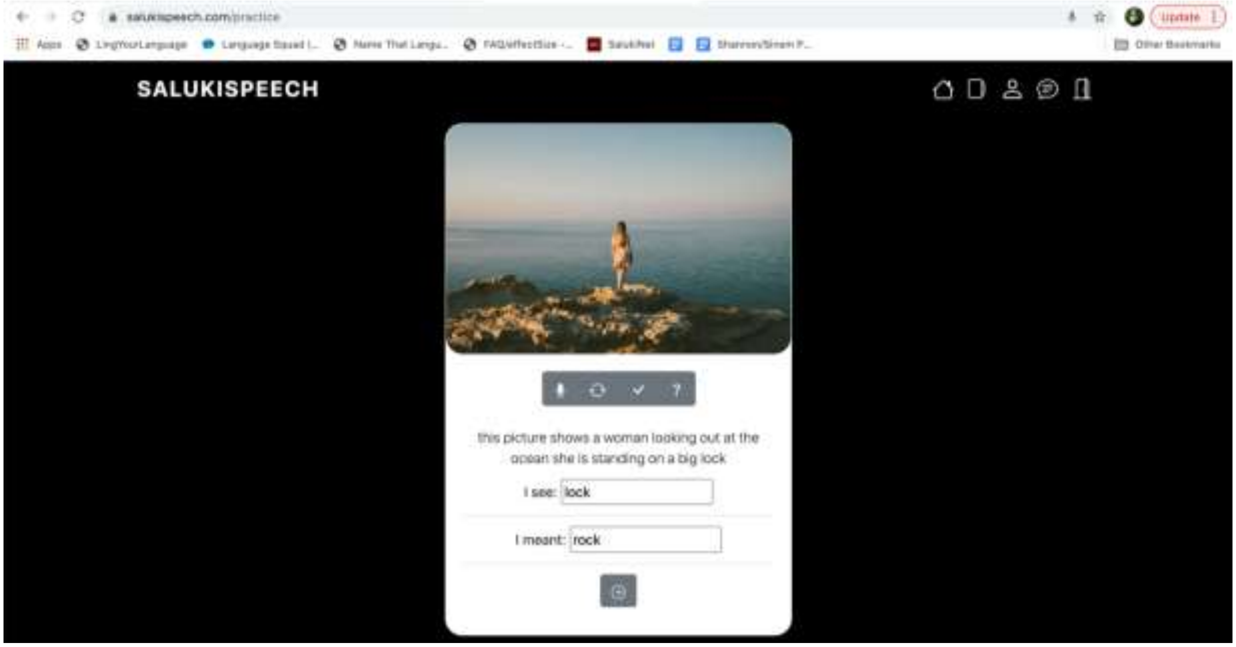


When you are ready, turn on the microphone and begin describing the picture. If you see that the program mistranscribes a word, click on it! The program will help direct you to an appropriate lesson to improve your pronunciation.

When the student is ready to describe the image, they click the microphone icon and begin speaking. The transcript populates in the space below the image, expanding the white space as needed. When students notice errors in the transcript, they can click on the mis-transcribed word. The system will then ask the learner to enter the intended word (see Figure 2) and compares the two words pulling from an IPA transcription library for Python, Pysle, in order to identify sounds in the intended word that may have been mispronounced.

Figure 2

Screen capture of SalukiSpeech featuring practice page with image from Unsplash.com.



this picture shows a woman looking out at the ocean she is standing on a big lock

I see: lock

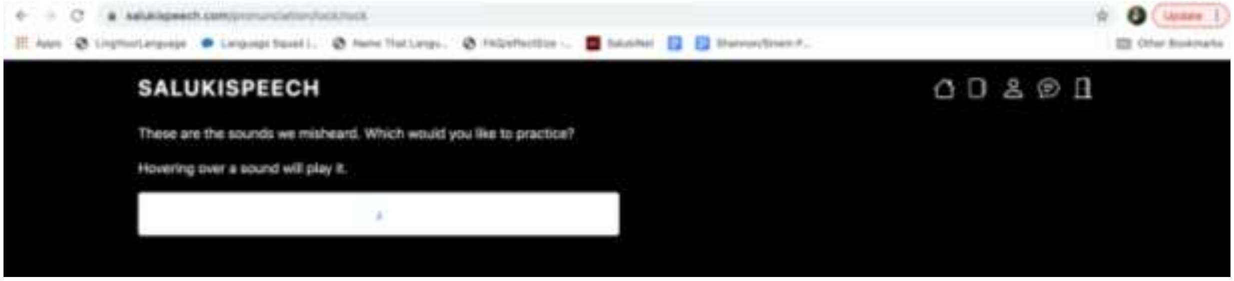
I meant: rock

Example transcript shown with error and prompt for intended word to initiate phonetic comparison

The website then provides the learner with an analysis of probable errors and, if more than one sound differed between the intended and transcribed word, learners are given a choice of lessons to pursue (see Figure 3). They can hover over the sound to hear the sound (currently displayed within an example word) and then click on the sound symbol to head to begin a lesson.

Figure 3

Screen capture of SalukiSpeech featuring analysis of probable error and link to lesson page



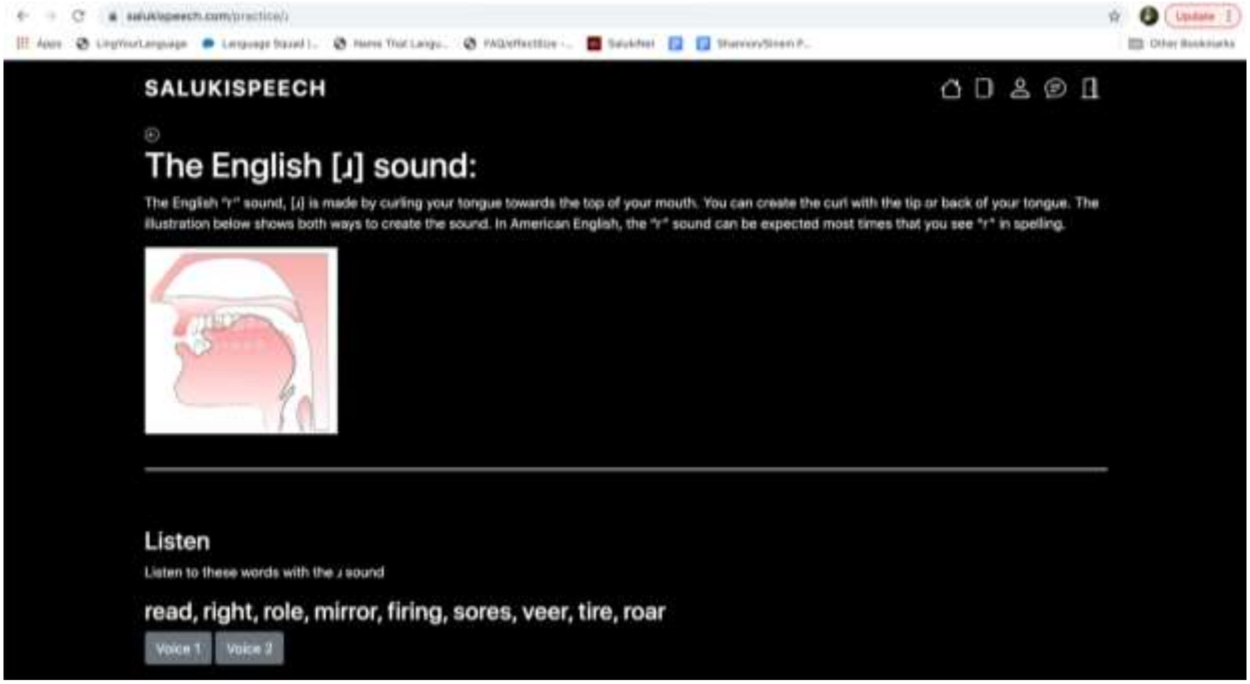
These are the sounds we misheard. Which would you like to practice?

Hovering over a sound will play it.

The segmental lessons each feature a similar structure. They start with an explanation of the sound and how it is made. Each lesson includes an articulatory diagram taken from either Pronuncian.com or *Pronunciation Contrasts in English* (2nd ed) (Nilsen & Nilsen, 2010). The explanation also includes basic information about spelling patterns that can help the learner predict the sound (see Figure 4).

Figure 4

Screen capture of SalukiSpeech featuring explanation and listening exercise for [ɹ]



The English [ɹ] sound:

The English "r" sound, [ɹ] is made by curling your tongue towards the top of your mouth. You can create the curl with the tip or back of your tongue. The illustration below shows both ways to create the sound. In American English, the "r" sound can be expected most times that you see "r" in spelling.

Listen

Listen to these words with the ɹ sound

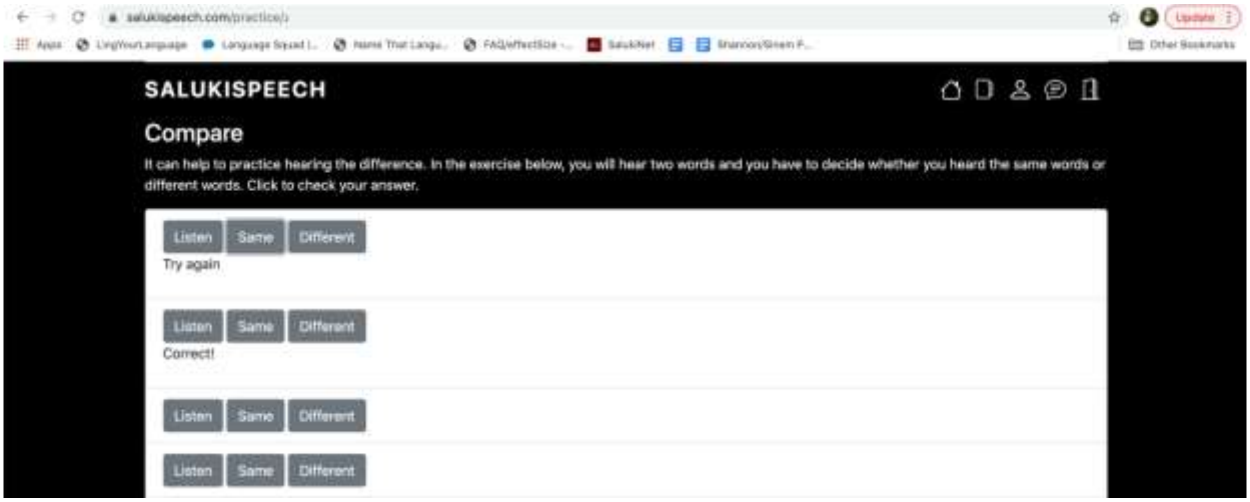
read, right, role, mirror, firing, sores, veer, tire, roar

**Note*. Source of articulatory diagram: https://pronuncian.com/pronounce-r-sound.

Following the explanation, lessons include two listening exercises. In the first activity (shown in Figure 4), learners can listen to a list of words that feature the segment in a variety of word positions (word initial, medial, and final, as applicable). These words are recorded in both a female and male voice. For the majority of lessons, the second listening activity (shown in Figure 5) focuses on discriminating minimal pairs. Learners can listen to a pair of words to decide if they hear the same word twice or two different words. The words used in these exercises are minimal pairs focusing on a similar sound that L2 learners often confuse or struggle to differentiate. An exception is the lesson for schwa [ə] for which learners choose whether the first or second syllable of a two-syllable word contains schwa instead of working with minimal pairs. Learners can check their answers as they work through examples and receive immediate feedback (correct or try again).

Figure 5

Screen capture of SalukiSpeech featuring the [ɹ] lesson (listening activity 2)



SALUKISPEECH

Compare

It can help to practice hearing the difference. In the exercise below, you will hear two words and you have to decide whether you heard the same words or different words. Click to check your answer.

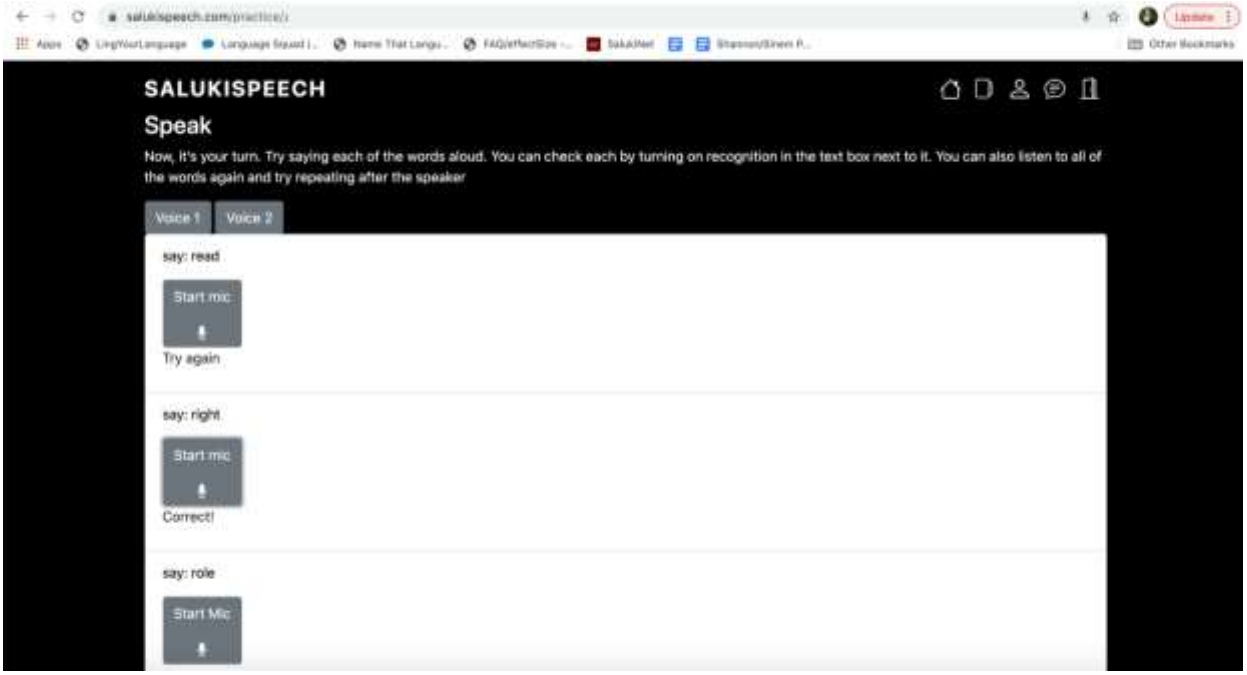
Try again

Correct!

Students can then work through the set of words provided in the first listening activity, now working to produce the words, which are checked with ASR (see Figure 6). For each word containing the target sound, students can activate the microphone and try pronouncing the target word. They are welcome to listen to the words in the audio recordings again as needed. Students then receive immediate feedback on their production (correct or try again).

Figure 6

Screencapture of SalukiSpeech featuring the [ɹ] lesson, controlled production activity



SALUKISPEECH

Speak

Now, it's your turn. Try saying each of the words aloud. You can check each by turning on recognition in the text box next to it. You can also listen to all of the words again and try repeating after the speaker

say: read

Try again

say: right

Correct!

say: role

Once students are satisfied with their practice in the segmental lesson, a link at the bottom of the page returns the student to the original practice room, with their saved image and transcript.

Students can continue practicing with that picture or submit the transcript (by clicking the checkmark icon below the practice image, shown in Figure 2) to move on to a new picture. Once submitted, the transcript is saved to the user’s account. From the profile page, learners can view all images they have worked with, their transcripts, and the sound lessons they practiced with. Below each picture and practice transcript, there is a link, “Send Practice Report.” If learners click this link, they are prompted to enter their instructor email. The system then emails the instructor a brief report of the time spent practicing and sound lessons practiced with.

Students will be occasionally prompted to leave feedback on the site through both short and long feedback forms when they submit their transcripts. The learners receive more prompts for feedback in the first few lessons, but the prompts are spaced and slow down after the first few sessions.

LIMITATIONS & FUTURE DIRECTIONS

While a testing version of the website is up and running and available to anyone to use for free, the authors are working to finish adding content and clear up glitches. At the time of writing, the most common glitches include trouble with initial access to the website and link failures leading to segmental lessons. For users having trouble getting to the site for the first time, two tips: 1) try writing out the entire website address https://www.salukispeech.com in the URL bar and 2) try clearing the cache from your browser history. For users struggling to get to a segmental lesson and receiving the message “Internal Server Error,” there is no current work around, but users can leave feedback on the site to help us identify issues that need to be addressed.

In addition to continuing to flesh out the current site and remove glitches, we hope to be able to expand the website substantially in coming years. Our current main areas of focus are augmenting the practice room spaces, adding content to segmental lessons, and increasing teacher control. Currently, the website only has a single practice space for the picture description tasks. An ultimate goal is to add in a wide variety of possible tasks, sentence reading, shadowing/mirroring, responses to questions, etc. each of which would have the same functionality allowing users to monitor their production with ASR-provided transcripts and jump to segmental lessons. Then, we hope to add in more segmental lesson content. The current plan is to add in instructional videos (likely drawing from existing videos on YouTube), along with additional listening and controlled production tasks. Ultimately, the goal would be that students could revisit sound lessons several times getting new content for each practice session. Finally, we know that to make this site useful as part of pronunciation classes, teachers may want to be able to assign work and receive more detailed reports of student practice with the site and we hope to be able to add this in the future.

Integrating SalukiSpeech into your Classroom

Despite these areas for growth, the website should allow students to practice and begin working on sounds they struggle with. Because the website uses automatic speech recognition, which is often sensitive to background noise, multiple students may struggle to use the website at the same time in a classroom unless there are sound attenuated spaces. Headset microphones may also help to lessen the impact of background noise. Teachers are encouraged, however, to consider ways that practice could be assigned to students as homework outside of class.

Although the practice activities are somewhat limited currently, teachers can get creative with the picture description tasks. Teachers could encourage students to find beautiful or fun images as part of homework that they share with the class next day. Students could vote on their favorite image and description as part of the class discussion. Teachers could also send students to find pictures relating to class content, perhaps even assigning particular vocabulary that students need to use in describing pictures, pushing students to connect vocabulary to pictures and try new words as part of their practice. Using the pictures, teachers could also ask students to critique images instead of simply describing them. Students could again bring their practice back to the classroom, drawing comparisons across the pictures. Teachers are also welcome to have students use the site for other practice (ignoring the pictures) as the website does not evaluate the content of the spoken text provided.

CONCLUSION

SalukiSpeech fills a current gap in the market by providing flexible dictation practice with the support of more traditional language learning programs. Although it is early in development, it is free and available to users. We hope that teachers and learners will explore the site and find it useful. We look forward to receiving feedback, either through the site or by email, in order to continue improving it.

ABOUT THE AUTHORS

Shannon McCrocklin is an Assistant Professor of Applied Linguistics/TESOL in the Department of Linguistics at Southern Illinois University. Her research focuses on the acquisition of second language phonology and computer-assisted language learning.

Email**:** shannon.mccrocklin@siu.edu)

Claire Fettig graduated from Southern Illinois University with a B.S. in Computer Science in May 2021.

Email: cfettig97@gmail.com

Simon Markus graduated from Southern Illinois University with a B.S. in Computer Science in December 2021.

Email: simon.markus@siu.edu

Additional Contact Information for Shannon McCrocklin

Address: 1000 Faner Dr., Rm 3228
Carbondale, IL 62901

Phone: 618-453-3428

REFERENCES

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

Daniels, S. & Taylor, K. (2021). Blue Canoe. https://bluecanoelearning.com/

Duolingo. (2021). Duolingo. https://www.duolingo.com/

Guskaroska, A. (2020). ASR-dictation on smartphones for vowel pronunciation practice. *Journal of Contemporary Philology, 3*(2), 45-61.

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.

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

Liakin, D., Cardoso, W., & Liakina, N. (2014). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal*, *32*(1), 1-25.

McCrocklin, S. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System*, *57*, 25-42.

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

McCrocklin, S. (2019b). Dictation programs for second language pronunciation learning: Perceptions of the transcript, strategy use and improvement. *Konin Language Studies*, 7(2), 137-157.

McCrocklin, S. (2019c). Learners' Feedback Regarding ASR-based Dictation Practice for Pronunciation Learning. *CALICO Journal*, *36*(2), 119-137.

Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals*, *51*(3), 1-21.

Mroz, A. (2020). Aiming for advanced intelligibility and proficiency using mobile ASR. *Journal of Second Language Pronunciation*, *6*(1), 12–38. https://doi.org/10.1111/flan.12348

Park, A. Y. (2017). The study on Automatic Speech Recognizer utilizing mobile platform on Korean EFL learners' pronunciation development. *Journal of Digital Contents Society, 18*(6), 1101-1107.

McCrocklin, Fettig, & Markus (2021). SalukiSpeech. https://www.salukispeech.com

Nilsen, D.L.F. & Nilsen, A. (2010). *Pronunciation Contrasts in English* (2nd Ed.). Long Grove, IL: Waveland.

Rosetta Stone. (2021). Rosetta Stone. https://www.rosettastone.com/

Pronuncian (2016). Pronuncian.com: American English Pronunciation. Seattle Learning Academy. https://pronuncian.com/

Siri. (2021). Siri. Apple. https://www.apple.com/siri/

Strik, H., Neri, A., & Cucchiarini, C. (2008). Speech technology for language tutoring. *Proceedings of language and speech technology (LangTech '08) conference* (pp. 73-76). Rome, Italy.

Unsplash. (2021). Unsplash. https://unsplash.com/

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.