

Wallace, L. (2016). Using Google web speech as a springboard for identifying personal pronunciation problems. In J. Levis, H. Le, I. Lucic, E. Simpson, & S. Vo (Eds). *Proceedings of the 7<sup>th</sup> Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Dallas, TX, October 2015 (pp. 180-186). Ames, IA: Iowa State University.

## USING GOOGLE WEB SPEECH AS A SPRINGBOARD FOR IDENTIFYING PERSONAL PRONUNCIATION PROBLEMS

[Lara Wallace](#), Ohio University

For students such as International Teaching Assistants who do not have heavily-accented speech but must improve their spoken English intelligibility, Google Web Speech (GWS), an ASR-based transcription tool, can be used to increase awareness of potential oral communication problems. As research has indicated, ASR (automatic speech recognition) serves as a tool to improve students' autonomy in learning, and ultimately is able to help students improve unclear pronunciation. GWS transcribes students' speech. Students then correct and mark the transcript. After analyzing the marked transcript for discourse intonation features and scrutinizing discrepancies between the transcriptions, students compare their speech to models. From this process, students may gain an understanding of where it is possible improve their delivery, including words they might not be pronouncing intelligibility. They then practice with the goal of recording an improved version. In this paper, the procedure, benefits, and limitations will be discussed.

### INTRODUCTION

Students report having difficulty identifying pronunciation weaknesses; however, without knowing what they should focus on, it is challenging for them to improve their pronunciation skills effectively. One tool to help students increase awareness of potential oral communication problems and to build self-monitoring skills is to listen carefully to a recording of themselves speaking and to analyze their speech in part by marking a transcript of this audio recording (Celce-Murcia, Brinton, & Goodwin, 1996). This type of activity is encouraged in *English Communication for International Teaching Assistants (ITAs)* by Gorsuch, Meyers, Pickering, and Griffiee (2013). For ITAs in particular, it is especially important that they work efficiently on improving their comprehensibility because not only do they contend with a heavy workload (teaching, studying, conducting research), they also must meet a higher standard of comprehensibility than their graduate counterparts who do not have teaching obligations. Unfortunately, ITAs who get by in their daily lives with few communication problems sometimes fail to notice when their spoken English skills result in communication breakdown, and may become discouraged when they must take a class to improve on a skill they thought they had mastered (Wallace, 2014). By listening to and analyzing an audio recording of their speech and seeing a transcript of it, ITAs can come to realize that there is room for improvement (Wallace, 2013, 2014). It is not only ITAs who can benefit from this activity, but other higher-level speakers as well.

Much of the usefulness of this activity comes from having this visual reference—a transcript of the student's speech; yet transcribing by hand is a time-consuming process and is not the

pedagogical goal. In this practitioner's experience, the quality of the transcriptions was variable and the completion rate was low when transcribing by hand was assigned as homework, and transcribing during class in order to improve quality and guarantee completion was far too time-consuming. Google Web Speech (GWS), ASR (automatic speech recognition)-based dictation software, shortens this process of transcription, taking less time to complete in class than transcribing by hand so that all students may benefit from the activity. Google Web Speech's transcription of the audio is only a beginning of students' increasing awareness. In order to gain maximum benefit, students must revise these transcriptions by correcting any word choice mistakes so that the transcription can serve as a mirror, reflecting the written form of their audio samples.

When making these corrections, it appeared that many of the discrepancies between what Google transcribed and what the students actually said illustrated some sort of pronunciation mistake. These occasional instances of low recognition could be helpful to learners. In her 2015 PSLLT presentation, Shannon McCrocklin reported on a research study that compared a fully face-to-face pronunciation workshop to one in which half of the work was completed with ASR. She found that both groups made statistically significant improvement. Although there was not a statistically significant difference between groups, the ASR group made slightly greater improvement on 5 of the 6 sounds/sound pairings investigated, indicating that ASR is a useful tool students can use to practice segmentals. Further, McCrocklin (2016) showed that introduction to ASR can significantly improve students' beliefs of their self-efficacy and autonomy in pronunciation learning. Thus, utilizing Google Web Speech as an ASR-based dictation tool not only saves time, but as McCrocklin's research indicates, it also can help students improve their pronunciation, beginning with the identification of potentially unclear pronunciation.

Using such a tool can save valuable time, thereby allowing students to focus on the analysis of the transcript so that they can discover for themselves areas they need to improve in order to speak more intelligibly. Nonetheless, teachers must realize that this tool may only be effective under two conditions: (1) that the student uses a headset microphone for maximum clarity of input, and (2) that the student's accent is not too different from the Web Speech ASR models. This practitioner has noticed that students with SPEAK Test scores of 42 and higher tend to benefit the most from these tools since the dictation tool has too much difficulty identifying what students with heavily-accented speech say. If there are too many discrepancies between the ASR-based dictation tool's transcription and what the student actually said, not only could it be an overwhelming task to identify what problems may have resulted in the miscomprehensions, but it could also be discouraging to the student, reducing the effectiveness of the learning experience. For this tool and the activity to work well, teachers and students should attempt it informally first to determine how easy each student's speech is to understand and to practice the computing skills since it requires some coordination.

## **PROCEDURE & PRACTICAL SUGGESTIONS**

This procedure is adapted from the transcription activity in Gorsuch et al. (2013, p. 173) where students are asked to transcribe exactly what they said and how, then make corrections where needed, and practice an improved delivery. It begins by having the students simultaneously record (e.g., on Audacity) two minutes of speech while Google Web Speech transcribes it.

Once the recording and transcription are complete, students copy and paste Google Web Speech's transcription into a document twice. They keep one as it is for comparison, and correct the other so that it is an exact reflection of the audio. On the original transcript, students highlight GWS' transcription errors in another color. They should also mark in another color the discrepancy in the revised transcription. In order for the corrected transcript to more closely reflect the audio recording, students should add in any punctuation, as well as any fillers ("eh," "uh"), hesitations (...), self-correction, and false starts that they may hear.

Depending on what pronunciation target(s) the students focus on, students can also do any of the following:

- Use ALL CAPS to indicate PROMINENT words (or the stressed SYLLable of prominent words).
- Indicate thought group division by marking pauses (/).
- Insert arrows to identify change in pitch movement ( $\nearrow \rightarrow \searrow$ ) and key choice ( $\uparrow \downarrow$ ).
- Indicate particularly fast sections with ">>" or indicate a slowing of speech rate with "<<" (this is not covered in Gorsuch et al. 2013).

Please see Figure 1 for a sample marked transcript and the original transcription side-by-side.

GWS Transcription	Student's Correction by Hand
<p>My major is <u>spanish</u> and the term I'm going to <u>the find</u> is what a cognate <u>iCall Knight</u> is <u>worth</u> it could be either a noun or an adjective and it's similar in meaning <u>on</u> the way it is written both in English I mean is the same both English and <u>spanish forest Humble any news</u> the word tolerance it's written <u>in</u> the same way <u>then</u> in <u>spanish</u>...</p>	<p><math>\uparrow</math>My MAJor / is SPANish <math>\nearrow</math>/ and the term, TERm I'm going / to <u>define</u> / is ...mmm ... / what is a COGnate<math>\searrow</math>. / A COGnate / is a <u>word</u><math>\rightarrow</math>, / it could be / either a NOUn <math>\nearrow</math>/ or/ an ADjective<math>\rightarrow</math> / AND it's similar <u>both</u> in MEANing <math>\nearrow</math> <u>and</u> the way it is written<math>\rightarrow</math> / both in English / I mean/ <u>it's</u> the SAME both in English and SPANish.<math>\searrow</math> // <u>For example,</u><math>\rightarrow</math> / in <u>English</u> / the word <u>TOLerant</u><math>\rightarrow</math> / it's written / the same way / <u>than</u> in SPANish<math>\searrow</math>... (<math>\downarrow</math>)</p>

Figure 1. GWS transcription and a student's revision of the transcription.

Once students have revised and marked their transcripts, they must analyze them for where they could make improvements. They can begin by looking at the highlighted discrepancies on the original transcript and listening again to the audio recording for these sections, making note of any words that seem to be mispronounced. If there is a string of words, the problem might have to do with linking, pausing, not pausing, stressing the wrong word or syllable, or there may be a grammar or word choice error. Students, perhaps with the help of their instructor or

pronunciation tutor, should try to determine what the problem might be. Apart from pronunciation, it is also valuable for students to correct any word choice or grammar errors that they notice. Again, they are encouraged to seek input from the teacher on this.

After students become aware of what they need to work on, they should begin practicing an improved delivery of the same content. Practice should be targeted and done on a small scale first. Students (perhaps with the help of the instructor or pronunciation tutor) start by isolating difficult vowels or consonants, then work on the pronunciation of these sounds in the targeted improvement of pronunciation at the word level, next the words in the context of a phrase, and so on. Practice of different word stress patterns, word choice, and grammar can be done in a similar fashion. It is important that students practice in such a way that they do not read from a script; rather, they speak extemporaneously.

One important way for students to know whether or not their speech is similar to target production is to record themselves imitating a model; they then listen again to the recording to get a sense of how similar or different their production is. To work on speech sounds, students can use University of Iowa's Sounds of Speech (<http://soundsofspeech.uiowa.edu/english/english.html>). To listen to the pronunciation of words in isolation, students can see a transcription and listen to an audio recording in Merriam Webster Dictionary (<http://www.merriam-webster.com/>), whereas for the pronunciation of words spoken in context, they can watch video clips in which the word or phrase was used (<http://youglish.com/>). If students have questions, they are encouraged to work with a teacher or pronunciation tutor for feedback.

Students should record their pronunciation and compare it to the models whenever possible. Even without models for comparison, students can benefit from listening to recordings of their speech for phrasal stress, prominence, thought group division, pitch movement, key choice, speech rate, and fluency. The ultimate goal with comparing their speech to a model is that students improve self-monitoring skills and practice improved pronunciation rather than repeat the same mistakes.

## **DISCUSSION**

### **Benefits**

Students have credited this activity with opening their eyes and their ears to what difficulties people could have understanding them (Wallace, 2013). When students see the corrected transcript of what they said, they often realize why they are asked to repeat themselves. Some comments students have made after completing this activity (when asked what they learned about their speech) include: "I pronounced the "th" sounds sometimes wrong," "I had [a] problem in pronouncing 'appearance'," "I [learned] I hesitate to speak. I think too much before speaking. And I usually repeat the words that I pronounce incorrectly."

Immediately evident to students are hesitations (...), speech rate being overly slow or fast, recasts, false starts, and fillers. Anecdotally, students have commented that when listening back to the audio to correct the GWS transcription, there were words that the students themselves had difficulty understanding. With the guidance of a teacher, students can also become more aware

of how they need to change their stress, rhythm, or intonation. As for pronunciation, Google Web Speech's transcription can often shed light on what words were unclear.

Returning to the sample transcription of Figure 1, it is possible to compare GWS' interpretation to what the student actually said. With the instructor's guidance, one can make inferences as to why the speech was not recognized, and perhaps give feedback on what can be done to improve recognition. Compare the following (Table 1):

Table 1

*Comparison of GWS's Interpretation with Corrected Transcript, Student A*

<b>Examples</b>	<b>Google Web Speech's Incorrect Interpretation</b>	<b>Corrected Transcript</b>
A	The find	Define
B	Worth	Word
C	Forest humble	For example

In examples A and B, a likely reason behind the incorrect interpretation was dentalization of the /d/. In C, it seems the speaker omitted the velar articulation of the consonant cluster, and perhaps the jaw was not open far enough. In these examples, GWS lack of recognition points to the speaker's pronunciation problems with consonants and vowels.

Looking at another student example, one can see that GWS not only points to segmental differences, but also to rhythmic differences (see Table 2).

Table 2

*Comparison of GWS's Interpretation with Corrected Transcript, Student B*

<b>Examples</b>	<b>Google Web Speech's Incorrect Interpretation</b>	<b>Corrected Transcript</b>
A	1952 2010	1950 to 2010
B	Ants y axis	And the y axis
C	The person page	The percentage
D	Lucas play score of	Look at this graph

Based on the difference in interpretation, it is plausible that in example A, the student might not have paused between "1950" and "to." In examples B and D, the student likely had difficulty

with articulating /ð/ in “the” and “this.” Example C seems to show that the student did not reduce the vowel in the final syllable of “percentage.” Example D seems to point to the student separating the consonant cluster /gr/. Regardless of whether or not the student or teacher can figure out reasons why GWS may not have recognized parts of a student’s speech, these discrepancies between GWS’ interpretation and what was actually said can shed light on what words the student pronounces differently.

### **Limitations**

As stated earlier, when using GWS to transcribe extended speech, it is important that students’ speech is not strongly accented, and that they speak into a headset microphone in order to reduce surrounding noise. A third limitation is that students should use GWS or another voice-to-speech app that does not “get to know” their voice. For this reason, it is preferable to conduct this activity in a computer lab, rather than have the students dictate something on their smart phones.

Regarding the activity itself, for maximum effectiveness, the teacher should dedicate class time to completing it. In this way, students can receive help both with the computing as well as in the analysis, correction, and practice sections (Hubbard, 2013). Of particular importance is giving students clear feedback on what they need to do to make their speech more intelligible since GWS can only indicate a different interpretation of what the students said.

### **CONCLUSION**

Using GWS to transcribe student speech is beneficial to students in a number of ways. First, it creates a safe speaking opportunity where students can be the only ones to judge their speech. Secondly, it saves time; instead of undergoing the arduous task of transcribing audio by hand, they check the transcription while listening to their audio recordings. Also, by marking and correcting GWS’ transcription, students practice their self-monitoring and analysis skills. Finally, discrepancies between GWS’ interpretation of students’ speech and what the students actually said can shed light on potential pronunciation problems at the word level, and sometimes with rhythm. Again, because GWS can only point to potential pronunciation problems, the instructor has the final say in what students must work on to improve their intelligibility. With these points in mind, it makes sense that an ASR dictation tool could be useful in helping students to improve their pronunciation. Future research could investigate which kinds of pronunciation errors GWS better detects and its accuracy rate in detecting those errors. Furthermore, to reduce computing and streamline the procedure, it would be helpful if an app were available that allows students to compare the original transcription to one that students can edit on screen. Taking it a step further, adding a pitch tracking option in such an app to the transcription would help students see their pitch movement and key choice since it can be difficult to hear. In this way, students would still gain the benefit of listening carefully to and analyzing and improving their speech, and the teacher’s work of giving meaningful feedback would remain relevant.

### **ACKNOWLEDGMENTS**

After PSLLT 2015, Jim Talley answered the call for a more streamlined experience. He created a simple web app (using the emerging Web Speech API) that allows for dictation of speech and

captures the recognition results in a text area that the students can directly edit (while preserving the original results for comparison). Thank you for comments, Jim, and thanks for the app. I would also like to thank Edna Lima and Shannon McCrocklin for their valuable input and observations on this topic.

## ABOUT THE AUTHOR

Lara Wallace is a full-time lecturer in the Ohio University Linguistics Department's English Language Improvement Program where she has been teaching oral communication courses to graduate students and coordinating the Pronunciation Lab since 2005. Her research interests include CALL, pronunciation, International Teaching Assistants, and learner autonomy.

## REFERENCES

- Celce-Murcia, M., Brinton, D., Goodwin, J. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. New York, NY: Cambridge University Press.
- Gorsuch, G., Meyers, C., Pickering, L., Griffiee, D. (2013). *English communication for international teaching assistants* (2<sup>nd</sup> ed.). Long Grove, IL: Waveland Press.
- Hubbard, P. (2013). Making a case for learner training in technology enhanced language learning environments. *CALICO Journal*, 30(2).  
<http://journals.sfu.ca/CALICO/index.php/calico/article/view/945>
- McCrocklin, S. (2015, October). *The effectiveness of ASR-based dictation for pronunciation improvement*. Paper presented at the 7<sup>th</sup> Annual Pronunciation in Second Language Learning & Teaching Conference, Dallas, TX.
- McCrocklin, S. (2016) Pronunciation learner autonomy: The potential of automatic speech recognition. *System*, 57, 25-42 [doi:10.1016/j.system.2015.12.013](https://doi.org/10.1016/j.system.2015.12.013)
- Wallace, L. (2013). Taking the first step: International teaching assistants' motivation to improve their spoken English intelligibility. *ITAIS Newsletter*. Retrieved from <http://newsmanager.commpartners.com/tesolitais/issues/2013-08-19/1.html>
- Wallace, L. (2014). *International teaching assistants' (ITAs') experiences with language learning, learner autonomy, and technology as students in a requisite oral communication course*. (Electronic Dissertation). Retrieved from <https://etd.ohiolink.edu/>