Garcia, C., Kolat, M. & Morgan, T. (2018). Self-correction of second language pronunciation via online, real-time, visual feedback. In J. Levis (Ed.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching conference*, ISSN 2380-9566, University of Utah, September, 2017 (pp. 54-65). Ames, IA: Iowa State University.

SELF-CORRECTION OF SECOND-LANGUAGE PRONUNCIATION VIA ONLINE, REAL-TIME, VISUAL FEEDBACK

Christina Garcia, Saint Louis University

Mark Kolat, The Ohio State University

Terrell A. Morgan, The Ohio State University

We have built a set of web-based tools that take learning far beyond the classroom for L2 pronunciation students. Among these is a user-friendly site where students record themselves and can instantaneously compare their own voiceprint to that of a native speaker. Students continue to practice specific sounds until their production "looks" like their native target. For each sound, the system displays native and non-native speech samples (i.e., clickable sound files) alongside spectrographic representations. A column to the left orients the user on how to interpret the spectrograms to analyze the sound. The interface enables students to produce live-generated spectrograms, re-recording until their productions both sound and look like the native output. Recordings are automatically submitted to the instructor for effortless monitoring of students' progress. This interface was originally developed for Spanish and has been extended to English and Quechua. The response from Spanish students is overwhelmingly positive, many citing that "seeing" their pronunciation allowed them to internalize the steps necessary to make improvements as never before. They report increased satisfaction at being able to work on their pronunciation outside of class and receive the necessary feedback to assure that they are pronouncing correctly, even without a language expert present.

INTRODUCTION

In her keynote speech at this year's annual conference, Isabelle Darcy (2017) highlighted the ways in which, as researchers and teachers, we can bridge the gap between research and practice in pronunciation teaching. How do we apply our research to the pronunciation classroom and vice versa? One such gap between teaching and research can be seen in the way we have our students practice. Despite considerable technological advances in phonetics research in the past twenty years, we often find ourselves employing the same pronunciation drills we have always used rather than harnessing the technology used in modern acoustic analysis for pedagogical purposes.

Some scholars have noted this gap, experimented with different types of visual feedback, and found these methods to have enormous potential. In his review of Computer Assisted Pronunciation Training (CAPT) techniques, Levis (2007) supports the use of visual feedback in CAPT systems, citing the best-known visualizations as: spectrograms, waveforms, and pitch tracings. Studies specifically looking at the efficacy of CAPT programs have shown increased gains when compared to control groups for features such as pitch, voicing, vowel length, and

geminate consonants (Hew & Ohki 2004); vowel height (Quintana-Lara 2014, Kartushina et al. 2015); and rhotics (Patten & Edmonds 2015), among many other features.[1]

Thus far there has been a focus on learners of English, and as is evident in all areas of linguistic research, our pedagogy can be significantly improved by trying to solve the "problems" of other languages. Additionally, despite these advances, students often report technical frustration with programs such as *Praat*, and are unable to see beyond the technology itself in order to truly appreciate what it has to offer. Visual feedback can be powerful, but it seems that with today's students, how this feedback is packaged is of utmost importance.

The motivations for using technology in pronunciation practice are evident: face time with students is limited and instructors are constantly left wishing they had more time for individual feedback. While listening to recordings outside of class commonly supplements in-person feedback, it can quickly become overwhelming. On the student side, there are learners who are unable to hear the difference between their pronunciation and the native pronunciation, rendering it impossible to identify how to "fix" their own. Thus, pronunciation instructors are left with a few problems to solve: how can students receive effective and timely feedback without the instructor present? How can cutting-edge technology be efficiently used to simultaneously help students improve their pronunciation *and* learn about acoustic phonetics?

We have designed a user-friendly site[2] that aims to resolve these issues by allowing students to record themselves in real time and receive instantaneous feedback via spectrographic representation. Upon logging in to the interface, students are presented with native and non-native speech samples of a particular sound accompanied by spectrograms of these speech samples. They can listen to the difference between the native and non-native productions while simultaneously noting how the two spectrograms differ, guided by an instruction panel. Students are then able to record their own production and compare it to the native and non-native, both visually and auditorily. Students can continue to practice and re-record the sounds until their production 'looks' like that of the native target. Crucially, students do not need to download a program or native speech samples in order to use this tool; everything is provided in a one-stop shop so that the technical difficulties mentioned earlier are eliminated and students can focus on the visual feedback itself. In the sections that follow, we describe in detail the design of the tool and the user experience (section 2), feedback from students and instructors (section 3), the benefits and limitations of this methodology (section 4), and areas for its future development (section 5).

Description of the tool

The online tool that we have developed makes practicing difficult sounds for the L2 learner of Spanish, English and/or Quechua a smooth and efficient experience. For now, each sound is housed on a separate webpage and thus the students are provided with a list of web addresses for

[1] This list of studies is necessarily limited due to the length restrictions of this paper. For an overview of other studies in CAPT, see Levis (2007) and O'Brien (2006).
[2] Inspiration for this online interface grew out of our collaboration with the OhioSpeaks project of the Department of Linguistics at The Ohio State University (Wanjema et al. 2013).

all of the target sounds. Ohio State students utilizing this tool for a course are asked to sign in with their credentials so that their usage and progress can be monitored by their instructor.

Upon reaching the webpage of a unique sound or phonological process to be practiced, the user finds sidebar instructions on the left that are tailored to the particular lesson. The target sound and its non-native counterpart appear in phonetic notation, and the instructions are broken down into phases, the first being "Listen & Observe" and the second "Practice & Record". Next to these sidebar instructions, we find a series of three interactive voiceprint boxes where the spectrographic representations appear. The first and third boxes correspond to the Phase 1 instructions and contain prerecorded examples of the native (box 1) and non-native (box 3) pronunciations of the target sound. The student is able to listen to these recordings by clicking the "Play" button below each box and, at the same time, see a real-time spectrogram of the recordings in the boxes. The student can repeat Phase 1 as many times as they desire, being led by the instructions to understand what they are hearing and viewing. These instructions ask the student comprehension check questions that also contain descriptions of the spectrograms, presented in such a way as to highlight the differences between the native and non-native versions. For example, if the student is trying to perfect their pronunciation of the Spanish approximant [β], which would be realized as the full stop [b] in English, the page would appear as in Figure 1.

The spectrograms here clearly show that there is a large gap of white space between the colored columns of the non-native version where the red formant lines break, but there is more continuity in the red formant lines of the native version. While a written description of this comparison is contained in the Phase 1 instructions, students also receive basic instruction in interpreting spectrograms in class.

In Phase 2 the student is given the chance to practice their own pronunciation of the target sound by clicking the "Record" button under the box labeled "Your Version". Upon clicking the "Record" button, they do their best to imitate the native pronunciation, and a spectrographic representation of their speech shows up in real time as they speak. Contained in the Phase 2 part of the sidebar are explicit instructions on how to execute their own recordings as well as the English translation of the word that they are practicing that contains the target sound, in this case *haba* 'fava bean'.

Figure 1. Screenshot of online interface for Spanish [β]

Once the student has completed their first recording, the options to "Play" or go "Back to Carmen" (Ohio State's learning management system) appear, and the student also sees that their first attempt has been saved. The student can then play their recording in coordination with the native and non-native ones to auditorily compare them as well as visually compare the spectrograms. They are able to rerecord themselves as many times as they wish until they are satisfied with their native-like audio and visual realization of the target sound. At this point they click a "Save" button that appears under their spectrogram to save the recording. Both the automatically saved first recording and their elected final recording are sent to the instructor so that they are able to track the student's progress and identify strengths and weakness among all students in their class.

So far, the target Spanish sounds for which we have adapted this tool are: approximants [β] (vs. English [b]): *haba* (screenshots in Figure 1), [ð] (vs. English [d]): *hada,* and [ɣ] (vs. English [g]): *haga*; unaspirated voiceless stops [p] (vs. English [p^h]): *pan*, [t] (vs. English [t^h]): *tan*, and [k] (vs. English [k^h]): *can*; and fricative voicing [zð]: *desde* vs. *de este*.

For Quechua, thus far we have focused on the differences between the unaspirated, aspirated, and ejective voiceless stops and affricates: [p] vs. [p^h] vs. [p'']; [t] vs. [t^h] vs. [t'']: *tanta* ('meeting'), *thanta* ('old'), *t'anta* ('bread'); [k] vs. [k^h] vs. [k'']; [tʃ] vs. [$tʃ^h$] vs. [$tʃ^{\prime}$]. Figure 2 is a screenshot of the comparison of *t'anta* (box 1) vs. *tanta* (box 3). The target sound is the ejective [t''] that does not appear in Spanish or English, and here it is compared to the unaspirated version *tanta* which contains the [t] that is present (in this context) only in Spanish (a language already spoken by most learners of Quechua). If the student speaks English but not Spanish, comparing

t'anta (with the ejective) to *thanta* (with aspiration) would be more appropriate since the default English pronunciation would be aspirated.

In this example, the native version contains three distinctive segments with white space between each, the first of which being an initial, high-intensity explosion representing the ejective [t']. In the non-native version, this ejective does not appear because the word *tanta* is being produced, which simply contains an initial unaspirated stop followed immediately by the vowel [a]. There is therefore no white space between the stop [t] and the appearance of the red vowel formants.

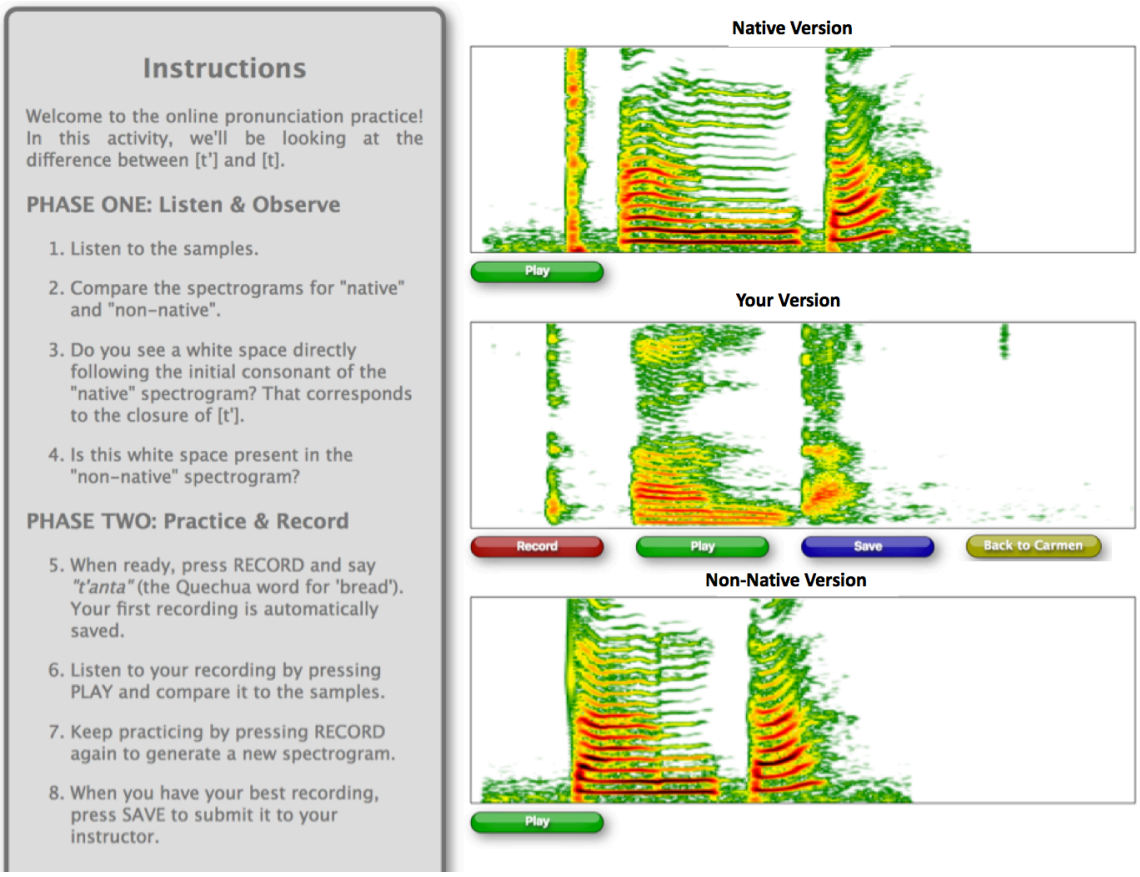

Figure 2. Screenshot of online interface comparing Quechua words *t'anta* 'bread' and *tanta* 'meeting'

Finally, for English, we have targeted the following: aspirated [p^h] (vs. [p] in many other languages: *pot*); aspirated [t^h] (vs. [t]): *top;* aspirated [k^h] (vs. [k]): *cop*; [h] vs. [ø]: *hall* vs. *all*; flapping /t, d/→[ɾ]: *kitty*, *buddy*; and bunched (postalveolar approximant) [ɻ̟]. Figure 3 displays a screenshot of [h] vs. [ø]: *hall* vs. *all*.

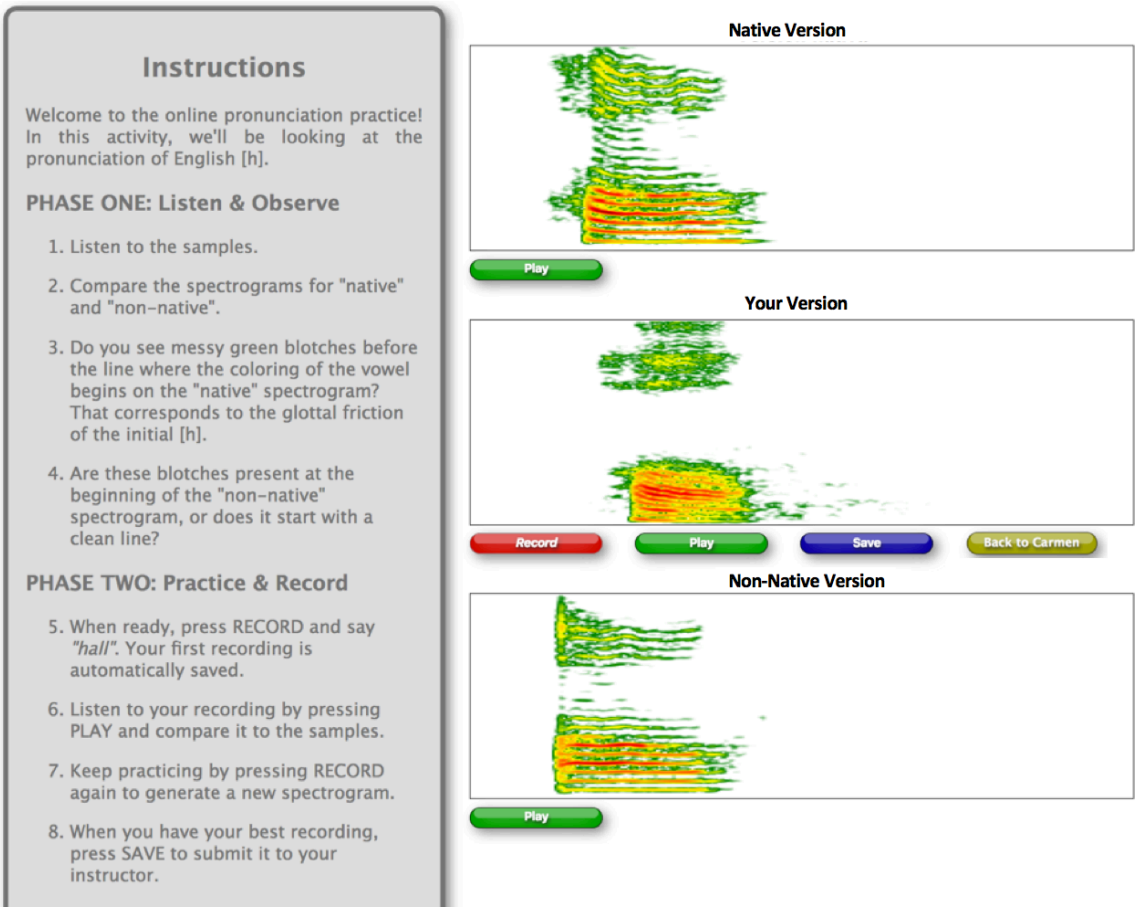

Figure 3. Screenshot of online interface comparing English words *hall* and *all*

In this example, the difference in frication is being focused on, with the native version containing frication of [h] as represented by the green friction before the yellow and red formants of the sound [a] begin. In the non-native version, there is no green friction before the yellow and red formants of the vowel [a] begin; it is just white. This represents the lack of the voiceless fricative [h], changing the lexical meaning of the word in English. This non-native version would be expected of students, for example, with L1s of French or Italian, which do not contain a voiceless glottal fricative in their phonetic inventory.

Having already implemented this tool into certain courses at Ohio State, we have received feedback from both students and instructors about their experiences. This feedback is outlined in the section that follows.

Student & Instructor Feedback

Thus far, this tool has been used in eight sections of Spanish 3404 (“Pronunciation”) at The Ohio State University since 2015. Quantitative survey results were collected in the first semester of its implementation (Autumn 2015), and since then qualitative impressions have been collected from both students and instructors. In this section we summarize this feedback, focusing on the areas most important for future development of the tool.

The quantitative survey results display the contrast between students' perceptions of online tools, in particular the pronunciation tool, at the beginning and end of the semester, before and after interacting with the actual tool. The survey included whether the online pronunciation activities *will be* (pre-survey, N=31) or *were* (post-survey, N=8) user-friendly, good preparation for class, worthwhile, enjoyable, and helpful in identifying and fixing pronunciation errors. At the beginning of the semester, students overall where somewhat skeptical about the utility of the online pronunciation tool, with rates of agreement vacillating between 45-55% agreement that it would be useful, worthwhile, helpful, etc. In the post survey, these agreement rates raised to 63-75% in the qualities of: user-friendly, easy to follow, enjoyable, teaching needed skills, helping to prepare for class, and identifying pronunciation errors.

In qualitative evaluations, students are quick to point out the advantage of receiving immediate feedback on their pronunciation. Pronunciation students and instructors alike are often frustrated with the small amount of class time in which feedback can be given. The tool by no means replaces invaluable instructor feedback; however, it offers students an *additional* way to receive quality feedback, without having to wait until the next class period or for their instructor to grade their recording. Instructors have recognized that this feature increases students' autonomy and sense of responsibility for their pronunciation gains. Interaction with the tool affords students a new sense of linguistic awareness that aids in meeting their pronunciation goals.

Students also repeatedly report that they enjoyed and appreciated the visual aspect of the tool's feedback. They like being able to see their voice visualized in a 'techy' way they have never seen before, especially those who are visual learners. Speaking to the need for multimodal learning (cf. Levis 2007: 14), several of these visual learners noted that, in class, they had difficulties hearing the difference between their own pronunciation and the native model of their instructor, but that the tool allowed them to understand and internalize their own pronunciation in a new way by giving them a visual target to work towards. More specifically, the fact that the interface includes both native Spanish and naïve native English productions gives students an idea of the two extremes on the spectrum and, even if they do not succeed in 'matching' the native, they might land somewhere in between. From the instructor's perspective, the visualization reinforces the Spanish versus English contrasts discussed in class and helps students to better conceptualize the phonological system. To put it succinctly, "seeing is believing" and this tool can convince students that the contrasts we talk about in class actually exist. One instructor also noted a feedback loop in which seeing the linguistic contrasts actually caused his students to be able to hear these contrasts better moving forward.

Finally, students like the interactive and creative aspect of the pronunciation practice. A couple of instructors had their students 'play around' with a super 'gringo' accent versus their best native accent, which encourages students to do the activity over and over again and see how the visual changes. When used properly, this tool can take one of the most tedious parts of language acquisition (pronunciation drills) and make it more fun. This feature of the system leaves students wanting more, which encourages them to continue taking linguistics, and specifically phonetics, as they further their studies.

Comparisons with other technologies

The online tool described here is just one of a number of emerging technologies with implications for L2 pronunciation instruction. Like all computer-assisted methodologies, it presents both virtues and limitations, which we will examine here in more detail.

To begin with, some articulatory distinctions can be clearly seen on a spectrogram that cannot be appreciated using other visual representations, and vice versa. While electropalatography illustrates nuances in the contact between articulators which are not easily discernable on a spectrogram, it is only able to do so if, in the production of the sound in question, the tongue makes contact with some part of the hard palate. Vowels and labial consonants (just to name two classes of sounds) are therefore irrelevant to the usefulness of electropalatograms as a teaching tool.

Similarly, the distinction between velar and uvular consonants (critical in Bolivian Quechua) can be illustrated neatly via ultrasound technology (Bird & Bliss 2017), but not so clearly on a spectrographic voiceprint. The spectrogram, however, is able to capture labials and other sounds whose articulators interact outside of the section of the vocal tract captured in the ultrasound image. And while relative tongue height and backness are visible in the formants on a spectrogram, the nuances of vowel quality are much clearer on a vowel plot, such as that produced by the computer program described in Lie-Lahuerta (2014).

Some studies have critiqued the use of spectrograms in pronunciation practice (cf. Neri, Cucchiarini, & Strik 2002), alleging that students are not likely to be able to pick up on the errors they are making via such feedback. However, we have shown that with a small amount of training in interpretation and an interface that allows for native vs. non-native comparison, the use of spectrograms can be beneficial to L2 pronunciation students. What is key is making sure the spectrographic representations are used in a motivated and targeted way and highlight particular features that are relatively easy to observe in this type of visualization.

In fact, one additional benefit that accrues to the learner is basic knowledge of acoustic phonetics—for example, how approximants appear different from stops, or the relationship between, say, aspiration or glottalization and the subsequent onset of voicing. These concepts, and their articulatory concomitants, are neither trivial nor impractical in the acquisition of native-like pronunciation, and it therefore makes sense to teach them in conjunction with the targeted practice in question. There is even reason to believe that students will better internalize such notions when presented on a need-to-know basis and as part of an interactive lesson with personalized impact.

Areas of future development

While student and instructor feedback has shown the usefulness of this pronunciation tool, there are of course ways in which is can be improved, especially at a time in which the advances in CAPT technology multiply by the day. The first area of development that we see regards content as opposed to technical features. We continue to work on the expansion of this tool to include Quechua and English pages, with the eventual goal of expanding to other world languages. This understandably requires the collaboration of instructors of other languages, but presents vast potential for several reasons, one being that the addition of new languages and sounds challenges the limits of the online interface, ideally making it more powerful and flexible. With the development of pages in other languages, the system should be flexible enough to allow for

different comparisons based on the student's L1, as mentioned above in the case of Quechua *t'anta*. Another area of content revision concerns which sounds are targeted. Currently the pronunciation tool targets those linguistic features (ex. aspiration) that are easiest to observe on a spectrogram; however, additional attention should be given to whether these are the sounds that are most difficult for L2 learners and whether there is a way to expand the system to other 'difficult' sounds. Finally, we intend to expand the database of native speaker recordings to allow for the students to choose the dialect and gender of the native speaker that their speech is compared to, given that recent developments in the 'golden speaker' technology have shown that learners benefit greatly from hearing a model voice that is similar to their own (Ding et al. 2017).

Beyond content development, there are several areas in which the tool itself can be improved in order to allow for an even more positive student experience. The quality of the spectrograms, although somewhat dependent on the students' recording environment, should be improved to facilitate the identification of visual differences between native and non-native spectrograms. Also, we have received several comments about having some instantaneous feedback beyond the visual comparison of spectrograms. For instance, for the aspiration of /p, t, k/, the interface could let users know what their VOT (Voice Onset Time) was in order to compare to the VOTs of the native/non-native samples. This would require the use of a *Praat* (Boersma & Weenink 2017) script or similar and potentially more linguistic training for the users; however, for some sounds it would resolve the issue of users not knowing what to look for in the spectrogram.

To that point, the main area of future development that we see at this time is figuring out how to make the interface even more user-friendly so that it could be used with novice and intermediate learners, who may or may not be receiving explicit pronunciation instruction in class. We feel that spectrographic representations have the potential to be used in this way if some adaptations were to be made. For instance, in addition to the instructions on the left side panel, there could be boxes or arrows that highlight the portion of the spectrogram that is relevant to the target sounds. There could also be a training module (cf. Levis 2007) added in order to guide the user through the ideal experience. Nevertheless, the question remains of how a novice or intermediate learner would apply the spectrographic information to their pronunciation. This presents a significant challenge, but one that might be overcome through the use of pronunciation podcasts to accompany the pronunciation interface.

Finally, we are currently working on making this site open to the public so that instructors and students at other institutions may take advantage of it. Ideally this version of the site will have a landing page that will allow users to peruse all of the sounds available for practice. It is our hope that by making the site open we are able to receive increased feedback that will allow us to continue to improve the tool's effectiveness. Look for it in the future at *pronounce.osu.edu*.

ACKNOWLEDGMENTS

We are grateful for the financial support of The Ohio State University's Impact Grant, and most especially the on-going and invaluable work of Henry Griffy and the technical expertise of Justin Slauson and Abhijit Varde. The student feedback reported here was collected under IRB protocol 2015B0261.

ABOUT THE AUTHORS

Christina García, Assistant Professor of Spanish and Linguistics at Saint Louis University, is a sociolinguist and phonetician interested in why Spanish speakers talk the way they do, and how to get students excited about the linguistic diversity of the Spanish-speaking world. She has studied and done fieldwork in Argentina and Ecuador, and her research examines how sounds are socially meaningful and contribute to the formation of regional identities. Christina bring her research into the classroom by challenging pronunciation students to use the same techniques she uses in her research (acoustic analysis) to internalize and improve their own pronunciation in new and interactive ways.

Christina García, Ph.D.
Assistant Professor of Spanish & Linguistics
Saint Louis University
1513 Morrissey Hall
3700 Lindell Blvd.
St. Louis, MO 63108
Phone: 513.505.2494
Email: christina.garcia@slu.edu

Mark Kolat is a first-year MA/PhD student in Hispanic Linguistics at The Ohio State University. He holds Bachelor degrees in Spanish and French from his present university, and has studied in Spain and Senegal. With previous studies of Spanish, French, Italian, Wolof, Swahili, and Quechua, his research interests include sociolinguistics, phonetics & phonology, and second language acquisition, particularly related to less commonly spoken/studied languages of Africa and Latin America. In serving as an L2 English/Spanish instructor in Senegal as well as spending a semester working to promote the Quechua language program at Ohio State, Mark hopes to expand the current applications of this pronunciation tool to other languages that he encounters through his graduate work.

Mark Kolat
MA/PhD student
Hispanic Linguistics
The Ohio State University
298 Hagerty Hall
1775 College Road
Columbus, OH 43210-1340 USA
Phone: 440.602.2781
Email: kolat.5@osu.edu

Terrell A. Morgan, Professor of Hispanic Linguistics at The Ohio State University, is a phonologist and dialectologist concerned with documenting linguistic diversity and finding new ways to put students, teachers, and fellow researchers in touch with the intimate details of the sounds and structures of Spanish. He has taught and done field work in Spain, Mexico, Guatemala, Honduras, Nicaragua, Costa Rica, Panama, the Dominican Republic, Puerto Rico, Colombia, Ecuador, Peru, Chile, Bolivia, Paraguay, Argentina, and Uruguay, as well as in

Hispanic communities in the United States. His Spanish phonetics textbook, *Sonidos en contexto: Una introducción a la fonética del español con especial referencia a la vida real*, was published in 2010 by Yale University Press.

Terrell A. Morgan, Ph.D.
Spanish & Portuguese
The Ohio State University
298 Hagerty Hall
1775 College Road
Columbus, OH 43210-1340 USA
Phone: 614.292.9555
Email: morgan.3@osu.edu

REFERENCES

Bird, S. & H. Bliss. (2017). Visualizing speech in a classroom setting using interactive ultrasound imaging. Paper presented at *Pronunciation in Second Language Learning & Teaching (PSLLT) 9th Annual Conference*, University of Utah.

Boersma, P. & Weenink, D. (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.35, retrieved 16 October 2017 from http://www.praat.org/

Darcy, I. (2017). "'A psycholinguist walks into a classroom…': A road-map for bridging research and practice in pronunciation teaching". Plenary talk at *Pronunciation in Second Language Learning & Teaching (PSLLT) 9th Annual Conference*, University of Utah.

Ding, S., C. Liberatore, G. Zhao, S. Sonsaat, E. Chukharev-Hudilainen, & Levis, J. (2017). Golden Speaker Builder: An interactive online tool for L2 learners to build pronunciation models. Paper presented at *Pronunciation in Second Language Learning & Teaching (PSLLT) 9th Annual Conference*, University of Utah.

Hew, S. H., & Ohki, M. (2004). Effect of animated graphic annotations and immediate visual feedback in aiding Japanese pronunciation learning: A comparative study. *CALICO Journal, 21*(2), 397-419.

Katrushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America, 138*(2), 817-832.

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics, 27*, 184-202.

Lie-Lahuerta, C. (2014). Fix your vowels: Computer assisted pronunciation training. Retrieved from Holt, D. E. & Lord, G., *IGNITE: CASPSLaP resources.* www.ignite-caspslap-2014.weebly.com, Retrieved 29 October 2017.

Morgan, T. A., & García, C. (2016). Seeing is believing: Student self-correction of Spanish pronunciation via online, real-time, visual feedback. Retrieved from Campos-Astorkiza, Rebeka, Terrell A. Morgan, and Holly Nibert. *IGNITE: CASPSLaP Resources.* Online: u.osu.edu/ignite2016/.

Neri, A., C. Cucchiarini, & H. Strik. (2002). Feedback in computer assisted pronunciation training: When technology meets pedagogy. *Proceedings of "CALL professionals and the future of CALL research" Conference* (pp. 179–188). Antwerp, Belgium. Retrieved October 27, 2017 from http://repository.ubn.ru.nl/bitstream/handle/2066/76206/76206.pdf

O'Brien, M. (2006). Teaching pronunciation and intonation with computer technology. In L. Ducate & N. Arnold (Eds.), *Calling on CALL: From theory and research to new directions in foreign language teaching* (CALICO Monograph Series, Vol. 5, pp. 127–148). San Marcos, TX: CALICO.

Patten, I., & Edmonds, L. A. (2015). Effect of training Japanese L1 speakers in the production of American English/r/using spectrographic visual feedback. *Computer Assisted Language Learning, 28*(3), 241-259.

Quintana-Lara, M. (2014). Effect of Acoustic Spectrographic Instruction on production of English /i/ and /I/ by Spanish pre-service English teachers. Computer Assisted *Language Learning, 27*(3), 207-227.

Wanjema, Shontael, Katie Carmichael, Abby Walker, & Kathryn Campbell-Kibler. (2013). Integrating teaching and large-scale research through instructional modules. *American Speech*, *88*(2), 223-235.