L2 Pronunciation in CALL: The Unrealized Potential of Rosetta Stone, Duolingo, Babbel, and Mango Languages

Joan Palmiter Bajorek
The University of Arizona

Abstract

Hundreds of millions of language learners worldwide use and purchase language software that may not fully support their language development. This review of Rosetta Stone (Swad, 1992), Duolingo (Hacker, 2011), Babbel (Witte & Holl, 2016), and Mango Languages (Teshuba, 2016) examines the current state of second language (L2) pronunciation technology through the review of the pronunciation features of prominent computer-assisted language learning (CALL) software (Lotherington, 2016; McMeekin, 2014; Teixeira, 2014). The objective of the review is: 1) to consider which L2 pronunciation tools are evidence-based and effective for student development (Celce-Murcia, Brinton, & Goodwin, 2010); 2) to make recommendations for which of the tools analyzed in this review is the best for L2 learners and instructors today; and 3) to conceptualize features of the ideal L2 pronunciation software.

This research is valuable to language learners, instructors, and institutions that are invested in effective contemporary software for L2 pronunciation development. This article considers the importance of L2 pronunciation, the evolution of the L2 pronunciation field in relation to the language classroom, contrasting viewpoints of theory and empirical evidence, the power of CALL software for language learners, and how targeted feedback of spoken production can support language learners. Findings indicate that the software reviewed provide insufficient feedback to learners about their speech and, thus, have unrealized potential. Specific recommendations are provided for design elements in future software, including targeted feedback, explicit instructions, sophisticated integration of automatic speech recognition, and better scaffolding of language content.

Keywords: pronunciation, software, CALL, automatic speech recognition L2

The ability to communicate clearly and effectively is a key to second language acquisition (SLA), and this includes speech and pronunciation. To best support student development of these crucial skills, *targeted feedback* can provide specific, evidence-based, and actionable information that can significantly improve learner pronunciation. Technology that provides students with instantaneous *targeted feedback* provides novel opportunities for students to improve their pronunciation in personalized and effective ways (Blake, 2013; Liakin, Cardoso, & Liakina, 2015). This article provides a snapshot of the current state of second language (L2) pronunciation technology through the review of prominent computer-assisted language learning (CALL) software for effective,

well-designed pedagogical materials for intelligible pronunciation (Lotherington, 2016; McMeekin, 2014; Teixeira, 2014), which includes Rosetta Stone (Swad, 1992), Duolingo (Hacker, 2011), Babbel (Witte & Holl, 2016), and Mango Languages (Teshuba, 2016). Currently, the topic of pronunciation is neglected in the L2 classroom due to the history of the second language acquisition (SLA) field, which directly informs classroom materials today (Krashen, 1982; Thomson & Derwing, 2014).

Although many believe that pronunciation is a superfluous skill only attained by a few, clear and intelligible speech is integral to language acquisition and use (Thomson & Derwing, 2014). While L2 acquirers may have differing goals concerning pronunciation (e.g., dialects, native-like accent, and fluency), the fundamental importance of pronunciation rests on intelligibility and the ability to communicate (Arteaga, 2000). Pronunciation is vital for L2 learners to progress in their communicative abilities and to avoid complete communication breakdowns (Morin, 2007; Morley, 1991). Morley writes, "intelligible pronunciation is an essential component of communicative competence" (1987, p. iv). However, for many L2 adult students, pronunciation does not improve, even over significant amounts of time with exposure to the L2 from native speakers. Data from several studies demonstrate that input alone is insufficient for pronunciation advancement in traditional language classrooms with time frames spanning from 12 weeks to 4 years (Elliott, 1995; Flege, 1980, 1981; Han & Odlin, 2006; Solon, 2016; Waniek-Klimczak, 2013).

To address this crucial skill of language learning, technology provides novel opportunities for L2 students to improve their pronunciation in personalized and effective ways (Blake, 2013; Liakin, Cardoso, & Liakina, 2015). CALL software has the ability to provide L2 learners with novel, sophisticated, inexpensive, and learner-centered tools (Blake, 2013; Chapelle, 2001; Eskenazi, 1999; Kennedy, Blanchet, & Trofimovich, 2014; Kissling, 2013; Lord, 2005). Specifically, for pronunciation, research indicates that *targeted feedback* can lead to significant improvement in L2 pronunciation development in hours to weeks of treatment (Derwing & Rossiter, 2003; Elliott, 2003; Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2015; Thomson, 2011). CALL software varies greatly and more can be learned about how these tools do and do not actively support L2 pronunciation development. To better understand how contemporary CALL software treats L2 pronunciation, this article considers the quality and quantity of TF provided to L2 learners in the most prominent CALL software available (Gass & Mackey, 2007; Lotherington, 2016; McMeekin, 2014; Teixeira, 2014).

Intelligibility: Being Understood

Speech and pronunciation are integral skills for L2 learners. Pronunciation is the phonetic and phonological realization of communicative competence (Morley, 1991) and includes continuity, intonation, pitch, rhythm, segmental and suprasegmental features, stress, and voice quality of spoken language (Arteaga, 2000; Morin, 2007; Morley, 1991; Szubko-Sitarek, Salski, & Stalmaszczyk, 2013). Poor pronunciation can render speakers

unintelligible and can result in a complete breakdown in communication (Morin, 2007).

While L2 acquirers may have differing goals concerning dialects, native-like fluency, etc., the fundamental importance of pronunciation rests on intelligibility (Arteaga, 2000). The *intelligibility principle* is concerned with "helping learners become more understandable" (Thomson & Derwing, 2014, p. 327). This differs from the *nativeness principle*, where it is considered desirable for L2 adult learners to produce native-like production (Levis, 2005; Thomson & Derwing, 2014). For the typical L2 learner, being understandable is crucial, whereas native-like speech might be a long-term goal of ultimate attainment (Bajorek, 2016; Colantoni, Steele, & Escudero, 2015). Native-like pronunciation acquisition in adult learners may be nearly impossible (Arteaga, 2000), even in contemporary, privileged environments such as university classes (Colantoni et al., 2015). Rare studies that show native-like pronunciation attainment in adult language learners exclusively focused on students who received extensive explicit phonetic and phonological training of the targeted language (Arteaga, 2000). Yet instead of this type of explicit training in pronunciation, the topic is generally neglected in modern classrooms.

SLA History and the Neglect of Pronunciation

Pronunciation instruction hardly exists in today's communicative language classroom due, in great part, to SLA theory and the evolution of the field. A Greco-Roman model that was solely *grammar translation* based, emphasizing pronunciation and speaking skills, came into vogue with the rise of technology and the *audio-lingual method*. In the 1970's, a movement arose that was united behind the rejection of the *audio-lingual method* (Arteaga, 2000). These "drill and kill" exercises where students mimicked and parroted input were found to inadequately prepare students for genuine communication outside of the classroom (DeKeyser, 2010, p. 156; Larsen-Freeman, 2011).

The term *communicative competence* demonstrated a shift in pedagogical approaches from *grammar translation* and *audio-lingual method* to a large amount of implicit input in a *communicative* environment (Hymes, 1971). As the pedagogy currently employed in a vast number of university language courses, the *communicative method* (CM) is one in which the instructor uses authentic language in appropriate contexts, and most communication is in the targeted language. Within the CM, pronunciation instruction exists in a context where instructors and students concede that it has value, but they are unsure of its place in the classroom or how to teach it (Morin, 2007).

Many assume that pronunciation will develop with time if given sufficient comprehensible input (Krashen, 1982; Thomson & Derwing, 2014). Krashen argued against the explicit teaching of pronunciation with the concept that native-like pronunciation acquisition is nearly impossible after a critical period (Jones, 1997; Krashen, 1982). This ultimately led to the "virtual disappearance of pronunciation work" from textbooks of the 1970's (Jones, 1997, p. 105). Thus, with the rise of the CM, "the acquisition of pronunciation has fallen to the wayside and has suffered from serious neglect in the communicative classroom" (Elliott, 1997, p. 95). Terms such as "stepchild" (Arteaga, 2000, p. 340), "neglected"

(Elliott, 1995, p. 530), and "casualty" (Thomson & Derwing, 2014, p. 326) described the representation of pronunciation in most adult language classrooms. It is ironic that a central goal of the CM, with such a focus on communication, would lead to the neglect of pronunciation development in the process.

This neglect is also notable in instructional material for language classrooms. In 2000, Arteaga's review of 10 Spanish textbooks found that pronunciation sections were invariably poorly designed, incomplete, inaccurate, relegated to the end of the chapter, exclusively within laboratory sections, underdeveloped, or omitted altogether (Arteaga, 2000; Terrell, 1989). Evidence would indicate that the situation has "remained relatively unchanged" as of 2013 (Lord & Fionda, 2013, p. 515).

Since Arteaga's review, there has been no recent study reviewing the state of pronunciation in language textbooks (Lord & Fionda, 2013). In academia, however, this is slowly changing. Research on second language phonological instruction is "in its infancy," and only recently has there been an uptick in conference proceedings, graduate student theses, and other scholarly work (Lord, 2005, p. 558; Thomson & Derwing, 2014). Nevertheless, changes in contemporary pedagogy practices across the board are small if existent. However, the neglect of pronunciation in the CM language classroom does not necessarily transfer to the realm of educational technologies.

L2 Perception and Production Theories

Many theories and evidence can be found surrounding L2 perception and production. For the purposes of this review, the focus is on the assumptions that impact adult learners learning a second language. When considering technology to improve learner speech, it is key to note that empirical evidence demonstrates that adult learners can improve their L2 oral production (Birdsong, Bohn, & Munro, 2007; Bongaerts, Mennen, & Slik, 2000; Colantoni & Steele, 2006; Colantoni et al., 2015). However, this L2 production is affected by the *filter* of the L1, and there might be interactions and transference between the two languages (Antoniou, Best, Tyler, & Kroos, 2011; Flege, 1987; Lee, Guion, & Harada, 2006).

Within this field, there is little consensus about directionality of the *interference* of the two languages languages, though much work has narrowed down how research is conducted, considering language modes (Grosjean, 2001); language exposure and context (Sancier & Fowler, 1997); and age of acquisition (Flege, 1987, 1991; Flege & Liu, 2001; Flege, Munro, & MacKay, 1995). These factors might affect learners who produce and perceive sounds in their L2.

Theoretical models that have been introduced to explain speech learning include the *perceptual assimilation model* (Best, 1995) and *speech learning model* (Flege, 1995). These models posit that there is an "'equivalence classification' at early stages of L2 learning for sounds that are similar in the L1 and the L2," such as the voiceless stop /p/ in Spanish in onset position that might be classified as equivalent to the English voiceless

stop in onset position that is aspirated /pʰ/ (Solon, 2016, p. 25). For L2 sounds that are not so easily mapped to L1 categories, novel categories for the new features are created (Antoniou et al., 2011). It may indeed be harder to produce L2 sound distinctions that are not present in the L1 and for those that were categorized as equivalent to L1 incorrectly.

Since learners map their L1 phonological and phonetic systems to those of the L2, it "is common for bilinguals to speak their L2 with a detectable foreign accent" and to "produce speech that is detectably different from that of native speakers of the language" (Antoniou et al., 2011, p. 558). In sorting through the new phonological inventory, literacy has been demonstrated to be a key component in L2 phonological and phonetic abilities (Duñabeitia, Orihuela, & Carreiras, 2014; Huettig & Mishra, 2014). This is especially vital for L2s who use different scripts in their L1, such those who switch between Roman, Cyrillic, Sanskrit, and character-based languages. (Gollan, Forster, & Frost, 1997; Mathieu, 2014). These factors are key when considering how technology mediates L2 speech acquisition and development.

Targeted Feedback and Effective Tools

Research demonstrates that students may need explicit instruction and targeted feedback to improve their pronunciation. *Targeted feedback* (TF) is defined as an intervention where the learner is provided with information about their utterances; it is specific, evidence-based, and actionable in respect to an L2 targeted production to further pronunciation development (Gass, 2013; Wiggins, 2012). This is significantly different from *binary feedback,* where the user is only told "wrong, try again" (Chapelle, 2001, p. 73). TF provides rich information that can inform how the L2 crafts a learner's next production. Building upon ideas from *focus on form* (Doughty & Williams, 1998; Long, 2000), the *noticing hypothesis* (Schmidt, 1994, 1995), and *explicit learning* (Hulstijn, 2005), the concept of TF is that SLA learners should pay attention to linguistic forms for explicit noticing of L1 and L2 contrasts (Larsen-Freeman, 2011). TF can come in many different forms such as oral feedback from an instructor or peer (Kennedy et al., 2014), scaffolded "self-analysis projects," (Lord, 2005, p. 557), and visual and auditory cues from software (Hincks, 2003).

Contrary to Krashen's theories that pronunciation need not be explicitly taught (Krashen, 1982; Thomson & Derwing, 2014), pronunciation acquisition may not occur without formal instruction. A large body of literature demonstrates that input alone is insufficient for pronunciation development with time frames spanning from 12 weeks to 4 years (Elliott, 1995; Elsendoorn, 1980; Flege & Hammond, 1980; Flege, 1981; Han & Odlin, 2006; Mitleb, 1981; Niemi, 1979; Solon, 2016; Waniek-Klimczak, 2013). While most classroom instruction and materials are inadequate for improving learner pronunciation (Arteaga, 2000; Morin, 2007), TF can lead to significant improvement (Chapelle, 2001; Elliott, 2003; Kartushina et al., 2015; Thomson, 2011).

Solon found no change in L2 Spanish production among university students spanning four years (2016). Examining allophonic patterning in the production of L2 Spanish

laterals, Solon found that fourth-year Spanish students with an average of 9 years of Spanish study demonstrated no statistical difference from their first, second, or third-year peers (2016).

Over the course of a 12-week semester, Elliott found that input alone resulted in no improvement of intermediate Spanish student pronunciation. The experimental group received explicit instruction and TF on their pronunciation and improved in their ratings by trained native Spanish speaking judges (Elliott, 1995). The data from approximately 30 hours of targeted native Spanish language input provide evidence that comprehensible input is not adequate for pronunciation development, while explicit instruction and TF result in improvement (Elliott, 1995). Many studies are similar in design to Elliott's (1995) work in that they provided L2 learners with TF and also found significant improvements for learners (Derwing & Rossiter, 2003; Elliott, 1995, 1997, 2003; Kennedy et al., 2014; Lord & Fionda, 2013).

Studies in the language classroom demonstrate that language software designed for L2 pronunciation development can also provide effective TF. Visual feedback from a computer elicited statistically significant improvement in learner vowel production within one hour of instruction (Kartushina et al., 2015); see also (Katz & Mehta, 2015). In this study, L1 French and L2 Danish participants were given explicit instructions about the position of their mouth in relation to visualizations. The experimental group visualized their production in comparison to native speaker targeted speech, where "articulatory feedback provided was based on an immediate, trial-by-trial acoustic analysis of the vowels produced by participants" (Kartushina et al., 2015, pp. 823-824). The control group was presented with the same visual field but was given no specific feedback about their production. Mean improvement was 17% for the experimental group, and there was no effect in the control group after 4 hours of training (Kartushina et al., 2015).

There is a pattern that learners who receive no instruction or inadequate instruction which does not feature TF may not improve in their production of the L2. In contrast, learners exposed to TF dramatically improved in brief timeframes (Kartushina et al., 2015) and intermediate timeframes (Derwing & Rossiter, 2003; Elliott, 1995, 1997, 2003; Gonzalez-Bueno, 1997; Kennedy et al., 2014; Lord & Fionda, 2013).

Computer-assisted Language Learning (CALL) Software

There is great potential in CALL software to improve L2 speech perception and production using novel, sophisticated, individualized, and inexpensive tools (Blake, 2013; Chapelle, 2001; Eskenazi, 1999; Golonka, Bowles, Frank, Richardson, & Freynik, 2014; Kennedy et al., 2014; Kissling, 2013; Lord, 2005). CALL can "make a huge difference" for language learners that should not be ignored (Duffy, 2015a, para. 2), even though use of CALL in the classroom lags greatly behind its ubiquity in social spheres (Lotherington, 2016). CALL can be successfully integrated into the language classroom (Blake, 2013), but for the purposes of this study, the primary focus is the potential of CALL software for

pronunciation development, regardless of whether it is employed inside or outside the classroom.

CALL's "powerful, inexpensive hardware and well-designed software" provide tools and opportunities for language learners that did not exist only twenty-four years ago (Rodman, 1999, p. 272; S.A.P., 2013). Even recently, technology that existed on the CALL market only a few years ago was "nearly unusable," especially in the domain of oral skills and voice recognition, leading some users to consider current tools somewhat "magical" (S.A.P., 2013, para. 3). Today, properly designed and scaffolded software can support learning in a "cost effective" and "efficient manner" (Blake, 2013, p. 272; Rodman, 1999). CALL software can provide individualized material with feedback and ample opportunity for repeated practice in loops. CALL software can also "offer examples, receive output from learner, evaluate response, give feedback, evaluate whether the output was sufficiently correct, provide correct pronunciation, repeat, and continue" (Rodman, 1999, p. 273). In many ways, CALL with speech capability has the potential to offer what no human can: unlimited stored knowledge, focused and personalized interactions, infinite patience and time, full attention, immediate feedback on each response, student-led pacing, and perfect consistency (Golonka et al., 2014; Rodman, 1999). Of course, the real application of any knowledge gained by CALL is the ability to communicate with actual speakers (Thomson & Derwing, 2014), but the tools to gain proficiency need not be facilitated by humans.

Thus far, the biggest limitation in the realm of speech recognition is technology's ability to gauge whether an utterance is "sufficiently correct" (Rodman, 1999, p. 273). Dependent on dialect, goals of the learner, and accent, oral production can vary greatly and be impacted by small and subtle shifts (Rodman, 1999). The ideal CALL speech software, the marriage of sophisticated speech technology and "implemented, commercially successful" tools, has not yet been realized and can be expected in the next few decades, if not sooner (Chen, 2011; Rodman, 1999, p. 273). Until the ideal tools are developed (Rodman, 1999), users must exploit current CALL software to its greatest extent.

Methodology

The choice of which software to analyze came from several sources, but most noteworthy is Lotherington's review of which software tools have won awards in recent years and which are prominently used by consumers (Duffy, 2015a; Lotherington, 2016). Rosetta Stone continues to be name brand language learning software in the United States. Lotherington found that Duolingo was featured in almost every "popular language teaching apps" reviewed between 2013-2015 (Lotherington, 2016, p. 7). In 2017, Duffy named Rosetta Stone and Duolingo the top 2 best language-learning programs on the market. Mango Languages and Babbel are up-and-coming companies who are less frequently cited but are making significant forays into the incorporation of speech technology in their software.

To evaluate how Rosetta Stone, Duolingo, Babbel, and Mango Languages address L2 pronunciation, the presentations of pronunciation were evaluated for the quantity and quality of TF provided to users. While other aspects of the CALL software might enhance L2 oral perception and production, this study solely examines parts of the software that specifically focus on pronunciation. Analysis was conducted through use of the software, evaluation of other reviews of the software conducted by researchers and non-linguists, and examination of company memos and communication about their software. Each product reviewed may have been updated since the technology was evaluated and have different features on different platforms (e.g., Android, iOS, Chrome, desktop versions, etc.). With the most up-to-date material possible, this paper serves as a snapshot of these products as of 2017.

Research Questions

1. Of the software reviewed (Rosetta Stone, Duolingo, Babbel, and Mango Languages), which use feedback and TF for L2 student pronunciation development?

2. Which of the four software programs provides the best TF for L2 learner development of pronunciation?

3. What extrapolations and conclusions can be made regarding the ideal L2 pronunciation software from the analysis of these products?

Building upon the idea that presentations of pronunciation are severely lacking in most L2 instructional material currently in circulation (Arteaga, 2000; Lord & Fionda, 2013; Thomson & Derwing, 2014), it was hypothesized that the CALL software reviewed will have great potential for learner speech development, but overall it does not fully exploit its capacities to use the most advanced research findings, pedagogical opportunities, and technology. Each software tool was analyzed individually and compared in an overall discussion.

Table 1.
Overview of Pronunciation in Prominent Language Learning Software

Logos	Rosetta Stone	(Duolingo)	Babbel	mango languages
CALL Software & Inception Year	Rosetta Stone, 1992	Duolingo, 2011	Babbel, 2007	Mango Languages, 2007
Rating of Presentation of Pronunciation	★★★☆☆	★★☆☆☆	★★★★☆	★★★☆☆

Price & Estimated Number of Users	$229 Frequent Sales Several Million Users	$0 Free Additional Features $$ Over 200 Million Users	$12.95/month or $83.40/year Cost Variable per Timeframe & Languages Over 1 Million Users	$20/month or $175/year Over 1 Million Active Subscribers
Languages for English Speakers	27 Languages	23 Languages	13 Languages	65 Languages
Targeted Feedback	Binary Feedback; Correct or Incorrect	Basic Binary Feedback; Correct or Incorrect Greater Feedback Provided on Chrome Platform	Basic Binary Feedback; Correct or Incorrect	No Feedback

The full overview of all the software analyzed in this review can be seen in Table 1. Information provided includes the year the company was founded, approximate cost of the software for American customers, approximate number of users, number of second languages supported for English speakers, and type of feedback provided. Star ratings (0 to 5) were given as a holistic evaluation of the TF provided to users. Material for Table 1 comes from sources cited within the paper and (Babbel, 2016; "Duolingo Language Courses," 2016; "History," 2016; "Language-Learning App Babbel Hits One Million Customers," 2016; Languages, 2016; "The Tenets of A/B Testing from Duolingo's Master Growth Hacker," 2017).

Rosetta Stone

"Crème de la crème" of L2 technologies (Duffy, 2015a, para. 7), one of the oldest and well-known CALL software is the sophisticated and pricey Rosetta Stone (RS) (S.A.P., 2013; Santos, 2011; Swad, 1992). Created in 1992, RS is a stable system that "continues to be the best full-featured software for learning a new language" on the market (Duffy, 2015a, para. 23). In RS's presentation of pronunciation, the learner listens to native speaker input and repeats the same utterance up to four times with a response from the interface whether the speech was correct or incorrect (see Figure 1). Users can slow down the native speaker speech to hear segments of speech more clearly (Santos, 2011).

A fascinating feature of RS's presentation is the display of waveforms and pitch contours. In a review of RS for Brazilian Portuguese, Santos writes that there is much lacking in this speech tool where "almost no explicit feedback is provided," "the speech recognition system is quite often unreliable," and the module accepts input as correct when it should not (Santos, 2011, p. 190, 192). When little to no feedback is given or pronunciation that

is accepted is faulty, the "obvious risk" is that reinforced poor pronunciation "might lead to entrenched pronunciation problems" (Santos, 2011, p. 192).

According to the RS official website, their "speech recognition technology is highly advanced" and gives learners "immediate feedback on [their] pronunciation" (Speech Recognition Talking Back Required," 2015, para. 1). While the technology might be advanced, these claims are not entirely accurate. Free software which offer similar displays and even more advanced capabilities than in previous years (see Audacity; Mazzoni, 2016) were first released in 1999; PRAAT (Boersma, 2015) was first released in 2001, and WASP (Huckvale, 2016) was first released in 2003. Therefore, while the material might be advanced for the average consumer, the technology to make waveforms and pitch contours has been widely available for decades.
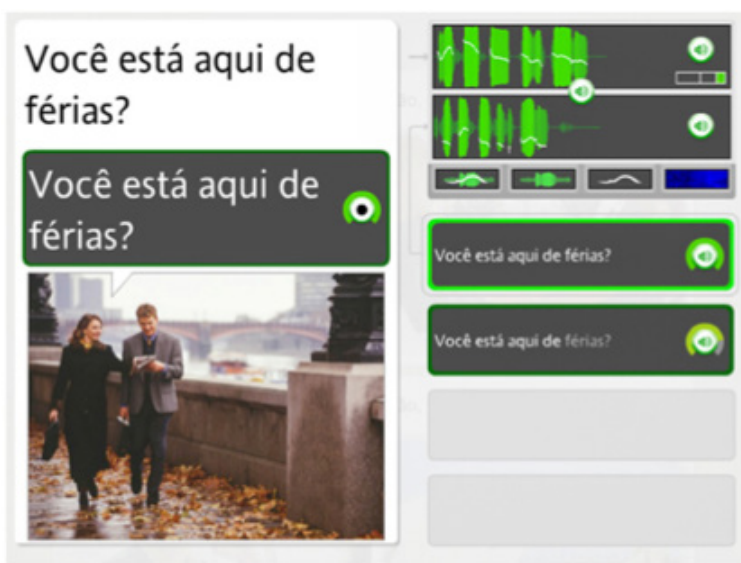

Figure 1. Rosetta Stone's presentation of pronunciation, Brazilian Portuguese. Reproduced from (Santos, 2011, p. 183)

Overall, "despite being a quite interesting technological feature," reviewer Santos could "not see how ordinary language learners could benefit from such graphic depictions (Figure 1)without proper linguistic or phonetic training" (2011, p. 183). Some research has suggested that waveforms and pitch contours can be beneficial for L2 speech development (Demenko, Wagner, & Cylwik, 2010; Hardison, 2005), but these studies provided scaffolding, explicit instruction, and feedback to learners in conjunction with visual displays. The biggest issue with RS for pronunciation instructions is that most learners might find these green waveforms aesthetically pleasing at best, and unhelpful at worst. While RS provides interesting displays, its binary feedback, poor recognition abilities, and poorly scaffolded, unexplained features leave much to be desired for the average, non-linguist language learner.

Duolingo

Rising star Duolingo is less well-known than Rosetta Stone (Duffy, 2015a; Von Ahn & Hacker, 2011), but this certainly will not be true for long. Duolingo is a crowdsourced text translation tool that gets its content from data mining the web (Garcia, 2013). Unlike the other software reviewed, Duolingo is free of cost to users, though users are giving the company ample amounts of user data. Winning first place in four recent independent reviews: Colour my learning (2014), fodors.com (2015), PC (2015), Tech Times (2015) (Lotherington, 2016, p. 7), Duolingo is considered by some reviewers to be "nearly as good" in content as Rosetta Stone (Duffy, 2015a, para. 10). Reviewers consistently consider the CALL software as "excellent" (S.A.P., 2013, para. 2), "pleasant, user-friendly" (Garcia, 2013, p. 20) and "a joy to use" due to its clean and easy to use design (S.A.P., 2013, para. 4).

Duolingo's presentation of speech pronunciation instruction through binary feedback is underdeveloped. Figure 2 demonstrates how Duolingo's speaking section prompts the learner to repeat the utterance a maximum of three times before moving on to the next section. Note that the movement of the line in the blue section of the left-most image is not a waveform but a moving graphic when the learner speaks so that the user knows their voice is being picked up.

Two major drawbacks of Duolingo are the inauthentic content and lack of speech support offered. Duolingo often offers "pragmatically absurd utterances" (Lotherington, 2016, p. 10) and the app does not "really teach conversational skills" (S.A.P., 2013, para. 4) or communicative competence (Lotherington, 2016). Duolingo is grammar and vocabulary heavy, but light on speech production.
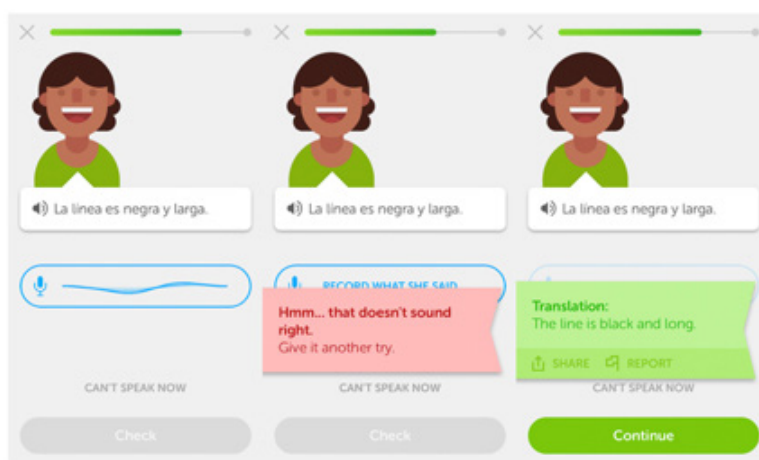

Figure 2. Duolingo's presentation of pronunciation, screenshots from iPhone 6. Left image is the prompt, middle image is an incorrect answer, right image is a correct answer response.

As one reviewer states it, "[i]f I didn't already know the basics of French conversation, I'd be helpless in France" (S.A.P., 2013, para. 4). Even worse for speech development, it is possible to turn off the speaking exercises completely in Duolingo's software. A

review from the New York Times Magazine discusses how users can use the software without ever having to speak the target language; "It took me about two weeks to make it through the Swedish course earlier this year — helped, of course, by never having to speak it out loud" (Fitzpatrick, 2017, para. 11). If learners truly never speak the language aloud and get poor feedback, it is unlikely that much of any progress is being made in spoken intelligibility. It is therefore of little surprise that students who use Duolingo and wish to be placed in university language classes based on this experience frequently lack crucial communicative abilities and are commonly sent back to the most basic introductory courses at the University of Arizona[1]. It will be interesting to see how this deficiency in the technology's communicative competence training will work as Duolingo is integrated more and more into language classrooms ("Duolingo: For Schools," 2016).

Duolingo has been working to improve on its speaking section for years and has succeeded to some extent on the Chrome platform. In a 2014 post, entitled "Vastly Improved Speaking Exercises in Chrome," Duolingo co-creator Louis Von Ahn unveiled new features in the technology that color-coded words that were pronounced correctly and incorrectly (Von Ahn, 2014). However, it should be noted that many nonsense utterances were accepted and the threshold of acceptability is low. In the 2014 post, Von Ahn wrote that the company had been working on ways to make the speech recognizer more accurate and give the user more specific feedback. Although the 213 comments follow this post report a wide array of bugs and issues, with users calling the feature "nonsense," "unusable," and "atrocious," others believed it was helpful and correct in evaluating speech (Von Ahn, 2014). Reviews specific to this feature on the Chrome platform were only found on the Duolingo website itself within the past two years (Von Ahn, 2014, 2015). Overall, Duolingo has room for improvement in its current binary, rudimentary presentation of pronunciation that exists on most platforms (Duffy, 2015a, 2016b; S.A.P., 2013).

Babbel

Little-known, relatively inexpensive CALL software Babbel exceeds expectations as up-and-coming language software (Duffy, 2015a), which is certainly seen in the ways it supports L2 pronunciation development. Babbel focuses directly on "building basic conversational skills" and speaking in a way that no other program does (S.A.P., 2013, para. 6).

---

1   Personal communication with Dr. Chantelle Warner, University of Arizona.
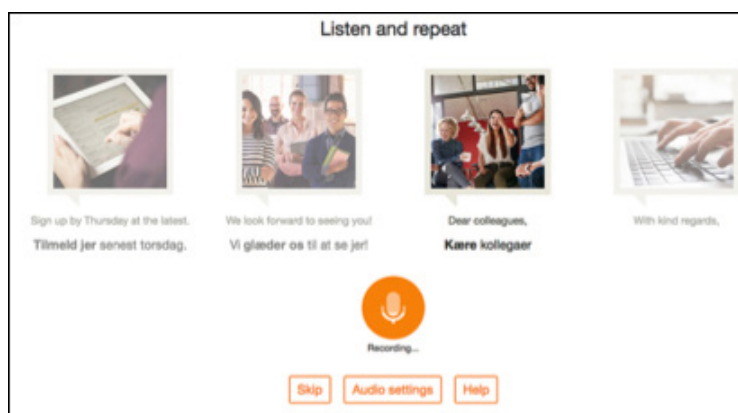
Figure 3. Babbel screenshot, Chrome Desktop, Swedish Lesson 1

Though it only grants users access to one single language when they sign up (Duffy, 2015b), this is the best pronunciation CALL software reviewed. With prompts to change the sensitivity and volume of microphones and speakers, Babbel integrates pronunciation instruction into almost all its lessons. In Figure 3, the vocabulary lesson about Swedish greetings is presented to the learner via a listen and repeat task where pronouncing the greetings and getting feedback about the production in the words is built into the activity (Duffy, 2016a). While sometimes "repeating yourself six or seven times to get it just right is wearying," there are small tutorials and popup explicit instruction throughout lessons (S.A.P., 2013, para. 7). In addition, there are sections specific to the explicit teaching of pronunciation that discuss form-pronunciation mapping. It is notable that these sections provide both pragmatic usage and cultural knowledge. Thus, the lessons give explicit instruction, feedback, and sociolinguistic knowledge, and integrate speaking into the global learning of the language. While the features of repeating utterances are "still a little buggy," reviewers commend the high quality of the material (Duffy, 2015a, 2016a; S.A.P., 2013, para. 7).

Babbel lessons "are more challenging than those of many other programs" and have a much higher a threshold for what constitutes a correct utterance (Duffy, 2015a, para. 26). Duolingo and Rosetta Stone might accept utterances that it should not, while Babbel has far higher standards for what constitutes "correct" production. If the standard is too low, learners might believe that their utterances are sufficient, such as in RS and Duolingo. If the standard is too high, the learner might get disheartened and find the tasks fatiguing.

Babbel's speech recognition technology is unique, powerful, and based on a rating system from 0-100 (Babbel, 2010; Crisi, 2010), but the learners are not privy to the score their utterance received. Users are frequently asked to repeat utterances where each "phrase you repeat is analysed and scored" and the ability to progress in the lesson is "conditional on intelligible repetition" (McKinnon, 2011, para. 2). As one reviewer writes, "the rating system is generous enough to make sure you don't become disheartened, but stringent enough to ensure that native speakers will be able to understand you" (McKinnon, 2011, para. 2).

In sum, by providing more linguistic and scaffolding support, Babbel heightens expectations for learner oral production in a cohesive way as they progress in L2 proficiency. Babbel's presentation of pronunciation goes above and beyond its competitors. Yet, more specific feedback about why production is incorrect, so as not to get users weary, would constitute a substantial improvement.

Mango Languages

An up-and-coming CALL program called Mango Languages (Mango) is not well known among consumers, but it is already receiving appreciative reviews by academics and reviewers (Duffy, 2015b; "Higher Education," 2016; McMeekin, 2014; "Public Libraries," 2016; Teixeira, 2014; Teshuba, 2016). Unlike its competitors, Mango's primary form of dissemination is through public and private libraries. "[M]ore than 300 academic institutions" have purchased licenses to their software and those with subscriptions have access to all the content of the 71 languages offered, while still tracking the progress of individuals (Duffy, 2015b; "Higher Education," 2016; "Public Libraries," 2016).

Figure 4. Mango Languages, Japanese, Chrome Desktop, Waveform comparisons

While some have noted that "Mango's core content is weak" (Duffy, 2015a) "tedious, dry, and ineffective" (Duffy, 2015b, para. 3), they use authentic, native speaker content in the form of films and TV shows as well as native speaker conversations ("Higher Education," 2016). This sharply contrasts the less authentic phrases that can be found on Duolingo (Lotherington, 2016).

In Mango's presentation of pronunciation, designed by PhDs and linguists ("Public Libraries," 2016), there are more sophisticated and well-thought out features than any of the other technologies combined. Using Mango, a learner can compare their waveforms with those of a native speaker (see Figure 4).

Learners can compare waveforms of their voice to that of a native speaker and play them at the same time. No feedback is provided, and the tool is "completely self-evaluating;" it can thus be said that the tool is "only as good as its user" (McMeekin, 2014, p. 203).

Figure 5. Mango Languages, Japanese, Chrome Desktop, Form-pronunciation mapping

Teixeira notes, "students can hear the pronunciation for any presented lexical or phrasal item, reinforcing form-pronunciation mapping" and "can record their voice and compare the shape of the subsequent sound wave to that of the recorded audio" (2014, p. 405), as seen in Figure 5. These abilities to map form-pronunciation content with explicit instruction are wonderful and uncommon among the software analyzed. However, the ability to create waveforms and compare them directly is misleading. It is unclear how non-linguist learners who get no feedback on their utterances could benefit from these features much more than hearing input, producing output, and seeing some attractive images (McMeekin, 2014; Santos, 2011; Teixeira, 2014). This is problematic for development and could be improved by TF.

Results and Discussion

This study provides a snapshot of the current state of CALL software in L2 pronunciation and its great potential for future development. An examination of the presentations of pronunciations in these four programs clearly demonstrates that the CALL software analyzed did not fully realize their innovative potential in providing TF to language learners.

Research Questions and Conclusions

1. Of the software reviewed (Rosetta Stone (RS), Duolingo, Babbel, and Mango Languages (Mango), which use targeted feedback (TF) for L2 student pronunciation development?

Rosetta Stone, Duolingo, and Babbel provided users with feedback about their utterances, whereas Mango provided no feedback to learners about their spoken production. Duolingo provided rudimentary binary feedback. Rosetta Stone provided binary feedback with some attractive waveforms and pitch contours. RS and Mango provided learners with attractive but cryptic waveforms with no explicit instruction of how they might be interpreted. It is unclear how non-linguist learners who get no feedback on their utterances could benefit solely from attractive, unexplained images (McMeekin, 2014; Santos, 2011; Teixeira, 2014). Babbel provided *binary feedback* with more explicit

pronunciation support in the software. None fully exploited the potential of using TF in the software to support language learning.

2. Which of the four provides the best TF for L2 learner development of pronunciation?

Of the four products, Babbel provides the best TF due to the explicit pronunciation instruction found in the software. This software has high-quality voice recognition abilities, explicit instruction of pronunciation, and integration of speech into vocabulary and grammar sections. Babbel included sociocultural information about dialectal variation, had the highest threshold for binary feedback for spoken utterances, and required users to speak to progress through the lessons. If the standard is too low, learners might believe that their utterances are sufficient, such as in RS and Duolingo.

3. What extrapolations and conclusions can be made regarding the ideal L2 pronunciation software from the analysis of these products?

Despite their innovative potential, these products, used by hundreds of millions of people worldwide, employ antiquated learning theories "that were thought to be buried a half century ago" (Lotherington, 2016, p. 9). They do not support the development of spoken skills, even though there is research that clearly demonstrates how TF can be effective even in short periods of time. As Duffy writes of contemporary tools, "Most software-based language programs will help you learn a base of vocabulary and grammar, but they won't turn you into a fluent speaker" (Duffy, 2015a, para. 21). Stemming from the analysis of these productions, below are specific recommendations from this study in relation to CALL software for L2 pronunciation.

Recommendations

- Targeted Feedback (TF): Learners should be provided wherever possible with TF so that they can act on this information and improve their speech. *Binary feedback* is poor support for learners if the only information that learners are getting about their production is that it is correct or incorrect, as defined by the software's thresholds and definitions.

- Explicit Instruction: Clear instructions need to be provided to learners about how to interpret their feedback so they can understand how to improve in specific, actionable ways.

- Automatic Speech Recognition (ASR): Can be helpful for learners in providing immediate TF, but this capability must be explained through explicit instructions rather than being used as an unexplained assessment tool as seen in RS and Duolingo.

- Visual Displays: If pitch contours and waveforms are included, explicit instruction and TF are essential. Otherwise, they are little more than attractive images and may introduce confusion.

- Better Scaffolding: Presentations of pronunciation should be clear and easily understandable by non-linguist learners (Santos, 2011).

The ability to speak clearly and effectively communicate with others is a core component of communicative competence. If software cannot provide support in these skills, they may need significant reimagining. Future software developers would be wise to consider the above recommendations in the design of L2 products for novel, learner-centered tools.

Scientific Images are Not Innovative Support

One of the major findings of this study was that many products are using visuals of waveforms and pitch contours that may seem like sophisticated tools (see Figure 2 of Rosetta Stone software and Figure 5 of Mango) but do not support pronunciation development. Users must be able to interpret graphics to have them provide innovative pedagogical support. However, visualizations of speech without feedback have been demonstrated to be useless in L2 pronunciation development (Kartushina et al., 2015). Especially since hundreds of millions of users and institutions are paying for this software (see Table 1), this usage of images is somewhat misleading and deceptive. If software is going to use this type of information to support learner development, it should not be as an afterthought or visually attractive addition to pronunciation modules (McMeekin, 2014; Santos, 2011; Teixeira, 2014).

Transfer from Textbooks: Low Standards

One of the reasons why RS, Duolingo, and Mango lag behind in their presentations of pronunciation may be in part due to material in textbooks. Recognizing that the current instructional material for classrooms is severely lacking when it comes to pronunciation development (Arteaga, 2000; Lord & Fionda, 2013; Morin, 2007), current CALL programs may be a result of software that is designed from the ineffective pedagogical materials already in circulation. Thus, as compared to what is available in textbooks, companies believe that their tools are "highly advanced" ("Speech Recognition Talking Back Required," 2015, para. 1). However, as demonstrated by the free tools available to anyone, such as Audacity (Mazzoni, 2016), PRAAT (Boersma, 2015), and WASP (Huckvale, 2016), this type of technology has been available free online since 1999. CALL software can move past poor, ineffective pedagogies and employ innovative strategies to support language learners.

CALL software can also work well as an integrated aspect of the language classroom (Blake, 2013). Technology is ubiquitous in social spheres, but when it comes to the language classroom, these tools are not always exploited to their fullest extent (Lotherington, 2016). Outside of the classroom, learners participate and are exposed to

multimodal environments where digital literacy is integral to their navigation of the world around them (Group, 1996; Reinhardt & Thorne, 2011). CALL in language classrooms commonly lags behind technology in the real world (Lotherington, 2016). The CALL software reviewed in this paper can be used in conjunction with in-person instruction for success in hybrid methods as is already being done with Duolingo and RS ("Duolingo: For Schools," 2016; Stone, 2016).

Expect More: The Role of Consumers and Clients

A common mindset amongst consumers is that "you get what you pay for" (Lotherington, 2016, p. 4), and thus if hundreds of dollars are not spent on RS, consumers expect that cheaper costs indicate lower quality. Customers of CALL software are frequently "grateful that [programs] exist at all" (S.A.P., 2013, para. 5), but learners could and should demand more of language products. Learners, instructors, and institutions have agency to tell companies what they want and need from L2 software. Additionally, they can design their own software, which is a concept that is more and more within the reach of faculty in the language field (Lindaman & Nolan, 2016).

Language technology companies are "competing quite vigorously with each other" (S.A.P., 2013, para. 8) in an industry worth over $8 billion dollars that has millions of users (Chen, 2015). It is no coincidence that the free of cost Duolingo is considered to be "nearly as good" in content as the pricey RS by some reviewers (Duffy, 2015a, para. 9). With this competitive market and high demand, learners can "expect some exciting developments" in these programs in the near future (S.A.P., 2013, para. 8). For example, the new tool, ELSA, is designed to target problematic pronunciation segments, like phonemes, for English learners (Van, 2016). ELSA analyzes L2 speech with sophisticated usage of automatic speech recognition in its software. This type of tool may be more common in the future.

Future Directions

There are other widely available tools that can be employed to enhance pronunciation improvement if properly scaffolded. Although PRAAT is most frequently used in research to analyze learners' production (Ladd, 2011; Ladefoged & Johnson, 2011; Simonet, 2010; Solon, 2016), PRAAT can also be used by language learners to evaluate their speech in terms of formants, intensity, and intonation, among other things (Boersma, 2015; Xu & Qiu, 2011). Yet explicit instruction and scaffolding would be necessary for learners to fully harness the potential of these powerful phonetic tools.

Automatic Speech Recognition (ASR) Software: Siri, Alexa, and Ok Google

Also widely available is Siri (Apple, 2016b), which has been recognized by non-academics to be helpful for language learners ("How to Use Siri to Practice and Master a Language," 2011; Simon, 2014). However, this tool is still not a commonly used resource in L2 classrooms. Used as a pedagogical tool, Siri is a "voice-activated personal assistant"

that answers "simple questions" and responds with "simple responses" (Lotherington, 2016, p. 15). If scaffolded correctly, Siri can be used to practice L2 pronunciation to see whether the software finds the L2 learner intelligible and gives immediate responses to the learner. Siri also has dictation capabilities that work with Chrome and Word Documents. Originally a component solely of iOS, Siri has also recently been added as a component to laptop and desktop software in macOS Sierra, version 10.12.1 (Apple, 2016a). Empirical studies could be designed to test how PRAAT, Siri, or any of the other technologies reviewed could be used in the language classroom and assessed as compared to a control group to aid pronunciation development.

References

20 Ways Siri May Forever Change Education. (2011). Retrieved from http://www.
onlinecollege.org/2011/11/27/20-ways-siri-may-forever-change-education/

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2011). Inter-language interference
in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-
switching. *Journal of Phonetics, 39*(4), 558-570.

Apple. (2016a). macOS Sierra. Retrieved from http://www.apple.com/macos/sierra/

Apple. (2016b). Siri: Apple. Retrieved from http://www.apple.com/ios/siri/

Arteaga, D. L. (2000). Articulatory phonetics in the first-year spanish classroom. *The
Modern Language Journal (Estados Unidos), 84*(03), 339-354.

Babbel. (2010). Practice your Pronunciation: New Speech Recognition Tool. *The
Babbel Blog.* Retrieved from http://blog.babbel.com/practice-your-pronunciation-
new-speech-recognition-tool/

Babbel. (2016). About Us. *Babbel.* Retrieved from http://about.babbel.com/en/about-us/

Bajorek, J. P. (2016). *L2 Spanish Learner Attitudes on Pronunciation: Spanish Heritage
and Non-heritage Speakers*. Master's Thesis. Linguistics. University of California,
Davis.

Best, C. T. (1995). A Direct Realist View of Cross-Language Speech Perception.
Speech perception and linguistic experience: Issues in cross-language research,
171-204.

Birdsong, D., Bohn, O., & Munro, M. (2007). Nativelike pronunciation among late
learners of French as a second language. *Language experience in second
language speech learning: In honor of James Emil Flege*, *John Benjamins
Publishing,* 99-116.

Blake, R. J. (2013). *Brave new digital classroom : technology and foreign language
learning / Robert J. Blake*: Georgetown University Press.

Boersma, P. W., David. (2015). Praat: doing phonetics by computer [Computer
program]. Version 5.4.22. http://www.praat.org/.

Bongaerts, T., Mennen, S., & Slik, F. v. d. (2000). Authenticity of pronunciation in
naturalistic second language acquisition: The case of very advanced late
learners of Dutch as a second language. *Studia linguistica, 54*(2), 298-308.

Celce-Murcia, M., Brinton, D., & Goodwin, J. M. (2010). *Teaching pronunciation: a course book and reference guide* (Vol. 2nd). New York: Cambridge University Press.

Chapelle, C. (2001). *Computer applications in second language acquisition: foundations for teaching, testing and research*. Cambridge: Cambridge University Press.

Chen, A. (2015). The Ever-Growing Ed-Tech Market. *Education.* Retrieved from http://www.theatlantic.com/education/archive/2015/11/quantifying-classroom-tech-market/414244/

Chen, H. H.-J. (2011). Developing and evaluating an oral skills training website supported by automatic speech recognition technology. *Recall, 23*(1), 59-78. doi:10.1017/S0958344010000285

Colantoni, L., & Steele, J. (2006). *Native-like attainment in the L2 acquisition of Spanish stop-liquid clusters.* Paper presented at the Selected proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages.

Colantoni, L., Steele, J., & Escudero, P. (2015). *Second Language Speech*: Cambridge University Press.

Crisi, B. a. (2010). Tech Background: Babbel Speech Recognition: Interview with Technical Director Thomas Holl. *The Babbel Blog.* Retrieved from http://blog.babbel.com/tech-background-babbel-speech-recognition/

DeKeyser, R. (2010). Practice for Second Language Learning: Don't Throw out the Baby with the Bathwater. *International Journal of English Studies, 10*(1), 155-165.

Demenko, G., Wagner, A., & Cylwik, N. (2010). The Use of Speech Technology in Foreign Language Pronunciation Training. *Archives of Acoustics, 35*(3), 309-329. doi:10.2478/v10168-010-0027-z

Derwing, T. M., & Rossiter, M. J. (2003). The Effects of Pronunciation Instruction on the Accuracy, Fluency, and Complexity of L2 Accented Speech. *Applied Language Learning, 13*(1), 1-17.

Doughty, C., & Williams, J. (1998). *Focus on form in classroom second language acquisition*. Cambridge, UK; New York: Cambridge University Press.

Duffy, J. (2015a). The Best Language-Learning Software of 2015. *Education*. Retrieved from http://www.pcmag.com/article2/0,2817,2381904,00.asp

Duffy, J. (2015b). Mango Languages.   Retrieved from http://www.pcmag.com/article2/0,2817,2489427,00.asp

Duffy, J. (2016a). Babbel. *PC Mag.* Retrieved from http://www.pcmag.com/article2/0,2817,2411398,00.asp

Duffy, J. (2016b). Duolingo. *PC Mag.* Retrieved from http://www.pcmag.com/article2/0,2817,2402570,00.asp

Duffy, J. (2017). The Best Language-Learning Software of 2017. *PC Mag.* Retrieved from https://www.pcmag.com/article2/0,2817,2381904,00.asp

Duñabeitia, J. A., Orihuela, K., & Carreiras, M. (2014). Orthographic coding in illiterates and literates. *Psychol Sci*, 0956797614531026.

Duolingo Language Courses. (2016). Retrieved from https://www.duolingo.com/courses/all

Duolingo: For Schools. (2016). Retrieved from https://schools.duolingo.com/

Elliott, A. R. (1995). Foreign Language Phonology: Field Independence, Attitude, and the Success of Formal Instruction in Spanish Pronunciation. *The Modern Language Journal, 79*(4).

Elliott, A. R. (1997). On the teaching and acquisition of pronunciation within a communicative approach. *Hispania*, 95-108.

Elliott, A. R. (2003). Phonology. Staking out the territory at the turn of the century : integrating phonological theory, research, and the effect of formal instruction on pronunciation in the acquisition of Spanish as a second language. In B. A. Lafford, and M. Rafael Salaberry. (Ed.), *Spanish second language acquisition: state of the science*. Washington, D.C.: Georgetown University Press.

Elsendoorn, B. (1980). Vowel durations as a cue to a Dutch foreign accent in English CVC-words. *Progress Report of the Institute of Phonetics, 5*, 64-86.

Eskenazi, M. (1999). Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype. *Language Learning & Technology, 2*(2), 62-76.

Fitzpatrick, A. (2017). Letter of Recommendation: Duolingo. *The New York Times Magazine, July 27, 20*.

Flege. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*(1), 47-65.

Flege. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America, 89*(1), 395-411.

Flege, J., & Hammond, R. M. (1980). Stereotypes in foreign accent- A preliminary analysis. Paper presented at the Southeast Conference on Linguistics, Memphis, April, 1980.

Flege, J. E. (1980). Phonetic Approximation in Second Language Acquisition. *Language Learning, 30*(1), 117-134.

Flege, J. E. (1981). The Phonological Basis of Foreign Accent: A Hypothesis. *TESOL Quarterly, 15*(4), 443-455.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, 233-277.

Flege, J. E., & Liu, S. (2001). The Effect of Experience on Adults' Acquisition of a Second Language. *Studies in Second Language Acquisition, 23*(04), 527-552.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America, 97*(5), 3125-3134.

Garcia, I. (2013). Learning a Language for Free While Translating the Web. Does Duolingo Work? *International Journal of English Linguistics, 3*(1), 19-25.

Gass, S. M. (2013). *Second language acquisition: An introductory course*: Routledge.

Gass, S. M., & Mackey, A. (2007). Input, interaction, and output in second language acquisition. *Theories in second language acquisition: An introduction, 175199*.

Gollan, T. H., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew–English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(5), 1122.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer Assisted Language Learning, 27*(1), 70-105. doi:10.1080/09588221.2012.700315

Gonzalez-Bueno, M. (1997). Voice-Onset-Time in the Perception of Foreign Accent by Native Listeners of Spanish. *IRAL, 35*(4), 251-267.

Grosjean, F. (2001). The Bilingual's Language Modes. One Mind, two Languages: Bilingual Language Processing. J. Nicol. *Blackwell, 1*, 1-22.

Group, T. N. L. (1996). A Pedagogy of Multiliteracies: Designing Social Features. *Harvard Educational Review, 66*(1), 60-92.

Hacker, V. A. (2011). Duolingo. Pittsburgh, US.

Han, Z., & Odlin, T. (2006). *Studies of Fossilization in Second Language Acquisition*: Multilingual Matters.

Hardison, D. M. (2005). Contextualized Computer-Based L2 Prosody Training: Evaluating the Effects of Discourse Context and Video Input. *CALICO Journal, 22*(2), 175.

Higher Education. (2016). Retrieved from http://info.mangolanguages.com/higher-ed

Hincks, R. (2003). Speech Technologies for Pronunciation Feedback and Evaluation. *ReCALL: The Journal of EUROCALL, 15*(1), 3-20. doi:10.1017/S0958344003000211

History. (2016).   Retrieved from http://www.rosettastone.com/history

How to Use Siri to Practice and Master a Language. (2011). Retrieved from https://www.englishclass101.com/2011/12/24/news-59-how-to-use-siri-to-practice-and-master-a-language/

Huckvale, M. (2016). WASP: Waveform, Spectrogram & Pitch: University College London. Retrieved from http://www.phon.ucl.ac.uk/resource/sfs/wasp.php

Huettig, F., & Mishra, R. K. (2014). How literacy acquisition affects the illiterate mind–a critical examination of theories and evidence. *Language and Linguistics Compass, 8*(10), 401-427.

Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in Second Language Acquisition, 27*(02), 129-140.

Hymes, D. (1971). The Contribution of Folklore to Sociolinguistic Research. *The Journal of American Folklore, 84*(331), 42-50. doi:10.2307/539732

Jones, R. H. (1997). Beyond "Listen and Repeat:" Pronunciation Teaching Materials and Theories of Second Language Acquisition. *System, 25*(1), 103-112.

Kartushina, 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 soundsa). *The Journal of the Acoustical Society of America, 138*(2), 817-832.

Katz, W. F., & Mehta, S. (2015). Visual feedback of tongue movement for novel speech sound learning. *Frontiers in Human Neuroscience, 9*. doi:10.3389/fnhum.2015.00612

Kennedy, S., Blanchet, J., & Trofimovich, P. (2014). Learner Pronunciation, Awareness, and Instruction in French as a Second Language. *Foreign Language Annals, 47*(1), 76-96.

Kissling, E. M. (2013). Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners? *MODL The Modern Language Journal, 97*(3), 720-744.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford; New York: Pergamon.

Ladd, D. R. (2011). Phonetics in phonology. *The Handbook of Phonological Theory, Second Edition*, 348-373.

Ladefoged, P., & Johnson, K. (2011). *A course in phonetics*. Boston, MA: Wadsworth/Cengage Learning.

Language-Learning App Babbel Hits One Million Customers. (2016).   Retrieved from http://press.babbel.com/en/releases/2016-02-17-Language-Learning-App-Babbel-Hits-One-Million-Customers.html

Languages, M. (2016). About Us.   Retrieved from http://mangolanguages.com/company/

Larsen-Freeman, D. A. M. (2011). *Techniques and principles in language teaching*. Oxford: Oxford University Press.

Lee, B., Guion, S. G., & Harada, T. (2006). Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition, 28*(03), 487-513.

Levis, J. M. (2005). Changing Contexts and Shifting Paradigms in Pronunciation Teaching. *TESOL Quarterly: A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Dialect, 39*(3), 369-377.

Lindaman, D., & Nolan, D. (2016). Mobile-Assisted Language Learning: Application Development projects within reach for language teachers. *IALLT Journal of Language Learning Technologies, 45*(1), 1-22.

Long, M. H. (2000). Focus on form in task-based language teaching. *Language policy and pedagogy: Essays in honor of A. Ronald Walton*, 179-192.

Lord, G. (2005). (How) Can We Teach Foreign Language Pronunciation? On the Effects of a Spanish Phonetics Course. *Hispania, 88*(3), 557-567.

Lord, G., & Fionda, M. I. (2013). Teaching Pronunciation in Second Language Spanish *The Handbook of Spanish Second Language Acquisition* (pp. 514-529): John Wiley & Sons, Ltd.

Lotherington, H. (2016). What's app? Negotiating the good, bad, and ugly of apps for (English and other) language learning. *Canadian Society for the Study of Education, CHSS, Calgary, AB,* 1-23.

Mathieu, L. (2014). The Influence of Unfamiliar Orthography on L2 Phonolexical Acquisition.

Mazzoni, D. (2016). Audacity: Audacity 2.1.2. Retrieved from http://www.audacityteam.org/

McKinnon, D. (2011). Babbel review. *Software.* Retrieved from http://www.expertreviews.co.uk/software/50392/babbel-review

McMeekin, A. (2014). Software Review: Mango Conversations—Japanese for English Speakers. *CALICO Journal, 32*(1), 199-210.

Mitleb, F. (1981). Timing of English vowels spoken with an Arabic accent. *Research in Phonetics, 2*, 193-226.

Morin, R. (2007). A Neglected Aspect of the Standards: Preparing Foreign Language Spanish Teachers to Teach Pronunciation. *FLAN Foreign Language Annals, 40*(2), 342-360.

Morley, J. (1987). *Current Perspectives on Pronunciation. Practices Anchored in Theory*: Teachers of English to Speakers of Other Languages.

Morley, J. (1991). The Pronunciation Component in Teaching English to Speakers of Other Languages. *TESOL Quarterly, 25*(3), 481-520.

Niemi, J. (1979). Noun and noun phrase stress: A phonetic study of English supplemented with an error analysis using Finnish speaker-hearers, I. Production. *Language and Speech, 22*(3), 221-235.

Mango for Public Libraries. (2016). Retrieved from http://info.mangolanguages.com/public-library

Reinhardt, J., & Thorne, S. (2011). Beyond comparisons: Frameworks for developing digital L2 literacies. *Present and future promises of CALL: From theory and research to new directions in language teaching*, 257-280.

Rodman, R. (1999). *Computer speech technology*. Boston: Artech House.

S.A.P. (2013). Language-learning software Review: Babbel and Duolingo. *The Economist*. Retrieved from http://www.economist.com/blogs/johnson/2013/06/language-learning-software

Sancier, M. L., & Fowler, C. A. (1997). Gestural Drift in a Bilingual Speaker of Brazilian Portuguese and English. *Journal of Phonetics, 25*(4), 421-436.

Santos, V. D. O. (2011). Software Review: Rosetta Stone Portuguese (Brazil) levels 1, 2, & 3 Personal Edition Version 4 (TOTALe). *CALICO Journal, 29*(1), 177-194.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *Consciousness in second language learning, 11*, 237-326.

Schmidt, R. (1995). *Attention and Awareness in Foreign Language Learning*. University of Hawaii at Manoa. Second Language Teaching Curriculum Center: Second Language Teaching & Curriculum Center, University of Hawaii at Manoa.

Simon, Z. (2014). *Can Your Phone Teach You a New Language?*, HuffPost. Retrieved from https://www.huffingtonpost.com/zach-simon/can-your-phone-teach-you-_b_4823704.html

Simonet, M. (2010). Dark and clear laterals in Catalan and Spanish: Interaction of phonetic categories in early bilinguals. *Journal of Phonetics, 38*(4), 663-678. doi:10.1016/j.wocn.2010.10.002

Solon, M. (2016). Do Learners Lighten Up?: Phonetic and Allophonic Acquisition of Spanish/l/by English-Speaking Learners. *Studies in Second Language Acquisition*, 1-32.

Speech Recognition Talking Back Required. (2015). *See All Languages.* Retrieved from http://www.rosettastone.com/speech-recognition

Stone, R. (2016). Language-Learning Solutions for K-12 and Higher Ed Students. *Higher Education Benefits.* Retrieved from http://www.rosettastone.com/education

Swad, S. a. P., Tom. (1992). Rosetta Stone. Arlington, Virginia, USA: Fairfield Language Technologies.

Szubko-Sitarek, W., Salski, L., & Stalmaszczyk, P. (2013). *Language Learning, Discourse and Communication: Studies in Honour of Jan Majer*: Springer International Publishing.

Teixeira, A. B. (2014). Software Review for Mango Conversations (Spanish). *CALICO*

Journal, 31(3), 401-408. doi:10.11139/cj.31.3.401-408

The Tenets of A/B Testing from Duolingo's Master Growth Hacker. (2017). *First Round Review.* Retrieved from http://firstround.com/review/the-tenets-of-a-b-testing-from-duolingos-master-growth-hacker/

Terrell, T. D. (1989). Teaching Spanish pronunciation in a communicative approach. In P. C. Bjarkman & R. M. Hammond (Eds.), *American Spanish pronunciation:theoretical and applied perspectives*. Washington, D.C.: Georgetown University Press.

Teshuba, J. (2016). Mango Languages. Retrieved from http://mangolanguages.com/index.html

Thomson, R. I. (2011). Computer Assisted Pronunciation Training: Targeting Second Language Vowel Perception Improves Pronunciation. *CALICO Journal, 28*(3), pp. 744-765. doi:10.11139/cj.28.3.744-765

Thomson, R. I., & Derwing, T. M. (2014). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics, 36*(3), 326-344.

Van, V. (2016). ELSA. Retrieved from https://www.elsaspeak.com/home

Von Ahn, L. (2014). Vastly Improved Speaking Exercises in Chrome. *Duolingo.* Retrieved from https://www.duolingo.com/comment/1880538

Von Ahn, L. (2015). Chrome speaking challenge experiment. Retrieved from https://www.duolingo.com/comment/8491737

Von Ahn, L., & Hacker, S. (2011). Duolingo. Pittsburgh, US. .

Waniek-Klimczak, E. (2013). Selected Observations on the Effect of Rhythm on Proficiency, Accuracy and Fluency in Non-native English Speech. In W. Szubko-Sitarek, L. Salski, & P. Stalmaszczyk (Eds.), *Language Learning, Discourse and Communication: Studies in Honour of Jan Majer*: Springer International Publishing, 167-181.

Wiggins, G. (2012). Seven Keys to Effective Feedback. *Feedback for Learning: ASCD Educational Leadership*, 10-16. Retrieved from http://www.ascd.org/publications/educational-leadership/sept12/vol70/num01/Seven-Keys-to-Effective-Feedback.aspx

Witte, M., & Holl, T. (2016). Babbel: Lesson Nine GmbH. Retrieved from https://www.babbel.com

Xu, Q., & Qiu, D. (2011). *Praat-assisted English pronunciation teaching*. Paper presented at the International Conference on Electrical and Control Engineering.