TECHNOLOGY REVIEW

Duolingo

Kevin Hirschi, Northern Arizona University

INTRODUCTION

Duolingo is likely the most used language learning app as evidenced by search results for *language learning* indicate by far the most popular (9,429,251 ratings in the Google Play Store as of this review). It has evolved greatly since its first unveiling in 2012 as a gamified translation app and has been the focus of numerous reviews (Munday, 2017; Savvani, 2018; Teske, 2017) and several intervention studies (Loewen et al., 2019; Vesselinov & Grego, 2012).

The present review focuses only on the pronunciation-related elements of the app. Several lessons were completed by the reviewer in a number of First Language (L1) and Target Language (TL) pairs (French and Turkish for speakers of English as well as English for speakers of Spanish). All were done on a recent iPhone and iPad; however, it should be noted that the app is available for a variety of other platforms and subtle differences exist (see Teske, 2017). After a description of the pronunciation-related features, an evaluation of the learning potential is conducted in terms of target constructs and feedback.

DESCRIPTION

Once the user chooses an available L1-TL pair, Duolingo's home screen presents options with minimal text and a number of icons for *Lessons*, *Stories*, *Audio Lessons*, *Profile* (user information and setting), *Team* (league or leaderboard), and *Shop*. Two of these sections (*Stories* and *Audio Lesson*) are available for only some L1-TL pairs. The bulk of the learning takes place in the *Lessons* which are labeled in a mixture of linguistic item categories and communicative functions (e.g., *Phrases, Travel, Requests, Narrative*). Each unit (or topic) is designed to be repeated several times and awards the user with points upon successful completion. Lessons are grouped into levels that begin with a checkpoint that can serve as a competency test to allow access to higher-level lessons. There is also a placement test option when starting a new L1-TL pair.

Within a lesson, the learning is implicit and repetitive using a variety of item types. Implicit instruction is achieved by skipping the *presentation* phase of instruction and immediately providing tasks with new vocabulary. In tasks that include the presentation of TL text, the user can see a translation by tapping on individual words. Additionally, each task is accompanied by a link to a user-driven forum that discusses the word or phrase with occasional peer answers to questions. It was through these forums that it became evident that items with translations are presented with different task types (e.g., comments about translations with different formality or alternate spellings were not relevant to the matching task at hand).

Figure 1. Home screen (left) and screen within a listening translation task (right).

There are six types of tasks that recycle vocabulary items with varying difficulty: listening, dictation, multiple choice, pair matching, translation, and speaking tasks. The most relevant to pronunciation are the listening, dictation, and speaking. The listening tasks at lower levels include individual phonemes or words, but in higher levels they tend to be complete sentences or phrases. While they do not require the learner to speak, they allow for connections to be made between listening and speaking through consolidating spelling-sound relationships. The spoken voice is computer-synthesized in all of the regular *Lessons* and several task types include a button with the image of a turtle to slow down the rate of speech. Listening tasks can elicit interaction through multiple choice responses, gapped or full translation, or dictation. A word bank is shown more often in early levels for open-ended responses and words are spoken by the computer-generated voice when tapping on them. Later levels require keyboard input (see Figure 1).

Speaking tasks that use the device's microphone seem sparse in comparison to other task types, however they increase in later laevels. They typically ask the learner to repeat a sentence of previously seen vocabulary items or phrases. In a similarly implicit approach, dichotomous feedback is given (i.e., correct or incorrect) for ASR-evaluation the learner's utterance. For phrases and sentences, the word turns blue if detected by the ASR, allowing for some element of visual feedback. In testing in speaking tasks for English-speaking learners of the three L1-TL pairs reviewed, the ASR seemed to give positive feedback for attempts with phonemes that were accurate or subtly different. For example, /ɪ/ and /i/ were equally accepted in one case and nasalization vowels in French were not always checked (see Figure 2).
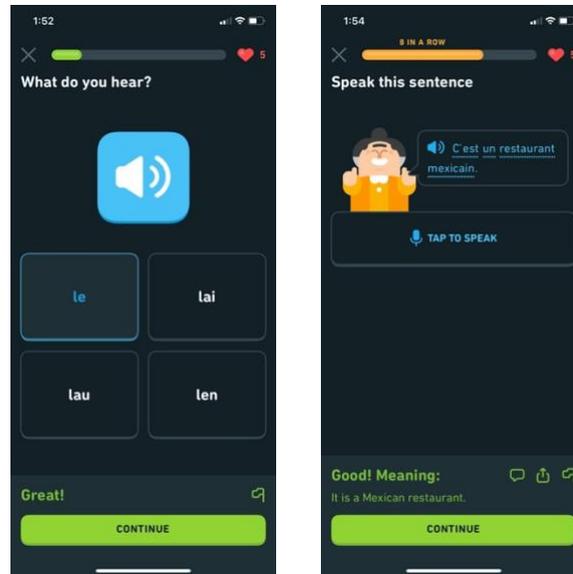
Figure 2. A listening task (left) and a speaking task (right) in a standard lesson.

Outside of the standard lessons, *Stories* are available for a limited number of L1-TL pairings. These depart greatly from the approach of the lessons in that they are comprised of locales or situations such as *A date, The Exam, The Reservation,* or *Happy Birthday*. Each story is a self-contained script that is read by native speakers (often a narrator, a male character, and a female character) and interwoven with tasks that check comprehension or vocabulary items. Many of them contain an interesting or fun plot twist such as a young girl seeing a photo in a museum that ends up being her great-great grandmother or a friend refusing an invitation to see a band because, unbeknownst to him, she is the guitarist in the band.

Audio lessons, available for an even fewer L1-TL pairs, resemble commercial audio cassette or CD programs of times past. These guide the learner through key phrases organized into categories such as *Greetings* or *Directions* in four lesson levels. True to their name, the lessons are entirely audio (i.e., no text is presented on the screen) during the 2-5-minute lesson. An introduction is done largely in L1 with a TL phrases sprinkled in with ample translation. The presentation stage includes friendly voices of a variety of native speakers of both L1 and the TL as well as background noises reminiscent of travel destination of the TL (i.e., Paris for learners of French). Typically, an anchor phrase is given and taught with a few alternative vocabulary items should there be a need to swap one out.

However, the Audio Lessons diverge greatly from the well-known static audio cassettes of the past when the learner is intermittently presented with an ASR-speaking task within the lesson. There are four or more speaking tasks within each audio lesson and as they begin automatically, they require constant attention to the lesson and require interaction. These are followed with positive or negative feedback, allowing for a second or third chance using a native speaker voice in the L1. In several attempts of using this feature, results were similar to speaking tasks in the regular *Lessons* in that slight segmental variations were accepted (see Figure 3).
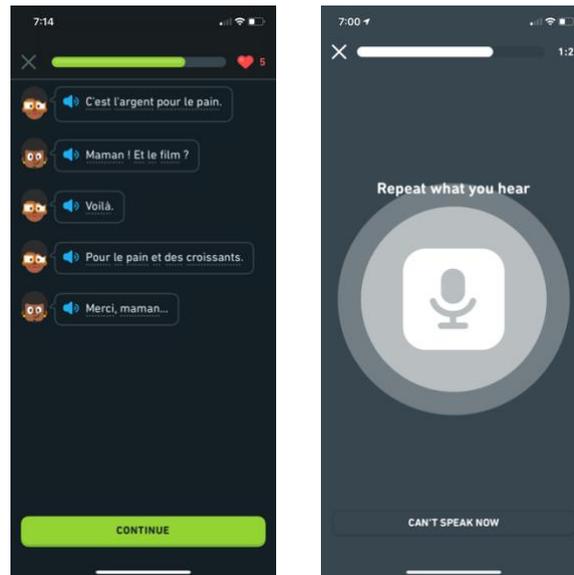
Figure 3. The end of a *Story* (left) and the screen during a speaking task of an *Audio Lesson* (right).

EVALUATION

Established target constructs for pronunciation teaching focus on comprehensibility and intelligibility over accent-reduction as realistic goals (Munro & Derwing, 1995). While Duolingo does not outline these targets explicitly, the use of synthetic voice as well as the ASR feedback that seems to accept slight segmental variations may not penalize accented speech that is otherwise intelligible. In minimal pair exercises, synthetic voice feedback has been found to help pronunciation in CAPT settings (Tejedor-Garcia et al., 2017). Additionally, as synthetic text-to-speech has been implemented in a variety of task types, a reinforcement of spelling-sound relationships for non-phonetic languages poses great advantages.

Duolingo *Stories* and *Audio Lessons* play a different role of TL examples as they feature a dialogue with recordings of native speakers. In the case of *Stories*, input is provided in suprasegmental features such as emotional prosody in cases of surprise, jest, or frustration; prominence to clarify a misunderstanding; and question intonation. Comprehension checking tasks may help learners to internalize these features, but they are not the focus of the lesson and always accompany by text that would not require the use of prosody for semantic distinction. As suprasegmental features are important for communication (Kang, 2010) and can be taught explicitly (Lee, Jang, & Plonsky, 2015), Duolingo *Stories* present an opportunity for learners to combine their prosodic knowledge with vocabulary to understand dialogue.

Audio Lessons and their ASR-enabled responses seem to prioritize segmental features and, in several attempts of TL utterances of a highly proficient non-native speaker, seem to reject accurate responses perhaps due to background noise or distance from a device microphone. These faults are

not uncommon in ASR-based feedback systems and will likely improve as technology continues to evolve. As the feedback is dichotomous, it can be difficult to know why the attempt is rejected.

However, *Audio Lessons* do provide a few advantages not found in other Duolingo activities. First, they are entirely audio with no visual input, simulating a realistic conversation where there are little to no textual clues to assist comprehension. Second, they promote and assist a second or third attempt in pronouncing a target item, sometimes through chunking. If a third attempt is not successful, the narrator says, "that's OK" and the lesson continues. Finally, *Audio lessons* seem to connect target items with a setting through a background track associated with the setting (e.g., the hum of conversations in a café, street noises, a metro station), allowing for additional connections to be made by the learner.

Throughout Duolingo, feedback is given dichotomously (i.e., a response is correct or incorrect). When applied to translating simple sentences in conjunction with gamified rewards, this could be effective. However, there is strong evidence of corrective feedback improving pronunciation across learning settings (Lee et al., 2015). As seen in vocabulary and grammar through user forums which allow multiple perspectives on a variety of issues, feedback-based pronunciation approaches would be desired. Perhaps future versions will allow for more detailed speaking feedback or a simple side-by-side playback buttons to compare a target form and the learner's utterance for comparison. As such, it is not surprising that the limited empirical research into the use of Duolingo found a positive relationship with app use and proficiency gains when combined with classroom learning as face-to-face interaction may allow for more detail in feedback (Loewen et al., 2019).

CONCLUSION

Duolingo has advanced considerably since its original gamified translation app. The ASR-enabled tasks and rich context provided through input in *Stories* and *Audio Lessons* allow for a variety of possibilities for learners to practice pronunciation. Much of highlights of this review focused on the available input for pronunciation that can build on communicative competence, potentially, though not explicitly, through a lens of comprehensibility and intelligibly. Even though users may have to make discerning interpretations of the ASR-based feedback, the elements of the app that consolidate learning through repetition are likely to promote proficiency. Additionally, the motivational aspects encourage learners to persist in their efforts within the app or leverage additional resources for pronunciation learning.

ABOUT THE AUTHOR

Kevin Hirschi is a Ph.D. student in Applied Linguistics at Northern Arizona University. He is interested in using corpus linguistics to describing phonological features of comprehensibility and intelligibility relevant to L2 learners of English and French. He is also collaborating with research on efficacy and interaction with Mobile-assisted Pronunciation Training applications.

REFERENCES

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language learning*, *45*(1), 73-97.

Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris & Ortega (2000) revisited and updated. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443-483). Amsterdam: John Benjamins.

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, *38*(2), 301-315.

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*(3), 345-366.

Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL*, 31(3), 293-311.

Munday, P. (2017). Duolingo. Gamified learning through translation. *Journal of Spanish Language Teaching*, *4*(2), 194-198.

Savvani, S. (2018, September). State-of-the-art duolingo features and applications. In *International Conference on Interactive Collaborative Learning* (pp. 139-148). Springer, Cham.

Teske, K. (2017). Duolingo. *CALICO Journal*, *34*(3), 393–401. https://doi.org/10.1558/cj.32509

Tejedor-Garcia, C., Escudero-Mancebo, D., González-Ferreras, C., Cámara-Arenas, E., & Cardenoso-Payo, V. (2017). Evaluating the efficiency of synthetic voice for providing corrective feedback in a pronunciation training tool based on minimal pairs. *Proc. SLaTE*, 26-30.

Vesselinov, R. and Grego, J. 2012. Duolingo effectiveness study. Final Report. Retrieved from https://static.duolingo.com/s3/DuolingoReport_Final.pdf