TECHNOLOGY REVIEW

ELSA Speak - Accent Reduction

Kimberly Becker, Iowa State University
Idée Edalatishams, Iowa State University

INTRODUCTION

ELSA Speak - Accent Reduction is an application (app) for reducing a non-native English accent. *ELSA* (an acronym for "English Language Speech Assistant") has a free version and a paid-for version for Android and Apple products. *ELSA*'s paid version can be purchased for $3.99 monthly, $8.99 for three months, or $29.99 for one year. As described on their website, this app employs "proprietary artificial intelligence" (AI) including automatic speech recognition (ASR) to provide feedback on users' pronunciation accuracy. This review will provide an overview of the basic features before moving to a critical evaluation of the free version of the app.

Description

ELSA's basic interface is simple and provides navigation between topics and skills, levels, reports, and other features. Users can set a display language (English, Vietnamese, or Japanese – good indicators of the target audience for this app). Depending on the proficiency level of the user, the skills include exercises such as beginning and ending sounds, minimal pairs, schwa, *th*-sounds, and consonant clusters. Applied linguists typically refer to these as segmental phonemes—individual sounds within a word. The topics include food, entertainment, technology, culture, daily conversations, and relationships, among many others. Figure 1 demonstrates the default skills-based page, which appears after the user chooses their proficiency level.

After choosing a skill (e.g., "/S/ /SH/" as shown in Figure 2), users can select a topic, under which they can choose from several lessons. Each lesson is labeled for proficiency level and includes various speaking and listening opportunities. Figure 3 illustrates an exercise at the *regular* level with the target sound underlined.
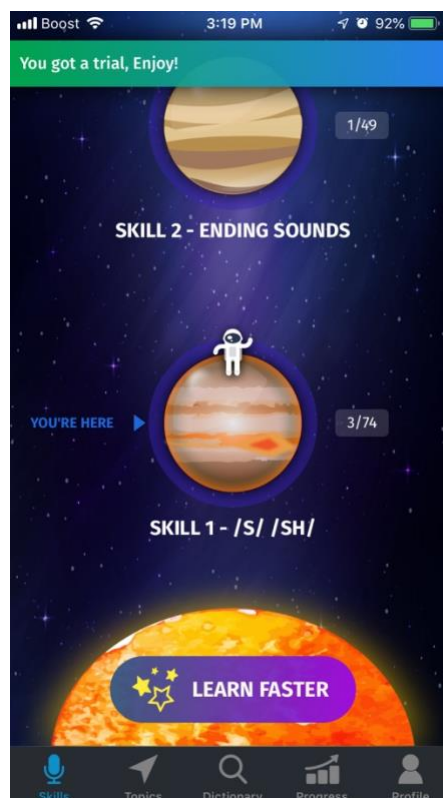

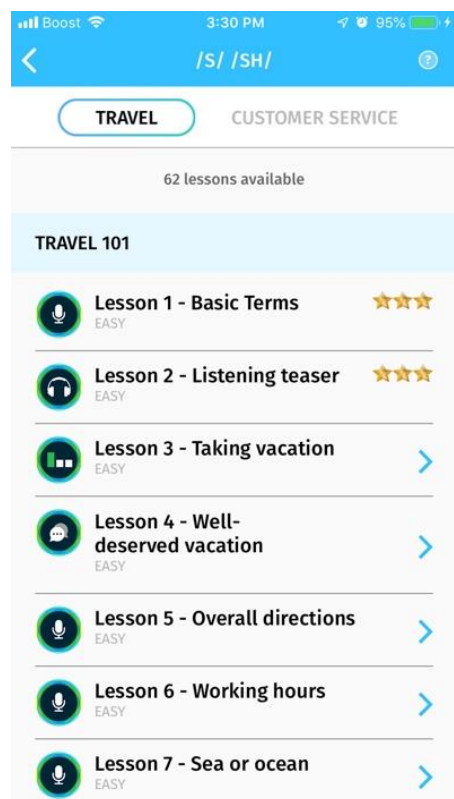*Figure 1*. The skills-based start-up page.

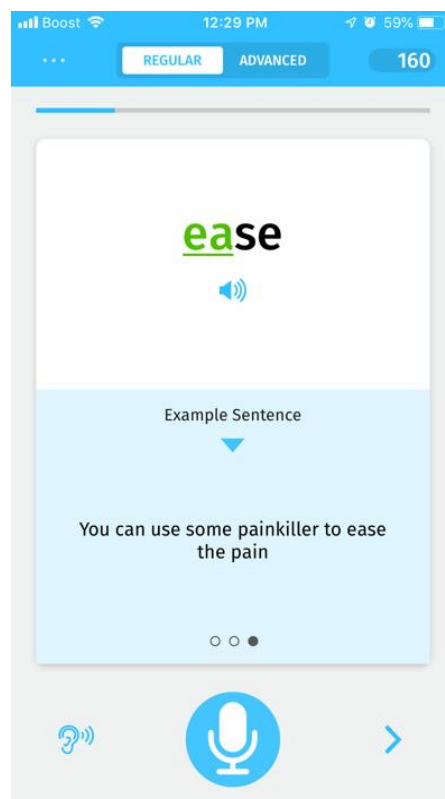*Figure 2*. Lessons for the /S/ /SH/ "skill".

Figure 3. An example activity at the *regular* level.

Users listen to the pronunciation of the target sound and then record themselves. *ELSA* determines the accuracy of the recording and provides feedback for correct and incorrect pronunciations. Correct pronunciation elicits a bell sound accompanying a score up to 100; incorrect pronunciation elicits a buzzing sound along with the correct pronunciation of the sound or word and an opportunity to review the recorded incorrect sound for comparison purposes. An example of feedback given for a user's incorrect pronunciation can be seen in Figure 4.

Besides organizing by skill, *ELSA* also provides "Topics" in the bottom menu, as shown in Figure 5.
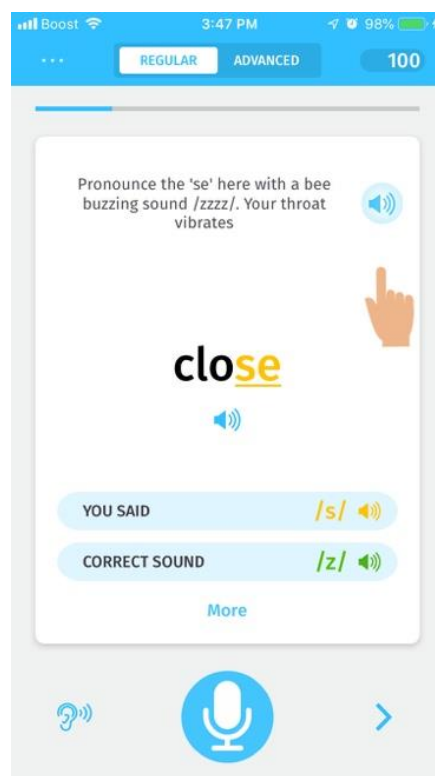

Figure 4. Incorrect pronunciation feedback.

Figure 5. The topic-based page.

The app also includes a dictionary that provides the pronunciation of any word plus the opportunity to connect to Youglish.com, a website offering YouTube clips with audiovisual emphasis on particular sounds, words, or phrases. *ELSA* also offers a multi-faceted feature called "Progress" containing a "Word Bank," "*ELSA* Pronunciation Score" (EPS), and "Assessment." "Word Bank" offers a progress summary in three sections: "Word Sound," "Word Stress," and "Conversation." Word sound specifies feedback about individual phonemes and word stress assesses syllable stress.

The second progress feature, EPS, is a collection of scores from *ELSA* activities. The third progress feature, the "Assessment," allows users to record 13 sentences and then gives feedback on segmental issues such as consonant clusters, aspiration sounds, and schwa. *ELSA* provides a percentage correct score, and users can also get a detailed report. Figure 6 illustrates an example of a detailed report.

The green letters indicate that the user correctly pronounced the sound, the yellow letters indicate that the sound was mostly correct, and the red letters means that the sound was incorrectly pronounced. Despite the fact that the test is marketed as a way to measure "speaking proficiency level" and prompts users to produce sentences, feedback is still focused on individual segmental sounds.

Figure 6. An example of detailed report.

Besides these features, *ELSA* offers a user "Profile," where users can customize settings about feedback, sharing, and notifications. Additionally, in the profile, users can identify goals, set a daily reminder, and view the app's terms and policies.

EVALUATION

As claimed on its website, *ELSA* has been featured on several digital media websites and magazines and has 4.5/5 ratings from 23,000 users. However, this pronunciation app could improve in many ways. This section will provide a critical evaluation of this pronunciation app in

relation to best practices for Mobile Assisted Language Learning (MALL), using Reinders and Pegrum's (2016) framework.

One major shortcoming is that *ELSA* focuses only on segmental aspects of pronunciation. The spotlight on individual sounds is not an issue in and of itself; however, the lack of balance between segmental and suprasegmental practice opportunities is problematic. Pronunciation research literature shows that focusing pronunciation instruction on suprasegmental features improves comprehensibility and fluency more than segmental focus (e.g., Derwing & Rossiter, 2003; Kang, Rubin, & Pickering, 2010). In fact, Kang et al. (2010) note that "listeners can tolerate a great deal of inaccuracy in pronouncing consonants and vowels," (p. 555) as long as suprasegmental features are somewhat accurate. Other research also demonstrates the value of suprasegmental features (e.g., intonation and speaking rate) over segmental features (Anderson-Hsieh, Johnson, & Koehler, 1992; Derwing & Munro, 1997). While two of the categories in Reinders and Pegrum's (2016) framework for MALL app evaluation relate app design to pedagogical approaches, *ELSA* has almost no focus on suprasegmental features that are likely to be the most beneficial in accent reduction.

Another category in Reinders and Pegrum's (2016) framework is the potential for educational affordances. *ELSA*'s developers have misrepresented the abilities of their proprietary AI system, which often mistakenly identifies incorrect sounds as correct. The documentation does not provide any information about its accuracy or piloting. Moreover, *ELSA* seems to have been developed with a focus on quantity over quality. This app has hundreds, if not thousands, of individual exercises about segmental phonemes, but with an inaccurate system for catching mispronunciations combined with inattention to suprasegmental features, it is not likely to be effective for English language learning experts or discerning learners. Even with the paid version of the app, users only get access to more lessons; it does not offer a substantially better version except in providing extra exercises about the same skills and topics.

Correspondence of app design to principles of second language acquisition (SLA) is another category in Reinders and Pegrum's (2016) framework. In a meta-analysis of studies in second language pronunciation instruction, Lee, Jang, and Plonsky (2014) found that an important aspect among many of the studies was the inclusion of both segmentals and suprasegmentals. Lee et al. (2014) further suggest three important aspects of SLA as related to pronunciation: (1) using segmental and suprasegmental approaches, (2) aligning lessons with needs analyses, and (3) considering demographic information such as learner backgrounds and/or first languages. *ELSA* implements neither a suprasegmental approach nor any needs analysis. Because users can input information about their language background in their user profiles, it can be argued that the third of these aspects has been considered in *ELSA*'s design, although it is not clear how the designation of a first language affects the interaction between the user and the app.

The final category of the MALL evaluation framework is related to affect, which Reinders and Pegrum (2016) define as engagement and attention to affective filter. As accurate evaluation of the *ELSA* app in this category would require assessments from non-native speaking users, it will not be discussed here but would be an appropriate consideration for further development of the app.

Several of these issues can be at least partially explained by the fact that the *ELSA* development team does not include any applied linguists or other English language learning experts. Although the 11 employees featured on the app's website have backgrounds in speech processing, software development, and engineering, a linguistics or an SLA perspective is necessary to make this app more useful for the audience.

Finally, a seemingly surface-level (but important) criticism about the *ELSA* Corporation website and app are the multiple typos. These include but are not limited to missing words, pluralized noncount nouns (e.g., "feedbacks"), and confusing if not conflicting references to numbers and other statistics without reference to any sources. One typo that stood out as particularly egregious is the misspelling of the word *diphthong [sic]* on the assessment page.

CONCLUSION

ELSA is an app that has made great strides in the world of AI for practicing the pronunciation of individual sounds. However, in order for it to break new ground in the world of accent reduction, it needs to be expanded to include suprasegmental aspects of pronunciation. The lack of attention to what many applied linguists consider the most important part of accent reduction combined with smaller issues such as typos cause the app to lose face with the population of experts (language teachers) who are best positioned to both evaluate and market it.

REFERENCES

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, *42*(4), 529-555.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*(1), 1-16.

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.

Kang, O., Rubin, D. O. N., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, *94*(4), 554-566.

Lee, J., Jang, J., & Plonsky, L. (2014). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*(3), 345-366.

Reinders, H., & Pegrum, M. (2016). Supporting language learning on the move: An evaluative framework for mobile language learning resources. In B. Tomlinson (Ed.), *Research and materials development for language learning* (pp. 219-232). New York, NY: Routledge.