



**HAL**  
open science

# Computer-aided pronunciation training in 2022: When pedagogy struggles to catch up

Sylvain Coulange

► **To cite this version:**

Sylvain Coulange. Computer-aided pronunciation training in 2022: When pedagogy struggles to catch up. Alice Henderson; Anastazija Kirkova-Naskova. Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices, pp.11-22, 2023, 10.5281/zenodo.8137754 . hal-04159763

**HAL Id: hal-04159763**

**<https://hal.science/hal-04159763>**

Submitted on 12 Jul 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Coulangue, S. (2023). Computer-aided pronunciation training in 2022: When pedagogy struggles to catch up. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7<sup>th</sup> International Conference on English Pronunciation: Issues and Practices* (pp. 11-22). Université Grenoble-Alpes. <https://doi.org/10.5281/zenodo.8137754>

## Computer-aided pronunciation training in 2022: When pedagogy struggles to catch up

Sylvain Coulangue  
Université Grenoble-Alpes

Numerous language learning applications and websites have recently introduced automatic pronunciation training features which are popular, particularly among young learners. However, language teachers may struggle to understand the capabilities, efficacy, and reliability of these programs. This chapter provides a critical overview of English pronunciation training features in mainstream applications such as DuoLingo, Memrise, Babbel, Busuu, Rosetta Stone, ELSA Speak, and IELTS Speaking Practice. It also surveys the rare studies investigating the diagnostic accuracy of these tools and whether learners make meaningful improvements by using them. We conclude that communication and collaboration need to develop among engineers, language teachers, linguists, pronunciation experts, and entrepreneurs to promote useful pedagogical outcomes for learners and to align with contemporary objectives in the field of English as a foreign language.

**Keywords:** CAPT, automatic pronunciation assessment, pronunciation training, language learning apps



This chapter is based on the oral presentation given by the author at the 7th International Conference English Pronunciation: Issues and Practices (EPIP 7) held May 18–20, 2022 at Université Grenoble-Alpes, France. It is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of the license, please go to: <http://creativecommons.org/licenses/by/4.0/>.

## 1 Introduction

In today's globalized world, more than three quarters of English speakers are non-native (Walker et al., 2021). It is likely that a given learner of English will have more opportunities to speak with fellow non-native English speakers than native ones. Learning pronunciation by focusing on intelligibility and comprehensibility has become key for successful communication; what the listener understands and how much effort they have to make is seen as more important than how native the speaker sounds.

Although the concept of English as a Lingua Franca (ELF) became much more widespread after Jenkins' book was published (2000), only recently has there been a paradigm shift in terms of realistic goals for pronunciation training, from native-likeness to comprehensibility. For instance, Walker et al. (2021) in their Oxford University Press position paper insist on the necessity to be easily understood rather than to speak with native-like pronunciation. They suggest concentrating on very specific features to effectively improve pronunciation, such as vowel length and word stress, along with most consonant sounds. On the other hand, aspects such as vowel quality or pitch movement are not considered a priority for good comprehensibility, but rather as a means to get closer to a given native accent if that is the learner's goal (Walker et al., 2021). The 2020 update of the Common European Framework of Reference for Languages acknowledges that the former focus on native-speaker accent was "detrimental to the development of the teaching of pronunciation" (Council of Europe, 2020, p. 134). Influence from the L1 is now accepted at C2 level as long as it does not hinder comprehensibility. Furthermore, half of the updated CEFR phonological control scale is now devoted to prosodic features.

Assessment scales such as that of the CLES<sup>1</sup> Certification emphasize that at B2 level pronunciation and intonation should be "clear enough to be easily understood, even if an accent subsists", and that at B1 level pronunciation should be "globally understandable despite a foreign accent and/or pronunciation errors"<sup>2</sup>. Isaacs et al. (2018) designed a comprehensibility scale for English formative assessment, helping teachers identify the aspects of L2 that they should prioritize with their learners in order to promote the production of comprehensible English and also to help L2 learners develop awareness of their strengths and weaknesses. Their scale puts the emphasis on word stress and hesitation markers' position more than on phoneme quality, and much more than on vocabulary and grammar accuracy. Word stress is also considered a priority in Frost and O'Donnell's (2018) prosody-based descriptors for assessing oral production in English, along with intonation, unreduced vowel quality and connected speech.

With ever-advancing AI technology, it seems potentially achievable to integrate immediate feedback focusing on intelligibility features into pronunciation practice. This chapter focuses on technology made for pronunciation training in language applications (apps hereafter) and websites. Research on automated pronunciation error detection started in the early 1970s and gained in popularity in the late 1990s, but results were rather limited, and few programs made it to the commercialization stage (Witt, 2012). In the last decade however, with the ever-growing degree of computation power, the democratisation of smartphones, and enhanced speech recognition technologies, the most popular language learning apps and websites now integrate automated feedback for pronunciation training. These systems are used by a large

1 The CLES (*Certificat de Compétences en Langues de l'Enseignement Supérieur*) is a national, government certified test of language proficiency in France (see <https://www.certification-cles.fr/>).

2 The CLES Oral Interaction assessment grids are available at: <https://www.certification-cles.fr/se-preparer/grilles-d-evaluation/grilles-d-evaluation-1196363.kjsp>.

number of people and are becoming more popular than ever, especially among young learners. However, how they function remains little known to students and teachers.

The following section of this chapter will give an overview of how recent mainstream CAPT tools tackle pronunciation practice. Then we will review some studies which have investigated the pedagogical efficacy of these tools, before discussing current systems limitations and suggesting potential solutions.

## **2 Pronunciation assessment in today's mainstream language apps**

According to the website Top10.com<sup>3</sup>, the five most widely used apps for learning languages in 2022 are DuoLingo, Memrise, Babbel, Busuu, and Rosetta Stone.<sup>4</sup> All of them provide automatic speaking evaluation tools as part of their contents. Apps and websites dedicated to English speaking practice, such as ELSA Speak or IELTS Speaking Practice, can also be found online. This section will describe how these apps tackle speaking production, going through their types of activities and how speech is elicited, as well as what kind of feedback is given to the learner. It is important to mention that apps constantly update, so this description covers the features available in spring 2022.

### **2.1 Types of activities, stimuli and focus**

In 2022, the most frequent English-speaking practice activity is an interface with a series of words or short phrases on the screen to be read aloud while pressing a record button. Most of them also offer immediate feedback, as shown in Figure 1. In most cases, audio is also played over the written form when it appears. All apps cited above provide an audio model of what is to be said as well as the written form, except for Rosetta Stone, which sometimes does not display text until after the recording. Words and sentences are mostly practiced out of context, generated from a vocabulary bank or sentences from the current lesson, with increasing difficulty. Student productions judged to be incorrect may reappear later – and repeatedly –, until the program considers the pronunciation to be correct.

All apps presented here are limited to this same basic read-and-repeat type of exercise, except for Rosetta Stone which provides more contextualised activities; each unit ends with a complex story activity presenting a succession of pictures, telling a story which involves the learners in the first person. The text for each character appears on the screen with a picture and is read aloud. When their turn comes, the students must read aloud the text appearing on the screen or speak spontaneously according to the illustration displayed. Rosetta Stone also provides another type of exercise in which the learners see two successive questions and answers, with the second answer being hidden until they say the target words. In both cases, there is only one target sentence, and the app recognises only words that match with it, even if the learners say something different that is also correct.

3 Top10.com <https://www.top10.com/language-learning-apps> (November, 2022)

4 Language learning apps (presented in alphabetical order):

Babbel (2022). <https://www.babbel.com/>

Busuu Online S.L. (2022). <https://www.busuu.com/>

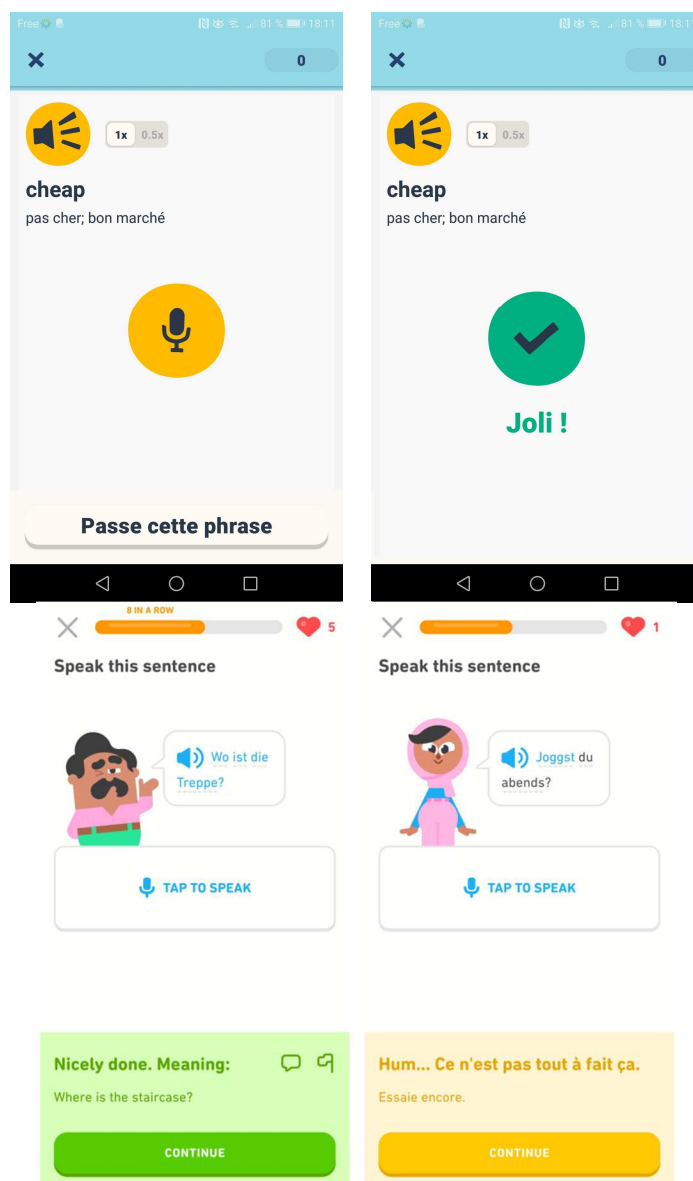
DuoLingo Inc. (2022). <https://www.duolingo.com/>

Memrise (2022). <https://www.memrise.com/>

Rosetta Stone Inc. (2022, v5.0.37). <https://www.rosettastone.com/>

**Figure 1**

*Examples of Speaking Practice Activity on Memrise (top) and Duolingo (bottom)*



Along with the written transcription and the audio, some apps provide a translation in the learner's L1, by default such as Memrise or Babbel, or on demand word by word on DuoLingo. The audio may come with a video or a picture giving contextual hints. ELSA also provides IPA transcription, but this was found in no other app. ELSA and Memrise also provide a feature to play the audio more slowly.

## 2.2 Types of feedback

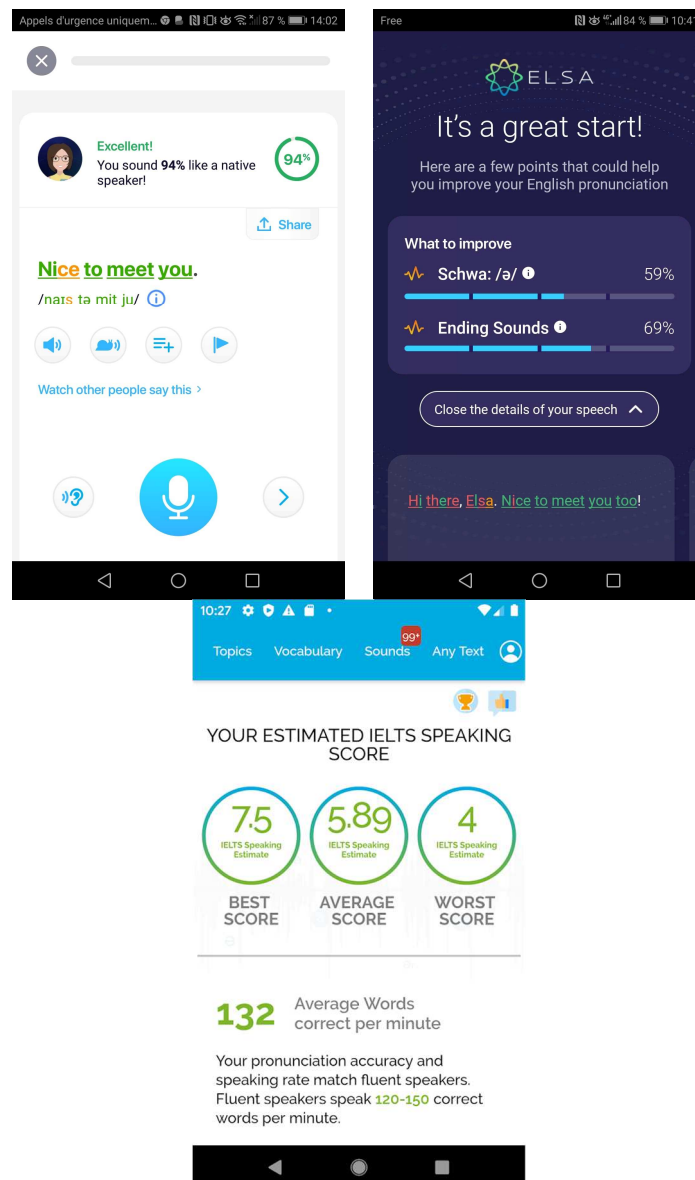
Adequate feedback provision is crucial in CAPT, with binary feedback (correct/incorrect) being the most common. In ELSA, for example, after recording, a green screen is displayed with congratulations if correct, or an orange screen with encouragement but no particular advice on how to pronounce it better (see Figure 1). Busuu and Duolingo also provide a translation when

the production is correct. Other apps show a success percentage indicating how well the words were recognised by the system.

ELSA and IELTS Speaking Practice go much further, displaying words or letters in colours: green (correct), orange (almost), red (wrong or missing). By clicking on a word, the learner can see what phonemes were expected, and which were recognized by the system. ELSA also gives explicit tips about how to pronounce incorrect segments and computes an overall English proficiency score (in percent), as well as scores for listening performance, fluency, word stress, intonation, and different categories of phonemes. ELSA and IELTS Speaking Practice calculate an estimated IELTS score, as well as the average number of correct words per minute. ELSA and IELTS Speaking Practice calculate an estimated IELTS score as well as the average correct words per minute (see Figure 2).

## Figure 2

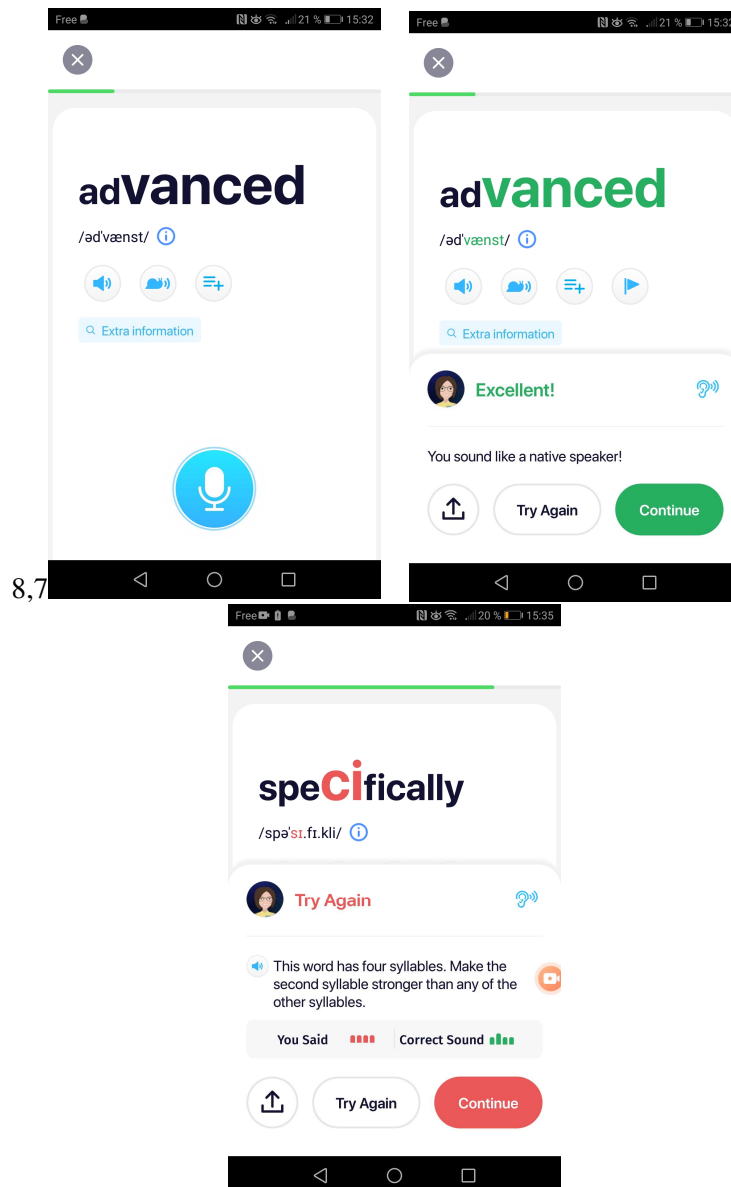
*Examples of Feedback on ELSA (left), and Overall Scores on ELSA (middle) and IELTS Speaking Practice (bottom)*



Feedback on lexical stress is also provided by ELSA in some specific exercises. In these activities, a single word appears on the screen with the expected stressed syllable written in larger characters. After recording, this syllable is coloured in green if the system detected the stress on that syllable in the learner's speech, or red otherwise. There is also a visual representation of the word's prosodic shape as a succession of bars, showing the syllable to stress and the syllable stressed by the student (see Figure 3).

**Figure 3**

*Examples of a Word Stress Detection Activity on ELSA*



In September 2022, ELSA launched a new premium feature on their website called Speech Analyzer<sup>5</sup>, allowing students to record free speech and get an overall speaking score, as well as

<sup>5</sup> ELSA's Speech Analyzer <https://speechanalyzer.elsaspeak.com/>

detailed scores for pronunciation, intonation, fluency, grammar and vocabulary. ELSA also provides global score predictions for IELTS, TOEFL, CEFR, and Pearson. In this feature, pronunciation feedback detects segmental errors and provides tips to pronounce them better. It also estimates intonation and gives average pitch variation, suggesting a range between 50 and 150Hz (typical of adult males), regardless of the learner's gender or age. Fluency estimations give a pace score (number of words per minute) and a pausing score as a so-called nativeness percentage.

### 2.3 How it works

Except for the word stress detection exercise of ELSA, all systems mentioned above work in a similar way: a speech recognition engine tries to match the target word or sentence, and compute a global, word or phoneme-level percentage of recognition confidence. The answer is considered correct when the confidence score is above a given threshold value. This value indicates how similar the user's speech is to the native model. In IELTS Speaking Practice or ELSA, non-constrained phoneme recognition allows the student to see which phonemes they actually pronounced, though this is limited to recognition of English phonemes only, and to the target number of phonemes.

In the case of ELSA's word stress detection exercises, an automatic speech recognition system (ASR) is probably used<sup>6</sup> to segment the word in syllables, and a stress classifier used to identify the stress pattern from acoustic measures such as duration, intensity or pitch.

## 3 Do the students actually improve their pronunciation?

There are very few independent papers reviewing the effectiveness of CAPT tools for pronunciation improvement. Most studies either were funded by the company that created the app, or have one or more authors working for it; both situations present a conflict of interest. For example, the final report of Duolingo Effectiveness Study by Vesselinov and Grego (2012) claims that the vast majority of participants in the study succeeded in improving their knowledge, with a statistically significant improvement of mean average points ( $M = 8.1$ ) on the WebCAPE placement test, per hour of use of the app. It would then take about 34 hours to reach 275 points on the test, 270 being the minimum to access the second semester of a university course. Based on this, Duolingo's home page claims that their app was scientifically proven to be more efficient than university courses. Krashen (2014) deplored this and took a closer look at the data. He found that there was great variability between participants ( $SD = 12.1$ , median = 3.9 points per hour), with participants studying Spanish for personal interest or school getting worse results ( $M = 5.7$ ) than those studying for business ( $M = 11.4$ ) or travel ( $M = 17.6$ ). As for the participants, of the 156 who volunteered (having responded to an online advertisement), only 58% of them continued until the end. Learning time ranged from 2–133 hours depending on the person. This great variability across users makes it hard to generalise any results to university students, who may be obliged to take a course and often are less motivated. Anguera and Van (2016), the two creators of ELSA, did a study on 50 regular users to see how many repetitions they needed to pronounce a word without any error. They then computed their relative nativeness improvement over time; their graphs show decreasing and increasing curves, but no axis label is provided, nor is there any further textual information about interpretation of the results. Doan et al. (2021), all employees of ELSA Corporation, carried out a study using ELSA at the private Chamakura Malla Reddy University, Bangalore, India. They had 206 students work with ELSA for six weeks. Improvement evaluation was done

<sup>6</sup> Details on ELSA's stress analysis are not publicly available.



by the app itself, through the English proficiency score percentage. The improvement found in this study was 10.44%, but the paper does not specify how this number was calculated. This also means that improvement is only measured for contents studied in the app. The authors insist on the fact that every student improved their pronunciation. Other papers about ELSA's effectiveness are limited to questionnaires given to students, showing that a majority of them subjectively find the app useful and motivating: 18 university students (Kholis, 2021), five junior high school students (Pangastuti, 2021), 12 college students (Samad & Aminullah, 2019), and 25 college students (Silaen & Rangkuti, 2022).

Loewen et al.'s (2020) study on Babbel in Spanish was funded by Lesson Nine GmbH, the third author being employed by Babbel. Eighty-three university volunteers used the app 10 minutes a day for three months. The attrition rate was 32%. The evaluation was done on the 58 students who had spent at least three hours on Babbel, using an ACTFL Oral Proficiency Interview. This involves assessment by two human raters of 15 responses recorded by each student during the test. The study showed that the more time spent learning with the app, the better the results, and that most improvements were observed in grammar and vocabulary knowledge, rather than speaking ability.

Jiang et al. (2021) carried out a study specifically on Duolingo's efficacy for teaching speaking skills. It was funded by Duolingo and involved six authors from the company and one associate professor in applied linguistics at Northern Arizona University. They contacted random users of Duolingo who had completed the beginning-level course in French and Spanish. They asked them to take the Pearson Versant Test (an automatic speaking test using ASR on read and listen-and-repeat tasks, as well as say-the-opposite or reorder-the-words tasks). One hundred seventy-five learners of Spanish and 155 of French did the test. Sixty-six percent of Spanish learners and 53% of French learners achieved speaking A2 level or above, according to the Versant test, which corresponds to Duolingo's expected proficiency objectives for speaking skills. Nonetheless, it is worth noting that 28% of learners of French did not even receive a score with the Versant test, probably because of too poor pronunciation, according to the authors.

Becker and Edalatishams (2019), two independent researchers, also carried out a study on ELSA. According to them, one major shortcoming is that the app focused only on segmental aspects of pronunciation. They do not mention the stress detection activities, which might have been launched later than their study. Even if global scores of fluency and intonation are provided in today's version, no exercise allows the user to practice these aspects specifically. Becker and Edalatishams (2019) regret the fact that quantity overrides quality; hundreds of phrases are available, but all the exercises are very similar, and the feedback is the same. The paid version only gives access to more exercises about the same skills and topics. In February 2022, ELSA was claiming to offer over 3000 lessons. The two authors also note that ELSA often mistakenly identifies incorrect sounds as correct, leading to wrong and confusing feedback. They also deplore the absence of any applied linguists or English language learning experts in the ELSA development team. In February 2022, the ELSA web site<sup>7</sup> indicated that one speech advisor had been added to the team and he was described as "a world-class accent reduction coach with more than 35 years of training executives and Hollywood stars". In November 2022, out of 40 open positions listed on their website, 21 concerned business development and growth marketing, eight product design, seven management and marketing, and four speech science – not a single one was listed for pedagogical engineering, linguistics, or pedagogy.

<sup>7</sup> <https://elsaspeak.com/en/>

## 4 Discussion

The CAPT tools reviewed here are easy to use, with a simple and intuitive design, and a great deal of graphic design rich in colours. Incorporated into gamified learning scenarios with various modalities of contents, everything is done to maintain learner motivation and to enhance the user experience. Even though most exercises are very similar, the contents and topics approached are varied. Providing this kind of automatic tool to learners allows them to get immediate feedback on their production, practice whenever and as much as they want without fear of being judged by others, and, above all, it gets them to practice speaking.

Notwithstanding, several limitations can be seen in the way these tools approach pronunciation training. These limitations can be divided into four categories: the stimulus, the focus, the model, and the feedback.

First, regarding the stimulus: with the exception of ELSA's brand new, (initially) free Speech Analyzer module, the speech stimulus of every type of exercise is a predetermined word or phrase the student has to read, repeat, or in rare cases to guess. Reading aloud may however lead to disfluencies or phonological errors, due to either pronunciation difficulties or to problems with decoding the written text before the reader's eyes. Repetition often leads to phonological imitation of the audio model, which makes it hard to diagnose student's spontaneous oral production proficiency. Furthermore, interactions and real communication situations are essential for practising speaking. Rosetta Stone was the only tool covered here to integrate student's production in a communicative situation, with much more context than the other apps (which tend to stick to micro-exercises out of context). This raises the question of the long-term effects of this predominance of micro-learning, which is observable to a larger extent in most language learning apps. Assessing spontaneous speech is a great challenge, but ELSA is proving that it is possible. Making pairs of learners speak to each other about a specific subject and providing individual feedback would be a very interesting type of activity, as well as integrating assessment features into an audio chatbot.

In terms of focus, assessment tools that go beyond mere word recognition mostly focus on phoneme quality and do not consider prosodic phenomena. ELSA Speak is the only app that seriously considers prosody in its scores and feedback. This, however, is only given as overall measures. In order to better suit students' needs, it seems important to target primarily those phenomena that hinder communication (such as speech rhythm and stress), and to give more weight to high functional load phonemes (Derwing & Munro, 2015), since their impact on speaker comprehensibility is widely accepted (Council of Europe, 2020).

All apps presented here compare students' speech production to a native model, using ASR or stress classification models. However, this tendency to base scores on the distance between one's production and a model only reflects how similar the users' production is compared to the model but not how acceptable their production is. Moreover, one might wonder how representative that model is of the diversity of voices, ages, accents, or types of discourse. This might also give more prominence to some hegemonic accents and ignore other less represented ones.

Concerning feedback, most systems still stick to binary response (good/bad) at phoneme, word or utterance level. Some apps also display a success percentage, often based on the ASR confidence score, which, in fact, can be hard to interpret. In the best cases, incorrect words or phonemes are displayed in colours along with what has been pronounced by the app user, and sometimes tips are provided on how to improve the pronunciation of incorrect phonemes. Yet, there is no error prioritization nor filtering depending on the learner's level or mother tongue, for example. Above all, the main issue is that too many mistakes are considered correct by the apps. Memrise especially, tends to validate every attempt by the student, even when the pronunciation is extremely poor. Encouraging any answer might be seen as a positive

reinforcement strategy to maintain learners' motivation, but it could become detrimental to the learning process. This issue of faulty feedback in CAPT systems was already considered a major issue by Levis (2007) and is still considered a real limitation of recent CAPT tools (O'Brien et al. 2018; Rogerson-Revell 2021). Basically, there is a need not only for more feedback, but also for more pedagogical feedback – and it needs to be prioritized instead of being displayed all at once with no hierarchy. Detey et al. (2016) call for a better consideration of diversity in learner profiles, providing adaptive feedback (as well as contents) according to intra- and inter-individual variables, such as observed difficulties, progress, and needs.

## 5 Conclusion

Research in CAPT seems to be focused on the technical challenges rather than the pedagogical ones. Very few studies analyse the actual impact of these technologies on language learning or how well these tools evaluate pronunciation. As for most of the commercial CAPT tools presented in this paper, it is likely that these are more often appealing technological creations intended to impress potential clients, rather than truly pedagogical tools that add value. This follows Rogerson-Revell's (2021) analysis that technological novelty “may temporarily disguise lack of pedagogic rigour” (p. 191) or even that “as technology progresses, pedagogy appears to regress”.

Contemporary language learning apps are based on artificial intelligence and there is a worrying tendency to uncritically praise AI and to use it for marketing reasons rather for true pedagogical plus-value. For the learner's sake, it is important to clearly indicate how accurate the automatic evaluation is and inform them that it might not always work properly. The lack of knowledge about how the automated scoring system works makes it hard for students and teachers to interpret the scores. This is probably one reason why many teachers doubt the effectiveness of these tools (Agarwal & Chakraborty, 2019). According to Evanini and Zechner (2019), engineers who make CAPT tools are not aware of the issues in the field, such as the need for reliability, validity and fairness of evaluation, as well as for a transparent description of how scores are computed, in order to encourage a positive wash-back effect. They join Agarwal and Chakraborty (2019), O'Brien et al. (2018), Detey et al. (2016), and many others in calling for more collaboration between teachers and engineers.

Another important issue to bear in mind is data privacy. None of the apps examined here indicate what data is used, which data is stored, where it is kept, for how long, and for what purpose. It does not seem to be a real concern for companies or users.

To conclude, CAPT tools are an opportunity to complement classwork, giving learners the chance to practice and get feedback as much as they like on aspects that may not be studied in class because of a lack of time or teacher training. As Levis (2007) points out, CAPT is “tireless”, “consistent”, and it “can meet varied individual needs”, and “promotes learner autonomy” (p. 197). The technology is here, we need to shape it according to our needs, rather than adapting our way of learning to available technology. In order to make pedagogically effective CAPT tools, engineers need to better understand the foundations of pronunciation acquisition and teaching.

## References

- Agarwal, C., & Chakraborty, P. (2019). A review of tools and techniques for computer aided pronunciation training (CAPT) in English. *Education and Information Technologies*, 24(6), 3731–3743. <https://doi.org/10.1007/s10639-019-09955-7>
- Anguera, X., & Van, V. (2016). English Language Speech Assistant. *Proceedings of InterSpeech*, 1962–1963.
- Becker, K., & Edalatshams, I. (2019). ELSA Speak – Accent Reduction [Review]. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 434-438). Ames, IA: Iowa State University.
- Council of Europe. (2020). *Common European framework of reference for languages*. Council of Europe.
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins.
- Detey, S., Fontan, L., & Pellegrini, T. (2016). Traitement de la prononciation en langue étrangère : approches didactiques, méthodes automatiques et enjeux pour l'apprentissage. *TAL Traitement Automatique des Langues*, 57, 15–39. ISSN 1248-9433 <https://hal.science/hal-01919021>
- Doan, T., Nguyen, Q., Nguyen, B., Anguera, X. (2021). How ELSA speaking app improves learners' speaking skills - A use case from the CMR University in India. [White paper]. ELSA Corp.
- English Language Speech Assistant (ELSA, 2022). <https://elsaspeak.com/>
- Evanini, K., & Zechner, K. (2019). Overview of Automated Speech Scoring. In Zechner, K., & Evanini, K. (ed.) *Automated speaking assessment* (pp. 3–20). Routledge.
- Frost, D., & O'Donnell, J. (2018). Evaluating the essentials: the place of prosody in oral production. In J. Volín & R. Skarnitzl (Eds.), *The Pronunciation of English by Speakers of Other Languages* (pp. 228–259). Cambridge Scholars Publishing.
- Isaacs, T., Trofimovich, P., & Foote, J. A. (2018). Developing a user-oriented second language comprehensibility scale for English-medium universities. *Language Testing*, 35(2), 193–216. <https://doi.org/10.1177/0265532217703433>
- Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press.
- Jiang, X., Rollinson, J., Chen, H., Reuveni, B., Gustafson, E., Plonsky, L. & Pajak, B. (2021). How well does Duolingo teach speaking skills? Duolingo Research Report DRR-21-02 June 1, 2021. <https://blog.duolingo.com/how-well-does-duolingo-teach-speaking-skills/>
- Kholis, A. (2021). Elsa Speak app: Automatic Speech Recognition (ASR) for supplementing English pronunciation skills. *Journal of English Language Teaching*, 9(1), 1–14. <https://doi.org/10.32332/joelt.v9i1.2723>
- Krashen, S. (2014). Does Duolingo 'trump' university-level language learning? *The International Journal of Foreign Language Teaching*, 9(1), 13–15. ISSN 2165-767X.
- Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, 27, 184–202. <https://doi.org/10.1017/S0267190508070098>
- Loewen, S., Isbell, D. R., & Sporn, Z. (2020). The effectiveness of app-based language instruction for developing receptive linguistic knowledge and oral communicative ability. *Foreign Language Annals*, 53(2), 209–233.
- O'Brien, M. G., Derwing, T. M., Cucchiari, C., Hardison, D. M., Mixdorff, H., Thomson, R. I., Strik, H., Levis, J., Munro, M., Foote, J., Levis, G. M. (2022). Directions for the future of technology in pronunciation research and teaching. *Journal of Second Language Pronunciation*, 4(2), 182–207. <https://doi.org/10.1075/jslp.17001.obr>
- Pangastuti, D. (2021). The effect of ELSA Speak application on students' pronunciation in English. *Prosiding Pekan Ilmiah Mahasiswa Unis*, 1(1), 127–133.
- Rogerson-Revell, P. M. (2021). Computer-assisted pronunciation training (CAPT): Current issues and future directions. *RELC Journal*, 52(1), 189–205. <https://doi.org/10.1177/0033688220977406>
- Samad, I. S., & Aminullah, A. (2019). Applying ELSA Speak software in the pronunciation class: Students' perception. *Edumaspul: Jurnal Pendidikan*, 3(1), 56–63. <https://doi.org/10.33487/edumaspul.v3i1.85>

- Silaen, T. N., & Rangkuti, R. (2022). Elsa Speak app usage in blended learning during the Covid-19 pandemic: Students' perspectives. *Journal of Basic Education Studies*, 5(1), 26–34. e-ISSN 2656-6702
- SpeechAce (2022). IELTS speaking practice  
<https://play.google.com/store/apps/details?id=com.ielts.speechace.ieltsace&gl=US>
- Vesselinov, R., & Grego, J. (2012). Duolingo effectiveness study. Final Report.  
[https://static.duolingo.com/s3/DuolingoReport\\_Final.pdf](https://static.duolingo.com/s3/DuolingoReport_Final.pdf)
- Walker, R., Low, E., & Setter, J. (2021). English pronunciation for a global world. Position paper. Oxford University Press.  
<https://elt.oup.com/feature/global/expert/pronunciation?cc=fr&selLanguage=en>
- Witt, S. (2012). Automatic error detection in pronunciation training: Where we are and where we need to go. *Proceedings of the International Symposium on Audiovisual Detection of Errors in Pronunciation Training (ISADEPT)*, Stockholm, Sweden, June 2012 (pp. 1-8). KTH, Computer Science and Communication. [http://www.m-a-arabia.com/wp-content/uploads/2014/11/Published-paper\\_isadept\\_ISADEPT-proceedings.pdf](http://www.m-a-arabia.com/wp-content/uploads/2014/11/Published-paper_isadept_ISADEPT-proceedings.pdf)

### About the author

**Sylvain Coulangue** is in the second year of his interdisciplinary PhD at the Université Grenoble-Alpes, affiliated with the LIDILEM research group (Linguistics & First and Second Language Education) and the LIG (Grenoble Computer Science Laboratory), as well as the Speech and Language Processing Laboratory (SLPL) at Doshisha University in Japan. His research focuses on the automatic diagnosis of L2 pronunciation in spontaneous speech. At LIDILEM he co-directs the working group in Didactics of Languages and Multimodality, and is part of the Language Acquisition Evaluation working group.

Email: [sylvain.coulangue@univ-grenoble-alpes.fr](mailto:sylvain.coulangue@univ-grenoble-alpes.fr)