UAB
Universitat Autònoma
de Barcelona

DEPARTMENT OF ENGLISH AND GERMAN STUDIES

Pronunciation Teaching in the Age of AI: Evaluating the Role of Automatic Speech Recognition Based Tools

Treball de Fi de Grau/ BA dissertation

Author: Nihade El Habbaj

Supervisor: Celia Gorba Masip

Departament de Filologia Anglesa i de Germanística

Facultat de Filosofia i Lletres

Grau d'Estudis Anglesos

June 2025

Statement of Intellectual Honesty

Your name: Nihade El Habbaj

Title of Assignment: Pronunciation Teaching in the Age of AI: Evaluating the Role of Automatic Speech Recognition Based Tools

I declare that this is a totally original piece of work, written by me; all secondary sources have been correctly cited. I also understand that plagiarism is an unacceptable practise which will lead to the automatic failing of this assignment.

Signature and date:

June 17th, 2025.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my TFG supervisor, Celia Gorba Masip, for her patience, guidance, and for always being there to resolve any doubts throughout the semester. I would also like to thank my professor, Juli Cebrian, for introducing us to many concepts related to teaching pronunciation; some subsections of my dissertation are inspired by his course Teaching and Learning English Pronunciation.

Furthermore, I want to express my deepest gratitude to my parents for always being so supportive of my studies - without their support, I would not have been able to complete my degree. I am also thankful to my three sisters and to my only and best friend for always being there, supporting and encouraging me.

TABLE OF CONTENTS

INDEX OF FIGURES

Abstract

This dissertation focuses on the role of Automatic Speech Recognition (ASR) tools in language learning. The main purpose of the study is to investigate the effectiveness and limitations of ASR tools in classrooms, pointing out that one of the main difficulties in English pronunciation is orthography. With the rise of AI, it is essential to investigate how these tools can enhance learners' pronunciation and motivate them. This study is based on an analysis of recent research on ASR tools functionalities and learners' experiences utilizing apps such as ELSA Speak. By analyzing these studies, it is shown that ASR tools are useful in improving pronunciation and enhancing motivation and students' engagement; they can work perfectly as a supplement for improving pronunciation in and outside the classroom. However, there are limitations regarding English variety recognition and lack of inclusivity of different accents and dialects. Overall, it is important to balance the integration of this technology into the classroom, with teachers' guidance.

Keywords: Automatic Speech Recognition, Pronunciation, Language learning, Language pedagogy, English varieties, ELSA Speak, Effectiveness, Limitations

1. Introduction:

Pronunciation instruction is often overlooked in teaching curricula, which can lead to difficulties that affect both intelligibility and comprehensibility. These challenges arise because pronunciation is difficult to learn and teach effectively (Sicola & Darcy, 2015). For this reason, integrating pronunciation instruction into language classrooms is essential.

With the rise of technology, several tools have been developed to support pronunciation teaching. One of these is Automatic Speech Recognition (ASR), a technology that can decode and transcribe oral speech (Levis & Suvorov, 2013). Various software and apps provide ASR-based feedback, which can be either global (general pronunciation assessment) or phonological and detailed. Some studies have demonstrated the effectiveness of the ELSA Speak app in engaging students and boosting their motivation (Kholis, 2021). However, other studies highlight limitations. Kim (2006) points out that while ASR tools are effective; they are not as accurate as a human instructor, and Muzzaki et al. (2024) note that they rely on reading written text rather than spontaneous speech.

As technology continues to shape the future of education, AI-powered tools are becoming increasingly relevant. Understanding how to integrate ASR into pronunciation teaching is crucial, as these tools have the potential to enhance learning while also presenting certain limitations. This dissertation aims to evaluate the effectiveness of ASR in pronunciation instruction and explore whether it can be used as a supplement to teachers. Furthermore, this study seeks to investigate whether ASR prioritizes intelligibility and comprehensibility over accentedness, since a strong accent

does not necessarily make speech unintelligible (Munro & Derwing, 1999). By analyzing the impact of ASR on learner pronunciation, motivation, and accuracy, this study will contribute to the discussion on how AI can be used to support language education effectively.

2. Pronunciation in language learning:

Pronunciation is one of the most important elements when learning a language because it enhances fluent communication among people. Mispronunciations can lead to confusion, misunderstandings, or even embarrassment. In most classroom settings, foreign languages are introduced through reading and writing, rather than listening and speaking, which makes orthography a significant challenge. One of the main factors to consider in terms of pronunciation is the inconsistency between English orthography and pronunciation. One of the most difficult features of pronunciation for students to learn is not the production of the sounds, but rather the orthography, which is different from the pronunciation. English spelling normally does not align with the pronunciation because of the different changes that the English language has suffered in the past (Teschner & Whitley, 2004).

The first factor that makes English orthography very different from the pronunciation is the Great Vowel Shift (GVS), in the late medieval and early modern times, English underwent major changes in the vowel system in terms of pronunciation, but it preserved its spelling. The second factor is the dissimilar orthographic conventions, in the past English absorbed numerous words from different languages such as Danish, Norwegian, French, Latin, Greek and so on, however it kept the original spelling rather than adapting them to English orthographic rules, which is what makes

the pronunciation of certain words from different origins unpredictable. The last factor is the limited alphabet; this limitation restricts the language to using a small number of graphemes (individual letters) and digraphs (two letters combined for one sound) (Teschner & Whitley, 2004). English has only five vowel letters to represent fourteen stressed vowel sounds; in English one vowel can have different pronunciations because there are not sufficient letters to represent those sounds (Teschner & Whitley, 2004). There are some rules to help learners predict the pronunciation, but in English, there are many exceptions and inconsistencies, especially with the most common words. These inconsistencies combined with the lack of enough oral input in classrooms highlight the potential role of Automatic Speech Recognition (ASR) tools in learning pronunciation. These tools can be very useful in contexts where the oral input is limited and can enhance pronunciation learning. Therefore, it is useful to know the different features of pronunciation to better understand how it is learned and taught.

2.1 Features of pronunciation:

By discussing the pronunciation's importance, it is also important to understand the specific features of pronunciation that learners must acquire. These are generally divided into two types of features: segmental and suprasegmental features.

2.1.1 Segmental features:

These refer to individual sounds consonants and vowels, their properties and how to produce them accurately. Segmental features are normally easier to teach because they can be learned through simple imitation and basic understanding of the articulatory properties such as place and manner of articulation (Avery & Ehrlich, 1992).

2.1.2 Suprasegmental features:

These features involve aspects like stress, intonation, and rhythm, which are more difficult to teach and learn. This is because they are concerned with individual sounds but with how sounds are combined and patterned across the speech. They require sensitivity to natural speech, considerable practice, and they are essential for accurate pronunciation (Avery & Ehrlich, 1992). However, they are often neglected or not explicitly addressed in teaching (Avery & Ehrlich, 1992).

2.2 Individual differences:

Learning pronunciation is often affected by several factors; they are called individual differences. These factors have a significant impact on the learning process, and it is essential to understand them because each student has a different profile and may have weaknesses in specific areas. Hence, it is important for the teacher to understand these factors. Moreover, ASR tools can help to address these differences. With their help, lessons can be customized to each student's needs. For instance, shy learners can practice in low-pressure environments, or students with low aptitude can be provided with repeated input. Therefore by acknowledging these differences, it can be assessed how ASR tools can be integrated effectively to respond to learners' individual needs.

To begin with, there are differences that originate from within the learner – internal factors, such as age. Age is a controversial factor. When we talk about age, the Critical Period Hypothesis (CPH) must be taken into account. The critical period refers to the time when children start to acquire their language before puberty, a stage during which the brain still retains its elasticity. Moreover, after this period, the functions of several of the brain's hemispheres are completed – this is known as lateralization (Celce Murcia et al., 1996). According to Krashen (1973, quoted in Celce et al., 1996) the brain starts to lose its plasticity around the age of five. This factor is relevant because a

child's ability to acquire a language differs from that of an adult, and it must be taken into account when adjusting ASR tool practices for certain learners.

In addition to age, one of the most influential factors is the learner's first language (L1) and its interference. When the target language has a phonological structure that differs significantly from the L1, it becomes more challenging for learners to acquire the new structure and separate from their native one without interference (Celce Murcia et al., 1996). It is important to know this because some ASR tools can personalize the learning process according to the learner's L1.

Another key individual difference is aptitude, which refers to the innate ability to pick up language features. Depending on the learner's aptitude, learners can be offered more or less repeated practice. Moreover, there are different types of learners depending on their background, personality (e.g., extroverts versus introverts) and goals. These elements can shape each learner's pronunciation profile. (Celce Murcia et al., 1996)

Building on this, it is equally important to explore external factors such as motivation and the teacher's role, which are also crucial. A teacher who possesses both phonetic knowledge and pedagogical training can significantly support learners. In addition, another essential factor is the amount of exposure to the target language. According to Celce Murcia et al. (1996), we acquire a language through the input that we receive and this input must be large and comprehensible in order for the speaker to be able to capture it before starting to speak. Finally, considering these individual differences can be key to adjusting lessons and practices in ASR tools to each student's needs in a way that helps them improve their pronunciation.

3. Pronunciation instruction:

Having mentioned the importance of pronunciation when learning a language, its complexity, and the different factors and elements involved, pronunciation instruction can be considered fundamental. However, it is usually overlooked in classrooms or even omitted in some cases (Sicola and Darcy, 2015). The first reason why pronunciation is overlooked in classes is the lack of teachers' pedagogical and phonological training, some teachers may be knowledgeable about the phonological system but they are not sufficiently trained to teach pronunciation effectively. As a result, they separate pronunciation from the components of the course, and they focus on teaching grammar and vocabulary (Sicola and Darcy, 2015). Some Non-Native English Speaking teachers (NNESTS) may not even feel confident teaching pronunciation because of their pronunciation or because they are not prepared enough .Hence, they prefer to deliver the grammar and vocabulary in their own language (Llurda, 2018). Moreover, pronunciation can be difficult to teach because teachers find it difficult to customize each lesson to every single student's needs. Therefore, even if pronunciation is taught in class, the focus is shifted to teaching minimal pairs and drills which are form-focused and lack contextualization and spontaneous interaction (Celce Murcia et al., 1996).

Furthermore, pronunciation assessment is time consuming. Teachers cannot assess pronunciation with simple questionnaires or multiple choice exercises. Pronunciation assessment needs individual exams, recordings, and phonological training to assess it. Besides, there is also the misconception that teaching pronunciation is only for advanced students because of phonetics and phonology, and the technical terms. Often, it is even considered less important than grammar and vocabulary, since they are the only components that are taught, they are given more

importance in class. Therefore, the lack of pronunciation practices and the little importance attached to it by teachers give students the misconception that learning pronunciation is less important; Avery and Ehrlich (1992) state : “Students didn’t notice its importance because a lot of teachers omitted teaching because they lack knowledge”.

This issue can be improved by integrating pronunciation into all components of the language, as mentioned by Sicola and Darcy (2015). Both meaning and form focused methods can be used to have the communicative framework and integrate pronunciation within the four components of a language. This includes the introduction of new vocabulary with accurate pronunciation, teaching some grammar rules such as the endings in present or past tense, reading and writing out loud, and listening and speaking activities. In this way, pronunciation will be integrated effectively within the existing components. Besides, by applying the communicative approach, which includes interactive tasks, negotiated interactions between students, repetition, and engaging in real conversation, pronunciation can be improved even to a geater extent (Avery & Ehrlich, 1992). Following this model, it is important to discuss the different approaches that were used historically and are used in the present days.

3.1 Approaches to teaching:

3.1.1 A brief history of methods and approaches used:

Historically, teaching pronunciation has been approached differently, and different approaches have been taken. Initially, pronunciation instruction was approached through the imitation of the sounds and the understanding of the phonetic system of the sounds. Those two approaches are called the Intuitive-imitative approach and Analytic-linguistic approach. The former refers to listening and mimicking certain sounds and patterns; it is less technical and does not require phonological training, and the latter

refers to studying and understanding the phonetic alphabet, the articulatory description of each sound (place and manner), the vocal apparatus, listening, and imitating; t is more scientific and analytical (Celce Murcia et al., 1996). Hence, traditionally, both approaches were used, aligning with certain traditional teaching techniques like imitation, phonetic training, and minimal pairs drills (Celce Murcia et al., 1996).

In the present days, the focus has shifted to the use of the communicative approach (Avery & Ehrlich, 1992). This approach has gained importance because of its focus on the speech as a whole, emphasizing both meaning and form, while leaving aside mimicking individual sounds. This approach consists of applying meaningful tasks beyond the word level, using full sentences while interacting with other students in the class, peer assessment, and feedback. It is beneficial for students because it can help them to become more aware of their own mistakes and correct them, because when they assess each other, they actively listen and compare pronunciation, which increases their metalinguistic awareness. By applying this approach, the practice will be student-centered and the aim will be to develop features learned in class outside the class (Avery & Ehrlich, 1992).

In addition to the different approaches that are used in present-day instruction, the rise of technology has also led to the inclusion of digital devices and several technological tools in classrooms. This development, leads us to the following point, which is Education 4.0, or the educational approach that aligns with the digital transformation and the fourth industrial revolution.

3.1.2 Education 4.0:

The fourth industrial revolution (industry 4.0) implies the inclusion of technology and automation within the industry. This improvement took place in education too, hence,

the focus of education has slightly shifted toward integrating technology devices and tools, including artificial intelligence, to personalize and adapt the learning to each student's needs in response to industry 4.0 (Hong & Ma, 2020). This shift implies both the shift in the skills and capability of managing the digital world and the shift in the methods applied and tools that are used in classes. Ultimately, there are some institutions that focus on the transformation of their teaching model to adapt to new ways of learning and teaching. One of those transformations is the integration of SMART campus environments, which offer the possibility of being both on and off campus, as well as simulated learning environments, blended learning, which is a mix of online and face-to-face teaching, and using digital apps and tools such as AI, data analysis, and automation to complete projects and tasks. This shift will allow teachers in language classes to personalize and enhance language learning, especially pronunciation. In language learning, there are many pronunciation Apps that are used as part of Education 4.0 tools. These apps can provide pronunciation models (mostly standard American or British), record students speech –they recognize the speech as automatic speech recognition tools, and provide feedback. Furthermore, learning language pronunciation would be a lot easier, as most ASR based apps provide learners with pronunciation models and enable them to assess their recordings in practice (Pham & Pham, 2025).

4. ASR technology:

Automatic Speech Recognition (ASR) is a technology that converts oral speech into written text. It is a "machine-based process of decoding and transcribing oral speech" (Levis & Suvorov, 2013). It appeared initially in the 1950s. The first ASR system,

developed at Bell Telephone Laboratories by Davis, Biddulph, and Balashek (1952), was based on pattern matching. Pattern matching is an early method that matches the speaker's speech with archived acoustic templates or patterns. At first, ASR was speaker dependent, meaning that it was trained based on just an individual speaker, leading to the recognition of just one variety. Later on, it was developed into a speaker-independent recognizer by Forgie and Forgie(1959). Then, it was approached by the modern statistical modelling methods such as hidden Markov modelling (HMM), which is a method used to predict sounds or words uttered by analysing patterns over time, however, the movement of the mouth or the steps of articulation were not directly observed (Levis & Suvorov, 2013).

With regard to speech continuity, there are four types of ASR systems. First, isolated or word recognition systems, which recognize individual words. Second, connected word recognition systems, which can identify words pronounced together even if there are no pauses. Third, continuous speech recognition systems, which recognize full sentences in natural speech. Fourth, word spotting systems, which identify and find certain words in a continuous speech (Levis & Suvorov, 2013). As a result of ASR capabilites, devolopers have created several educational apps to enhance pronunciation learning.

There is a large number of language learning tools such as Elsa Speak and Duolingo, which can help students by providing them with vocabulary, grammar, comprehension exercises, and pronunciation drills. With the rise of Artificial intelligence (AI), language learning has become more personalized and adapted to each student by providing instant feedback and adjusted exercises. The recent improvements in technology have led to improved language training, resulting in more benefits and

ease of use (Rusmawaty et al., 2024). It is argued that learning pronunciation with the help of mobile apps can be more beneficial and effective than using physical books, because in this way learners have access to audios and videos and are able to hear how sounds are pronounced accurately in context, which is essential for learning how to identify and produce sounds (Nguyen & Van Tuyen, 2024).

Therefore, ASR can be a very helpful and useful tool for learners who want to enhance their pronunciation; they can develop their spoken-language skills. With ASR-based systems, learners of a foreign language can be exposed to hearing the most accurate version of pronunciation and practice in a low-pressure environment, and they can still have individual, relevant, and instant feedback on pronunciation, which can enhance self-correction and improvement.

4.1 ASR based apps and their effectiveness:

ASR technology has become a key component in language learning. As a result, creators have developed several apps for pronunciation learning as a response to the challenges faced in classes, such as the lack of personalized feedback and practices. In the following sections, the benefits of four apps that have gained importance in pronunciation learning will be analysed.

4.1.1 ELSA Speak app:

ELSA Speak is a Mobile-Assisted Language Learning (MALL) tool powered by artificial intelligence (AI) that can be downloaded on smartphones. It helps students improve their pronunciation, as it records learners' words and sentences, transcribes them, and gives them immediate feedback to improve their pronunciation and sound more natural (in a neutral American accent). In this app, there is the possibility to choose a theme and start a conversation with the app, and if learners run out of words or

ideas, the app provides them with a text to read out loud, and when they finish, learners can have the assessment summary, a personalized feedback about tone, engagement, pronunciation level and vocabulary. Moreover, like in Duolingo, learners can also decide their daily streak whether they want it to be five/ten minutes daily.

One of the most significant benefits of ELSA Speak is its contribution to pronunciation improvement. ELSA Speak app deals with different features of pronunciation; one of them is allophonic variation. It has been demonstrated that students can perceive a more accurate pronunciation because they are exposed to audio recordings by native speakers (Saragih et al., 2021). Furthermore, learners receive detailed feedback on individual phonemes, word stress, and segmental features such as consonants, vowels and reduced vowels. One student reported in Rusmawaty et al.'s study (2024), "I can see my report as ELSA Speak gave me detailed and comprehensive information. I am assessing my learning; it is good." (S26) . Overall, students seem to be satisfied with this type of feedback.

In addition, the study by Rusmawaty et al. (2024) showed a significant improvement in participants' post-test scores after using the app in the production of the sounds /p, k, (s)/, although (/s/) involved the lowest degree of improvement. This is illustrated in a graph (Figure 1) from Rusmawaty et al., (2024) study showing pre-test and post-test result, where the red line indicates a significant improvement.

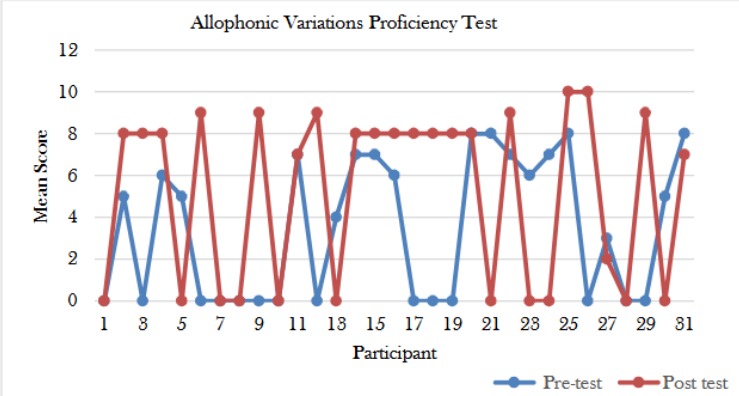

Figure 1. Mean score of allophonic variation (Rusmawaty et al., 2024)

Similarly, the findings in the study conducted by Kholis (2021) showed that students significantly improved their pronunciation after using Elsa Speak. Comparing it to a previous study by Elimat and AbuSeileek (2014, quoted in Kholis, 2021), which included a control group, this experiment showed that ASR technology was more effective than traditional classroom instruction.

Another important advantage of ELSA Speak is the ability to enhance aptitude, engagement and motivation, all of which play an essential role in the improvement of language learning. Students who were tested in the study conducted by Nguyen & Van Tuyen (2024) reported that the app had been effective. The study was conducted at a university in Vietnam; fifty college students from an introductory English class were tested. They were split into two groups: a control group (which did not use ELSA Speak) and an experimental group (who practiced using ELSA Speak as part of their learning routine). They were tested using a pre-test and post-test – before and after the experiment – to measure their English speaking skills. The study showed a significant difference in post-test scores between experimental and control groups. Specifically, it was demonstrated that "the mean score of the control group was above the experimental group (M: 3.56 > M: 2.88) in the pre-test, the results reversed since the Mean score of

experimental group (M= 5.48) was higher than the control group (M= 4.64) in the post test"(Nguyen & Van Tuyen, 2024). The following graph from Nguyen and Van Tuyen's study illustrates the difference in scores between control and experimental groups in the pre-test and post-test.

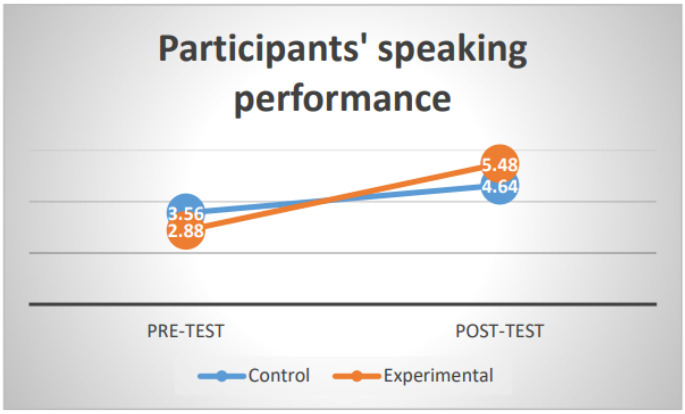

Figure 2. Speaking performance of the control group and experimental group in the pre-test and post-test (Nguyen & Van Tuyen, 2024).

Students seem to have positive attitudes towards the use of the app, they seem motivated and more engaged with the app in terms of learning English speaking skills. They admitted to "favour ELSA Speak as their main English-learning tool outside classroom" (Nguyen & Van Tuyen, 2024). They were motivated to practice even after class and in their free time with this app. In addition, they were very proud of using the app and their progress, and they considered it as their "optimal English learning application" (Nguyen & Van Tuyen, 2024).

Furthermore, the gamification of the app was also a key factor to encourage students to use it and engage with the app. Students found it joyful and entertaining. The app attracted students to it to keep learning. One student stated in Rusmawaty et al.'s study (2024): " I like how ELSA Speak blends games with learning. Enjoyable and intriguing pronunciation games keep me playing." (S12).

The effectiveness of ELSA Speak has also been demonstrated in learners' satisfaction. In a study that focused on student satisfaction rather than on ELSA Speak app itself, researchers used two technical models. The first is Technology Acceptance Model (TAM), which explains why people choose to use new technology. It was shown that users are more likely to use an app or tool if they think it is easy to use. The ease of use influences their decision to use it or not, which means that the ease of use equals usefulness. The second model is Expectation-Confirmation Theory (ECT), which is used to understand how users' expectations before using the tool affect their opinion, satisfaction, or assessment of the tool, indicating that usefulness equals effectiveness. The third model is the proposed research model, which combines the two previous theories that are mentioned above. Taken together, these four elements—perceived usefulness, ease of use, confirmation, and satisfaction—are key to understand why users use certain apps, in this case ELSA speak (Pham & Pham, 2025).

As a result, the study showed high learners' satisfaction rates with ELSA Speak in terms of improving their English pronunciation. Students were also satisfied with the variety of content and levels, and the immediate feedback. Feedback accuracy and independence in practice were key satisfaction factors. (Pham & Pham, 2025). Similarly in Kholis's study (2021), student interviews revealed that students felt more motivated to learn pronunciation due to the immediate feedback and the engaging nature of the app. The study concludes that ASR apps like Elsa Speak can be a useful tool in pronunciation instruction, as they increase student engagement, provide personalized feedback, and adapt to individual needs (Kholis, 2021).

4.1.2 NOVO Learning and ILI:

Novo Learning and ILI are also two apps similar to Elsa Speak; however, they are not as global as Elsa Speak, as they are mostly used in Indonesia. In a comparative study by Muzakki et al. (2024) of both apps (Novo Learning and ILI), it was shown that I Love Indonesia (ILI), which provides global corrective feedback (overall feedback without correcting every single error), and NOVO Learning (NOVO), which offers phonetic feedback, both provide individualized and personalized feedback. A total of 128 participants from four class groups, aged 14 to 17, took part in the study. There was no control group; they all participated in the study. Two class groups used ILI, while the other two used NOVO. To evaluate the results, participants were tested through a pre-test and post-test. The results showed that students using NOVO demonstrated greater improvement in word-level pronunciation compared to those using ILI. On a sentence level, both groups showed progress, but the improvement was more significant with NOVO. Raters also found a strong correlation between accentedness and comprehensibility. At the word level, most words showed improvement, except for some target words like “black”, which were influenced by the participants’ first language. Overall, students reduced their foreign accent and increased comprehensibility. Phonetic feedback (as provided by NOVO) proved more effective because it was more explicit. This study also Highlights the effectiveness of ASR tools and suggests that they can be useful for students looking to improve their pronunciation.

4.1.3 Google Read Along app:

Google Read Along App is a web-based AI media application owned by Google. It uses AI Speech Recognition technology to decode and transcribe speech and assess it. The App uses automatic speech recognition (ASR) and the read-aloud method combined. It

has been demonstrated that this App is very effective as well. In a study by Abimanto & Sumarsono (2024) that included an experimental group and a control group, it was shown that Read Along app achieved an average N-Gain score of 65.73% which is classified as effective, indicating that the use of the app achieved success to satisfactory degree. In contrast, the control group, which did not use Google Read Along app, showed an average N-Gain score of 50.39% , which was considered less effective. By saying so, this study shows a clear performance gap between the two groups favouring, the group using ASR app which showed a higher progress (Abimanto & Sumarsono, 2024). This study tested a control and an experimental group. There were 70 participants, with 35 of them forming the experimental group. This group used Google Read Along app with the Read Aloud method to practice, and they completed questionnaires about their experience using the app, while some of them were interviewed. Both the experimental and control groups took a pre-test and post-test to evaluate their pronunciation improvement. The maximum possible improvement was measured by calculating N-Gain scores from pre-test and post-test results based on Hake's percentage formula . This is the parameter table that was used to categorize the N-Gain scores: not effective, less effective, effective enough and effective.

In this app, instant feedback and personalized learning were significantly helpful and useful for the students. Additionally, the element of gamification features –such as earning rewards, competing with peers– was very engaging; it increased motivation. Students also appreciated practicing pronunciation at their own pace, and the easy access to the app. The support that students received from teachers also helped to guide their focus of study and personalize their exercises, which motivated them more to

continue using the tool (Sun, 2023). ASR technology combined with read-aloud practice significantly enhanced learners' pronunciation (Sun, 2023).

Overall, all the apps showed a significant improvement in language learning and heightened students' engagement and satisfaction. With clear instructions and flexible use, providing features such as daily streaks, several topics, different levels and an adjusted practice pace, and independent practice, students found it both easy and useful to use this technology. Moreover, all the apps mentioned before provide instant, personalized, and detailed feedback; some provide a phonetic feedback more focused on the word-level, while others provide general feedback on the pace, intonation, and rhythm. When it comes to students, they were highly satisfied, particularly with the use of Elsa Speak App (Matysiak, 2023), since the App was the most frequently used by participants among the other Mobile Apps. One hundred percent of the students who participated in Matysiak's study (2023) recommended ELSA Speak; they noted the value of IPA (international phonetic alphabet), phonetic transcription and listening and recording features. It is confirmed by the Expectation-confirmation theory (ECT) that satisfaction occurs when an app meets or exceeds expectations, and in this case, it mostly exceeded expectations. Furthermore, with gamification features, motivation and engagement increased.

5. Integration of ASR technology into classrooms:

ASR based tools are increasingly gaining importance in the field of education; they are sophisticated and accessible for students, which makes them easier and more suitable to use in class. By giving immediate feedback to students and allowing repeated practice, students can improve their pronunciation. Moreover, with the property of gamification,

which provides students with interactive games to practice with, their motivation and engagement are heightened. Building upon this, there are studies that support the idea of integrating ASR-based tools into classrooms, mainly to enhance students' pronunciation. (Kholis, 2021; Abimanto & Sumarsono, 2024; Pham & Pham, 2025).

Additionally, ASR tools are considered to be best used to stimulate motivation and encourage independent practice. Since AI-based pronunciation tools and AS technology have increasingly improved, it is important to consider integrating them into classrooms to enhance motivation and engagement. Hence, by using ASR tools, students can practice pronunciation and get immediate feedback, which would increase their autonomy by allowing them to practice more outside the class at their own pace, and increase their motivation by offering them engaging games and visual feedback. Moreover, it was confirmed by Pham & Pham (2025) that tools like ELSA Speak improve pronunciation and fluency, and curriculum designers must take ASR tools into account to integrate them into classroom lessons and choose the tools that align with specific course goals, focusing only on the relevant high –quality components for the course.

Nevertheless, these tools are not supposed to replace teachers; the teacher's role will remain the same, and ASR tools will work as a support to students to practice more. Research done in this area emphasizes that those tools serve as a supplement and not a substitution for teachers (Pham & Pham, 2025). Kholis (2021) also emphasizes the value of these tools in enhancing classroom activities and not replacing teachers. Teachers offer the theoretical explanation of pronunciation concepts, act as a guide for students, and give them instructions to follow. In this way, students will have the opportunity to practice more outside the class what they learned in class and internalize

it. Human interaction is fundamental to clarify doubts and misunderstandings or responding students' questions. Moreover, these tools will allow teachers to follow students' progress.

Furthermore, teachers have an essential role in interpreting feedback provided by the ASR tools; they are supposed to have the phonetic training to understand feedback that is too technical for students. Teachers can also provide special review sessions to discuss students' performance and progress; they can redirect students' focus based on app feedback focusing on specific phonemes or phonetic features that are problematic for each student. They also can help students to be more aware of their errors, how to identify them and make progress. So the role of the teacher is still essential in the class to interpret the results from apps and help students reflect on app feedback and direct their focus for more personalized activities, and set realistic goals. So, these apps will be just a tool for more engagement and independent practice. Aligning with the app activities, students can get guidance from teachers to direct their focus of study (Abimanto & Sumarsono, 2024), it was stated by Abimanto and Sumarsono (2024) that "special sessions with their teachers to discuss results from Google Read Along app were very helpful."

Additionally, combining ASR feedback and teachers' guidance with peer correction would create a more dynamic learning environment. Students can compare their performances among each other, and they can develop critical listening skills as they analyze their peers' pronunciation. As found by Sun (2023); "the utilization of ASR technology with peer correction can be a robust approach in teaching and learning pronunciation and speaking skills among EFL learners".

6. Focus and limitations of ASR based tools:

The main limitation of ASR tools is the lack of recognition of different English varieties. ASR tools seem to support native speaker norms, which contradict the Word Englishes paradigm by Kachru. For instance, ELSA Speak recognizes only General American English, which can frustrate learners from different backgrounds or those learning other varieties. Learners using British English were assessed as if their pronunciation was incorrect, leading them to feel confused and, in some cases, demotivated (Matysiak, 2023).

This may reflect the lack of inclusivity of different varieties and the reinforcement of the idea that English belongs to native speakers. Moreover, mentioning that teachers still demand that their students be consistent with a variety means that English is the property of the native speaker (Llurda, 2018). Hence, these apps may not be reliable or realistic because it is almost impossible to achieve native-like pronunciation, especially with adult learners. This focus on phonetic precision can lead to the creation of unrealistic expectations. However this has shifted the focus towards communication and intelligibility, which are more practical concerns.

Since one of the main goals of this dissertation is to investigate whether ASR prioritizes intelligibility and comprehensibility over accentedness, it is important to note that a strong accent does not necessarily make speech unintelligible (Munro & Derwing, 1999). However, Muzzzaki (2024) points out that there is a correlation between accentedness and comprehensibility. He confirms that NOVO Learning helps learners to reduce their foreign accent and enhance comprehensibility. Moreover, these apps seem to be focusing and prioritizing only on pronunciation and phonetic accuracy over intelligibility, two factors which may be difficult to set apart in real life, resulting in

unreliable measure of successful communication. As pointed out: “One main drawback is that it only focuses on pronunciation and ignores chances for comprehensive language acquisition, including training in grammar and vocabulary. Students who must develop these skills simultaneously may find this problematic”(Rusmawaty et al., 2024). Since the focus of pronunciation instruction is to increase intelligibility and comprehensibility (Munro & Derwing, 1999), the pedagogical goal is not to sound native. Thus the goal must be intelligibility (Celce-Murcia et al., 1996), making ASR tools seem biased toward native speakerism and failing to distinguish between being understandable and being native.

ASR tools such as Dragon Naturally Speaking, which was tested by Derwing, Munro and Carbonaro (2000), recognize native speech with “90% accuracy, but only 71-72% for non-native speech”. This occurs because most ASR apps are designed to rely on limited native data, which can lead to less accuracy for non-native accents (Eskenazi, 1999). As a result , ASR can be considered a biased technology; the application’s limitations in voice recognition can lead to struggles with underrepresented dialects and being less helpful in providing feedback, and not reliable with regional variations, which limits its pedagogical reach. (Pham and Pham, 2025) .

Moreover, these apps may be overly critical of learners’ minor errors, which can lead to self-doubt, it was stated that“… a few students reported that the apps are being too critical with accent or little pronunciation errors, such a feedback can lead them to demotivation and self-doubt” (Rusmawaty et al., 2024). Sometimes, this kind of feedback can seem harsh on students who already feel insecure. It was reported that these apps can be also repetitive or overly rigid, as mentioned by Rusmawaty et al. (2024) “some students may find it repetitive and not intuitive as they thought at the

beginning so it may reduce their engagement the content could be redundant; therefore, some users might find it challenging to find ways to go beyond the same feedback patterns". Furthermore, these apps often fail to identify idioms, slang, and contextual speech, which limits advanced learners' use (Pham & Pham, 2025). Moreover, it was stated that ASR is a tool that is meant for the voice recognition and not assessing and giving feedback in order for the speech to make progress (Levis & Suvorov, 2013)

Another limitation pointed out by Muzzaki et al. is that apps like ELSA SPEAK rely on written speech read by participants rather than spontaneous speech representing real life conversations. Kim (2006) also points out that ASR tools are effective but not as accurate as a teacher. Overall, almost all of the studies have showed that automatic measures or ASR tools are never as good and accurate as human ratings; however, the combination of both ASR tools and human input is better than a single rating and more reliable (Levis & Suvorov, 2013).

Some Studies have shown that ELSA Speak is not beneficial for all students and learners. Rinaepi et al. (2022) point out that ELSA Speak may improve students' pronunciation by 17%. Due to individual differences, not all students will find it beneficial. For instance, students who have difficulties in identifying and learning the phonetic system, and not all children respond positively to digital devices, it was reported that,"these findings suggest that while ELSA Speak may benefit many students, it may only be a panacea for some learners' pronunciation challenges" (Rusmawaty et al., 2024).

Overall, the main limitation of ASR tools consists in the lack of the recognition of different varieties, reinforcing native speaker norms, which can lead to demotivation

and self-doubt. Additionally, it can be ineffective equally for each student due to their individual differences. It can also be over criticizing minor errors, ignoring idioms slang, recognizing only General American Englsih, and prioritizing phonetic accuracy over intelligibility. Moreover, the feedback may seem repetitive at times and the content may appear poor in free versions.

7. Conclusion:

Overall, by mentioning the main difficulties of English pronunciation, the different approaches to teaching pronunciation, the rise of education 4.0, and analyzing the existing research and data, it has been demonstrated that ASR-based tools have several benefits in language learning. ASR-based tools can provide immediate and personalized feedback to improve pronunciation. They can enhance motivation through interactive games and independent practice. Moreover, this technology can be a good supplement for teachers in classrooms. In this context, teachers can interpret the feedback provided by the apps and redirect the focus of students. Thus, teachers' role is essential in this case to clarify any doubts and help students set realistic goals.

Despite their benefits, ASR- based tools have certain limitations. They can be biased towards native speaker norms, not recognizing other varieties than General American English, and focusing only on phonetic accuracy, leaving aside intelligibility and comprehensibility. Their overcorrection of minor errors and critical feedback can impact students negatively, leading to demotivation and self-doubt. Moreover, the effectiveness of these tools cannot be assured equally to all students due to their individual differences, and their lack of recognition of idioms, slang and context can result confusing.

Therefore, ASR-based tools can be powerful and useful if they are integrated thoughtfully into classrooms. Aligning with teachers' guidance and peer feedback, they can have a positive impact on students and their process of language learning. However, there is a need to include more inclusivity and flexibility to suit different learners' profiles and English varieties. It would be relevant for further research to consider finding effective ways to integrate ASR tools effectively into classrooms, and to combine ASR tools feedback along with teachers' clarifications and peer feedback. The implementation of different strategies for effectively integrating of this technology can be investigated, as well as the complementary role of ASR tools and human feedback on segmental and suprasegmental features.

References

Abimanto, Dhannan., & Sumarsono, Wasi. (2024). Improving English Pronunciation with AI Speech-Recognition Technology. *Acitya Journal of Teaching and Education, 6*(1), 146–156. https://doi.org/10.30650/ajte.v6i1.3810

Avery, Peter & Ehrlich, Susan. (1992) *Teaching American English Pronunciation*. Oxford: Oxford University Press

Celce-Murcia, Marianne, Donna Brinton, Janet M. Goodwin. (1996). *Teaching pronunciation: a reference for teachers of English to speakers of other language.* New York: Cambridge University Press.

Derwing, Tracey M., Munro, Murray J. & Carbonaro, Michael (2000). Does Popular Speech Recognition Software Work with ESL Speech?. *TESOL Quarterly, 34*(3), 592-603. Retrieved May 27, 2025 from https://www.learntechlib.org/p/91560/.

Eskenazi, Maxine. (1999). Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype. *Language Learning & Technology*, 2(2), 62–76. http://dx.doi.org/10125/25043

Hong, Christina & Ma Wai Kit, Will. (2020). *Applied Degree Education and the Future of Work: Education 4.0* (1st ed. 2020.). Springer Singapore.

Kim, In-Seok. (2006). Automatic Speech Recognition: Reliability and Pedagogical Implications for Teaching Pronunciation. *Journal of Educational Technology & Society*, *9*(1), 322-334.

Kholis, Adhan. (2021). Elsa Speak App: Automatic Speech Recognition (ASR) for Supplementing English Pronunciation Skills. *Pedagogy: Journal of English Language Teaching, 9*(1), 01-14.

Kim, In-Seok. (2006). Automatic Speech Recognition: Reliability and Pedagogical Implications for Teaching Pronunciation. *Journal of Educational Technology & Society*, *9*(1), 322-334.

Levis, John & Suvorov, Ruslan. (2013). Automatic Speech Recognition. *In The Encyclopedia of Applied Linguistics*, edited by C. Chapelle, NJ: Blackwell Publishing.

Llurda, Enric. (2018). English language teachers and ELF. In Jenkins, Jennifer, Baker, Will & Dewey, Martin (eds.), *The Routledge Handbook of English as a Lingua Franca*. (pp. 518-526). Routledge.

Matysiak, Aleksandra (2023). Students' perceptions on the use of selected mobile apps in the process of acquiring L2 pronunciation – a preliminary study. *Językoznawstwo*, *2/19*, 209–233. https://doi.org/10.25312/j.6840

Munro, Murray J. & Derwing, Tracy M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 49*, 285-310.

Muzakki, Bashori, Roeland, van Hout., Helmer, Strik. & Catia, Cuccharini. (2024). I Can Speak: improving English pronunciation through automatic speech recognition-based language learning systems. *Innovation in Language Learning and Teaching, 18*(5), 443-461.

Nguyen, Dong, Thi Thao & Van Tuyen, Le. (2024) The Effects of Using "Elsa Speak app" on the Enhancement of College Students' English-Speaking Skills.

(n.d.). *International Journal of English Literature and Social Sciences*. https://doi.org/10.22161/ijels.91.4

Pham, Vi Thi Tuong., & Pham, Anh Tuan. (2025). English major students' satisfaction with ELSA Speak in English pronunciation courses. *PloS One*, *20*(1), e0317378-

Rinaepi, Rinaepi, Triwardani, Henni Rosa & Azi, Raysal Nur (2022). The Effectiveness of Elsa Speak Application to Improve Pronunciation Ability. *Jurnal Fakultas Keguruan Dan Ilmu Pendidikan*, 3(1), 28–33.

Rusmawaty, Desy, Limbong, Effendi, Ahada, Ichi, F. Hafizh, M. Indra &Rahmatullah, Achmad. N. (2024). Unlocking Phonological Proficiency: Exploring Allophonic Variation Using ELSA Speak App in Early Semester EFL Students at Mulawarman University. *Indonesian Journal of EFL and Linguistics*. https://doi.org/10.21462/ijefl.v9i2.828

Saragih, Enni, Erawati, Tabrani, Nur'aini, Putri & Muthmainnah, Nur. (2021). The Use of Digital Feedback on Elsa Speak in Learning Pronuncitaion for Seventh Grade of Junior High School. *JEELL (Journal of English Education, Linguistics and Literature) English Department of STKIP PGRI Jombang*, 8(1), 48. https://doi.org/10.32682/jeell.v8i1.1979

Sicola Laura & Darcy Isabelle. (2015). Integrating Pronunciation into the Language Classroom. In Reed, Marnie & John M. Levis (Eds.), *The Handbook of English Pronunciation* (pp. 471-487). Oxford: Wiley Blackwell.

Sun, Weina (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: a mixed methods investigation. *Frontiers in Psychology*, *14*, 1210187–1210187. https://doi.org/10.3389/fpsyg.2023.1210187

Teschner, Richard V. & M. Stanley Whitley. (2004) From Orthography to Pronunciation. In *Pronouncing English: a stress-based approach*. Washington, D.C.: Georgetown University Press.

Websites:

ELSA Speak. https://elsaspeak.com/en/