UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS TRINDADE
CENTRO DE COMUNICAÇÃO E EXPRESSÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INGLÊS

Daniel Reschke Pires

An investigation of an ASR-based mobile application and its effects on speech intelligibility

Florianópolis
2022

Daniel Reschke Pires

An investigation of an ASR-based mobile application and its effects on speech intelligibility

Tese submetida ao Programa de Pós-graduação em Inglês: Estudos Linguísticos e Literários da Universidade Federal de Santa Catarina para o Exame de Qualificação, como requisito parcial para a obtenção do título de Doutor em Inglês: Estudos Linguísticos e Literários

Orientador: Prof. Dr. Celso Henrique Soufen Tumolo.

Florianópolis
2022

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Reschke Pires, Daniel
An investigation of an ASR-based mobile application and its effects on speech intelligibility. / Daniel Reschke Pires ; orientador, Celso Henrique Soufen Tumolo, 2022.
188 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Comunicação e Expressão, Programa de Pós Graduação em Inglês: Estudos Linguísticos e Literários, Florianópolis, 2022.

Inclui referências.

1. Inglês: Estudos Linguísticos e Literários. 2. Automatic Speech Recognition. 3. L2 pronunciation instruction. 4. Speech intelligibility. 5. Mobile Assisted Language Learning. I. Soufen Tumolo, Celso Henrique . II. Universidade Federal de Santa Catarina. Programa de Pós Graduação em Inglês: Estudos Linguísticos e Literários. III. Título.

Daniel Reschke Pires

An investigation of an ASR-based mobile application and its effects on speech intelligibility.

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Celso Henrique Soufen Tumolo, Dr.
Universidade Federal de Santa Catarina (UFSC)

Profa. Caroline Chioquetta Lorenset, Dra.
Instituto Federal de Santa Catarina (IFSC),

Profa. Rosane Silveira, Dra.
Universidade Federal de Santa Catarina (UFSC)

Prof. Walcir Cardoso, Dr.
Concordia University

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado adequado para obtenção do título de doutor em Estudos da Linguagem.

Documento assinado digitalmente
Celso Henrique Soufen Tumolo
Data: 16/12/2022 10:13:30-0300
CPF: ***.275.298-**
Verifique as assinaturas em https://v.ufsc.br

Celso Henrique Soufen Tumolo, Dr. - Coordenação do Programa de Pós-Graduação

Documento assinado digitalmente
Celso Henrique Soufen Tumolo
Data: 16/12/2022 10:14:49-0300
CPF: ***.275.298-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Dr. Celso Henrique Soufen Tumolo
Orientador

Florianópolis, 2022

I dedicate this thesis to those who believed in me, and to those who believe in science.

ACKNOWLEDGEMENTS

There are many people who contributed, directly or indirectly, to the development of this piece of research. Their trust, dedication and experience were invaluable for my growth as a researcher and as an educator, so I take this opportunity to acknowledge their support.

I would like to express my gratitude to my advisor, Professor Celso Henrique Soufen Tumolo, who was also my advisor during my master's degree. Since 2016, we have worked together in various projects and publications, and, during this time, Professor Celso has supported me with his insights and partnership. I thank you for the sincere conversations we have had about academic and non-academic issues, and also for the liberty you have given me throughout the development of this thesis, which has allowed me to carry out research on challenging topics and to explore new possibilities for L2 instruction. Finally, I thank you for the horizontal relationship we have had.

I would like to thank all the teachers from the foreign language department from Colégio de Aplicação, whom I have had the pleasure to work with since 2019. Most of all, I am thankful for the support and friendship of my English teacher colleagues, Maristela, Nadia and Renata, who substituted me for three months so that I could focus on the development of this thesis. Dear colleagues, without you, this would have been a much more difficult process. It has been an honor to work with such dedicated and empathetic people.

My special gratitude goes to the members of the defense committee: Caroline Chioquetta Lorenset, Rosane Silveira, and Walcir Cardoso. Professor Caroline Chioquetta Lorenset (IFSC - Florianópolis), who was also advised by Professor Celso Tumolo, shares an interest in digital technologies and vocabulary learning, and was an important colleague in my years as a PhD candidate. Professor Rosane Silveira (UFSC) has been a very important professional and researcher for me. Having taken three courses at graduate level with her, I can confidently state that she has taught me a lot about second language acquisition, L2 pronunciation learning and instruction, and statistics. In fact, it was during one of her courses (*Produção e Percepção da Fala em L2*) that I decided to investigate ASR-based instruction, which was a topic in one of our readings for the course. Professor Silveira was also a member of my qualifying committee, and helped me at various stages of my PhD journey. Finally, I would like to thank Professor Walcir Cardoso (Concordia University, Montreal), whom I have not yet had the pleasure to meet in person, but whose research on phonology, sociolinguistics,

and digital technologies for L2 have contributed significantly for this piece of research. Dear members of the defense committee, I thank you for dedicating the time and energy to read my research and for the contributions you have made to it.

I am very thankful for the contributions that current and former members of PPGI (UFSC) have throughout my journey as a PhD candidate. A special thanks to Professor Hanna Emilia Kivisto de Souza, who was part of my qualifying committee and who made important suggestions in the early stages of the research. I thank Valdete Billotta, who worked at PPGI office until July 2022, and whose work helped me present my research in places such as Colombia and Portugal. I must also thank Davi Alves Oliveira for reviewing the statistical tests used in this study, and for giving me valuable advice regarding the more quantitative aspects of this piece of research. Finally, I would like to thank my colleague and friend, Bruno de Azevedo, for sharing so many moments of learning and joy since 2016. Thank you for being a resourceful colleague, but also for being there as a friend in the moments of doubt and anxiety, which, as you know, tend to be plentiful among graduate students.

My gratitude also goes to the participants of this study. First, I would like to thank the students from *Colégio de Aplicação*, who dedicated their time to using the application among their very busy study routines, especially during the havoc caused by the COVID-19 pandemic. Second, I acknowledge the people from all over the world who participated as listeners and voluntarily made the effort to transcribe the students' sentences. Thank you very much for that!

Carrying out scientific research in Brazil has been notoriously difficult in the last years due to the multiple cuts in funding and continuous attacks to public universities. In this context, I must thank the National Council for Scientific and Technological Development *(Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq)* for the two years of scholarship, which enabled me to dedicate exclusively to my PhD during its first two years. This type of funding is essential for the scientific and technological growth of any country, and should be a priority for future governments.

Finally, I must thank those who were my bridge over troubled water when times were rough. First, I thank my partner, Mariana Rudek, who heard every research idea I had and heard me talk endlessly about it, even when she did not understand much of what I was saying. I must also thank my family, who have always believed in the value of education, and, most of all, believed in me. I'm specially thankful for my mother, Lia, who has always encouraged me to

study, and to be curious and resilient. Last but now least, I must thank myself for attempting to carry out this piece of research, and for persevering.

Autonomia é algo que se constrói por dentro, com projetos e com expectativas, com diálogo e com interação. E nada disso se fará sem esperança. Somente no coração cotidiano da escola poderão ser instituídas novas formas de legitimação do ato de ensinar, com ciência, com arte, e certamente com muito tato pedagógico. (BOTO, 2017, p. 293-294).

ABSTRACT

Nowadays, numerous commercially available mobile applications employ Automatic Speech Recognition (ASR) technologies to provide immediate and individualized feedback on the pronunciation of L2 learners. However, it is still unknown whether these applications can help learners become more intelligible speakers. This study investigated the effects of an ASR-based mobile application on the speech intelligibility of Brazilian learners of English. It also sought to investigate the self-regulated use of the application and the participants' perceptions of it. The participants were 13 students from a Brazilian public school, who used an application named ELSA Speak to complete 38 lessons focused on the instruction of the English vowels /i/ and /ɪ/. They received instructions on how they were expected to study with the application, but were allowed to choose when, where and how they used it. To measure their intelligibility before and after the instruction, they recorded short sentences containing /i/ and /ɪ/ target words, which were then orthographically transcribed by native and non-native listeners. The results from the intelligibility tests revealed significant differences between the pre and posttests for overall and target-word intelligibility, with participants with lower levels of intelligibility in the pre-tests having more gains than those with higher levels. Although the participants were asked to complete the lessons in one month, a few of them failed to do so, likely due to a lack of motivation to maintain the use of the application. Overall, the participants had positive perceptions of the application and believed that it was able to recognize their speech and provide appropriate feedback, even though they reported instances in which they struggled to be recognized. They appreciated the possibility of learning without the judgement of others and of choosing when to learn, but believed that the lack of a teacher and their lack of organization were major disadvantages in this type of learning. This study has relevant implications for L2 pronunciation instruction as it provides empirical data on the potentials and limitations of ASR-based instruction and Mobile Assisted Language Learning.

Keywords: Automatic Speech Recognition; L2 pronunciation instruction; Speech intelligibility; Mobile Assisted Language Learning; Self-regulated learning.

RESUMO

Existem atualmente diversos aplicativos móveis que utilizam tecnologias de Reconhecimento Automático de Fala (RAF) para fornecer retorno individualizado e imediato à fala de estudantes de segunda língua. Todavia, ainda não se sabe se estes aplicativos podem ajudá-los a se tornarem falantes mais inteligíveis. Este estudo buscou investigar o uso e os efeitos do uso de um aplicativo móvel baseado no RAF na inteligibilidade da fala de aprendizes de inglês. Além disso, o estudo buscou investigar o uso autorregulado do aplicativo e as percepções dos participantes a respeito dele. Os participantes da pesquisa foram 13 estudantes de uma escola pública de educação básica (Colégio de Aplicação – UFSC), os quais utilizaram o aplicativo *ELSA Speak* para completar 38 lições cujo foco era a instrução das vogais /i/ e /ɪ/ da língua inglesa. Eles receberam instruções sobre como utilizar o aplicativo, mas puderam escolher quando, onde e como o utilizariam. Para medir a inteligibilidade antes e depois do uso do aplicativo, eles gravaram sentenças curtas contendo palavras-alvo com os sons /i/ e /ɪ/, as quais foram transcritas ortograficamente por 13 ouvintes nativos e não-nativos da língua inglesa. Os resultados dos testes de inteligibilidade revelaram diferenças significativas entre os pré e os pós-teste em termos de inteligibilidade geral e inteligibilidade das palavras-alvo, sendo que os participantes com menor nível inicial de inteligibilidade tiveram os maiores ganhos. Embora os participantes tenham sido instruídos a completar as lições durante o período de um mês, alguns não conseguiram cumprir este prazo e tiveram um uso inconsistente do aplicativo, provavelmente porque não estavam motivados para continuar seu uso. Em geral, os participantes tiveram percepções positivas sobre o aplicativo e acreditam que ele foi capaz de reconhecer sua fala e dar feedback apropriado, embora alguns tenham relatado momentos em que enfrentaram dificuldades para ter sua fala reconhecida. Eles apreciaram a possibilidade de aprender sem se sentirem julgados pelos colegas, e de poderem escolher quando aprender, mas acreditam que a falta de um professor e a falta de organização para utilizar o aplicativo foram as maiores desvantagens envolvidas na sua experiência com a aprendizagem móvel. Este estudo traz implicações relevantes para a instrução de pronúncia em segunda língua à medida que fornece dados empíricos sobre as potencialidades e limitações da instrução de pronúncia baseada em RAF e da aprendizagem de línguas assistida por dispositivos móveis.

Palavras-chave: Ensino de Pronúncia em L2; Reconhecimento Automático de Fala; Inteligibilidade da fala; Aprendizagem de Línguas Assistida por Dispositivos Móveis; Aprendizagem Autorregulada.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ASR – Automatic Speech Recognition

CA – Colégio de Aplicação

CALL – Computer Assisted Language Learning

CAPT – Computer Assisted Pronunciation Training

CEFR - Common European Framework of Reference for Languages

EFL – English as a Foreign Language

ELSA – English Language Speaking Assistant

ESL – English as a Second Language

FL – Foreign Language

IPA – International Phonetic Alphabet

L – Listener

L1 – First Language

L2 – Second Language

LFC – Lingua Franca Core

MALL – Mobile Assisted Language Learning

MAPT – Mobile Assisted Pronunciation Training

NNS – Non-native Speaker

NS – Native Speaker

P – Participant

RQ – Research Question

SLR – Self-Regulated Learning

UFSC – Universidade Federal de Santa Catarina

SUMMARY

1 INTRODUCTION

The number of people who use mobile devices for learning has grown dramatically in the last five years. In the first quarter of 2017, approximately 522 million educational mobile applications were downloaded, while in the first quarter of 2020 this number jumped to 936 million (CECI, 2021). What is more, in 2022, mobile education apps were the second most downloaded type of application worldwide, amounting 10.4 % of the total downloads (CECI, 2022). Multiple factors have contributed to this, including the increased accessibility and affordability of smartphones, the lack of time and resources to attend face-to-face courses, the COVID-19 pandemic, and, in the case of applications for English learning, the increasing academic and professional demands for the language (PALMAI; SMALE, 2021). To give it some context, the three most popular applications of 2018, namely Duolingo, Hello English and Memrise, had more than 390 million users in that year alone (PIRES, 2018). In 2020, fueled by the lockdowns imposed by the COVID-19 pandemic, Duolingo had 67% more users than in 2019 (PALMAI; SMALE, 2021). While these and other popular mobile applications have predominantly focused on vocabulary instruction (ANDRADE, 2017; DUMAN; ORHON; GEDIK**,** 2014; HEIL; WU; LEE, 2016; PIRES, 2020), the incorporation of Automatic Speech Recognition (ASR) technologies has allowed some applications to recognize learners' speech and provide immediate feedback for their pronunciation (NERI; CUCCHIARINI; STRIK, 2006). However, the potential of digital technologies such as ASR for pronunciation instruction has not been extensively exploited (MUNRO; DERWING, 2015).

Researchers have argued that pronunciation instruction should help learners become intelligible speakers (DERWING; MUNRO, 2005; FIELD, 2005; KAUR, 2018). This argument is based on the understanding that speakers can maintain aspects of their first language (L1) as long as they meet the minimal phonological requirements to achieve successful communication (SAITO, 2012). Considering that English is commonly used as a Lingua Franca[1], these speakers need to be comprehended not only by native speakers of the language, but also by non-native ones. In light of this, pronunciation instruction should not be guided by the nativeness principle, wherein learners are taught to emulate the speech of native speakers, but should adhere to the intelligibility principle and assist learners to become more

[1] In essence, English acts as a Lingua Franca when it serves as "a means of communication between people who come from different first language backgrounds" (JENKINS, 2012).

intelligible speakers, focusing on the instruction of phonological features that can enhance their intelligibility (LEVIS, 2005).

In spite of the growing interest in intelligibility research, pronunciation teaching is often overlooked in foreign/second (L2)[2] language classes (GILAKJANI; SABOURI, 2016; NAIR et al., 2017). In the case of the English language, teachers have reported a number of challenges involved in pronunciation teaching. For one, many do not feel prepared to teach pronunciation because they have not been adequately trained for that (HENDERSON et al., 2012; THOMSON, 2012), an issue that has also been identified in Brazil (BUSS, 2013; COSTA, 2016). The lack of time and the lack of resources have also been pointed as limiting factors for teaching pronunciation (DERWING; MUNRO, 2009; GILAKJANI; SABOURI, 2016). Considering the limited time learners have in L2 classrooms, ASR-based applications have been seen as potential tools to develop L2 pronunciation, as they can be used for as long as learners need (KAISER, 2018). Moreover, they have the potential to promote pronunciation practice with reduced levels of anxiety (BASHORI et al., 2021), and with the possibility of choosing when and where to study (HINCKS, 2005; KAISER, 2018). However, this often means that learners need to choose adequate places and times to study, and need to be able to self-regulate, that is, they need to regulate their learning process to accomplish learning goals independently. Currently, investigation is necessary to understand if learners can actually benefit from the affordances of ASR-based mobile instruction, and if this type of instruction can help English learners become more intelligible. In light of this, the objectives of the current study are presented in the following section.

1.1 OBJECTIVES

This study has the general objective of investigating the effects of the use of an ASR-based mobile application on the intelligibility of Brazilian learners of English. More specifically, it seeks:

1) To investigate the effects of the use of an ASR-based mobile application on the overall intelligibility of Brazilian learners of English, using orthographic transcriptions as measures of intelligibility.

[2] The term L2 is henceforward adopted to refer to both foreign and second languages. The terms ‘foreign language’ and ‘second language’ were used when a distinction between them was necessary, or when the referenced studies dealt with specifically with one of them.

2) To investigate effects of the use of the ASR-based mobile application on the intelligibility of the English high front vowels (/ɪ/ and /i/).

3) To investigate how the participants engaged in self-regulated learning with the mobile application.

4) To investigate the participants' perceptions of the ASR-based application and of mobile learning.

1.2 RESEARCH QUESTIONS

Based on its objectives, four research questions (RQs) guided the current study:

RQ1) Does the use of an ASR-based mobile application improve the participants' overall intelligibility in English?

RQ2) Does the use of an ASR-based mobile application improve the participants' intelligibility of the English high-front vowels?

RQ3) How did the participants engage in self-regulated learning with the mobile application?

RQ4) What were the participants' perceptions of the application and of mobile learning?

1.3 SIGNIFICANCE OF THE STUDY

The literature on the field of Mobile Assisted Language Learning (MALL) has grown considerably in the past decades, with studies related to vocabulary learning being the most proliferate ones in the field (DUMAN et al., 2014; KAMASAK et al.; 2021). There have also been several studies that explored the potential of ASR-based systems (ELIMAT; ABUSEILEEK, 2014; LIAKIN; CARDOSO; LIAKINA, 2015; NERI; CUCCHIARINI; STRIK, 2006). In these studies, the effects of pronunciation instruction were measured with perception tasks, and with production tasks in which the productions of the adult learners were judged by native speakers in terms of pronunciation accuracy. Although there have been studies that investigated the effects of ASR-based mobile instruction by using intelligibility measures (BURLESON, 2007; VAN LIESHOUT; CARDOSO, 2022), investigation is still necessary to

understand the feasibility of using ASR-based mobile applications to promote intelligibility in an L2, especially when they are used by young learners.

The COVID-19 pandemic has created the largest disruption in the history of education, affecting 94% of the world's students (UNITED NATIONS, 2020) and causing learners all around the globe to rely on digital technologies to have access to education (KAMASAK et al., 2020). In this context, smartphones became essential tools for remote learning because of their increased accessibility and portability, with ownership rates rapidly growing, especially in developing countries (OROZCO, 2021). In the specific case of Colégio de Aplicação, the public school where this study took place, 76% of the students reported using smartphones to access the remote classes during the pandemic, with 14% of them having smartphones as their only option of access (SILVA, 2021). The importance of these devices is even higher among less favored socioeconomic groups in Brazil. For example, 78% of the families with incomes lower than one minimum salary depend exclusively on mobile phones to access the internet (MARTINHÃO, 2019). Thus, it becomes relevant to explore the potential that such prevalent devices have to promote MALL and provide insights on how they can be used by language learners and teachers.

Due to the ubiquity of smartphones and the proliferation of applications for language learning, more and more people have used their mobile devices with the intent of learning an L2. More often than not, this learning takes place in independent fashion, with learners themselves downloading applications and exploring their functionalities. Consequently, learners become responsible for establishing what, when, where and how to study (SHA et al, 2012) which means that, for optimal results, they should be able to self-regulate their learning process. However, literature on the self-regulated use of MALL applications is still scarce, with most of MALL research being conducted in formal, classroom-based contexts (PERSSON; NOURI, 2018). In light of this, the current research investigated the self-regulated use of MALL wherein learners could choose where and how they studied, and offers insights on the benefits and limitations of MALL that takes place in these settings.

1.4 ORGANIZATION OF THE STUDY

This thesis is organized into five chapters. Following the introductory chapter, Chapter 2 presents a literature review on topics relevant to this research, namely: a) Mobile Learning;

b) Self-regulated learning; c) Mobile Assisted Language Learning (MALL); c) Automatic Speech Recognition (ASR) and ASR-based systems; d) Intelligibility; and e) The English high-front vowels.

Chapter 3 describes and explains in detail the method used to achieve the objectives of this research, including the research questions, the study design, the research context, the participants, the materials and procedures for data collection, and the procedures for data analysis.

Chapter 4 presents and discusses the results of this study according to the four research questions that guided it. Finally, Chapter 5 presents final remarks based on the findings of the study, with its first section presenting a summary of the findings, followed by its pedagogical implications and limitations.

2 LITERATURE REVIEW

This chapter presents a review of literature of fields related to the current study, whose objective was to investigate the effects of the use of an ASR-based mobile application on the intelligibility of Brazilian learners of English. Thus, the sections of this chapter are divided according to the main fields that constitute the scope of this study. Section 2.1 presents the field of Mobile Learning and what the literature has claimed to be its strengths and challenges. Section 2.2 presents the field of Self-Regulated Learning, and discusses the role of self-regulation in mobile learning and the challenges involved in it. Section 2.3 introduces the field of Mobile Assisted Language Learning and examines its current trends and future challenges. In Section 2.4, the fields of Automatic Speech Recognition (ASR) and ASR-based pronunciation instruction are presented, along with a review of studies that investigated the use of ASR-based systems for L2 pronunciation instruction. Section 2.5 introduces the field of speech Intelligibility, reviews intervention studies, discusses pronunciation features that can affect intelligibility, and reviews research that has been conducted with Brazilian participants. Finally, section 2.6 is concerned with the English high front vowels, as they were the instructional focus of the lessons the participants completed in the application.

2.1 MOBILE LEARNING

The COVID-19 pandemic has affected education all over the globe. With schools closed in an effort to reduce the spread of the virus, teaching and learning were emergently adapted to take place remotely[3]. Consequently, large-scale, national efforts were made to make use of technology in support of remote learning[4], with remote learning tools becoming more urgent than ever. Considering the relevance of smartphones to learning in the current world scenario and the fact that smartphones are mobile devices, it is important to define and understand what mobile learning is and where it is situated within the different types of learning.

Because mobile learning is a relatively young field which is still evolving, defining it is an ongoing issue that is attached to the "technological attributes available and the pedagogical opportunities provided by those technologies" (CROMPTON, 2013b, p. 47). In other words, as digital technologies are constantly evolving, new possibilities arise for mobile learning, which in turn can alter the understanding of what mobile devices are and what mobile learning is. For instance, the mobility of tablets was once questioned because of their lack of portability and slow start-up times. Yet, tablets quickly became more portable and faster, making it hard to argue against their mobility (CROMPTON, 2013b).

Although definitions of mobile learning are not perpetual, Winters (2006) argues that mobile learning generally can be broadly categorized into four categories: a) technocentric, which was once the prevailing perspective in the literature and sees mobile learning takes place whenever a mobile device is used for learning; b) relationship to e-learning, which characterizes mobile learning as an extension of e-learning without regarding the unique aspects of mobile learning; c) augmenting formal education, in which mobile devices serve to augment formal, traditional learning; d) learner-centered, which poses that the focus of mobile learning should not be on the devices, but on the mobility of the learner[5].

---

[3] This type of education is also referred to as Emergency Remote Education, and is necessary during emergency situations when students are unable to attend school in person (CROMPTON et al., 2021).

[4] Remote learning, also known as Distance Learning, refers to learning that requires no in-person interaction between teachers and students (Stauffer, 2020). Nowadays, the use of digital forms of communication is paramount for this type of learning.

[5] The term 'learner' is employed to refer to anyone engaging in learning activities. To Leask (2013), learners absorb, analyze, create and discuss anytime, anywhere". The term 'student' is used when to refer to those who are students in educational institutions, such as schools and universities, usually under the guidance of teachers. The distinction is relevant as mobile learning often takes place without teachers and is not restricted to any learning environment.

Mobile learning has predominantly been seen as technocentric (WINTERS, 2006), and has been heavily underpinned by behaviorist models of instruction[6] [7] where knowledge is teacher-centered, and is transmitted, not constructed (BURSTON, 2014; PIRES, 2020). Nevertheless, there have been arguments in favor of more learner-centered approaches (CROMPTOM, 2013; HAAG; BERKING, 2019; O'MALLEY et al., 2005). For example, in O'Malley et al. (2005), mobile learning is defined as "any sort of learning that happens when the learner is not at a fixed, predetermined location, or learning that happens when the learner takes advantage of learning opportunities offered by mobile technologies" (2005, p. 7). Under this definition, the mobility of the learner is also a factor, not only the mobility of the technology. In this regard, O'Malley et al. argue that mobile learning can take place without the use of mobile technologies as long as learners are not at established, predetermined locations. On the other hand, they believe that mobile learning also happens whenever learners take advantage of mobile technologies, even if the learners are at their usual learning environments.

A more encompassing view is provided in Crompton (2013a), in which mobile learning is defined as "learning across multiple contexts, through social and content interactions, using personal electronic devices". The word 'context' has an important role in this definition:

> (…) the word context in this definition encompasses mobile learning that is formal, self-directed, and spontaneous learning, as well as learning that is context aware and context neutral. In other words, the learning may be directed by others or by oneself, and it can be an unplanned, spontaneous learning experience; learning can happen in an academic setting, or any other non-academic setting; and the physical environment may or may not be involved in the learning experience (2013a, p.4).

In light of this, the author exemplifies that mobile learning takes place when mobile devices are used in more traditional lessons, but also when people choose their own approach to achieving a learning goal i.e., in self-regulated learning. What is more, mobile learning may be connected to the environment where it takes place, such as when learners scan codes to learn more about the place where they are, but it may also have a neutral role, as when learners read

[6] Eisner (1964) defines instruction as "those activities that are consciously planned and executed by the teacher which are intended to move the pupils toward the attainment of the educational held by the teacher" (p. 118). In the context of mobile learning, the role of teaching is often performed by the software that is providing the instructional content to the learners.

[7] The terms 'instruction' and 'teaching' have been widely used in mobile learning literature, sometimes interchangeably. In this study, we adopt the term instruction when referring to the content and the activities provided by smartphone applications with the goal of promoting learning, while teaching is used to refer to classroom-based learning.

an article at home or while commuting. In sum, Crompton believes that "the essence of m-learning is not in the learning or in the technology, but in the marriage between the two entities" (2013, p.10).

This study adheres to the definitions of mobile learning brought by O'Malley et al. (2005) and Compton (2013), in acknowledging the array of contexts in which current mobile learning can take place, the mobility of the learners and of the technologies, and encompass the self-regulated nature of the study.

2.1.1 Strengths of mobile learning

One of the most appealing advantages of mobile learning is that learners can choose where and when to study (HSU, 2017; KHADDAGE et al., 2016), which allows them to "exploit small amounts of time and space for learning" (TRAXLER, 2007, p.8). This is especially true in the case of smartphones, which nowadays are carried virtually everywhere and are ready to be used when these unexpected small amounts of time appear (STOCKWELL, 2013). Because of this possibility, smartphones are often referred to as ubiquitous devices. Regarding this, Peng et. al concluded that the word ubiquity refers not only to the "idea of 'anytime, anywhere' but to the 'widespread', just-in-time', and 'when-needed' computing power for learners" (2009, p. 175).

Another advantage afforded by mobile devices is the possibility of using them in authentic environments[8] and of promoting situated learning[9] (NAISMITH, 2004; SHADIEV, 2020). Here, mobile devices serve to enhance the learning that is taking place in such environments. For example, when visiting museums and galleries, learners can use their smartphones to get additional information about exhibits (NAISMITH, 2004). Likewise, mobile devices have been explored to enhance field trips (PFEIFFER et al., 2009; ROGERS et al., 2010). Pfeiffer et al. (2009) investigated the suitability of mobile devices for situated learning by having university students use mobile DVD players during a field trip to study Mediterranean fish and found that these devices helped learning about biodiversity. In Rogers et al. (2010), students acted as researchers seeking to restore urban riverbanks, using

[8] An authentic environments are places where "the focal point and goal of all activities reflect the real-world situation in which mastery of the subject matter occurs" (CASEY, 1996, p. 79).

[9] Situated learning sets forth that learning can be enhanced if it takes place in an authentic context (NAISMITH, 2004)

smartphones to access contextual information about the trees they saw during the field trip. In both studies, mobile devices served to enhance the learning that was taking place in-situ.

Mobile devices have been regarded as potential tools to promote access to learning, especially in developing countries (BROWN, 2003; GASKELL, 2007; WEST, CHEW, 2014). According to Brown (2003, p. 1), wireless and mobile technologies "make it possible to provide learning opportunities to learners that are either without infrastructure for access (e.g., rural or remote learners) or continually on the move (e.g., business professionals)". Likewise, West and Chew (2014) found that one of the main reasons why people in developing countries use their mobile phones to read is that, as e-books are often free or inexpensive, they become more affordable and accessible options, especially in remote areas where there are no libraries and books. Because mobile devices such as cellphones and smartphones tend to be cheaper than laptops and computers, they have become the most used device to access the internet, and are often the only device available for this. According to Martinhão (2019), in 2018, 56% of the Brazilian population accessed the Internet[10] exclusively through mobile phones. Among less favored socioeconomic groups, 78% of the population relies exclusively on cellphones for this access (MARTINHÃO, 2019). In light of this, the accessibility of mobile devices can be regarded as one of its strengths, especially when compared to other Information and Communication Technologies[11] (ICTs). Paradoxically, access to these devices is one the major challenges in the implementation of mobile learning. This and other challenges of mobile learning are discussed in the following section.

2.1.2 The challenges of mobile learning

In spite of the aforementioned strengths, such as mobility, ubiquity and accessibility, mobile learning faces many challenges. While some of them are specific to this type of learning, others may be part of any type of learning that involves DICTs. In this regard, Souza Neto (2020) mapped what Brazilian teachers considered to be the challenges of implementing DICTs and found that the word 'lack' was the most prevailing one in their responses. They mentioned the lack of: technological devices, technical and pedagogical support, time, maintenance,

[10] Nowadays, accessing and using the internet involves more than accessing website. Rather, it involves any type of activity that requires internet connection (MARTINHÃO, 2019)

[11] Alaim (2020) defines ICTs as Technologies that emerge in the process of social digitalization, and include a variety of devices, such as computers, games, tablets, cellphones and smartphones.

internet access, guidance and formation. These answers reveal that the mere ownership of a mobile device is not enough for mobile learning, which typically requires access to all of the things the teachers in Souza Neto (2020) reported lacking. For instance, in a classroom where all students own a smartphone, mobile learning will still demand adequate internet connection most of the times, which can be a challenge on its own. In Brazil, although 82% of the schools have some sort of internet connection, it is only accessible in 68% of the classrooms. However, if learning is expected to have any type of mobility, internet must be wireless, and in this case only 45% of the Brazilian schools with Wi-Fi allowed students to access the network (TIC EDUCAÇÃO, 2020).

On top of the technological challenges, teachers often lack pedagogical and technical support for mobile learning. Although they may be familiar with the use of digital technologies for things such as accessing the internet, paying bills and using social media, many feel they are not prepared to use them with educational purposes. This was evidenced in 2020, as 84% of the basic schoolteachers from Brazil believed they were not prepared for remote learning, and 75% felt they needed support and training for it (TIC EDUCAÇÃO, 2020). Learning how to use mobile technologies effectively requires educational and technical formation, but, having limited time and pedagogical support, many teachers resist their implementation, especially into tasks that were once done without them (SOUZA NETO, 2020; CROMPTON, 2013c).

While some educators resist the implementation of mobile learning, students seem to be eager to explore the functionalities of the technologies they use so frequently in their everyday lives, and do not want to "power down as they enter the school" (CROMPTON, 2013c, p. 39); that is, they want to stay connected at all times, and at all places. In this regard, Souza Neto (2020) says that students are not more prepared than teachers for the pedagogical integration of DICTs in the school culture, but acknowledges they are more inserted the digital culture as they explore it with more curiosity, agility, and without many fears. Conversely, the author found that teachers put themselves in a comfort zone when they say they have no time to learn about new technologies among so many other assignments, or when they claim that, even if there were time, there would be no real pedagogical or financial incentive to learn about them (SOUZA NETO, 2020, p. 97). Thus, a difficult scenario emerges in which learners, especially younger ones, are keen to use digital devices, while this can represent a costly challenge for teachers.

Learners' enthusiasm to use technologies does not mean they will engage in effective mobile learning. For instance, Asabere (2013) noticed that mobile devices may facilitate cheating during tests and create gaps between students who own better devices and are more technological savvy, and those who do not have or do not know how to operate them. Another issue is that many learners use their smartphones during classes when they are not supposed to, or for unintended purposes and, being personal and portable devices, they can do this in a concealed manner. This is yet another problem for teachers, who often feel that they lose time with students' misbehaviors when mobile technologies are in use (CROMPTONc, 2013). Yang, Asbury and Griffiths (2019) found that main reasons for problematic smartphone usage are the fear of missing messages, boredom in class and poor self-regulation. They also learned that smartphones can negatively impact learners as they serve as a source for distraction, time wasting and procrastination in and outside classes.

While the possibility of learning anywhere and anytime is an undeniable advantage of mobile learning, it is also one of its challenges. The fact that learning can take place virtually anywhere and at any time also means that it can take place in environments that do not favor concentration such as noisy, busy places, which can lead to suboptimal outcomes. Because of this, Stockwell (2013) argued that activities for mobile learning should be "short and succinct, with short start-up times and short-segmented sections that can be completed individually as a single unit" (p. 211). In fact, mobile learning has been associated with behaviorist, teacher-centered models of instruction, which is arguably a consequence of the restraints imposed by their use in unfavorable learning conditions (BURSTON, 2014; PIRES, 2020). Mobile learning thus faces the challenge of exploring its communicational and multimedia affordances to support types of learning that are more collaborative, learner-centered, and task-based (BURSTON, 2014).

Another challenge faced by mobile learning is that the autonomy to choose where, what and when to learn leaves a gap as to who should be responsible for choosing appropriate places, times, and strategies to learn. Sha et al. (2012) wondered whether this should be a role for teachers, learners, instructional designers, or if should be a role shared by all of them. In this regard, the authors believed that:

> A relevant issue is that effective mobile learning is assumed to be dependent not only on learners' capabilities in rightfully determining what, when, where, and how to learn but also on their willingness to behaviorally and cognitively engage in learning whenever/wherever they themselves realize it is needed. This suggests that a student-centered mobile curriculum involves more than technological or pedagogical

> considerations, and learner characteristics such as motivation and abilities in monitoring and controlling one's learning in different settings need also to be considered (2012, p. 367).

Thus, successful mobile learning requires not only appropriate technology and pedagogy, but also engagement of the learners themselves. In other words, if learners have access to adequate mobile devices with well-designed instructional material but lack motivation and struggle to control and monitor their learning, then learning outcome is likely to be hindered.

This section discussed some of the challenges[12] that teachers and learners face with regard to mobile learning, such as the lack of access to adequate technologies, the scarcity of technical and pedagogical support, the problematic usage of mobile devices, and the challenges that arise from the autonomy that is part of mobile learning. The following section focuses on self-regulated learning, which has been argued to be a key factor and a predictor of successful mobile learning (SHA et al., 2012).

2.2 SELF-REGULATED LEARNING

As stated, one of the key features of mobile learning is the possibility of a more autonomous type of learning that allows learners to choose where, when and what to study. In Cotterall (1995), autonomy in learning is defined as "the extent to which learners demonstrate the ability to use a set of tactics for taking control of their learning" (p. 195) in which are included tactics for establishing goals, selecting materials, planning opportunities for practice, and monitoring and evaluating progress. Tumolo and Santibanes (2011) mention that having organization strategies are also a key factor for autonomous learning, among which are the administration and the efficient use of time, and the establishment of adequate places to study.

Because mobile learning affords a more autonomous type of learning, the ways and the goals of the learning process are often directed by the learners themselves. According to Knowles (1975), self-directed learning (SDL) happens when individuals evaluate their learning needs, formulate learning goals, identify the necessary human and material resources, and then choose and implement learning strategies for these goals. This happens, for example, when

[12] The challenges presented in this section are related to the scope of the current study, and represent only a fraction of the challenges faced by mobile learning. For more a more encompassing discussion, please refer to Asabere (2013).

people decide to use smartphone applications to learn a language they deem necessary for their lives.

Mobile learning can also take place when the goals are not decided by the learners, as when teachers tell students to use their mobile devices to accomplish a given task or goal. In this case, Saks and Leijen (2014) argue that self-regulated learning (SRL) is taking place. Even though the terms self-regulated and self-directed learning have been used interchangeably as both involve self-regulation, there is a key difference between them. In SDL, learners have control over the whole process, as they determine their learning goals and decide how they are going to achieve them. In SRL, however, the emphasis is on independently accomplishing a learning goal which was set by someone else. Because of this, SRL tends to be practiced predominantly in school environments, with teachers setting the directions and goals for learning and students being in charge of their study routines and strategies (SAKS; LEIJEN, 2014).

To Zimmerman (1990), SRL is different from other forms of learning and instruction because it emphasizes how students select, organize, and create advantageous learning environments, and how they plan and control the form and amount of their own instruction. In light of this, self-regulated learners should be able to: a) improve their ability to learn through the use of metacognitive and motivational strategies; b) establish optimal learning environments; c) choose the form and amount of instruction they need; and d) regulate and maintain their motivation (AN et al, 2021; SHA et al, 2012; ZIMMERMAN, 1990).

Garrison (1997) provided a comprehensive approach to self-directed and self-regulated[13] learning that encompasses three dimensions, namely self-management, self-monitoring, and motivation. The first dimension, self-management, is concerned with task control issues, and involves the management of the learning resources and the support available. Garrison (1997) points out that, in SRL, learners should be provided with choices in terms of how they want to reach the learning goals, and for that, it is important that different approaches are suggested and that learners with distinct paces are accommodated. In other words, as SRL entails having options and resources to achieve goals, it is crucial that learners know how to manage these resources and make choices that suit their learning styles and needs.

[13] Garrison (1997) uses the term self-directed learning to refer to both self-directed and self-regulated learning, as noted in Saks and Leijen (2014).

The second dimension, self-monitoring, is a process that requires monitoring the repertoire of learning strategies and the ability to think about how we learn i.e., metacognition. This dimension involves taking responsibility for the construction of meaning, and this is only possible if learners are willing and able to monitor the learning process. Self-monitoring is dependent on internal feedback, that is, the learner's own cognitive and metacognitive processes. As internal feedback may not be accurate and explicit enough, external feedback is also necessary because it can provide effective feedback to improve self-monitoring, as long as learners can integrate the external feedback to their internal assessments.

The third dimension, motivation, plays a vital role in the initiation and maintenance of the learning goals. In basic terms, to be motivated is to be moved to do something. People who feel no impetus or inspiration to act are characterized as unmotivated, while those who are can be considered motivated (RYAN, DECI, 2000). Garrison believes it is important to distinguish entering motivation from task motivation. Entering motivation is the initial process of deciding and showing intent to act, which the author sees as the "motivational reserve or fuel which the student possesses when initiating a learning experience" (GARRISSON, 1997, p. 27). He argues that learners will have higher entering motivation when they believe that the learning goals are of their interest, and when they think there is a good chance of achieving them. However, having high entering motivation does not mean learners will sustain their motivation throughout the learning process. For that, they must stay committed to their goals and persist until they are achieved, that is, they must have what the author refers to as task motivation. Task motivation is fundamentally linked to self-management because it demands learners to manage and sustain their motivation.

Moreover, he argues that motivation can be influenced by intrinsic and extrinsic factors, i.e., intrinsic and extrinsic motivation. Ryan and Deci (1985) defined intrinsic motivation as the doing of activities for their inherent satisfactions. They say that when people are intrinsically motivated, they are moved to act for the fun or challenge involved in the activity, and not because of external pressures or rewards. In contrast, extrinsic motivation refers to activities that are done with the goal of attaining separable outcomes, that is, they are done because they have an instrumental value, not because of the enjoyment of the activities themselves. For example, when students are moved to do their homework solely because they fear their parents' reprimands if they do not do it, their motivation is extrinsic because their actual goal is a separable one i.e., avoiding the reprimands. Likewise, students are extrinsically

motivated when they do activities because of their perceived instrumental value for their careers (RYAN, DECI, 2000).

Although extrinsic motivation has archetypally been thought as lesser type of motivation when compared to intrinsic motivation, Ryan and Deci (2000) point out that there are different types of extrinsic motivation:

> Students can perform extrinsically motivated actions with resentment, resistance, and disinterest or, alternatively, with an attitude of willingness that reflects an inner acceptance of the value or utility of a task. In the former case - the classic case of extrinsic motivation - one feels externally propelled into action; in the latter case, the extrinsic goal is self-endorsed and thus adopted with a sense of volition. Understanding these different types of extrinsic motivation, and what fosters each of them, is an important issue for educators who cannot always rely on intrinsic motivation to foster learning (p. 55, 2000).

Based on this, it seems like the challenge of formal education is to offer a curriculum that learners perceive as more than a summary of their obligations. Thus, typically extrinsic motivators such as assignments and tests should ideally be designed so that students eventually do them because they are interesting or enjoyable, not because they were forced to. As Garrison (1997) puts it "the challenge is to have students internalize external goals and rewards which are often more dominant during the entering stages of learning" (1997, p. 13).

SRL has been associated with learning performance and academic achievement (PINTRICH, DE GROOT, 1990; ZIMMERMAN, 1990). For example, in Pintrich and De Groot (1990), SRL was observed to affect learners' cognitive engagement, and academic performance in the classroom, which was measured with classroom tasks and assignments. In their study, students who reported more self-regulating abilities were more likely to persevere when they faced difficult or uninteresting tasks. On the other hand, students with lower self-regulating abilities had worse academic performance, and, as a result, lost their motivation to learn and their focus on their goals.

More recently, researchers have paid particular attention to the role of SRL in learning that involves the use of digital technologies (VAN LIESHOUT; CARDOSO, 2022; PALALAS, WARK, 2020). For example, Van Lieshout and Cardoso (2022) investigated the use of Google Translate in a self-directed[14] learning setting. In their study, thirty adults with no previous knowledge of Dutch were given a list with 10 learning goals (e.g., saying "I don't understand"

[14] Van Lieshout and Cardoso use the term "self-directed" learning because it is the most frequently used in the field of adult education, but acknowledge that "self-directed" and "self-regulated" are semantically similar terms.

in Dutch), which they attempted to learn autonomously by using Google Translate for 60 minutes. The results from the posttest revealed were significant vocabulary gains in the immediate posttest (9.5 of the 10 target sentences), but diminished results for the delayed posttest (5.4 of the 10 target sentences). There were also improvements to all pronunciation components assessed ((comprehensibility, accentedness and intelligibility) in the posttest. The authors found that the participants interacted with Google Translate in different ways, which indicates that the tool can be used by learners with different approaches to self-directed learning.

As Volga, Khalil and Baars (2020) point out, even though SRL was once thought to be relatively stable across learning situations, studies have observed that it varies according to the settings, i.e., online versus physical ones. Considering this, and taking into account the scope of the current study, the following section focuses on the role of SRL in mobile learning, and reviews a few studies in this regard.

2.2.1 Self-regulation in mobile language learning

The increased accessibility of mobile devices and applications has allowed many to use their devices as learning tools. Because mobile learning often takes place autonomously, that is, with learners themselves deciding what to study and how to study, SRL abilities can critically affect learning performance (CHEN, C., CHEN, L., YANG, 2018; PALALAS, WARK, 2020). However, self-regulation in mobile learning can be challenging because devices such as smartphones and tablets are not designed exclusively for learning. As Stockwell and Hubbard put it, “the primary function of mobile devices has been for personal and/or social purposes, as opposed to work or study purposes” (2013, p. 4). In fact, the literature in the field of mobile learning and MALL has shown that smartphones are predominantly for non-academic purposes, as texting and checking social media, and that these devices tend to be used when students are bored or when they are waiting for someone or something (ATAŞ, ÇELIK, 2019; DASHTESTANI, 2016).

Important aspects of self-regulation such as establishing optimal learning environments and maintaining motivation can be remarkably complex when learners can access a myriad of other media through the same device that is used for studying. In light of this, researchers have started to investigate how mobile learning takes place when students are not

in formal academic environments and need to self-regulate their learning. One of these studies is Botero et al. (2019), who sought to examine how university students engaged in self-directed, mobile language learning. To do so, they invited 574 modern language students at a Colombian university to use Duolingo[15]. Of the 149 students who reported using it after the invitation, 118 agreed to have their use monitored. After a year, the authors categorized the students into three categories. There were 73 try-and-quit users, who used the application for less than ten days of activity, 31 sporadic users, who used the application for more than ten days but less than 50, and 14 balanced users, who used Duolingo for more than 50 days. ctive users (n=44) and dropouts (n=35) o

The results from Botero et al. (2019) revealed overwhelming positive perceptions with regard to the application and to their experience learning with it. There were particularly high agreement rates when learners were asked if they would like Duolingo to be integrated as a component of independent learning, and when they were asked if they liked the gamification[16] of the application. In spite of the positive perceptions of various features of Duolingo, the number of active users decreased significantly over time. The authors report that the participants had high entering motivation but lacked the sustained motivation, self-management, and self-monitoring necessary for self-directed learning, and refer to this as a novelty effect because what initially motivated learners was the exploration of a new technology, not the desire for learning. They found that many students struggled to maintain their use of Duolingo because it was not attached to their formal learning obligations, which were often prioritized over the application. These students mentioned that if teachers had guided or graded their progress, they would have felt more motivated. In this regard, Haluk (2020) found that reduced motivation and lack of guidance were among the most common reasons to cause Duolingo to stop using the application. Botero et al. (2019) believe that students need training and support to use MALL effectively, and suggest exploring "whether a more active role of the teacher in relation to students' self-management and self-monitoring can lead to a more engaged student MALL participation in the informal, out-of-class context" (p. 90).

As shown, students may struggle with SRL in mobile learning if they lack guidance from educators. Aware of this, Chen, C., Chen, L., and Yan (2018) sought to investigate whether a technology-supported SRL mechanism could help learners with self-regulation and

[15] Duolingo is a free application for language learning, and is one the most popular applications available worldwide (PIRES, 2018).
[16] In brief, gamification refers to the use game elements outside games (RODRIGUES et al., 2022).

therefore improve their vocabulary learning. To do so, one class of EFL students from a Taiwanese elementary school acted as an experimental group and used a mobile application for vocabulary learning that contained an SRL module, in which learners could set goals for their learning, such as number of activities and words to be practiced daily. Another class from the same school used the same application but without the SRL module, and acted as a control group. The participants' vocabulary knowledge and motivation were measured before and after a two-week period of instruction, during which they could use the applications at their own pace, but were expected to practice for at least five hours a week. The participants from both groups were also classified as either field-dependent or field-independent[17] learners, according to their learning styles. The results from the study revealed that the participants from the experimental group significantly outperformed the control group in terms of vocabulary gains and learning motivation. However, no significant effects between the groups were found for field-independent learners. This finding suggests that SRL mechanisms can be especially important for field dependent learners, that is, for those who are more reliant on explicit directions and guidelines.

As mobile technologies evolve, more research becomes necessary to understand the relations between the fields of SRL and mobile learning. In their review of literature, Palalas and Wark (2020) found that most of the studies encompassing these two fields has been carried out in the area of Mobile Assisted Language Learning (MALL), which is the focus of the following section.

2.3 MOBILE ASSISTED LANGUAGE LEARNING: CURRENT TRENDS AND FUTURE CHALLENGES

Mobile Assisted Language Learning is a subset of mobile learning that uses mobile technologies to enhance language learning (LOEWEN et al, 2019). MALL is also considered a subset of Computer Assisted Language Learning (CALL), as mobile devices are also capable of computing data (CHINNERY, 2006). Although definitions of MALL may vary, the key components consistently involve the possibility to choose when and where to study, and the

[17] In the field of learning styles, field dependent learners learn better when they receive explicit directions and assignments, while field independent learners usually prefer more autonomy and like to direct their own learning (HICE, 2022).

adaptability to personal study habits (LOEWEN et al., 2019). In this regard, MALL does not differ from mobile learning, except that, in MALL, mobile devices serve to assist language learning.

Current mobile devices such as tablets and smartphones are equipped with many tools that can aid language learning. Learners can access online content wirelessly via a wi-fi modem or via mobile networks i.e., 4G connections; they can use their cameras to interact with other learners or to access content using QR[18] codes; they can use their microphones to record their speech and to communicate with other learners, with teachers and with the devices themselves. Finally, as these tools are built in a single, portable device, they can promote situated language learning, in that they can be used while learners work on authentic tasks that take place in the real world (GEORGE, SERNA, 2011). For example, learners can take their smartphones to a zoo and learn the names of the animals in the target language while also learning about them. In spite of the mobility of these devices, Persson and Nouri (2018) found that 70% of MALL research explored the use of mobile devices in indoor contexts such as in classrooms, with outdoor exclusive usage accounting being the focus of only 4% of the research reviewed.

Mobile applications[19] have predominantly focused on vocabulary instruction (DUMAN, ORHON, GEDIK, 2014), especially instruction of vocabulary in isolation (ANDRADE, 2017; HEIL, WU, LEE, 2016; PIRES, 2020). For example, out of the 50 applications analyzed in Heil, Wu and Lee (2016), only six focused on pronunciation, while 42 worked with vocabulary learning. In similar fashion, Andrade (2017) analyzed ten of the most popular applications for language learning and drew attention to the fact that in seven applications the instruction of vocabulary was done in a decontextualized manner, with isolated words and expressions presented in isolation and disassociated from communicative practices from the real world.

In their analysis of 69 studies on MALL published from 2000 to 2012, Duman, Orhon and Gedik (2014) found that 28 of these studies investigated vocabulary, while only six dealt with issues related to speaking and pronunciation. The reason why studies that tackled vocabulary in MALL have prevailed over other types of study is likely an effect of the number of applications that are designed for vocabulary instruction. In other words, because of the many

[18] QR codes (Quick Response codes) are black squares arranged in a square grid on a white background, which can be read by imaging devices such as cameras. Currently, they are widely used as they provide a fast, simple way to access links.

[19] Mobile applications, also referred to as a mobile apps or simply ‘apps’, are computer programs or software applications designed to run on mobile devices such as smartphones or tablets.

applications for learning vocabulary available, researchers tend to investigate this type of application.

Researchers have found that the majority of applications for MALL still work under a behaviorist paradigm that is based around drilling (HEIL, WU, LEE, 2016; LOWEN et al., 2019; PIRES, 2018). One reason for this is that developing an application whose main type of activities are decontextualized drills is arguably less pedagogically challenging than developing applications based on more communicative, contextualized resources. To Burston (2014), the predominance of vocabulary and grammar drills also has practical reasons as they are relatively easy to program and can be self-correcting. Moreover, and perhaps more relevantly, given that one of the advantages of mobile learning is the possibility of exploring short amounts of time and space for learning (TRAXLER, 2007), developers may see short and succinct activities as more well-suited for mobile learning. Drilling vocabulary activities can certainly be short, allowing learners to start and finish them quickly. In the words of Burston (2014, p. 346), "anytime, anywhere most often equates with brief time periods squeezed between other activities". Heil, Wu and Lee (2006) argue that one of the challenges MALL faces is moving away from the "behaviorist-style instruction that has long been abandoned in language classrooms" (p. 43), while implementing a more communicative model that is more in line with the present understanding of what language learning should be about. On top of that, another challenge is having a more communicative, holistic type of instruction that is also suitable for mobile learning environments.

As shown, there has been a trend to focus on vocabulary instruction, which indicates a disparity between the multiple affordances of mobile devices and their actual use for language learning. In this sense, MALL faces the challenge of developing other language skills, such as listening, speaking, and pronunciation, which have been underexplored so far. Considering this, one of the goals of the current study is to investigate the use of a mobile application that is designed for pronunciation development. Given that this application uses Automatic Speech Recognition (ASR), the following section reviews and discusses this technology and its use for pronunciation instruction.

2.4 AUTOMATIC SPEECH RECOGNITION AND ASR-BASED PRONUNCIATION INSTRUCTION

L2 teachers face a myriad of professional challenges on a daily basis. One of them is the limited time in class to develop all the necessary language skills, e.g., speaking, writing, listening and reading. These, along with all the other responsibilities of teachers, leave little time for pronunciation instruction, which some consider to be overwhelming and time-consuming, especially when it involves drilling practices (PI-HUA, 2006). In fact, Gilakjani and Sabouri (2016) found that 76 of the 100 EFL teachers who participated in their study believe they did not have enough time for pronunciation instruction.

When there is time for it, teaching pronunciation in classrooms poses challenges of its own nature. To date, numerous studies have reported that ESL and EFL teachers believe they lack the necessary training and confidence to instruct pronunciation (BUSS, 2013; COSTA, 2016; GILAKJANI; SABOURI, 2016; HENDERSON et al., 2012; THOMSON, 2012). For example, in Gilakjani and Sabouri (2016), 78% of the participants said they did not have enough knowledge to teach pronunciation, with 70% reporting to know more about grammar and vocabulary than about pronunciation. Another challenge involves personal feelings with regard to pronunciation. For one, learners tend to feel high levels of anxiety when speaking in front of their colleagues or teachers (PRICE, 1991). In this regard, Baran-Lucarz (2011) found that learners with lower pronunciation levels tend to feel more anxiety to speak in front of others, and that learners fear they will be negatively evaluated because of their accented speech. In light of such challenges, Automatic Speech Recognition (ASR) has been regarded as a technology that has the potential to assist pronunciation instruction.

In simple terms, Automatic Speech Recognition is the process of converting speech signals into words and phrases using digital technologies (HUANG et al., 2001). The advancements in ASR technology have enabled computers to successfully recognize and assess human speech, allowing pronunciation instruction to be individualized and maximized (HINCKS, 2005) and allowing pronunciation learning to be more autonomous (MCCROCKLIN, 2016). When it comes to L2 learning, CALL systems have incorporated ASR technologies to evaluate the pronunciation of language learners and to provide them with immediate feedback. Neri, Cucchiarini and Strik. (2003) referred to these systems as ASR-based Computer Assisted Pronunciation Training (CAPT) systems.

Broadly speaking, CAPT is the “use of technology for learning and teaching the segmental and suprasegmental features of a sound system” (ELIMAT; ABUSEILEEK, 2014, p. 2). CAPT can take place without the use of ASR if computers are used to teach and learn

pronunciation without the aid of ASR technologies. For example, in Bataineh and Al-Qadi (2014), authentic videos were used to develop learners' prosodic competence in English, without using ASR to do so. Due to its scope, the current study focuses on CAPT that uses ASR.

According to Neri, Cucchiarini and Strik (2003), an ideal ASR-based CAPT[20] system has five phases, namely: 1) speech recognition, when the ASR engine translates the speech signals into a sequence of words using phonetic and syntactic models; 2) scoring, when the system first evaluates the quality of the pronunciation by comparting temporal and acoustic properties of the learners' utterances to models which are usually based on native speakers; 3) error detection, when the system locates errors in the utterance; 4) error diagnosis, when the system identifies the specific types of errors of learners' utterances, since they may not be able to understand the nature of their mistakes on their own; and 5) feedback presentation, when the data gathered in phases 2,3 and 4 is presented to the learners, who will only benefit from this information if it is presented in a meaningful way, that is, feedback that can be interpreted by the learners.

ASR-based systems[21] that successfully run these 5 phases can have several benefits for learners. For one, they can practice pronunciation without the pressure of speaking in front of a whole classroom, reducing or eliminating the anxiety that students feel when doing so (BASHORI ET AL., 2021; KAISER, 2018). What is more, they can do so at their own pace and at their own time, receiving real-time feedback (MAHDI; KHATEEB, 2019). However, as observed in Neri, Cucchiarini and Strik (2003), developing ASR-based systems is a complex and challenging job, and problems have been found in different ASR phases, with one of the most pressing ones being their ability to recognize non-native speech (DERWING, MUNRO, CARBONARO, 2000). Problems of a more pedagogical nature have also been found. For instance, there have been ASR-systems that provided feedback by showing waveforms of what the learners pronounced and what they were expected to pronounce. While this and other types of complex feedback were once seen as innovative, they led to poor results since learners were unable to understand the feedback (NERI, CUCCHIARINI; STRIK, 2003).

[20] While these five phases were established in the context of ASR-based CAPT, they are still part of any ASR-based system, regardless of the device that is used to run them.

[21] The term 'ASR-based system' is henceforth adopted to describe any system that uses ASR in the context of second language (L2) learning, encompassing systems designed for stationary devices, such as desktop computers, and for mobile devices, such as smartphones. From now on, the term 'ASR-based CAPT system' is employed to refer specifically to systems designed for stationary devices.

In spite of the problems discussed, ASR-based systems have the potential to complement the learning that takes place in classrooms, which, as stated, is limited in time. In order to comprehend if the recent developments in speech recognition and assessment translate into gains for L2 learners, the following section reviews research which investigated ASR-based systems and their use for pronunciation instruction.

2.4.1 Studies investigating ASR-based systems for L2 pronunciation instruction

While ASR technologies may be perceived as a recent technological trend, their potential for L2 pronunciation instruction has been investigated since the 1990s. This section presents a literature review of studies that examined ASR-based systems and their suitability for L2 pronunciation instruction. The studies are reviewed in chronological order to evidence the advancements and changes in the field. With this in mind, this section (2.3.1) focuses on studies that involved ASR-based CAPT systems, while the following section (2.3.2) focuses on those that involved mobile devices.

In the late 1990s, early studies on ASR, such as Egan (1996), Franco et al (1997), Byrne et al. (1998), and Ehsani and Knodt (1998), sought to describe how speech recognition worked, and provided suggestions on how to create learning environments that took advantage of the strengths of speech technology, while also acknowledging and discussing their limitations. At the time, one of these limitations was that speech recognition was mostly domain specific, that is, it could only do what it had been programmed to do. For example, if an ASR-system was trained to recognize business news dictation under laboratory conditions, it would not be able to recognize spontaneous conversational speech. This raised the question of whether ASR-systems were able to recognize non-native speech when they were trained to recognize native speech. In this regard, Ehsani and Knodt (1998) noted that ASR technology performed well when it was trained to recognize the voice of specific speakers, but struggled to recognize non-native, disfluent and conversational speech. To solve this issue, they suggested either adapting native acoustic models so that they recognized non-native input, or creating models that were entirely based on non-native data.

In the early 2000s, the suitability of ASR for L2 learning was under debate. Derwing, Munro, Carbonaro (2000) believed that, in order to be useful for ESL speakers, ASR-systems should be able to recognize their speech and identify their speech errors with the same quality

and precision as they recognized native listeners. Seeking to provide empirical data on the issue, Derwing, Munro, Carbonaro (2000) set out to assess the recognition accuracy of an ASR software named 'Dragon NaturallySpeaking Preferred'. To do so, they first trained the software to recognize the speech of ten Cantonese speakers, ten Spanish speakers who were highly proficient in English, and ten native speakers of Canadian English, which was done by having these speakers read three chapters from a book to the software. Then, they asked the speakers to read 60 short sentences containing high-frequency words and simple syntax. These sentences were orthographically transcribed by the ASR software as well as by 41 native speakers of Canadian English. The results showed that the software was able to recognize 90% of the native speech, which was its advertised recognition level. However, it did not perform as well as the human listeners when non-native speech was involved. More specifically, its recognition scores were 24% lower for the Cantonese speakers, and 26% lower for the Spanish speakers. The authors believe that such recognition levels would probably disappoint those hoping to receive reliable feedback from the software, and argued that, until ASR-based systems have humanlike recognition levels, they should not be considered as beneficial to ESL speakers.

In the 2000s, researchers began to examine their actual pedagogical use of CALL and CAPT tools. One of such researchers was Hincks (2005), who set out to examine the effectiveness of an ASR-based CAPT product named *Talk to Me*. This product was a computer software that, in simple terms, asked users questions and offered them three choices of answers, which they were supposed to pronounce. If the users' responses were recognized by the ASR, the conversation would continue. This created a more naturalistic interaction with the software, and at the same time facilitated recognition as the ASR system had been trained to recognize all the possible sentences. The participants, who were middle-aged students at an intensive English course for immigrants in Stockholm, had their oral proficiency, reading fluency and pronunciation evaluated before and after the experiment. An ASR-based system named PhonePass was used to evaluate them in terms of phonetic accuracy, on the premise that ASR is an appealing alternative for testing pronunciation, as human rating can be time consuming, costly, and inconsistent. What is more, PhonePass was found to provide results as reliably as human raters.

Hincks (2005) had a pre-test, posttest control group design, with the participants from the experimental group receiving ASR-based instruction, while the control group received

regular phonetic training[22]. Although the students in the experimental group were supposed to use Talk to Me for one hour a day, five days a week for period of ten weeks, their actual use varied immensely, going from 2 up to 46 hours of use, with a mean of 12 hours of use. They believed Talk to Me was fun to use and that it had benefited their English, but reported they did not have the time to use it as much as they desired because of the number of assignments they had in their formal language course. The results from the pronunciation tests showed no significant improvements for either group. However, when the results were sorted by pronunciation scores, Hincks found that the participants from the experimental group with lower scores improved much more than those with intermediate scores. Based on these results, she believed that, with the level of technology available at the time, ASR-based instruction had the potential to assist less proficient learners, but could be detrimental to more proficient ones.

As stated, ASR can provide immediate, individualized feedback for language learners. With the goal of investigating the effectiveness of this type of feedback, in Neri, Cucchiarini and Strik (2006), learners of Dutch as an L2 were assigned to three groups. The experimental group used an ASR-based CAPT system that provided feedback on the learners' pronunciation of Dutch, whereas the first control group used a reduced version of the same system that did not provide feedback, meaning that they themselves had to evaluate the quality of their production. The second control group did not use the CAPT system, but attended their regular, teacher-fronted Dutch classes. All the participants had their segmental features of Dutch evaluated by six raters, who rated their pronunciation on a 10-point scale in terms of segmental quality.

The authors found that all three groups improved their global segmental quality after the training, with the group that received ASR-based corrective feedback showing the largest improvements, followed by the CAPT group with no feedback. Nonetheless, the difference in the improvements was not significant. The authors also analyzed whether the use of the CAPT system reduced the number of errors the participants made on elven Dutch phonemes that are known to be problematic to learners of the language. When only the target sounds were considered, the group receiving ASR-based feedback improved significantly in comparison to the other groups, while no significant differences were found for untargeted sounds. In light of these findings, the authors concluded that:

[22] In the field of education, the term 'training' is usually employed to refer to the development of abilities through practice with instruction or supervision (MacRAE, 2017). In the current study, the term was kept when it was used in the referential studies.

> Although it is undeniable that global ratings of pronunciation quality are an appropriate dependent variable, because at the end of the day CAPT should improve overall pronunciation quality, it is also clear that when evaluating systems that address specific pronunciation problems, a type of analysis with higher resolution may be required to assess the ultimate effect of the training. In our case this more detailed analysis has shown that the ASR-based feedback was effective in proving the errors addressed in the training, but the results of the overall pronunciation ratings have made clear that this is not enough to get a significant difference in improvement with respect to the control groups (NERI; CUCCHIARINI; STRIK, 2006, p. 4)

As the authors reported, the ASR-based instruction available in 2006 significantly improved the pronunciation ratings for the target sounds, but not for overall pronunciation ratings. However, considering that current ASR systems have better recognition rates for non-native speech (MCCROCKLIN; HUMAIDAN, EDALATISHAMS, 2019), ASR-based instruction may also promote overall intelligibility. With this in mind, the current study measured the intelligibility of both target and non-target words. By doing so, it was possible to evaluate if current ASR-based instruction can lead to an overall more intelligible speech, or if it is still more suitable for the target-words.

The studies reviewed so far had adult participants. Neri et al. (2008) noticed that the use of computers to help children learn pronunciation was rapidly increasing, and sought to investigate the effectiveness of *Tell me More, kids*: *The City,* an ASR-based CAPT system that offered feedback at word and sentence level. To do so, a group 11-year-old learners of English from Italy used the CAPT system for four weeks, while the children in the control group received traditional, teacher-led instruction. Three native speakers of British English rated the learners' speech before and after they received pronunciation instruction, with ratings for isolated words and for the speakers.

The results from Neri et al. (2008) show that the pronunciation scores for children who used the CAPT system were comparable to those achieved by the children who received traditional instruction, even though the children in the control group had 60-minute sessions compared to the 30-minute spent in each CAPT session. These similar results were attributed to the fact that the children using the CAPT system had 30 minutes of undivided attention during which they received individual feedback, while the children training with the teacher rarely had the chance to practice individually. However, the authors acknowledge that the pronunciation improvements for isolated words may not translate to sentences or to more spontaneous, connected speech, and suggest that future CAPT research addresses pronunciation aspects beyond the word level. Six years later, Elimat and AbuSeileek (2014) investigated a very similar

system, named *Tell Me More English*, a sign that there were not as many ASR-based systems available at the time. However, unlike Neri et al. (2008), the tests in Elimat and AbuSeileek included questions that required participants to pronounce whole sentences and read dialogues, which was one of the suggestions for future research in Neri et al. (2008).

Elimat and AbuSeileek (2014) argued that ASR-systems could be useful for L2 teachers and learners alike. They pointed out that students can discover their pronunciation weaknesses even if they are afraid to speak in public, and teachers can access the logs with the students' pronunciations to verify their improvements and their difficulties. With these potential benefits in mind, the authors conducted a study with third grade EFL students from Jordan with the goal of investigating whether instructing pronunciation with an ASR-based CAPT system yielded better results than regular pronunciation instruction. They also sought to evaluate the effects of ASR-based instruction when students worked in groups, in pairs or individually, and whether the effects of this instruction differed at the word, sentence and dialogue level. Thus, the participants were divided into four groups. The participants from the three experimental groups completed lessons from the CAPT system that contained pronunciation exercises at the word, sentence and dialogue levels, and received feedback that was based on native speaker models. The difference between the experimental groups was that the first group worked individually, with each participant having access to one computer, the second group worked in pairs and shared a computer, and the third group worked in groups of four learners for a single computer. The control group received the similar instructional material but was instructed without ASR, that is, a teacher was responsible for observing and correcting their mistakes. Data was collected with a pronunciation test devised by the authors, with both production and identification tests.

The results from to Elimat and AbuSeileek (2014) show that the experimental group significantly outperformed the control group in the production and identification tests. This finding was attributed to the fact that ASR enabled pronunciation practice to be more independent, which prevented foreign language classroom anxiety and helped those who were afraid of speaking in front of their colleagues. As for the type of work, the participants from the experimental group who worked individually had better results than those from the pair and group work, respectively. This result was partially attributed to the fact that the participants were only nine years old and were more easily distracted when practicing pronunciation with other participants. Even though this was not stated by the authors, it is possible to hypothesize

that the inferior results for the pair and group-work were a result of the sharing of computers, i.e., the pairs and groups used the same computer for the same amount of time than the individual work. In fact, this meant that these students had less time for the ASR to provide feedback on their pronunciation, reducing the chances of effective results. Finally, the authors found that the best results were displayed at the dialogue level of production, followed by the sentence and word level, although no statistically significant differences were found between them. They also argue that the participants' intonation helped develop their pronunciation at the dialogue level, which did not happen at the word level.

Mahdi and Khateeb (2019) conducted a meta-analysis of experimental and quasi-experimental studies to examine the overall effectiveness of using computers in pronunciation training. They reviewed 20 studies, with a total of 1014 participants, and used Cohen's *d* to represent the effectives of the training. They found that CAPT had positive effects on foreign language pronunciation, with a medium effect size ($d = 0.68$). The authors also tried to establish relationships between CAPT and other variables, such as learners' age, proficiency and the type of pronunciation features (segmentals or suprasegmentals) involved in the studies.

Out of the twenty studies examined in Mahdi and Khateeb (2019), eight dealt with segmental features of pronunciation, three with suprasegmental, and nine had both features. The effect sizes varied according to the type of instruction, with CAPT training on segmental features having an average effect size ($d = 0.47$), while suprasegmentals had a large one ($d = 0.89$). Although the authors did not discuss possible reasons for this difference, the smaller sample of suprasegmental studies may have been a factor. Regarding proficiency, effects sizes were larger for beginner and intermediate learners ($d = 0.68$ and 0.74, respectively) than for advanced learners ($d = 0.21$). The authors argue that beginner and intermediate learners acquire FL pronunciation faster than the advanced ones, and that the advanced learners might suffer from the fossilized pronunciation issues.

Mahdi and Khateeb (2019) learned that the largest difference in effect sizes was found in relation to the settings where the studies were carried out. The 13 studies that were conducted in the learners' classrooms had a large effect size ($d = 0.91$), whereas the seven studies where CAPT was used outside classrooms had a small one ($d = 0.13$). Although this is not discussed in the study, this finding indicates that L2 learners struggle to practice pronunciation with CAPT in more autonomous settings outside of their regular classrooms. However, the disparity in the effect sizes between these settings should not be interpreted as a sign that outside settings are

unsuitable for pronunciation or language learning. Rather, they show that this type of setting requires systems to be designed with clear instructions and appropriate feedback for autonomous learning. Furthermore, as Mahdi and Khateeb (2019) acknowledge, because different CAPT systems were used in different settings, it is unclear whether the diminished effects were caused by the outside settings or by the CAPT systems. In light of this, they believe that more investigation is necessary to understand CAPT in such settings.

Because learners can practice pronunciation on their own and without feeling judged by others, ASR-based instruction has been associated with reduced levels of speaking anxiety. This was mentioned as a positive aspect of ASR-based instruction in publications such as Aulia (2018), Elimat and Abuseileek (2014) and Kaiser (2018), but these studies did not focus on investigating the relationship between ASR-based instruction and speaking anxiety. This was done in Bashori et al. (2021), in which the participants of the experimental groups used two ASR-based websites (one for each group), while the participants from the control group attended regular classes about the same learning topic but did not receive any ASR-based instruction. All participants responded questionnaires to measure their Foreign Language Speaking Anxiety levels before and after the intervention, but only the experimental groups experienced reduced levels of speaking anxiety after the intervention. Participants commented that because they did not feel nervous while practicing through the websites, their confidence was boosted, which helped them to read texts in front of their EFL classmates.

The studies reported in this section had the goal of analyzing ASR-based systems that were developed for stationary platforms, such as desktop computers. More recently, however, these systems have been incorporated to mobile devices. Thus, the following section presents a few studies which investigated mobile ASR-based pronunciation instruction.

2.4.1.1 **Studies investigating ASR-based instruction using mobile devices (MAPT)**

In the past decade, ASR technologies have been integrated to mobile devices such as smartphones, which now have enough processing power to run complex software, and incorporate the functions of other devices, such as cameras and microphones. Kaiser (2018) understands that a new framework has emerged: Mobile Assisted Pronunciation Training (MAPT). According to Kaiser, MAPT is different from CAPT in that “it addresses a new market with a different development process both in terms of content and technology” (2018, p. 38).

This section reviews studies in which mobile technologies were used in conjunction with ASR for pronunciation instruction. As was done in section 2.3.1, the studies are presented in chronological order.

Liakin, Cardoso and Liakina (2015) investigated the production and the perception of the French vowel /y/ via a commercially available smartphone application named Nuance Dragon Dictation, which they tested and found to be able to recognize a hundred percent of the utterances pronounced by five native speakers of French. Forty-two adult learners of French as an L2 participated in the study and were randomly assigned to three groups. The ASR group practiced French pronunciation using the mobile system. They did a 20-minute pronunciation session once every week, in which they read target words and phrases aloud and received immediate feedback from the application consisting of a visual display of the recognized word. If the displayed word matched the target word, this meant that the pronunciation was correct, while if it did not, this meant the pronunciation was incorrect and another attempt was required. The Non-ASR group completed the same activities as the ASR group, but the feedback was orally provided by a French teacher by means of recasts and repetitions. The control group took part in weekly meetings with the same duration as the non-ASR group, but practiced conversation skills in French instead of having pronunciation training. For the production pre and posttests, the participants read words and phrases with the target /y/ sound. Their productions were assessed by two bilingual francophones, who listened to the participants' recordings and determined if their pronunciation of /y/ was correct or incorrect. When there were conflicting opinions, one of the responsible researchers decided if the utterance was correct or not. For the perception tests, the participants listened to 45 monosyllabic pseudowords in French, with 15 using the target vowel and the others being distractors. After listening to the pseudowords, they had to choose the alternative that matched what they had heard.

The results from Liakin, Cardoso and Liakina (2015) revealed that the only group to improve significantly in the production tests was the ASR-based group ($p < .001$). The Non-ASR group showed some improvement, but this was not significant. For the perception tests, none of the groups showed significant improvements, which suggests that the production gains were not transferred to perception. The authors attributed the production gains to factors such as immediate feedback, multiple opportunities for learning, explicit focus on the target form and game-like approach to learning. Moreover, they believe that the application used in the

study benefited from having comprehensible and visual feedback and allowing students to learn L2 features on their own outside their regular classrooms. The lack of gains in the perception tasks was attributed to the short instruction time (1.5 hours), and acknowledge that the expectation that a focus on production would translate into perception gains was optimistic. Regarding qualitative results, the authors reported that the participants had very positive views on their learning experience with the application. For example, participants enjoyed the possibility of learning with no one else around as they felt more comfortable and less nervous. Taking into the account the improvements found and the participants' positive evaluations, the authors argue that ASR should be further explored to complement pronunciation activities conducted in language classrooms, and suggest that future research investigates "the impact of ASR-based training on overall pronunciation skills (e.g., the development of intelligibility)" (LIAKIN; CARDOSO; LIAKINA, 2015).

In Park (2017), a distinction is made between ASR-based mobile applications and ASR-based mobile platforms. The author argues that they are different in that platforms, unlike applications, offer free and accessible features that can be used without internet connection and downloads[23]. He noticed that the focus has been on investigating commercially available applications, as was done in Liakin, Cardoso and Liakina (2015), and argued that it was important to explore the effectiveness of ASR-based platforms as well. With this in mind, the author carried out a quasi-experimental study in which 117 Korean ESL students at university level were assigned to three different groups. The first group received ASR-based pronunciation instruction, the second received conventional face-to-face pronunciation instruction and the third was instructed with a mixture of conventional and ASR-based instruction. Park conducted a needs' analysis to identify the participants' needs and difficulties with pronunciation, and based on it, selected six English consonants as the focus of instruction. The participants' pronunciation was assessed before and after the intervention with adapted versions of the production and perception tests used in Elimat and AbuSeileek (2014).

The results of Park (2017) revealed that, even though all groups showed significant improvements in the posttests for production, the best scores were achieved by the group that received hybrid instruction, which was the only one to have significant gains in both production

[23] In spite of the validity of this argument, it is relevant to point out that, nowadays, there are commercial applications that have these characteristics. For example, as shown in Baldissera (2020), the application 'English Pronunciation Tutor' has most of its content available offline, and 'EnglishPronunciation' offers all of its content for free.

and perception. The author believes that the feedback from the platform and the feedback from the language instructors acted in complementary ways, that is, while the platform helped identify the pronunciation difficulties of the learners, the instructors helped to improve the identified problems. In light of this findings, Park argues that ASR-based instruction can complement traditional instruction, but should not be seen as a substitute for it. He also points out that L2 teachers should be trained so that they can integrate ASR technologies in their syllabi, which can reduce their teaching burden and increase the autonomy of the students.

In similar fashion to Liakin, Cardoso and Liakina (2015), Fouz-González (2020) explored the potential of a smartphone application to assist EFL learners' pronunciation. The participants of the study were 52 Spanish EFL learners who were taking an English phonetics course at the University of Murcia. They used a commercially-available application named English File Pronunciation over a period of two weeks, wherever and whenever they wanted, but followed guidelines on how to use it so that the instruction they received was not so unequal. The focus of the instruction was on four English vowel sounds (/æ, ɑː, ʌ, and ə/) and the /s – z/ contrast. The design of the study was unorthodox, as it used both of its groups as control and experimental groups, so that the participants of the control group were not deprived of instruction. More specifically, the participants were randomly assigned to two groups, named G1 and G2. Both groups took perception pre-tests (an oddity discrimination task an and an identification task) and production pre-tests (an imitation reading task, a sentence-reading task and a timed picture-description task) at the beginning of the study, but only G1 used the application in the first two weeks. During this period, G2 was labelled as G2C because it acted as a control group for G1. After these two first weeks, both groups took perception and production post-tests. Then, G2 became the experimental group (G2E), while G1 received no instruction. After using the application for two weeks, G2E took a second round of perception post-tests, but did not do the production post-test again to avoid imposing excess demands on the group as compared to G1. All of the participants' productions were assessed by three non-native, experienced EFL teachers, who rated the productions as either 'pronounced adequately' or 'pronounced inadequately'.

In terms of perception, Fouz-González (2020) found significant differences between discrimination pre and post-tests, but not between the experimental and control groups. For the identification test, there were significant differences between pre and post-tests and between G1 and G2C. As for the production tests, the differences between G1 and G2C were not

significant for every task and sound: in the imitation task, only /æ/ reached significance; in the sentence reading task, only /ɑː/, /z/ and /ə/; and for picture description, only /ʌ/ and /ə/. Overall, the instruction provided by the application had positive effects on the participants perception and production of the target sounds. However, considering the lack of statistical significance for some of the sounds, it is unclear whether the improvements were caused by the use of the application or by the explicit phonetic instruction the participants received in a phonetics course which took place at the same time as the study.

In Liakin, Cardoso and Liakina (2015) and Fouz-González (2020), ASR-based instruction yielded positive effects in terms of L2 pronunciation. Yuan and Liu (2020) were cognizant of these results, but wondered how learners with different proficiency levels would benefit from this type of instruction. To investigate this, 50 Chinese high school students were assessed in terms pronunciation, reading, vocabulary, grammar and writing. The 25 students with the lower scores were classified as the low-proficient group, and the 25 with higher scores became the participants of the high-proficient group. All participants used an application named *Oral English Drill & Test* for 28 days, in which they read texts aloud and received feedback on their pronunciation. The results showed that the overall pronunciation accuracy of the low-proficient group increased significantly, while the high-proficient group did not. In spite of this, they found that instruction from the application helped to improve specific pronunciation errors from both groups, such as the deletion of sounds.

In Brazil, the possibility of ASR-based mobile instruction has begun to be investigated, with studies such as Baldissera (2020) and Gottardi, Almeida and Tumolo (2022) exploring the issue in a more qualitative, bibliographical nature. Baldissera (2020), for example, analyzed four MAPT applications designed specifically for pronunciation learning, namely ELSA (English Learning Speaking Assistant) Speak, EnglishPronunciation, English Pronunciation Tutor, and Juna. These applications were rated in terms of their content, pronunciation teaching methods and their features and usability. The author found that all of the applications instructed segmental and suprasegmental features, with the former receiving more focus than the latter. ELSA Speak and EnglishPronunciation were the top-rated applications, as they included all the content[24] which Celce-Murcia, Brinton and Goodwin (2010) postulated to be necessary for resources for pedagogical instruction. In spite of this, the author believes that the applications

[24] According to Celce-Murcia, Brinton and Goodwin (2010), these contents were the development of sound awareness, phonological perception, guided production, controlled production, and communicative practice.

failed to support more communicative pronunciation practice, as when there is meaningful exchange of information. The type of analysis is relevant because with the proliferation of MALL applications, it becomes difficult to understand the differences between them, and what pedagogical uses they may have.

To date, only one[25] study has attempted to investigate the use of ASR by Brazilian learners of English, a case study by Moura and Lopes (2021). In this study, two elementary school students from a public school in Pernambuco used ELSA Speak in autonomous fashion for 10 days. The goal of this case study was to investigate the participants' learning process with the application, which took place in the context of pandemic-related remote learning. They found that one participant, who self-rated his pronunciation level as low before using the application, self-rated his pronunciation level as average after using the application. The other participant did not report improvements and maintained his self-rated level as average. In light of this, the authors inferred that learners in beginner levels may have more immediate results, but believe that if they had used the application for more time, different results could have been achieved. They also believe that without supervision and guidance, MAPT is unlikely to achieve its goals satisfactorily.

Overall, the studies reviewed in sections 2.3.1 and 2.3.2 tackled the potential that ASR-based systems have to assist in the instruction of pronunciation for L2 learners. Although early studies such as Ehsani and Knodt (1998) and Derwing, Munro, Carbonaro (2020) reported serious issues in their ability to recognize non-native speech, promising results have been reported in more recent studies (AULIA, 2018; ELIMAT; ABUSEILEEK, 2014; FOUZ-GONZÁLEZ, 2020; LIAKIN; CARDOSO; LIAKINA, 2015; NERI et al., 2008). To assess the effects of ASR-based instruction, many of the studies reviewed had native speakers judge the segmental quality of the participants' productions, which was often assessed to be either 'correct' or incorrect'. Thomson and Derwing (2014) alert that this type of measurement is problematic when the definitions of 'correct' or 'incorrect' are unclear. For example, in Fouz-González (2020), the participants' productions were assessed as either 'adequate' or 'inadequate', but it was not clear whether 'adequate' meant 'pronounced as a native speaker would' or 'pronounced well enough to be recognized by a proficient speaker'. In this regard, orthographic transcriptions of speech samples are a good option to measure pronunciation because they do not require the definition of what constitutes adequate or inadequate

[25] As revealed by searches in the electronic database Portal de Periódicos Capes, in May 2022.

pronunciation, since if the transcribed word matches the intended message, then that word is an intelligible one.

Given the limited of studies that assessed ASR-based pronunciation instruction using intelligibility measures, evidence about the applicability of ASR-based systems to promote intelligibility is limited. This study sought to address this gap by investigating the effects of an ASR-based, mobile application on the intelligibility of L2 learners by using listeners' transcriptions as a measure for pronunciation improvement. In light of this, the following section discusses the concept of intelligibility and reviews literature on the field.

2.5 INTELLIGIBILITY

In its everyday sense, people are intelligible if what they try to communicate in written or spoken form is understood. Simply put, we expect that that the meaning of our messages is correctly received, that is, we strive to be intelligible. Being unintelligible, on the other hand, is not desired as it can lead to innumerous problems. Recognizing this, fields such as communication technology, audio engineering, speech language pathology, and audiology have conducted research involving intelligibility (MUNRO; DERWING, 2015).

In the field of L2 pronunciation, the importance of intelligibility has been widely recognized (JENKINS, 2000; MUNRO; DERWING, 2015; PENNINGTON, 1996). Despite this recognition and the simplicity of its everyday concept, defining intelligibility in this field is a complex matter as there is no agreed definition of it and no consensus on the best way to measure it (ISAACS, 2008; MUNRO; DERWING, 2015). Essentially, pronunciation research has been guided by two conflicting principles: the nativeness principle and the intelligibility principle (LEVIS, 2005, p. 370).

The nativeness principle assumes that learners can and should achieve native-like pronunciation in an L2. Under this principle, intelligibility is a unilateral process in which non-native speakers strive to make themselves understood by native speakers (NSs) who have the prerogative to decide what is and what is not intelligible (BAMGBOSE, 1998). Although there is abundant evidence against this principle, it still permeates and affects language teaching and learning around the globe (HOLLIDAY, 2006; LEVIS, 2005).

The intelligibility principle holds that the goal of L2 learners is to be intelligible and recognizes that "communication can be remarkably successful when foreign accents are

noticeable or even strong" (LEVIS, 2005, p. 370) and that communication depends on and is affected by both speakers and listeners. In light of this, definitions of intelligibility have encompassed both speakers and listeners. In the field of L2 pronunciation, the definition presented in Munro and Derwing (1995) is commonly adopted (KAUR, 2009). They have defined intelligibility as "the extent to which a speaker's message is actually understood by a listener" (MUNRO; DERWING, 1995, p. 76). This is the definition adopted in the current study.

Intelligibility has been assessed in a number of ways. In L2 studies, having listeners hear and orthographically transcribe speakers' utterances has been one the most common ones, with the number of correctly transcribed words representing an index of the intelligibility of a speaker (MUNRO; DERWING; MORTON, 2006). Other approaches have involved the use of comprehension questions, picture selection, identification tasks (BURLESON, 2007), summary elicitations (PERLMUTTER, 1989) and true or false statements (MUNRO; DERWING, 1995; DERWING et al., 2014). In Munro and Derwing (2014), for example, speakers read a list with obvious true statements, such as *the sky is blue,* or false statements as *dogs have wings.* The role of the listeners was to hear the sentences and determine if they were true or false. In theory, if a message was intelligible, the listeners should be able to extract its truth value, even if there are words that are hard to comprehend or are misunderstood, thus providing a more global measurement of intelligibility. Conversely, orthographic transcriptions enable a finer understanding of how words or utterances are recognized, as they allow researchers to verify how each word was transcribed (MOYER, 2013). All things considered, there are advantages and disadvantages to each approach, although none is likely to provide, on its own, a complete understanding of the intelligibility of a particular speaker (DERWING et al., 2014).

Research has shown that there is no direct correlation between accentedness[26] and intelligibility. In other words, an L2 speaker may have a heavily accented speech, but still be intelligible (MUNRO; DERWING, 1995). In this regard, the goal of achieving native-like accent becomes an unrealistic and unnecessary one, since native-like pronunciation is very rarely achieved and is not necessary for effective communication (MUNRO; DERWING, 1995; MUNRO, DERWING, 2015). Moreover, such goal has little relevance in our current world as a large proportion of interactions in English involve non-native speakers communicating with

[26] Derwing and Munro (2005) define accentedness as "a listener's perception of how different a speaker's accent is from that of the L1 community (p. 385).

other nonnative speakers of the language (JENKINS, 2000). In light of these facts, this study adheres to the intelligibility principle because it understands that the participants, who are Brazilian learners of English, can become intelligible users of the language and successfully communicate with speakers from around the globe without seeking to achieve 'native-like' pronunciation.

According to Munro and Derwing (2015), research concerned with speech intelligibility in L2 can be broadly placed into three categories: a) intervention studies that investigated the effectiveness of certain methods or techniques on improving pronunciation; b) studies that attempted to identify factors that affect intelligibility; and c) acquisitional studies that sought to trace learners' development of pronunciation intelligibility. Considering the scope of the current study, the following two sections reviews intervention studies and studies that sought to identify factors that affect intelligibility.

2.5.1 Intervention studies

Under the intelligibility principle, learners should seek to achieve a pronunciation level that allows them to communicate with ease, even when nonnative characteristics of their accent are maintained (MUNRO, DERWING, 2015). Because achieving such level is a not a simple task, researchers have investigated how effective different approaches are in promoting changes to learners' pronunciation. This section includes studies in which the effectiveness of pronunciation instruction was assessed with intelligibility measures.

One of first studies to do so was Perlmutter (1989), which was motivated by an US state law that required International Teaching Assistants to have a minimum level of oral proficiency in English. The speakers were 24 graduate students from eight different countries, who had recently arrived in the US and received six months of language instruction with a focus on pronunciation. The listeners were 21 undergraduate students who were native speakers of English. Although the results from the study show that intelligibility improved significantly after training, it is difficult to establish if the improvement was caused by the instruction because of the participants' recent arrival to an English-speaking country, where their pronunciation could be improved regardless of the training.

Another study that assessed the effects of instruction with intelligibility measures was Derwing, Munro, and Wiebe (1997), in which suprasegmental pronunciation features were

instructed to ESL learners over a 12-week classroom program. They found significant improvements to the intelligibility of the participants, which was measured with ortograpic transcriptions from native speakers. In contrast to Perlmutter (1989), the participants in Derwing, Munro, and Wiebe (1997) had been living in an English-speaking country for around ten years. According to the authors, this compensates for the lack of a control group because, considering the learners' extensive length of residence, spontaneous improvements to their speech should not be expected over a short period of training.

As commercially available CALL systems became more available in the late 2000s, studies such as Derwing, Munro, Carbonaro (2000) and Hincks (2005) set out to investigate their applicability to instruct pronunciation. One study to do so was Burleson (2007), which investigated the effects of a computer-based pronunciation training program on the intelligibility of ten English learners whose L1 was Mandarin Chinese. Five of these learners received 15 hours of computer-based training on consonantal errors that are commonly found in the speech of Chinese speakers of English. Another five learners received the same training but focused on vowel contrasts. One participant did not receive training at all and acted as a control. In both vowel and consonantal groups, the training consisted of hearing samples of how native speakers pronounced the target minimal pairs, then pronouncing lists of isolated minimal pairs for the computer, which provided feedback in two ways. If the computer recognized the utterance as target-like, as when sick was pronounced as /sɪk/, it returned a chime sound and the text message "yes, the computer heard sick". If the word was pronounced more closely to the expected error, as when sick was pronounced as /sik/, it made a buzz sound and showed the message "no, the computer heard seek". In similar fashion to Perlmutter (1989) and Derwing, Munro, and Wiebe (1997), the listeners of Burleson (2007) were native speakers of English. However, intelligibility of individual words was assessed with forced-choice identification tasks in which listeners were shown two words thar formed a minimal pair, heard a pronunciation token from a participant, then chose which of the two words they believed to have heard.

The results from Burleson (2007) revealed similar improvements for the vowel and consonant groups, and showed significant gains to words that were part of the training, which were 46% intelligible in the pretest and 87% in posttest. Significant effects were also found for words that were phonologically related but were not part of the training. Among the vowels, the /ɪ - i/ contrast was the one with the least improvement, with two participants having

considerably lower scores in the posttest. The author believes that these results served to encourage teachers and students of English to incorporate CALL systems in their pronunciation courses, as they are “limited in opportunities for lengthy one-on-one and face to face tutoring” (BURLESON, 2007, p. 16).

Being an intelligible English speaker can be especially important for immigrants living and working in English speaking countries, where the workforce is commonly made up of native and nonnative speakers of the language. One of the challenges of investigating the effectiveness of pronunciation for this population is that they often lack the time and resources to attend traditional classroom-based ESL programs. One alternative is to offer courses that take place at their workplaces, as was done in Derwing et al. (2014), in which a 17-hour pronunciation training program was conducted for seven non-native English speakers from Vietnam, who had been living in Canada for an average of 19 years. Native speakers of Canadian English assessed the participants in terms of comprehensibility, accentedness, fluency and speech intelligibility, before and after the training. In this study, the listeners heard a list of 14 true or false statements recorded by the speakers, and were asked to tell whether they were true or false. For example, the sentence *many people like to drink coffee* should be marked as true, while *spaceships use hot dogs for fuel* should be false. Derwing et al. (2014) found that the participants’ comprehensibility improved significantly after the training, but their fluency did not. With regards to intelligibility, their scores went from 75% in the pretest to 80% in the posttest. The authors argue that the fact that the participants enhanced their pronunciation with the training, even after living for almost 20 years in Canada, indicates that “fossilized productions may not be permanent after all” (DERWING et al., 2014, p. 542).

In a review of literature on the effectiveness of pronunciation instruction, Thomson and Derwing (2015) found that 75% of the human-assessed studies measured discrete pronunciation features such as the accuracy of individual phonemes. More global measures, as intelligibility transcriptions and comprehensibility ratings, were adopted in only 18% of the studies, while 7% used both discrete and global measures. Because of that, Chau et al. (2022) believe that there is less evidence for the effects of instruction on the development of global features in spontaneous speech. Seeking to address this gap, Chau et al. (2022) investigated the effects of segmental and suprasegmental instruction on the fluency and intelligibility of 45 Vietnamese EFL learners. In the pre and posttests, the speakers answered wh- questions so that their speech was recorded in a more spontaneous manner, in contrast to the more commonly

used and controlled methods for sampling speech, such as reading aloud and sentence imitation. The authors point out that exploring spontaneous speech is valuable because "speaking spontaneously in daily communication is likely the goal for many learners" (p. 2). The listeners of the study were 11 native speakers from the USA, who were undergraduate students enrolled at linguistic courses.

Chau et al. (2022) found no significant effects of the instruction on either intelligibility or fluency, and found no clear relationship between these measures. These results diverge from the general results of the effectiveness of pronunciation instruction for discrete features, as reported in review studies such as Thomson and Derwing (2014). Chau et al. (2022) partially attribute the lack of improvements to a discrepancy between the controlled practices that took place during the instruction and the spontaneous speech that was required in the pre and posttests. Moreover, they argue that the lack of significant gains may be related to what they refer to as a ceiling effect, as the participants' speech was 83.4% intelligible in the pretest, which was deemed to be quite high, especially considering their low proficiency levels. They mention that the more proficient participants from Munro and Derwing (1995) had similar levels of intelligibility, and argue that the relationship between proficiency and intelligibility may be weak. This was also noted in Thomson and Derwing (2015), who pointed out that pronunciation is different from other L2 aspects because "it is not tied to proficiency - a beginner can have excellent productions and an individual with a superb grasp of L2 syntax and vocabulary can be difficult to understand" (2014, p. 339).

The studies reviewed in this section revealed that pronunciation instruction can develop L2 intelligibility (BURLESON, 2007; DERWING et al., 2014; DERWING; MUNRO; WIEBE, 1997; PERLMUTTER, 1989). Although these results may be interpreted as a sign of the effectiveness of instruction for intelligibility, the number of studies is limited, and additional investigation is still necessary. Furthermore, to date, few studies have explored the effects of computer-based L2 pronunciation instruction on speech intelligibility (BURLESON, 2007; VAN LIESHOUT; CARDOSO, 2022). Considering the growing expansion of this type of instruction, one of the goals of the current piece of research was to explore the extent to which an ASR-based smartphone application could improve the intelligibility of L2 learners, adopting intelligibility as a measure of pronunciation. As is discussed in the following section, a key factor in this type of research is that the pronunciation features instructed are relevant in terms of intelligibility, that is, these features should be likely affect ones' intelligibility in an L2.

4.5.1 Pronunciation features that affect intelligibility

Under the intelligibility principle, learners can maintain phonetic features of their L1 and still be intelligible. However, different features affect intelligibility in distinct ways, and identifying those that contribute the most to intelligibility has been one of the most pressing issues in L2 pronunciation research (FIELD, 2005). Derwing and Munro (2009) explain the importance of this type of research:

> We have to know where to put the focus. If not, there is a risk of teaching things that are salient, but which will not result in actual improvement in communication for the speaker. In other words, we might modify accent without improving intelligibility or comprehensibility (p. 482).

In addition to this, time for pronunciation instruction is limited in L2 classes, so if it is spent on features that do not affect intelligibility, then those that could really help learners may be left behind (GILAKJANI; SABOURI, 2016; DERWING; MUNRO, 2009). Therefore, identifying what affects intelligibility is a necessary yet challenging task, especially because the literature on the field is not always convergent.

Jenkins (2000) was one of the first to pursue the identification of such features in the context of communication between non-native speakers, with the proposal of the Lingua Franca Core (LFC), which considers that the core features for international intelligibility are: a) most consonant sounds; b) consonant clusters and appropriate consonant cluster simplification; c) vowel length distinctions (short vs long vowels); and d) nuclear[27] or tonic stress (JENKINS, 2000). The author reminds us that, in spite of the value of these sounds, further investigation is still required to validate their importance in ELF contexts.

The LFC emphasized the role of segmental features i.e., individual sounds, over suprasegmentals ones i.e., features that extend over more than an individual sound, such as stress and intonation. Out of its core features, only one involved a suprasegmental one, namely nuclear stress. Jenkins (2000) argues that the value of suprasegmentals for intelligibility was highlighted in studies which had native speakers as listeners, and points out that these speakers may process speech differently from non-native speakers. The author claims that the

[27] Stressing a particular word in an utterance to signal a particular meaning. This type of stress is also referred to as tonic, primary, or contrastive stress (ZIELINSKI, 2015).

intelligibility of ELF speakers is mostly affected by “deviant core sounds in combination with misplaced and/or misproduced nuclear stress” (p. 155).

Aiming to establish hierarchies for the instruction of pronunciation, attempts have been made to predict which phonetic features will cause the most trouble for L2 learners from particular L1 groups, with the rationale that the most difficult features should be given priority. Munro (2021) investigated if a hierarchy could be established for the difficulty of the English tense and lax close vowels (/i ɪ u ʊ/) for Cantonese speakers living in Canada, and if such hierarchy could be pedagogically useful. Even though a general order of difficulty did emerge (/i/ > /u/ > /ʊ/ > /I/), with the two lax vowels being the most difficult, the author found a high degree of idiosyncrasy between the participants. For example, some participants had high levels of intelligibility for /I/ in sick and kick, but had low levels for sit and hit, with some participants exhibiting an opposite pattern. More importantly, no correlations were found between the participants’ intelligibility and socio-linguistic variables as use of English and length of residence in Canada. In light of these findings, Munro (2021) argues that establishing a hierarchy of vowel sounds would not be pedagogically useful, and defends the implementation of programs that are customized to individual needs. However, he acknowledges that doing so requires suitable assessment tools to identify the most challenging sounds, and believes that recent ASR-based applications for pronunciation instruction offer “a great deal of promise for individualized instruction that includes direct feedback on production” (2021, p. 13).

Under the intelligibility principle, successful communication is dependent on both speakers and listeners. As they can have different L1s, socioeconomical backgrounds and proficiency levels, different phonological and non-phonological features can influence what affects intelligibility and what does not. In this regard, Becker (2013) believes that is very difficult to generalize findings to claim that one phonetic feature hinders intelligibility to speakers and listeners with different backgrounds. Given that this study has Brazilian learners of English as speakers, it is necessary to look at what factors have found to influence the intelligibility in Brazilian contexts. With this in mind, the following section reviews studies that investigated pronunciation intelligibility that were carried out with Brazilians, either as speakers or listeners of English.

2.5.1 Intelligibility studies in Brazil

To Gonçalves and Silveira (2015, p. 52), “speech intelligibility has been little investigated in Brazil”. In spite of this, studies such as Cruz (2003), Delatorre, Silveira and Gonçalves (2017), Gonçalves e Silveira (2014; 2015) and Silveira and Silva (2018) have allowed us to have a better understanding of the pronunciation features that affect the intelligibility of Brazilian learners of English, and studies such as Becker (2013) have discussed the factors that influence the intelligibility of English speakers to Brazilians listeners. This section presents a review of these studies, which is done in chronological order.

Cruz (2003) sought to investigate the extent to which pronunciation errors of Brazilian Portuguese Speakers of English affected their intelligibility by native speakers of English. To do so, the speech of six Brazilians who were advanced learners of English was recorded while they tried to convince an Englishman to live in the city where they lived. The speech samples were then selected so that they did not contain grammar or vocabulary errors, ensuring that the raters would only deal with pronunciation ones, namely consonant, vowel, epenthesis, and word stress errors. The samples were sent to three native speakers of English who were familiar with Brazilian Portuguese, as they had lived in the country from one to four years. In the first listening section, the raters indicated, on a scale from 1-6, the degree of intelligibility[28] of the thirty speech samples. In the second session, the raters listened to the samples again and reported the words they found impossible or difficult to understand, the words that helped them understand other words, and the ones that they believed to be accented.

The analysis of the results shows that stress errors were the ones that most affected the ratings, with six out of seven words containing this type of error being considered impossible or difficult to understand. These included inappropriate stress placement in words such as ‘interview’, ‘efforts’ and ‘sometimes’. Regarding the words that helped raters understand other words, Cruz reports 5 instances in which the intelligibility of a mispronounced word depended on other words. For example, the word ‘my’ in ‘when I visit my dad (pronounced as [dɛd]) helped ‘dad’ to be intelligible, probably because of the unlikeliness of one saying, ‘when I visit my dead’. As for high-front vowels, pronouncing /i/ instead of /ɪ/ caused the sentence ‘at least where I’m living’ to be understood as ‘at least where leaving’, and ‘sit with my friends and talk’ to be understood as ‘sleep with my friends and talk’. In light of these findings, the author

[28] Although Cruz employs the term ‘intelligibility’, using a rating scale to judge how easy or difficult it is to understand an utterance is considered a measure of comprehensibility according to Derwing and Munro (2005). For more on the use and on the distinction between the terms intelligibility and comprehensibility, please see Gonçalves and Silveira (2015).

suggested investigating how much segmental errors affect the intelligibility of Brazilian learners of English, and how far the occurrence of words with stress errors affects these learners' intelligibility to native speakers who are not familiar with the Brazilian accent (CRUZ, 2003, p. 170).

Based on previous studies that investigated the intelligibility of Brazilian learners of English, Cruz (2012) formed a corpus with 12 categories of words and sounds that were found to hinder the intelligibility of these learners, which included:

a) misplaced word stress: stress falling on the second or third syllable instead of the first

b) orthographic influences: the grapheme <u> produced as [u] instead of /ʌ/, such as when 'public' is produced as [public]

c) inappropriate consonant sounds

d) inappropriate vowel sounds, with the neutralization of the differences between lax and tense vowels being the source of unintelligibility, such as when the high-front vowels /i/ and /ɪ/ are pronounced as [i].

e) vowel insertions, especially the epenthesis of [ə], [i] and [^{i}]

Cruz (2012) points out that English teachers tend to define their priorities for pronunciation instruction based on intuition, not on research results. In light of this, she believes that this research-based corpus can help English teachers define pedagogical priorities more appropriately, even though the corpus is limited in size.

Becker (2013) investigated how intelligible the speech of English speakers from other countries were to Brazilian listeners. As the study sought to explore intelligibility under an ELF paradigm, the speech samples were produced by speakers from major trading partners from Brazil, namely the United States, Germany, China and Japan, and consisted of paragraphs taken from the Speech Accent Archive. The listeners, who were undergraduate language students from Brazil, performed three tasks. In the first, they listened to the whole text and indicated, on a scale from 0-100, how much they understood from them. In the second, they attempted to transcribe the words from the samples, and in the third they indicated the items they believed to be detrimental to intelligibility.

Becker (2013) found that the Japanese speakers were the least intelligible of the four, with 61% of their words judged as intelligible by the listeners. The author also measured the

percentage of intelligible content and function words. For the German and US speakers, there were no significant differences in intelligibility between content and function words, while the content words from Japanese and Chinese speakers were more intelligible than their function words. Another important finding from the study is that word frequency plays a role in intelligibility, as less frequent words were also less intelligible.

Another study which tackled the intelligibility of Brazilian learners of English was Gonçalves and Silveira (2014). However, the listeners from their study were not native speakers of the language. Rather, the 32 listeners had 11 different language backgrounds, with the majority of them speaking several L2s. One of the goals of the study was to identify which of the high front vowels, namely the tense /i/ and the lax /ɪ/, produced by the Brazilian speakers caused more intelligibility problems. To do so, the listeners orthographically transcribed the missing words from the sentences recorded by the speakers. The missing words were all monosyllabic words containing the target vowels, such as 'beat' and 'bit'.

The results from Gonçalves and Silveira (2014) revealed that the tense vowel /i/ posed more difficulty to the listeners as its durational cues were not salient enough to be differentiated from the lax vowel. Phonetic transfers from Brazilian Portuguese were also found to hinder intelligibility, as when initial stops were produced without aspiration and final alveolar stops were palatalized (e.g., 'pit' being transcribed as 'beach'). In light of these findings, the authors argue that duration is an important acoustic cue for those who speak English as an L2, but that other segmental errors also influence speech intelligibility. To help learners become more intelligible, they argue that pronunciation instruction should involve a variety of language instruction techniques, including controlled and communicative practices.

Delatorre, Silveira and Gonçalves (2017) investigated the intelligibility of English verbs in the simple past. In contrast to the studies reviewed so far, the speakers were both native and non-native speakers of English. This allowed the researchers to investigate if the listeners, who were Brazilian, would find the pronunciation of the verbs more intelligible if they were produced by the two native speakers from the United States and Australia, or by the six non-native speakers, who were from Brazil, Spain, Argentina and Germany.

The researchers found that the intelligibility of the verbs in the past simple was higher for speakers whose L1 was Portuguese or typologically similar to Portuguese, as in the case of Spanish, although this result was not statistically significant. Understanding the native speakers of English proved to be more difficult in comparison to the other non-native speakers because

the listeners had had little experience talking to native speakers of the language. It was also found that the pronunciation of regular verbs whose past tense morpheme is pronounced as [d] hindered intelligibility more than those that ended in [t]. Based on these results, the authors believe that more contact with spoken English, either with native or non-native speakers, in conjunction with pronunciation instruction, should help improve how intelligible the speakers of the language are to Brazilians.

More recently, Silveira and Silva (2018) sought to investigate how different types of coda modification affected the intelligibility of Brazilian speakers of English. In additions, it explored how the degree of semantic information, the listeners' L1, proficiency level, familiarity with English spoken by Brazilians and length of residency in Brazil affected intelligibility. Based on previous studies, their intelligibility test included 30 utterances recorded by Brazilian speakers, with six types of word-final coda modification that are believed to hinder the intelligibility of the English spoken by Brazilians, namely: a) the vocalization of nasals; b) vocalization of /l/; c) vowel insertion; d) palatalization of alveolar stops; e) voicing of alveolar fricatives; and f) devoicing of alveolar fricatives. Eighteen of these utterances were presented with limited semantic context, as in 'a little doll', and 12 were presented with substantial semantic context, as in 'a little doll with colorful dress'. They were presented to 38 listeners, who had been living in Brazil for different amounts of time, had various L1s, different levels of familiarity with the English spoken by Brazilians, and diverse levels of English proficiency. In the listening tasks, they first heard the utterances and orthographically transcribed the missing words. Then, they rated how easy or difficult it was to comprehend the sentences and wrote about what made it easy or difficult to do so.

Concerning the influence of the degree of semantic context on intelligibility, Silveira and Silva (2018) found that utterances with substantial semantic context were more intelligible than those with limited semantic context for five of the six types of coda modification. Vowel insertions, such as when 'sick' was produced as ['sɪkɪ]', and the affrication of alveolar stops were the mispronunciations that most impeded intelligibility, likely because these modifications made the target words sound as other English words. For example, when 'bed' was produced as ['bedʒ], many listeners believed the word 'badge' had been pronounced, which happened even in the presence of substantial semantic context. As for the other variables involved in the study, the authors found that listeners whose L1 had different origins (e.g., English and Swedish) had higher intelligibility scores than listeners whose L1 is a Romance language, such as Spanish.

This was attributed to the fact that speakers of Romance languages are less likely to be familiar with the English spoken by Brazilians because they are less likely to use English to communicate with them. Likewise, listeners who were more familiar with the features of the English spoken by Brazilians had better intelligibility scores than those with less familiarity. In contrast, the length of residency in Brazil and the proficiency level of the listeners did not predict their performance on the intelligibility test.

In light of these findings, Silveira and Silva (2018) believe that while some types of coda modification seem to cause less intelligibility issues, others should be addressed in classroom contexts because they are more likely to hinder intelligibility, as was the case of the affrication /t/, which can cause 'cat' to be understood as 'catch', for example. They point out that while the degree of semantic information may help intelligibility, it does not guarantee it, and in light of this suggest classroom practices that involve listening and producing words with different context levels so that learners can see how detrimental certain modifications can be for communication.

The studies reviewed in this section investigated how intelligible Brazilian speakers of English are to native and non-native speakers of the language, and how intelligible speakers whose L1 is not Portuguese are to Brazilians. Moreover, these studies attempted to identify the phonological aspects most affect intelligibility between speakers of English. In spite of these efforts, Becker (2013) and Gonçalves and Silveira (2015) acknowledge that establishing a common core for pronunciation teaching is a complex and challenging task. Even though the studies reviewed had different foci, there is a general consensus that more research investigating the intelligibility of English in Brazil is still required. Moreover, pronunciation instruction involving different language instruction techniques and technologies is necessary if the goal is to improve the intelligibility of learners (GONÇALVES; SILVEIRA, 2014; MUNRO, 2021).

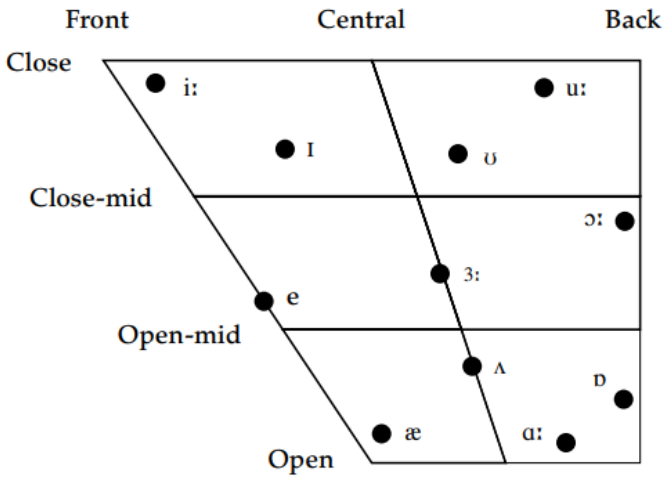
One current and promising technology that can be used for pronunciation practice is ASR, which, as stated, can provide automatic, immediate, and limitless feedback to learners. Due to the novelty of the technology, few studies have been carried out investigating the use of this technology with the goal of improving intelligibility. This research seeks to address this gap by investigating the effects of the use of an ASR-based smartphone application. The focus of instruction of this application will be on the high-front vowels of the English language, namely /i/ and /ɪ/. Being able to identify and produce this pair of vowels is relevant because

they serve to distinguish many[29] high frequency minimal pairs i.e., they have a high functional load. Furthermore, many words formed with the pair that can be easily confused because of identical syntactic functions, e.g., *leave* and *live* can both function as verbs, and *sheep* and *ship* can both function as nouns (MUNRO, 2021). Overall, these vowels tend to be challenging for L2 learners to perceive and produce (KIVISTÖ-DE SOUZA, 2011; SILVA, 2012), and have been found to hinder the intelligibility of Brazilian speakers of English (BECKER, 2003; CRUZ, 2003; GONCALVES; SILVEIRA, 2014). Considering the focus of this study, the following section presents relevant information and literature on the English high-front vowels.

2.6 THE ENGLISH HIGH FRONT VOWELS AND THEIR INSTRUCTION TO BRALIZIAN SPEAKERS OF ENGLISH

English vowels are commonly described according to three basic variables: open/close; front/back; and rounded/unrounded. When the tongue is high in the mouth, this vowel is described as close, but if it is low in the mouth, it is called an open vowel. If the tongue is towards the front of the mouth, it is a front vowel, while if it is at the back of the mouth, it is a back vowel (DETERDING, 2015). As Figure 1 shows, both /i/ and /ɪ/ sit at the front, high (close) section of the monophthong vowel chart.

Figure 1 - The monophthong vowels of British English.



Source: DETERDING, 2015.

[29] A total of 446 according to Higgins (2013).

Because /i/ and /ɪ/ are pronounced with the tongue high in the mouth, many linguists prefer to call the high front vowels, while the International Phonetic Association prefers the terms "close" and "open" for vowel height. In this study, we adopt the term high front vowels to refer to /i/ and /ɪ/ as the term has been used in many referential studies, such as Gonçalves and Silveira (2014) and Kivistö-de Souza et al. (2017).

Although /i/ and /ɪ/ are both produced with the tongue in a high, frontal position of the mouth, they differ in that /i/ is considered to be a tense, long vowel, while /ɪ/ is lax and short. This means that /i/ is a longer sound that is produced with more tenseness[30] than /ɪ/. To show that /i/ is a long sound, a length diacritic is commonly used, so /i/ is often phonologically transcribed as /iː/. However, this is not a consensus, as the diacritic is frequently omitted on the basis that this vowel is tense rather than long, and that the tense vowels may actually be shorter in duration than the lax vowels, depending on the phonological environment and speaking rate (DETERDING, 2015). Acknowledging the lack of consensus in this regard, Silva (2012) points out that, in an analysis of the American English, the distinction between tense and lax vowel is more common, which suggests that such distinction is more relevant than that of length.

There are a number of reasons why the English high front vowels tend to pose challenges for EFL learners, both in perception and production. For Brazilian learners, this difficulty may be because the tense/lax distinction is not distinctive feature in Portuguese, nor is the duration of the vowel. For example, if the Portuguese word ‘li*’* /li/ were pronounced as /lɪ/ or /liː/, its meaning would remain the same, albeit with an unusual accent. In English, however, pronouncing the word ‘beach’ as /bɪtʃ/ instead of /bitʃ/ alters the meaning of the word and can lead to embarrassing misunderstandings. Also, the letter ‘i’ in Portuguese is always pronounced as /i/, while in English it is often pronounced as /ɪ/, as in /kɪs/ and /mɪs/.

Even though Portuguese does not use duration contrastively, Brazilian and other EFL learners from various L1 backgrounds have been found to weight temporal cues over spectral ones to discern the /i/ and /ɪ/ pair (FLEGE; BOHN; JANG, 1997). What is more, Kivistö-de Souza et al (2017) found that the reliance on durational cues appears to be related to ones’ vowel inventory size, with speakers of languages with smaller vowel inventories over-relying more on duration clues than those with larger inventories. Native speakers of English, on the other hand, tend to rely on spectral differences, not temporal ones (ESCUDERO; BOERSMA,

[30] Tenseness refers to the force of constriction involved in the production of vowels. In general, tense vowels have greater force of constriction than their lax counterparts (GIEGERICH, 2005).

2004). Bohn (1995, p. 294-295) attempted to explain this with what he called the Desensitization Hypothesis, which states that "whenever spectral differences are insufficient to differentiate vowel contrasts because previous linguistic experience did not sensitize listeners to these spectral differences, duration differences will be used to differentiate the non-native vowel contrast". For example, native speakers of Spanish, who have only one high front vowel in their native inventory, should be linguistically desensitized to spectral differences between vowels from other languages that occupy the same area of the acoustic vowel space (BOHN, 1995). The same should happen with Brazilian speakers of Portuguese, as their linguistic experience as L1 speakers of the language has not sensitized them to the spectral differences between /i/ and /ɪ/.

The lack of experience of English teachers with phonetics and phonology may be a contributing factor for English L2 learners' reliance on temporal clues. As these cues tend to be more salient than spectral ones, it is arguably easier and simpler to teach that one vowel is simply longer than the other, especially when teachers themselves often claim to have had inadequate training to teach phonetics (BOHN, 1995; ESCUDERO, BOERSMA, 2004; COSTA, 2016). However, Flege, Bohn and Jang (1997) learned that experienced non-native learners produced and perceived English vowels more accurately than inexperienced ones, and tended to use spectral clues better.

The fact that the English high front vowels tend to pose challenges for Brazilian learners has motivated studies such as Junior (2010), Lima (2005) and Nobre-Oliveira (2007) to investigate if phonetic instruction and phonetic training can help improve pronunciation. Lima (2005) sought to test the efficacy of an intervention on the production of the English vowels /i/, /ɪ/, /ɛ/ and /æ/. The participants were six students from an undergraduate course in Languages, who were divided into a control and an intervention group. The three participants from the intervention group received perception and production training on two minimal pairs during 23 encounters, which lasted an average of one hour and a half each. To assess the participants' pronunciation before and after the intervention, they had their speech recorded during an informal talk, and also read a sequence of words and sentences containing the vowels /i/ and /ɪ/, and /ɛ/ and /æ/. Their speech was analyzed based on an error analysis model in which Lima himself was responsible for determining whether the pronunciation of a given word was correct or not. Overall, the results show that, for the intervention group, 54,8% of the target words were incorrect in the pretest, while in the posttest this percentage was of 33,6%, that is,

there were 21% less errors. Nonetheless, considering only the /i/ and /ɪ/ vowels, this reduction in errors was of only 0,9%, indicating a much smaller positive effect of the training for this pair of vowels.

The goal of Nobre-Oliveira (2007) was to investigate the effects of perceptual training on the learning of the English vowels /i/, /ɪ/, /ɛ/, /æ/, /ʊ/ and /u/. The author also sought to understand if this type of training could lead to improvements in the production of these vowels. To do so, 29 Brazilian EFL learners trained with either natural stimuli from native speakers, or with computer-generated, synthesized stimuli that had enhanced spectral cues and did not vary in duration. The results revealed improvements in the perception tests for /ɛ/ and /æ/ or /ʊ/ and /u/, but not for the /i - ɪ/ pair. However, this was the only pair in which significant improvements were found in terms of production. She argues that, because the /i - ɪ/ pair already had good perceptual identification rates in the pretest, it was difficult to improve their perception even further, but the higher identification rates paved the way to the enhancements in the production. On the other hand, the pretest revealed that the participants had more difficulties perceiving the other four vowels, and thus had to learn more in the perception domain until these gains could be translated to production. Nobre-Oliveira (2007) argues that these findings support the claims of the Speech Learning Model (FLEDGE, 1995) in that perception precedes production.

Another study that tackled the effects of instruction on the pronunciation of the English vowels was Junior (2010), who recorded the speech of 28 EFL learners from Brazil, and then analyzed the percentage of errors in their production. In similar fashion to Lima (2005), the speech of the participants was analyzed by the researcher himself, who judged whether their productions were correct or incorrect. The /i/ vowel had the fifth highest percentage of error (81%) in the pretest, but after 4 hours of communicative pronunciation instruction administered at the end of the regular classes, this percentage was 18% lower in the posttest (63% of all /i/ productions). However, there were better outcomes for other sounds: the percentage of errors was 43% lower for /ð/ and /θ/, and 29% lower for /æ/ after the intervention. These results indicate that Brazilians can improve their pronunciation of the /i/ vowel, but that it may require more instruction and attention than other challenging sounds.

The results of the studies reviewed in this section suggest that instruction and training can have positive effects on the production of the English high front vowels by Brazilian learners. However, it is still unclear whether such improvements can be achieved when the instruction is delivered by a smartphone application. In light of this, the current study sets out

to investigate the use of an ASR-based smartphone application and its effects on the intelligibility of Brazilian learners. The method employed to achieve this objective is described in the following chapter.

3 METHOD

3.1 RESEARCH QUESTIONS

This study sought to investigate the use of an ASR-based mobile application and its effects on the intelligibility of Brazilian learners of English. Based on the objectives presented in Chapter 1, four research questions guided this investigation, namely:

RQ1) Does the use of an ASR-based mobile application improve the participants' overall intelligibility in English?

RQ2) Does the use of an ASR-based mobile application improve the participants' intelligibility of the English high-front vowels?

RQ3) How did the participants engage in self-regulated learning with the mobile application?

RQ4) What were the participants' perceptions of the application and of mobile learning?

3.2 STUDY DESIGN

This study follows a quasi-experimental, mixed-methods, one group[31] pre-test-post-test design (DÖRNEY, 2007). Because of the lack of random assignment and of a comparison or control group, which are integral parts of experimental studies, this study has a quasi-experimental design. According to Dörney (2007), quasi-experimental studies have strong external validity because they "take place in authentic learning environments using genuine class groups" (p. 120), but they subject the study to a number of new threats caused by the inequality of the participants and of the initial treatment.

[31] Initially, the participants of the study were to be assigned to two groups, one experimental and one control, but because the number of students who accepted to participate in the study was small, and having a control group would result in fewer participants in each group, a decision was made to have only one experimental group. Having only one group meant that every student who participated in the research would have access to application, which was a motivating factor for their participation.

This research involved a combination of qualitative and quantitative methods in complementary and interactive ways, which is commonly referred to as mixed method-design research (BROWN, 2014; DÖRNEY, 2007). To Mertens (2005), this design is particularly valuable if the issue under examination involves complex socio-educational contexts, which is the case of the current study, as its participants have different backgrounds and language learning trajectories. In this research, quantitative data were primarily collected through the intelligibility tests (described in section 3.5.4) and the application's dashboard (described in section 3.5.6), while data of a more qualitative nature were primarily collected through the questionnaires on the participants English learning background (described in section 3.5.2), and on their use and perceptions of the application (section 3.5.7) The combination of these two types of data enabled a multi-level analysis of mobile and pronunciation learning in the self-regulated context in which it took place.

This study is also an exploratory one. To Swedberg (2020), exploratory studies are necessary when there is a need to know more about a topic, but because there is little information on it, the study has to be exploratory in nature. Although there have been previous studies involving ASR-based mobile instruction, few studies have investigated its effects on speech intelligibility. Moreover, this study was conducted in the unprecedent context of the COVID-19 pandemic, which required the use of remote learning tools that had never been explored by most teachers and learners. Ultimately, this study is exploratory in the sense that it entails the use of digital technologies that are altered and updated on a constant basis, which means that unexplored and unknown factors are likely to be involved in their investigation. The exploratory nature of the study is emphasized when considering the novelty of some of the fields encompassed in this research, with fields such as MALL and MAPT often involving multiple definitions and unsettled terminology.

3.3 RESEARCH CONTEXT

The current study took place with students from Colégio de Aplicação from the Federal University of Santa Catarina[32] (UFSC), a public school in the city of Florianópolis that works with primary, middle, and high school students. As part of the university, the school offers

[32] The Federal University of Santa Catarina (in Portuguese: Universidade Federal de Santa Catarina) is a public university in the city of Florianópolis, the capital city of the state of Santa Catarina, in southern Brazil.

many research projects and extracurricular activities for students and the local community. The students at the school come from different neighborhoods in Florianópolis as well as other cities nearby and constitute a diverse group in terms of socio-economical, educational, and racial backgrounds (FARIAS, 2018).

Four foreign languages are taught at Colégio de Aplicação: English, French, German and Spanish. In the first grade of middle school (Grade 6), students take the four languages so that in Grade 7 they can choose the language which they will study until the end of high school, with English being the language chosen by most students. Foreign language classes take place twice a week in middle school and once a week in high school. Their English classes do not follow one specific language teaching method. Rather, the language is seen as a way to act in the world, with communicative goals and intercultural knowledge having a crucial role on communication. In this sense, the classes provide opportunities for the students' linguistic, interdisciplinary, intercultural, and critical development. Because of the limited time in class and the various objectives of the classes, time spent on pronunciation teaching in class tends to be limited.

From April 2020 to October 2021, all of the classes and research projects from Colégio de Aplicação took place exclusively online due to the COVID-19 pandemic. This was very challenging as the school had never worked remotely before, and there were a number of issues involved in the adaptation process, such as the lack of access to adequate internet connection, computers, and suitable study environments (SILVA, 2021). As the students were not used to this modality, they struggled to keep track of their school activities and deadlines. Moreover, many students had to help their families at home by doing chores and taking care of siblings or relatives. All of these challenges resulted in multiple cases of school dropout throughout 2020 and 2021 (SILVA, 2021).

As this study involved students from the school, all the instruments and procedures for data collection had to be adapted so that they did not involve any sort of physical contact with the participants.

3.4 PARTICIPANTS

There were two types of participants involved in this study, namely: a) students from Colégio de Aplicação, who took intelligibility tests, used the application, and participated as

speakers in the study; and b) the listeners, whose role was to orthographically transcribe the sentences produced by the speakers.

3.4.1 Students from Colégio de Aplicação

Middle and high school students from Colégio de Aplicação (UFSC) were invited to participate in this study in the months of June and July of 2021. To do so, the researcher accessed one of the English classes of each grade of middle and high school, and shared links where they could access the consent and assent forms.

Initially, 35 students from Colégio de Aplicação (UFSC) agreed to participate and signed the assent and consent forms to participate in this piece of research. Eighteen students dropped out before they had started using the application, while four dropped out after starting its use. One of them completed the use of the application but had issues recording the intelligibility post-test and did not wish to record them again, so his data was not included in the results. Thus, from the 35 participants who signed the consent forms, 13 successfully completed their participation in the study. The data presented henceforth was collected with these 13 participants. Although there were two types of participants in the study i.e., the students from Colégio de Aplicação and the listeners, the term 'participants' (abbreviated as 'P[33]') is from now on used to refer to the 13 students who concluded their participation, while those who participated as listeners (abbreviated as 'L') will be addressed as such.

Before starting the use of the application, the participants answered a questionnaire[34] to gather information on their profiles and English learning trajectories. They were nine females and four males, and their ages ranged from 12 to 17 (M:14.5; SD 1.6). There were five middle school[35] students, and eight high school[36] students. They had been studying English from one to 10.5 years (M:5; SD 2.1), with four of them taking private English courses in addition to their classes at CA. An online test[37] revealed that their proficiency levels ranged from A0 to B2 in the Common European Framework of Reference for Languages (CEFR). Only one

[33] This abbreviation was used because the participants stated they preferred this type of identification rather than choosing pseudonyms.
[34] For more information on this questionnaire, please refer to section 3.5.2.
[35] In Brazil, middle school encompasses grades 6 to 9.
[36] There are three grades in Brazilian high schools: 1st, 2nd and 3rd grade.
[37] For detailed information on the proficiency tests applied, please refer to section **3.5.3.**

participant had been to a country where English is spoken as a first language, where he stayed for less than a year. Detailed information on these participants is displayed on Table 1.

Table 1 - Information on the students who participated in the study

Participant	Proficiency level	Proficiency level (numerical[38])	Years of Eng. Instruction	Age	Grade
P1	A1	2	1.5	13	8th
P2	B2	5	6	14	1st
P3	A2	3	6	16	1st
P4	A1	2	5	12	7th
P5	B1	4	6	16	2nd
P6	A2	3	3	16	3rd
P7	B2	5	5	17	2nd
P8	A2	3	5	14	9th
P9	A0	1	6	15	1st
P10	B2	5	5	14	9th
P11	A1	2	3	13	8th
P12	B1	4	10.5	16	2nd
P13	B1	4	3	12	7th
Mean (SD)	A2	3.3 (1.3)	5 (2.1)	14.5 (1.6)	-

Regarding the participants' trajectories as learners of English, it was found that, besides studying English at school or private courses, most of them practiced the language by watching videos, movies and TV shows in English without subtitles, by listening to podcasts and songs in English with or without screen lyrics, and by reading news and internet posts in English. They also mentioned computer and smartphone games and applications as important sources for practicing the language.

Reading was the skill they believe to have developed the most during their trajectories as English learners, followed by listening, writing and speaking. None of them believed that pronunciation instruction had a central role in their language courses in comparison to other

[38] Given that the results from the proficiency tests were given in the CEFR scale (from A1 to C2), their results had to be converted into numerical values so as to run statistical tests. More specifically, the lowest level (A0) was coded as 1, and the highest (B2), was coded as 5.

language skills. Seven participants believed their pronunciation was well developed during their school years, and three believed it was only partially developed. When asked to mention activities or practices that they believed to be good for pronunciation, their answers mentioned: a) speaking and conversation activities in class; b) activities in which the students read texts aloud and teachers asked questions about it; c) talking to other people who speak English; d) shadowing, which is similar to traditional repetition with the exception that speakers try to repeat sentences with as little delay as possible; e) oral presentations to their classmates; f) listening to the same word multiple times; and g) oral tests.

One participant believed not to have developed pronunciation at all, and two participants could not tell whether they had learned it. In spite of that, they all believed that studying and practicing pronunciation was either important or very important to learn English. Likewise, they stated to be eager or very eager to study to practice pronunciation in the language.

3.4.2 Listeners

English speakers were invited to participate as the listeners of the current study. Their role was to orthographically transcribe the sentences produced by the students from Colégio de Aplicação, before and after using the application. Considering that, nowadays, learners of English are more likely to use the language to communicate with other non-native speakers than with native ones (JENKINS, 2012), the listeners were both native and non-native speakers of the language. As this study did not have the goal of investigating possible differences in intelligibility between native and non-native listeners, the listeners' L1s were not treated as an independent variable.

A total of 13 (eight native, five non-native[39]) listeners participated in the study, meaning that each listener only transcribed the sentences from one speaker. This was done to avoid repetition effects, because if listeners carried out the sentence transcription task multiple times, they would become familiarized with the target sentences, which could result in inflated scores.

After reading and signing a consent form, the listeners answered a questionnaire to identify their nationality, their first language, the frequency they communicated with NNSs of

[39] Although the goal was to have equal numbers of native and non-native speakers, there were more native speakers

English and their knowledge of Portuguese. They were also asked to rate their familiarity with the English spoken by Brazilians, as this is one of the aspects that can affect intelligibility scores (CRUZ, 2003, SILVEIRA; SILVA, 2018). The data obtained with this questionnaire is summarized in Table 2.

Listener	First Language	Place of birth	Frequency[40] of communication with NNSs	Portuguese knowledge[41]	Familiarity[42] with English spoken by Brazilians
L1	Chinese	Hong Kong	5	0	2
L2	Danish	Denmark	5	1	2
L3	Dutch	Belgium	5	0	3
L4	English	USA	5	1	4
L5	English	USA	2	1	4
L6	English	USA	5	1	4
L7	English	Canada	5	1	2
L8	English	USA	4	1	2
L9	English	UK	1	0	1
L10	English	USA	4	0	4
L11	English	Canada	5	0	1
L12	Spanish	Argentina	4	1	2
L13	Spanish	Argentina	1	1	1
Median (IQR[43])			5 (2)	1 (1)	2 (2,5)

Table 2 – Listeners background information

On average, the listeners of this study: a) frequently used English to communicate with non-native speakers; b) had low levels of Portuguese knowledge, with five of them knowing no Portuguese at all and eight knowing just a few words; c) were mostly unfamiliar with the English spoken by Brazilians, but four listeners believed they had a good level of familiarity with it. The following section describes ELSA Speak, the mobile application used in the current study.

[40] On a scale from 0 to 5, with 0 being never and 5 being every day.
[41] On a scale from 0 to 5, with 0 meaning no knowledge of Portuguese and 5 meaning an advanced knowledge.
[42] On a scale from 1 to 5, with 1 meaning completely unfamiliar and 5 meaning completely familiar.
[43] IQR stands for Interquartile Range

3.5 THE ASR-BASED APPLICATION, ELSA SPEAK

Launched in 2015, ELSA Speak is a mobile application for English pronunciation development which, according to its developers, "allows users to practice and improve their pronunciation and intonation skills through a set of exercises that are evaluated on our servers" (ANGUERA; VAN, 2016). Because it uses ASR to evaluate and give feedback on the users' productions, the application is an example of an ASR-based mobile application[44].

ELSA Speak is a freemium application, which means that it can be downloaded for free, but most of its content can only be accessed by 'premium' users. More specifically, there are two versions of the application: ELSA Free and ELSA Pro. For this study, the participants received licenses to access ELSA Pro for three months. The licenses were acquired by the researcher, with funding provided by CNPQ[45].

ELSA Speak was chosen for this research for the following reasons: a) its ASR technology was developed for the application[46] and, according to the developers, is capable of recognizing and providing feedback to non-native speakers (ESLA API, 2022); b) it has standardized units, allowing all participants to receive the same instructions and do the same activities; c) it has a specific unit for the development of the high-front vowels, named "Skill 8 – /i/, /ɪ/"; d) the feedback it provides appears to be easy-to-understand, even for unexperienced learners; e) it offers a teachers' dashboard for companies and schools, allowing teachers to keep track of the students' progress; f) its content is available in Brazilian Portuguese[47]; g) the application has all the content expected to be found in pedagogical resources for pronunciation instruction, and covers the most relevant segmental and suprasegmental features (BALDISSERA, 2020); and h) it is a commercially available application, that is, it can be purchased or downloaded by the general public, in contrast to applications or platforms that are developed specifically for research.

Upon accessing the application for the first time, users are asked what their native language is. Below this question, there is a message explaining that this is done because "ELSA

[44] Although there are numerous mobile applications for English pronunciation available nowadays, not all use ASR to promote pronunciation instruction.

[45] This funding was provided by what is known as "taxa de bancada" a monthly stipend which CNPQ scholars can use to fund the expenses related to their research.

[46] Many mobile applications use more generic speech recognition technologies, such as Google Speech-to-Text, which are not designed with non-native speakers in mind.

[47] This was important because it allowed less proficient learners to navigate and use the application more easily.

adapts to your mother tongue to provide accurate feedback" (ELSA SPEAK, 2022). Users can also choose one of the nine languages available for the presentation of its content, one of them being Brazilian Portuguese. If they choose to use it in Brazilian Portuguese, all the instructions will be presented in Portuguese, and most the target words and sentences will be followed by their translations into the language. Once this is done, users are asked to choose an option that best describes: a) their goals as English learners; b) their proficiency level, with the options being beginner, intermediate and advanced; c) the amount of time they want to practice each day, which ranges from 10 minutes to 20 minutes a day; d) an optimal time to study so that the application can send daily reminders at that specific time.

Once these questions have been answered, the application suggests that users take a test to discover their "English speaking proficiency level" and identify their "top pronunciation challenges, like schwa and diphthong" (ELSA SPEAK, 2022). This test asks users to pronounce a number of sentences in English and then provides results with the scores obtained for each sound analyzed. To check how the pronunciation of a proficient speaker would be scored, I took the test, and received a score of 90%, as shown in Figure 2.

Figure 2 - Results from the Assessment Test

Source: ELSA Speak (2022)

As can be seen in the leftmost part of Figure 2, my level was judged to be ‘native’, although no further information was provided as to what this meant. As shown in the rightmost part of Figure 2, the application also shows the results according to a number of pronunciation features, including the high front vowels /i/ and /ɪ/. ELSA Speak utilizes these results to suggest learning plans for the users, with the lowest scoring features receiving more attention. However, taking the test is not mandatory. Users can simply skip it and go directly to the lessons they are interested in. In the current study, the participants were asked to go directly to unit 8, named “Skill 8 – /i/, /ɪ/”.

Unit 8 begins with five YouTube videos on the pronunciation of the high-front vowels /i/ and /ɪ/. The first video, named “how to make the EE vowel”, is from a YouTube channel called Rachel’s English, and provides information on tense high-front vowel /i/. It states that, in order to make this sound, “the jaw drops just a little bit, and the tongue tip stays behind the bottom front teeth” and that “the corners of the lips pull a little wide, they’re not quite relaxed” (RACHEL’S ENGLISH, 2016). Figure 3 shows how this is visually presented.

Figure 3 - Videos on the pronunciation of the high-front vowels

Source: ELSA Speak (2021)

In the second video, users are asked to watch how a woman pronounces the words ‘see’, ‘tea’ and ‘key’, and to repeat them accordingly. The third video, named ‘how to make the IH vowel’ and also provided by Rachel’s English, gives information on the lax vowel /ɪ/, and begins by stating that “this vowel can be a challenge for non-native speakers, as the tendency is to replace it with the /i/ vowel, but for the /i/ vowel, the jaw drops more, so the tongue isn’t as close to the roof of the mouth” (RACHEL’S ENGLISH, 2016). As in the first video,

examples are provided so learners can visualize the articulation of the sound. In similar fashion to the second video, the fourth one asks learners to repeat the words ‘sit’, ‘hit’ and ‘kick’.

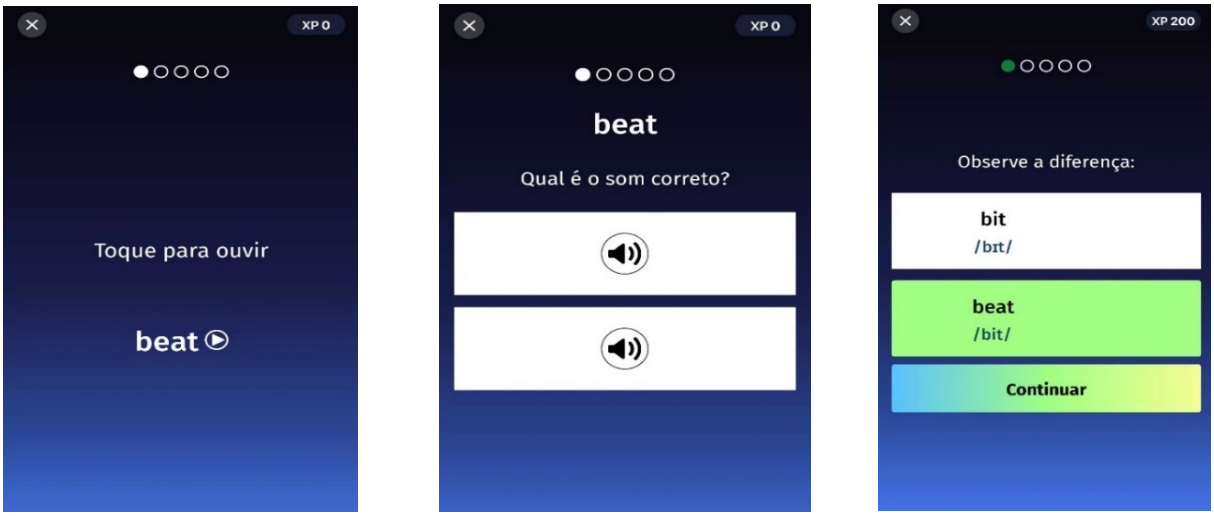
Building on the first and third video, the last one draws a comparison between the two high-front vowels. The video begins by stating the following:

> Now, I just did a video on the IH vowel versus the EE vowel, a comparison. And I didn’t talk about length at all. That’s because I don’t think length really comes into play when we’re talking about a vowel on its own. But since a lot of people teach length when they teach these two vowels, I thought I should talk about it. A lot of people will say that ‘ee’ is a long vowel, and ‘ih’ is a short vowel. But I feel like vowels themselves don’t have a length. To me the length of a syllable depends on: is it a stressed syllable, or is it unstressed? So, if the IH vowel is in the stressed syllable of a content word, then it will be long, even though some people would call it a ‘short’ vowel (…) So don’t think about the length of the vowel being tied to the vowel itself (RACHEL’S ENGLISH, 2016).

In other words, the video is telling learners not to rely on duration cues to distinguish between /i/ and /ɪ/, as the length of /ɪ/ can vary based on other factors that are not tied to the vowel itself. As discussed, native speakers of English have been found to use vowel quality, not vowel duration, to differentiate this vowel pair (ESCUDERO; BOERSMA, 2004), while L2 learners, especially inexperienced ones, have been found to weight temporal cues over spectral ones due to their L1 vowel inventory size and instruction that oversimplifies these vowels as being either ‘long’ or ‘short’ (FLEDGE, 1997; KIVISTÖ-DE SOUZA ET AL., 2017).

Unit 8 is divided into two main sections: ‘sound introduction’, which has 14 lessons focusing on introducing and practicing target vowels, and ‘entertainment’, which has 24 lessons for users to practice the vowels in the contexts related to entertainment such as movies, reading, and sports. Both sections have four types of lessons. In the first type, named ‘Can you hear the difference’, the application first pronounces one word from a minimal pair, then plays the pronunciation of the two words that make up the pair, and finally asks users to identify which one was pronounced. An example of this type of lesson is shown in Figure 4.

Figure 4 - 'Can you hear the difference' lessons

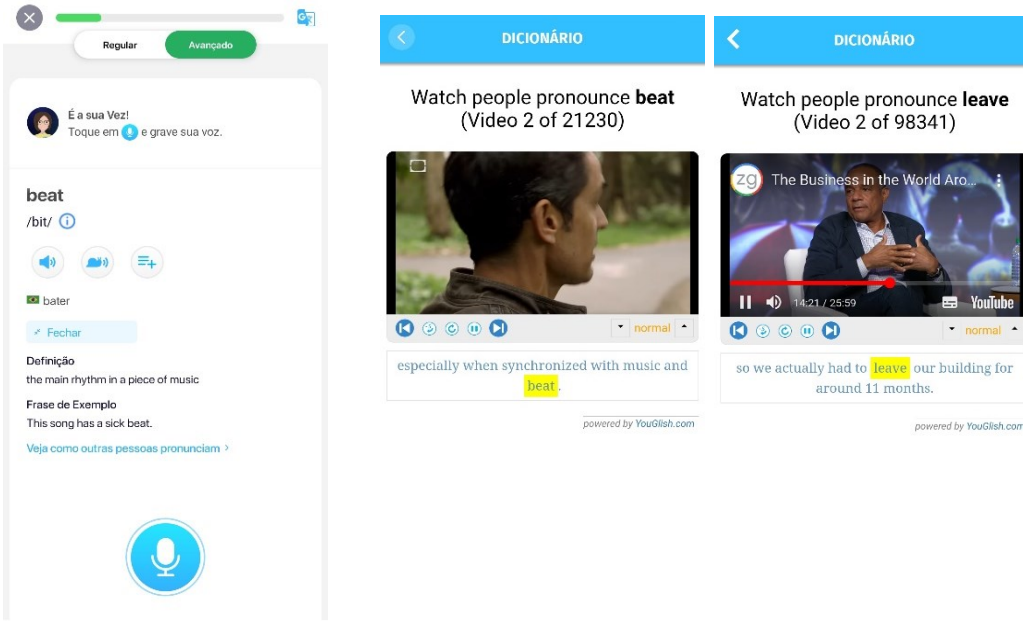


Source: ELSA Speak (2021)

As can be seen in the leftmost part of Figure 4, users first hear the pronunciation of ‘beat’, and then are asked to choose between the pronunciation of two words, which in this case were ‘bit’ and ‘beat’. After choosing, they receive feedback on whether they chose the correct sound, and are shown the spelling and the phonetic transcription of the words. In this example, green feedback signaled that the correct answer ‘beat’ was chosen. This type of activity resembles identification tasks from speech perception studies, in which listeners label a sound as belonging to one category or another (CHENG et al, 2021).

In the second and most common type of lesson in the application, the users pronounce words and sentences that contain the target vowels of the unit. An example is shown in Figure 5 with the word ‘beat’. Before pressing the microphone button to record their pronunciation, users can listen to the pronunciation of the words and sentences at regular and slower pace, check its translation into their L1, read example sentences and watch videos that show examples of the words in use. These are YouTube videos that are found using YouGlish.com, a tool which is able to locate specific parts of videos where the searched words are pronounced, as Figure 5 shows.

Figure 5 - Information provided for users before they pronounce the words



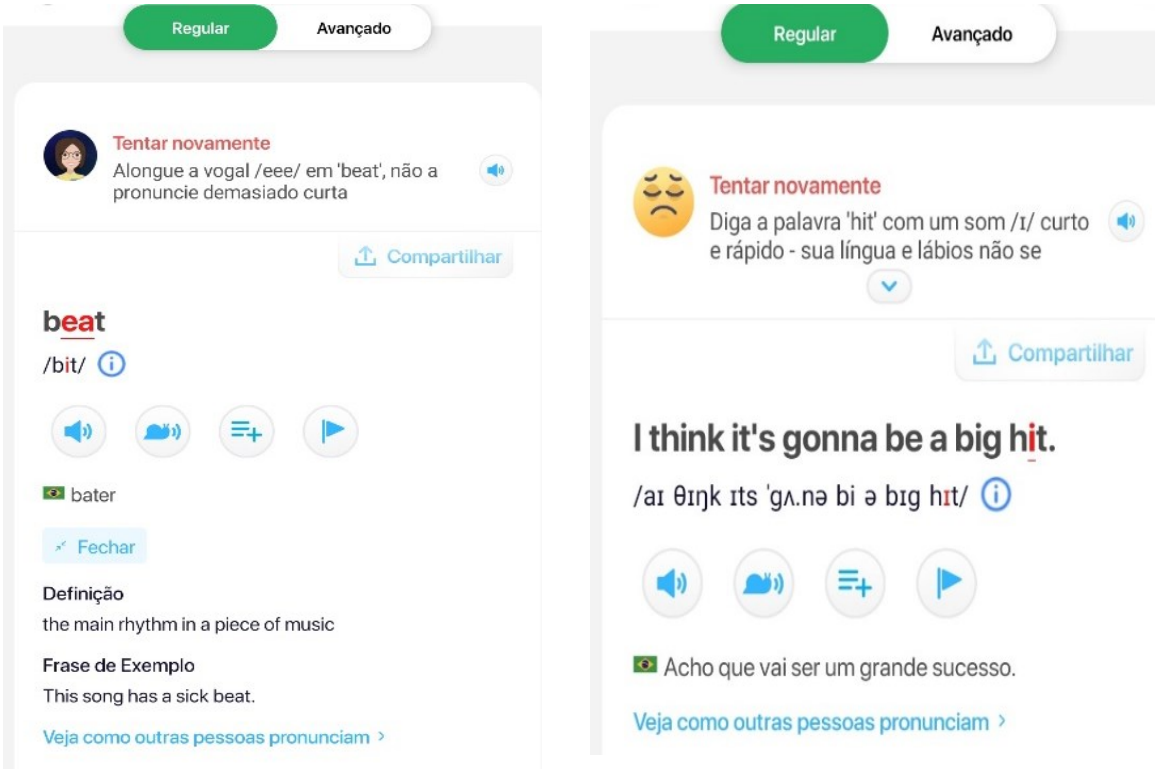
Source: ELSA Speak (2021)

To access videos such as the ones in Figure 5, users need to select the option ‘see how other people pronounce this word’. This will open a screen in which the section of the video where the word is pronounced can be repeated as many times as the users want. They can also alter the speed of video, and have access to subtitles of what was said, with the target word being highlighted in each of them. As can be seen in Figure 5, 21230 videos were found with the pronunciation of the word ‘beat’, and 98341 were found with the word ‘leave’, allowing learners to see plenty of examples of the usage of the words.

3.5.1 *Feedback in ELSA SPEAK*

After pronouncing the words and sentences, ELSA Speak provides two levels of feedback: regular and advanced. In the regular level, the feedback is focused on the target sounds i.e., /i/ and /ɪ/ in the current study. Examples of the regular feedback are shown in Figure 6, at the word and sentence level.

Figure 6 - Examples of regular feedback.



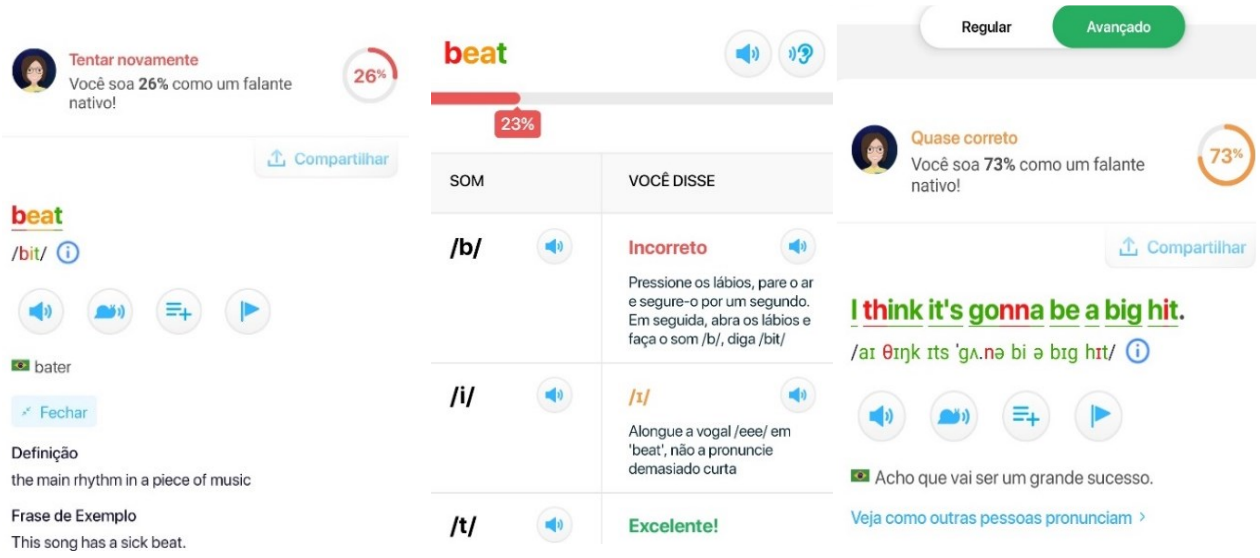
Source: ELSA Speak (2021)

In this instance, I intentionally mispronounced the word ‘beat’ with the lax /ɪ/ and with an unaspirated /b/. In the regular level of feedback, the application told me to “stretch the vowel /eee/ in ‘beat’, don’t pronounce it too shortly[48]”. The application uses ‘/eee/’ to represent the tense vowel /i/, probably to emphasize the length of the sound in comparison to the lax /ɪ/.

In the advanced level, the feedback is more complete, encompassing all the sounds produced, as can be seen in Figure 7.

[48] The feedback was originally provided in the user’s L1: “alongue a vogal /eee/ em ‘beat’, não pronuncie demasiado curta”.

Figure 7 – Examples of the advanced feedback



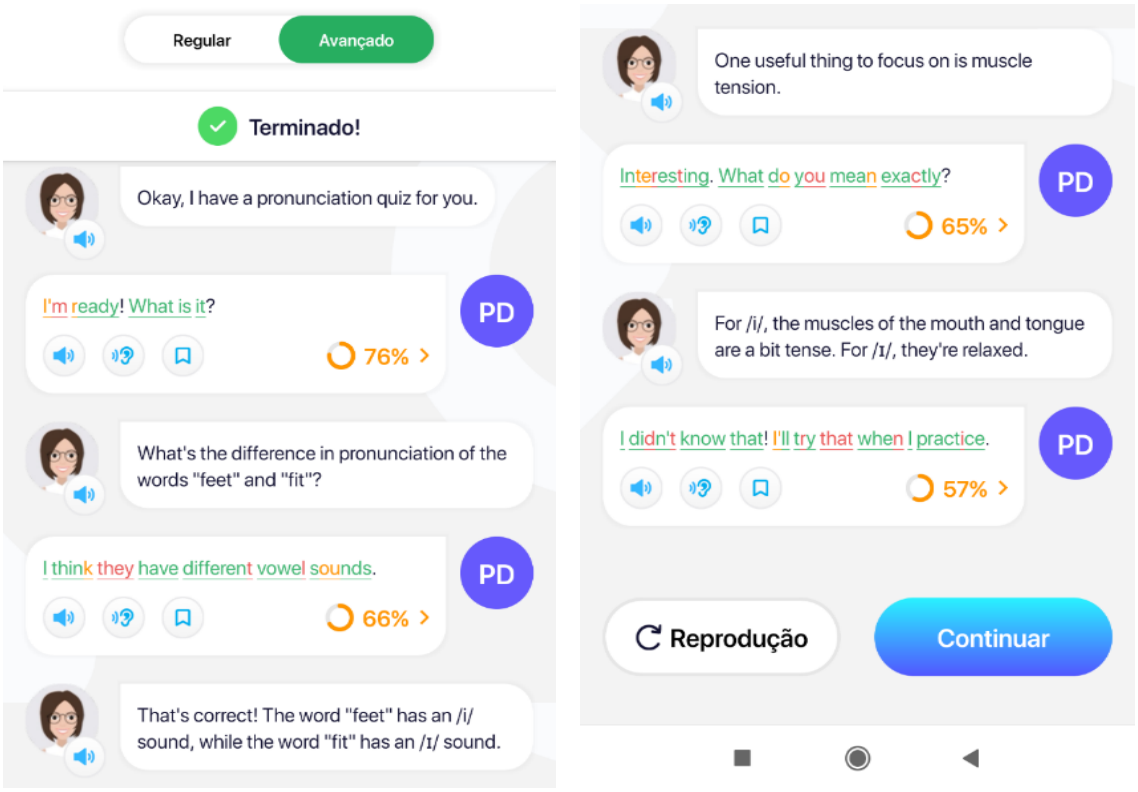
Source: ELSA Speak (2021)

As shown in Figure 7, in the advanced feedback, the application provided feedback for the vowel and for unaspirated /b/. Users can also get more detailed feedback by tapping the target word. In this case, it attempted to help the user to pronounce the /b/ by saying "press the lips, stop the air and hold it for a second, then open the lips and make the sound /b/[49]". The rightmost part of Figure 6 shows the advanced feedback for the sentence "I think it's gonna be a big hit", with many words signaled in red, in contrast to the regular feedback shown in Figure 5, where only 'hit' was signaled.

In the third type of lesson, users read a set of sentences to form a conversation with Elsa, the application's avatar. In the example of Figure 8, the conversation begins with Elsa saying 'Okay, I have a pronunciation quiz for you', to which the user has to read the response 'I'm ready! What is it?". The application gives feedback for each sentence recorded. Once users have read all the sentences, they can listen to the whole conversation.

[49] The feedback was originally provided in the user's L1: "Pressione os lábios, pare o ar e segure-o por um segundo. Em seguida, abra os lábios e faça o som /b/, diga /bɪt/".

Figure 8 - Example of a conversation lesson

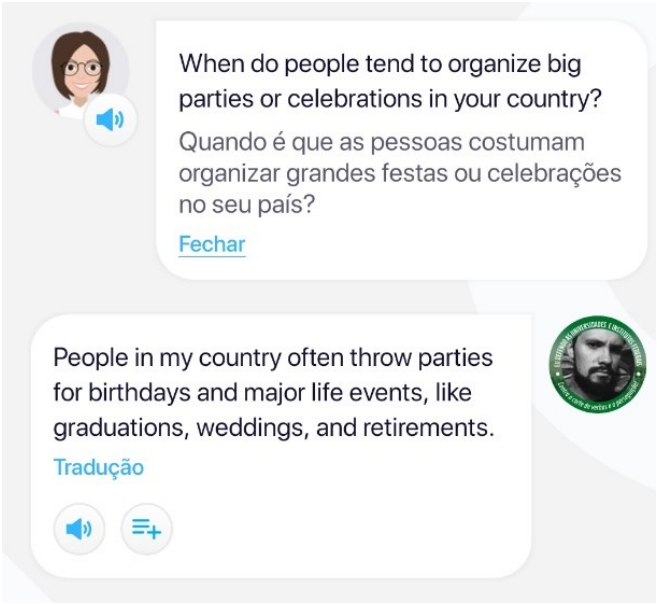


Source: ELSA Speak (2021)

In this conversation, explicit instruction regarding the pronunciation of /i/ and /ɪ/ is provided, with ELSA stating that for the /i/ sound, “the muscles of the mouth and tongue are a bit tense”, whereas “for the /ɪ/, they’re relaxed” (ELSA, 2021), which is an important distinction between these two vowel sounds.

It is relevant to point out that, in 2021, there were no translations available for the conversation lessons during the period of use of the application in this research, as can be seen in the leftmost part of Figure 8. Translations became available for this type of activities in 2022, which can be accessed by tapping on the ‘translation’ icon, as Figure 9 exhibits. It is also possible to do these activities without reading the translations.

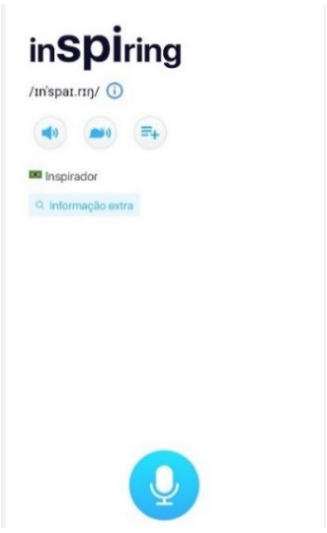
Figure 9 - Availability of translations for the conversation activities



Source: ELSA Speak (2022)

In the fourth type of lesson, users pronounce polysyllabic words and receive feedback focused on stress placement. In the case of Figure 10, the word ‘inspiring’ is shown with the second syllable in larger font, signalizing it should be stressed.

Figure 10 - Example of a stress placement lesson



Source: ELSA Speak (2021)

Having described the mobile application used in the current piece of research, the following section describes the materials and procedures for data collection.

3.6 MATERIALS AND PROCEDURES FOR DATA COLLECTION

The following instruments were used to collect data for this study: 1) consent and assent forms for all participants i.e., the speakers and the listeners; 2) a questionnaire on the participants' English learning background; 3) an English proficiency test; 4) an intelligibility pre-test; 5) unit 8 from the mobile application ELSA Speak; 6) an intelligibility post-test; 7) reports from ELSA Speak's dashboard, with data on the participants' use of the application; and 8) a questionnaire on the participants' use and perceptions of the application. These instruments, and the procedures taken to collect data with them are detailed below, in the order they were administered. For further reference, these materials are available in the appendices section.

3.6.1 Consent and assent forms

In compliance with Resolution 510/16, prior to data collection, a Portuguese version of the project of this study was submitted to the Ethics Review Board (Comitê de Ética em Pesquisas com Seres Humanos da Universidade Federal de Santa Catarina – CEPSH/UFSC). The project and all the consent and assent forms were approved on May 5th, 2021, under the serial number 4.835.478.

Consent forms were sent to all the participants of the study, that is, the students from Colégio de Aplicação and the listeners. To invite students from the institution to participate in the study, I visited one of their online classes in 2021 and shared the links where they could read the consent and assent forms in their Moodle courses. Consent forms (see Appendix A) were sent to the parents or guardians of the students who were interested in participating in the research but had less than 18 years of age. After their parents signed the consent forms, these students received and signed assent forms (see Appendix B). Consent forms (see Appendix C) were also sent for the listeners, who were invited via email to participate in the research.

3.6.2 Questionnaire on the participants' English learning background

The participants answered a questionnaire on Google Forms to gather data on their background as learners of English (see Appendix D). In its first section, there were questions on how long they had studied English, and on the type and amount of contact they had had with English inside and outside of the school. The second section asked four questions to gather data on their experience with pronunciation learning and their interest in learning pronunciation.

3.6.3 Proficiency test

The participants took a proficiency test (see Appendix E) to check for possible relationships between their proficiency levels and their performance on the intelligibility tests. Initially, the Oxford Placement was selected, but due to COVID-19 restrictions, an online placement test became necessary. After careful consideration, a proficiency test from ESL Language Studies Abroad website was deemed appropriate for the study because: a) it assessed grammar, listening and vocabulary knowledge; b) it was free and required no registration; c) it took from 10 to 15 minutes to complete, which meant a smaller burden on the participants; d) its design was intuitive, with questions that involved simple actions such as dragging and dropping words into matching photographs, or matching sentences halves, all of which could be completed with basic smartphones.

The participants were asked to write my email on the last page of the test so that I received their results as soon as they had finished the test. Individual results were sent via email to each participant. This was done so that they could check their detailed results and perhaps learn from them, thus making the completion of the test something useful for them as well as for the current study.

3.6.4 Intelligibility tests

The participants took intelligibility tests before (pre-test) and after (post-test) the use of the application. Intelligibility was measured by having listeners orthographically transcribe the speakers' utterances, which has been a common way to assess it (MUNRO; DERWING; MORTON, 2006).

In these tests (see appendix F), the participants read a list with 24 sentences, with each sentence composed of five words. There were: a) ten sentences in which the last word (target word) had either /i/ or /ɪ/, such as *sequel* or *listeners*; b) ten sentences in which the last word (target word) was a /i - ɪ/ minimal pair, such as *ship* or *sheep*; and c) four sentences that acted as distractors, that is, sentences in which the last word that did not have any /i/ or /ɪ/ sound. In total, 1560 words were recorded for the pre-test and 1560 for the posttest, 260 of them being target words with either /i/ or /ɪ/. The sentences were read in randomized order, as can be seen Appendix G. To check the sentence sorted according to these three categories, please see Appendix H.

With the exception of the words 'ship', 'sheep', 'fit' and 'feet', all of the target words and all of the distractors presented in the tests were instructed in unit 8 of the application. Likewise, most of the non-target words also appeared in unit 8. Because most of the words the application instructed in Skill 8 contained /i/ or /ɪ/, the sentences also had some words with /i/ and /ɪ/ in positions other than the final, but these were not considered target words in the tests.

Initially, the intelligibility tests were going to be recorded face-to-face, using standardized equipment and procedures. However, due to COVID-19, the intelligibility tests had to be adapted so that they could be taken remotely. Thus, JotForms[50], an online form platform, was used since it allowed the speakers to read and record the sentences using their own computers or smartphones. The speakers were given instructions via Moodle on how to take the test and could also access these instructions in the test itself. They were told to listen to their recordings before submitting them to avoid the submission of empty or inaudible audios.

3.6.4.1 Sentence transcription task

Once all the participants recorded their pre and posttests, 26 .MP4 files (13 pretest and 13 posttest) were created by merging the sentences each participant recorded. The .MP4 files were then uploaded to YouTube[51]. The YouTube videos were only accessible to those who received their links, and did not have any information about the participants. The listeners were sent a link to a Google Forms questionnaire where they carried out the sentence transcription tasks (see Appendix I). To avoid repetition effects, each listener received the pre and posttests

[50] Available at https://www.jotform.com.

[51] This was done because Google Forms does not support .MP3 files, with YouTube videos being the only way to reproduce the recordings appropriately in this platform.

recordings from only one speaker, with some listeners transcribing the pretest before the posttest, while others transcribed the posttest before the pretest. They were not aware that the recordings were from the same participant[52], and that the recordings were from a pre and a posttest, which was important so that they did not transcribe one or another test more favorably.

Before beginning the transcription tasks, the listeners were instructed to: a) type a question mark for words they could not understand; b) type the words they think they heard, even if they were not sure about them; and c) type 'I didn't understand this sentence' for sentences they did not understand at all. The task was set so that once they clicked on the play button, the 24 sentences would be played in sequence, with each sentence being repeated once, and with a 10 second pause between every sentence to allow time for their transcription. This was done to prevent listeners from hearing the sentences more than twice, which could increase the intelligibility scores for some speakers. After transcribing the sentences, the listeners were asked to comment on the pronunciation aspects that impeded or facilitated comprehension.

3.6.5 Use of the application ELSA Speak in the current study

After the participants answered the background questionnaire, the proficiency test, and the intelligibility pre-test, they were sent the codes to activate the application and a file with instructions to do so. Before activating the application, they read a file with instructions and suggestions on how to use it appropriately (see Appendix J). More specifically, they were instructed to: a) watch introductory videos on the pronunciation of the high-front vowels; b) complete the 38 lessons from "sound introduction" section within one month, with a recommendation[53] to complete 10 lessons each week; c) do the lessons from the 'entertainment' section after finishing the 'sound introduction' ones; d) access both the regular and the advanced levels of feedback; e) pay attention to the aspects of the application that they liked or disliked; and f) take notes of words that the application struggled to recognize or that they struggled to pronounce. This file also described and exemplified the feedback provided by the application and highlighted how important it was to pay attention to it. Finally, the file had a Frequently

[52] Although each listener only heard the recordings from one participant, many listeners commented on differences between the participants, which indicates that they believed to be hearing different participants, and not pre or posttest recordings of the same person, as was the case.

[53] This recommendation was made in light of the spacing effect, which postulates that "for a given amount of study time, spaced presentations yield substantially better learning than do massed presentations" (Ellis, 1995, p. 16).

Asked Questions section with questions the participants could have and the answers to them. They were reminded that, if they had any questions regarding the use of the application, they could contact me via Moodle, e-mail or WhatsApp, and that the file with instructions was always available via Moodle.

It is important to point out that, even though they received instructions on how to use the application, the participants were free to use the application as they found appropriate, that is, when, where and for as long as they wanted. Since the ELSA Speak licenses acquired for the study were valid for three months after activation, the participants could also explore the application freely after the completion of the assigned lessons. As stated, a key aspect of self-regulated learning is that learners have some degree of freedom and autonomy to control and monitor their learning process, which involves selecting tasks, strategies and study times (SHA et al., 2012). However, whenever participants spent more than two weeks without accessing the application, I reached them via email or Moodle messages to offer the necessary support. This was done since, as Garrison (1997) pointed out, in SRL, internal feedback may not be accurate and explicit enough, so external feedback may also be needed to improve self-monitoring.

Two measures were taken in an effort to limit their pronunciation instruction with other sources than the application. First, I asked the participants not to access other sources to study pronunciation until their participation in the study was concluded, that is, until they had completed all lessons and submitted the intelligibility posttest. Second, I read their English teachers' syllabi[54] for the second semester of 2021, and found that there were no plans to work with the explicit pronunciation instruction, nor to develop the pronunciation of the high-front vowel.

3.6.6 Reports from ELSA Speak's dashboard

Schools and companies that purchase more than 20 licenses of the application ELSA Speak get access to a dashboard that monitors the use of the application. Because more than 20 licenses were acquired, the dashboard could be used in the current study. With it, it was possible to observe**:** a) the number of lessons the participants completed each day; b) the amount of time they spent on the application on a given day/week/month; c) the dates when the application was used or not used; and d) the average time dedicated to completing each lesson.

[54] These teachers read and signed a consent form (see Appendix L), in which they allowed me to read their syllabi.

The data from the dashboard helped investigate the self-regulated use of the application as it allowed us to understand how the participants controlled and monitored their learning process.

3.6.7 Questionnaire on the participants' use and perceptions of the application.

After completing the assigned lessons, the participants answered a Google Forms questionnaire with close and open-ended questions about their use of ELSA Speak, and their perceptions of the application and mobile learning (see Appendix K). These questions were informed by previous research on MALL, SRL and ASR-based pronunciation instruction.

The participants used an ASR-based application to independently achieve the goal of improving their pronunciation, which means they had to self-regulate their learning process. To collect data on how they engaged in self-regulated learning with the mobile application, questions were designed based on three[55] basic assumptions of SRL theories. More specifically, the participants answered questions about: a) the places where they used the application and the role of mobility in their use; b) the activities they performed while using the application i.e., whether they multitasked during the use of the application; c) their motivation to use the application at the beginning of the study, and whether they maintained this motivation; and d) their organization to use the application, with an organized user defined[56] as one who used the application in appropriate times and places, read and followed the instructions the researcher provided, and established weekly/daily goals to complete the lessons. These questions served to understand the self-regulated use of the application, and also provided an insight of how the participants engaged in mobile assisted language learning, which has been closely associated with the possibility of choosing where, when, and how to study (HSU, 2017; KHADDAGE et al., 2016; SHA ET AL.; 2012), with learners being able to "exploit small amounts of time and space for learning" (TRAXLER, 2007, p.8), often while multitasking.

To unveil the participants' perceptions of the application, there were questions about: a) its ASR system; b) the quality of the feedback provided and their experience with it; and c) what they perceived as advantages and disadvantages of learning with the application. The

[55] In SRL, learners should be able to a) select, organize, or create advantageous learning environments for themselves; b) plan and control the form and amount of their own instruction; and c) regulate and maintain their motivation (AN et al, 2021; SHA et al, 2012; ZIMMERMAN, 1990).

[56] This definition was presented in the description of the question about organization, so that the participants were aware of what an organized user was.

answers from this questionnaire should help understand the effects of the use of the application through a more qualitative perspective, which, in combination with the results from the intelligibility tests, will allow a multi-level investigation of ASR-based pronunciation instruction and of mobile learning.

3.7 PROCEDURES FOR DATA ANALYSIS

3.7.1 Analysis of the quantitative data

To analyze the participants' intelligibility in the pre and posttests, the listeners' transcriptions were copied to Microsoft Excel spreadsheets which, using "COUNTIF" functions, compared the transcriptions to the original words and scored the sentences. Three different spreadsheets with specific goals were created.

In the first spreadsheet, every word of each sentence was scored. As each sentence had five words, the maximum score for each sentence was five, and the maximum score for each participant in a given test was 120 (24 sentences with five words each). This scoring provided a measure of their overall intelligibility before and after using the application and helped understand the effects of the use of the application on words that were not specifically targeted on unit 8, but that still received ASR feedback. This helped to provide an answer to RQ1.

In the second spreadsheet, only the 5th i.e., last word of every sentence was scored, with the exception of the distractors, meaning that only words with high front vowels were scored. In total, there were 20 of these words in each test. With this, it was possible to analyze the effects of the application on words that were targeted on unit 8. Finally, only the 5 /i-ɪ/ minimal pairs were scored, more specifically: bit, beat, hit, heat, live, leave, fit, feet, ship, sheep. With this, it was possible to analyze whether the use of the application allowed the participants to produce words with /i/ or /ɪ/ and /i-ɪ/ minimal pairs in a more intelligible manner, thus providing an answer to RQ2.

Because the "COUNTIF" function could only score perfect word matches, the words were analyzed individually to verify if the listeners had misspelled any word. For example, if "beat" was misspelled as "beatt", its spelling was fixed so it was properly scored. The same was done to abbreviations with missing apostrophes, such as "didnt" and "dont". These

corrections were done so that possible typing mistakes from the listeners did not affect the intelligibility scores.

To check if the results of the intelligibility tests were normally distributed, Shapiro-Wilk tests were run. With the verification of the normal distribution of the data, two-tailed paired-samples t-tests were run to verify if the scores from the intelligibility pre and posttests were significantly different. Pearson correlation tests were used to investigate possible associations between the average time spent on a lesson and the intelligibility gains, as well as to check for relations between the participants proficiency levels[57] and their intelligibility. An alpha level of .05 was used for all statistical tests included in the current study.

Considering that the current study sought to investigate the effects of an ASR-based application using a mixed methods design, the quantitative data were analyzed in conjunction with data from ELSA Speak's dashboard, and with data of a more qualitative nature, such as the participants' perceptions on the application, and their self-perceived motivation and organization to use the application. This enabled the investigation of the participants' mobile learning experience in multiple but complementary ways, which Dörney (2007) considers to be a strength of mixed methods research.

3.7.2 Analysis of qualitative data

The qualitative data obtained with the questionnaires was analyzed with the goal of unveiling how the participants perceived the ability of the application to recognize their speech and the quality of its feedback. Their perceptions of the positive and negative aspects of the application and of their MALL experience were also analyzed. This analysis was done at the group and individual level, and involved their answers to open and close-ended questions of the questionnaire. Comparisons with the findings of studies reviewed in Chapter 2 were drawn to establish the affordances of current ASR-based systems, indicating in what ways they have evolved and what seem to be ongoing issues.

[57] As stated, the results from the proficiency tests were converted into numerical values so as to run the Pearson correlation tests.

4 RESULTS AND DISCUSSION

This chapter presents and discusses the results of this study based on the research questions presented in Section 3. Section 4.1 discusses RQ1, i.e. whether the use of an ASR-based mobile application improved the participants' overall intelligibility in English. Section 4.2 discusses RQ2, i.e. whether the use of the mobile application improved the intelligibility of the English high-front vowels. Section 4.3 discusses RQ3, i.e. the participants engagement in self-regulated learning with the mobile application. Finally, section 4.4 discusses RQ4, i.e., the participants perceptions of the application and of their mobile learning experience.

4.1 RQ1 - DOES THE USE OF AN ASR-BASED MOBILE APPLICATION IMPROVE THE PARTICIPANTS' OVERALL INTELLIGIBILITY IN ENGLISH?

The first research question of this study sought to investigate whether the use of an ASR-based mobile application improved the overall intelligibility of Brazilian learners of English. To do so, each participant recorded 24 5-word sentences before using the application and then recorded the same sentences after its use. The sentences were orthographically transcribed by the listeners and then scored so as to provide a measure of their overall intelligibility. Descriptive statistics for the overall intelligibility pre and posttest results are exhibited in Table 4.

Table 3 - Descriptive statistics for overall intelligibility

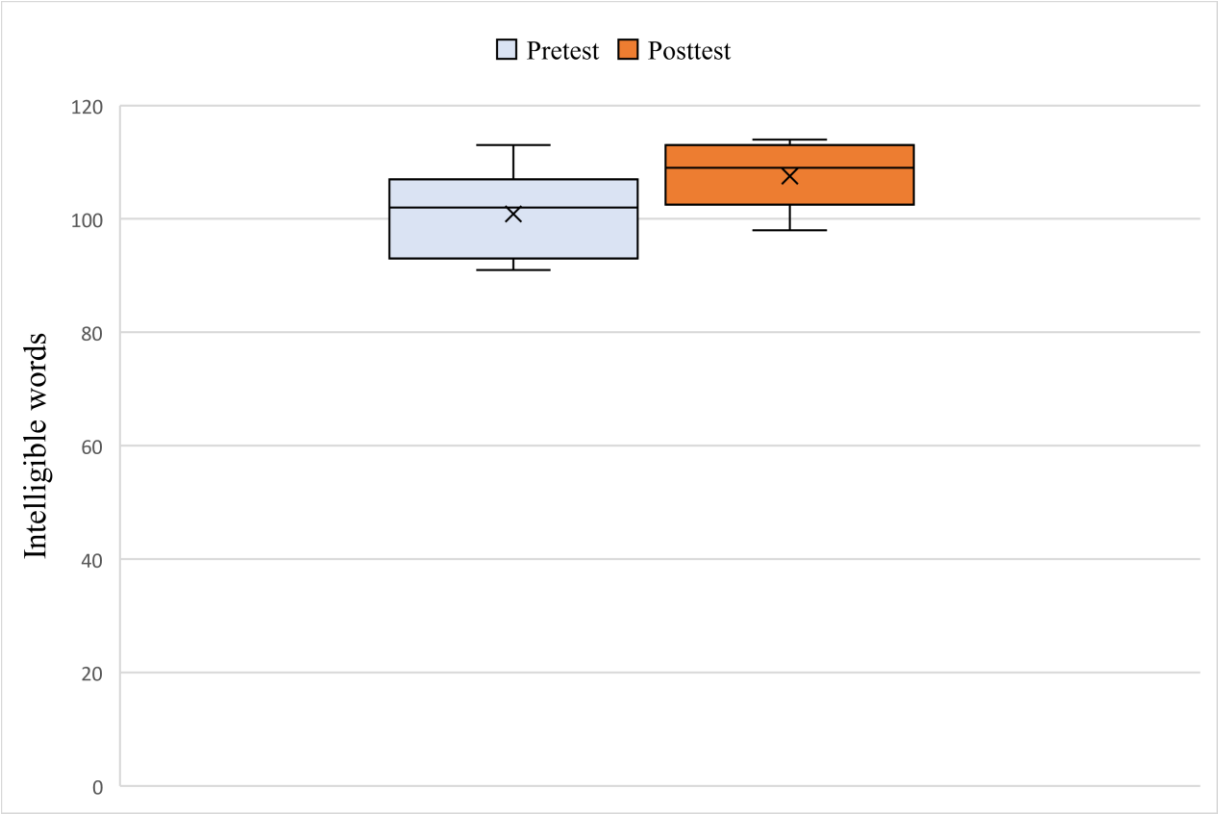
		Pretest		**Posttest**	
	Max	**M**	**SE**	**M**	**SE**
Overall intelligibility	120	100.8	2	107.5	1.47

Considering the small sample size, a Shapiro-Wilk test was performed and did not show evidence of non-normality (W = .932, p = .357). Based on this, a paired-samples t- test was run and showed that there was a significant difference between the pretest scores and posttest scores; t(12)=3,54, p = .003[58], 95% CI [2.67, 10.7]. This indicates that the use of the

[58] As stated, an alpha level of .05 was used for this and for all statistical tests included in the current study.

application improved the speech intelligibility of the participants. The differences between pretest and posttest scores can be visualized in Figure 11.

Figure 11 – Overall intelligibility in the pre and posttest.

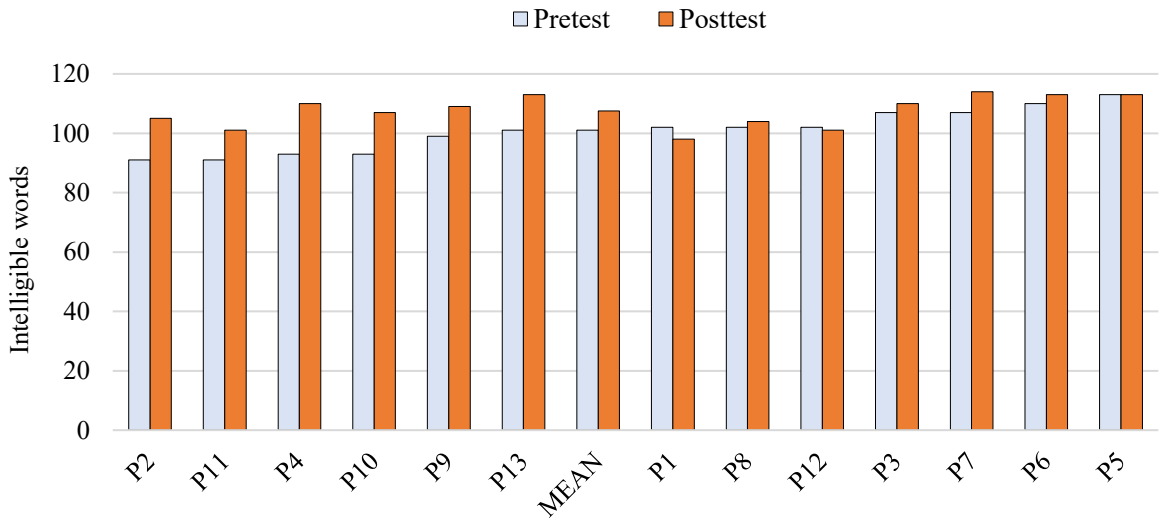


As illustrated in Figure 11, scores in the posttest were generally higher than in the pretest, suggesting that the use of an ASR-based system helped the participants of this research improve the intelligibility of their pronunciation in English. These results are in line with previous studies in which ASR-based instruction led to significant improvements in pronunciation (e.g., NERI, CUCCHIARINI, STRIK, 2006; LIAKIN, CARDOSO, LIAKINA, 2015; PARK, 2017), and signals that this type of instruction can also lead to gains in intelligibility. In contrast to Neri, Cucchiarini and Strik (2006), in which an ASR-system only reduced the pronunciation errors directly addressed in the CAPT training, the current study found significant gains for overall intelligibility, that is, there were gains for both target and non-target words. This is likely because the ASR-based CAPT system used in their study provided feedback exclusively for phonemes deemed to be problematic for L2 Dutch learners,

while ELSA Speak provides feedback for every word the participants pronounced, at segmental and suprasegmental level.

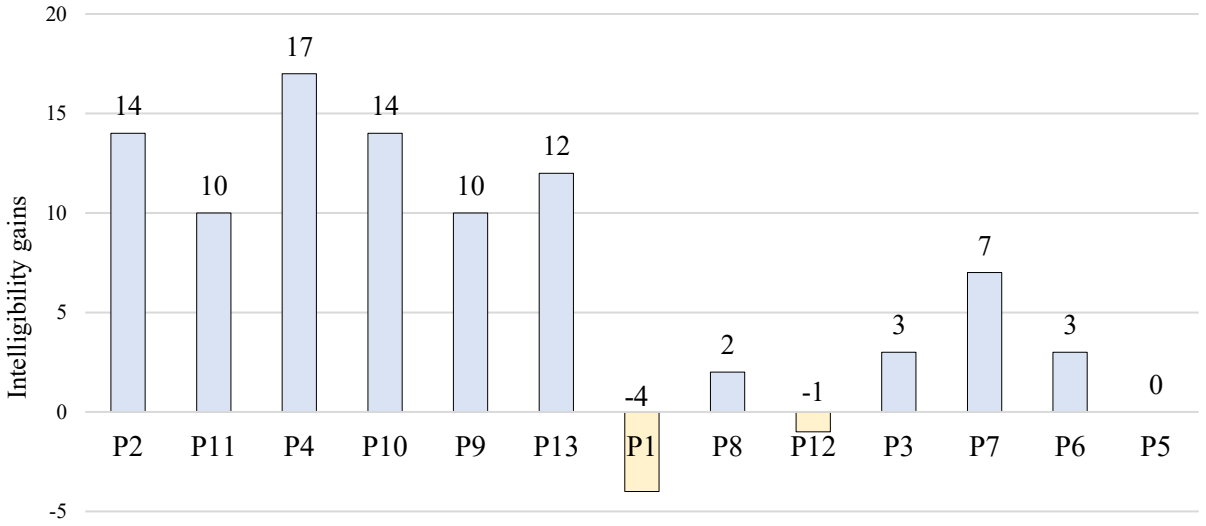
In spite of the significant differences between the pretest-posttest scores, the participants did not have uniform intelligibility gains. Figure 12 displays the individual results for overall intelligibility in the pre and posttests, with the lowest pretest score (P2) appearing in the leftmost column and the highest pretest score (P5) appearing in the rightmost column.

Figure 12 - Individual results for overall intelligibility



As can be seen in Figure 12, ten participants became more intelligible in the posttests. Participants 3, 7, 6 and 5, who already had higher levels of intelligibility in the pretest, seem to have been less affected by the use of the application. On the other hand, participants 2, 11, 4, 10, 9 and 13, who had lower scores in the pretest, had higher intelligibility gains. This can be better visualized in Figure 13, which displays their intelligibility gains, with participants with the lowest pretest scores appearing in the leftmost columns and participants with the highest pretest score appearing in the rightmost columns.

Figure 13 – Intelligibility gains, sorted by pretest scores.



As Figure 13 evidences, a pattern emerged when the intelligibility gains were sorted according to the scores in the pretests. The six participants (P2, P11, P4, P10, P9 and P13) who scored below the mean score (M=100.8), had a combined intelligibility gain of +77 (88.5% of the total gains) and a mean improvement of +12.8, while the participants who scored above the mean score had a combined gain of only +10 (11.5% of the total gains) and a mean improvement of 1.42. A strong[59] negative correlation was found between the pretest scores and the intelligibility gains, $r(11) = .-70$, $p < .007$, 95% CI [-0.9, -0.25], that is, as pretest scores increased, the effects of the application tended to decrease. This indicates that the application is more suitable for learners with lower intelligibility levels, a finding that is in line with previous CAPT studies such as Hincks (2005), Mahdi and Khateeb (2019) and Yuan and Liu (2020). For example, in Hincks (2005), an ASR-based CAPT system was found to be more beneficial for the participants who had lower pronunciation scores in the pretests and had an "intrusive foreign accent" (p.38), and in Yuan and Liu (2020) reading aloud practices supported by ASR were only able to improve the pronunciation accuracy of low-proficient learners. It is also possible that the diminished improvements for participants with higher intelligibility levels reflect a ceiling effect, that is, there was little room for improvement as they had already reached a ceiling in their learning, as was also the case in Chau et al. (2022). For example, P5 had 113 intelligible words in the pretest, so there were only 7 words which she could have improved.

[59] According to Cohen (1988), correlation coefficients larger than .50 are deemed to be strong ones.

In Chau et al. (2022), comparable results were found regarding the diminished gains for the participants with high intelligibility pre-tests scores. In their study, the authors argued that the participants' scores did not improve because they were already highly intelligible (83.4%) before receiving instruction. However, their scores were negatively skewed, meaning that there were few participants with lower scores to be analyzed. In the current study, a very similar mean pretest intelligibility of 84% was observed, but given the fairly symmetrical distribution of scores, it was also possible to examine the effects on both lower and higher levels of intelligibility.

No significant relationships between the participants' proficiency levels and their pre-test intelligibility scores were found $r(11) = .09$, $p = .76$, 95% CI [-0.48, 0.61]. This suggests that their English proficiency levels did not seem to predict the intelligibility of their pronunciation, a finding that corroborates Thomson and Derwing's (2015) claim that intelligibility is not tied to proficiency, that is, less proficient learners can be highly intelligible, while those with great knowledge of L2 syntax and vocabulary may have speech that is difficult to understand.

Although a relation was found between the level of intelligibility at the beginning of the study and the intelligibility gains, there are other factors that may have influenced the negative or positive outcomes of the use of the application, especially in the cases of participants 1 and 12. These factors are discussed in sections 4.3 and 4.4. The following section presents and discusses the results of the intelligibility tests with regard to the targeted high-front vowels.

4.2 RQ2 – DOES THE USE OF AN ASR-BASED MOBILE APPLICATION IMPROVE THE INTELLIGIBILITY OF THE ENGLISH HIGH-FRONT VOWELS?

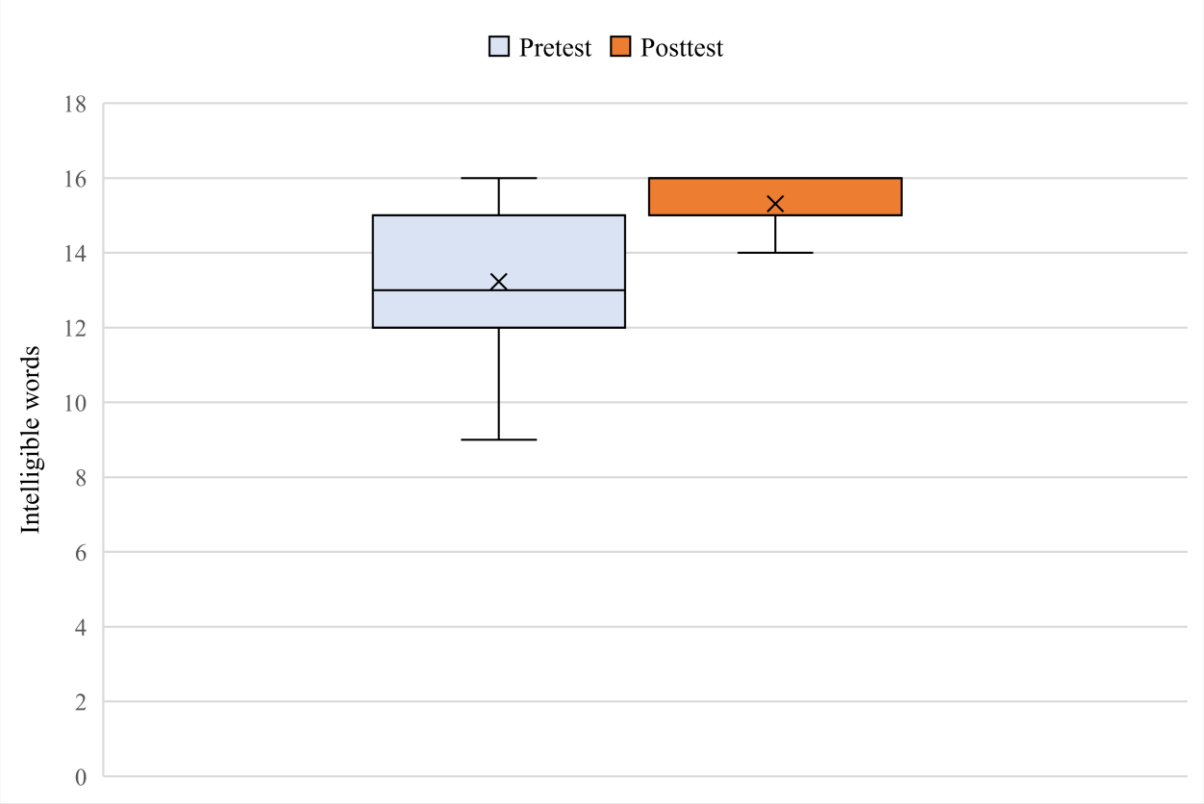
To investigate if the use of the application improved the intelligibility of the target /ɪ/ or /i/ words, the participants recorded 20 words with /ɪ/ or /i/ sounds, more specifically ten words that formed minimal pairs, such as *ship* and *sheep*, and ten words which had either /i/ or /ɪ/ but were not minimal pairs, such as *sequel* and *listeners*. Table 5 shows the descriptive statistics for the target words analyzed.

Table 4 - Descriptive statistics for target word intelligibility

		Pretest		Posttest	
	Max	**M**	**SE**	**M**	**SE**
Target word	20	13.2	0.53	15.3	0.44
Minimal pairs	10	4.8	0.51	6	0.26

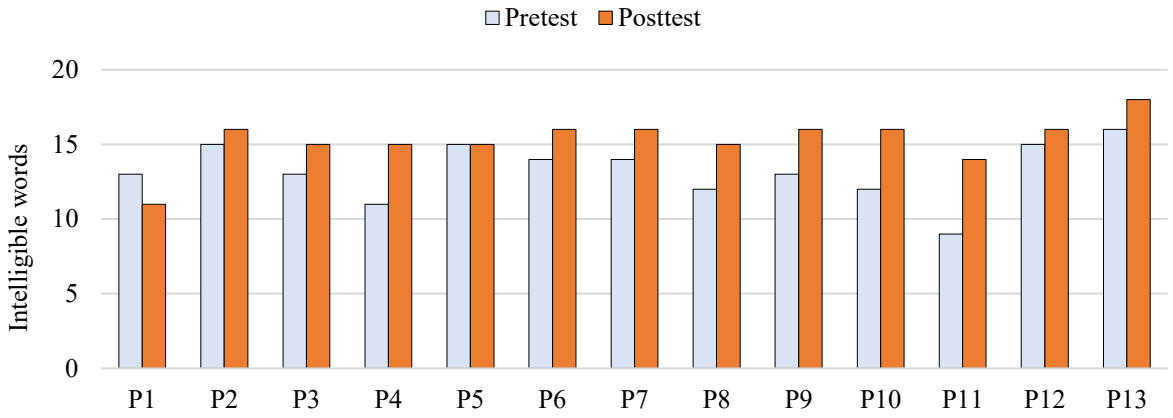
A paired-samples t-test was conducted and showed a significant difference between the pretest and the posttest scores; t(12)=4.1, p = .002., 95% CI [0.96, 3.19], which suggests that the use of the application had positive effects on the speech intelligibility of the English high-front vowels. Figure 14 displays these results.

Figure 14 - Target word intelligibility



When considering the individual results, only P1 did not score higher in the posttests, as can be seen in Figure 15.

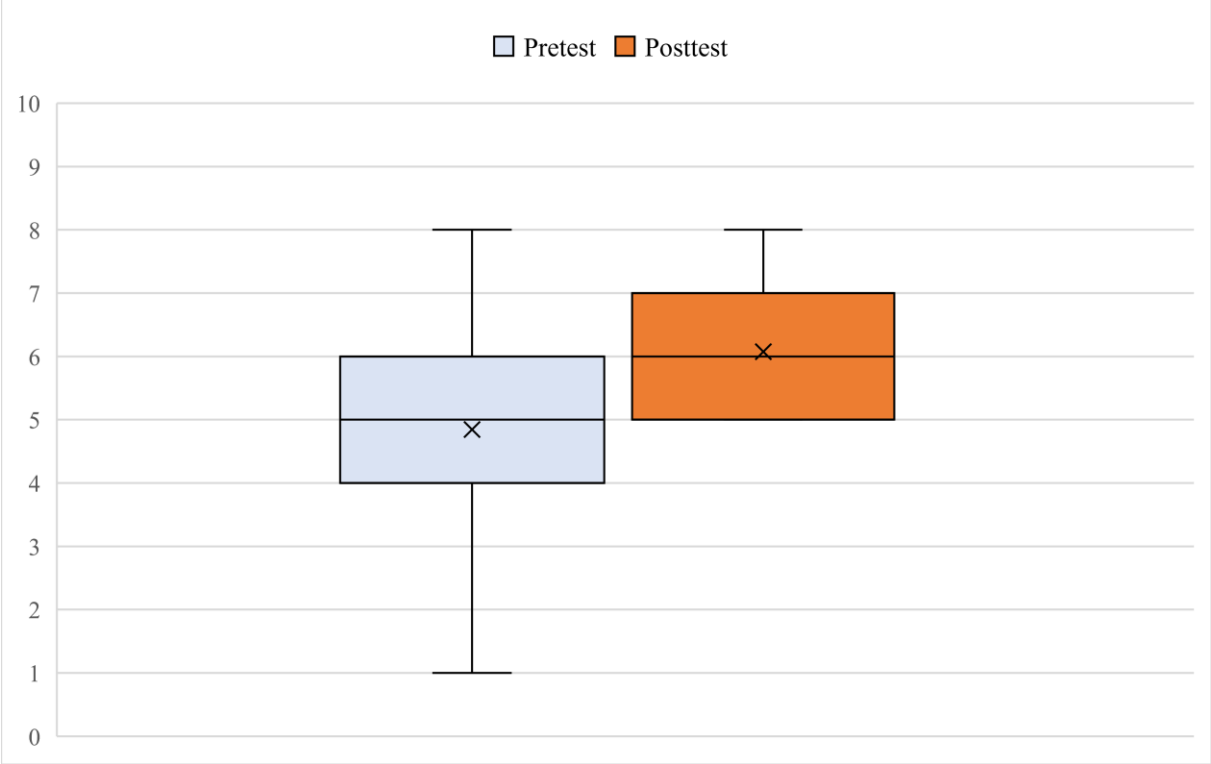
Figure 15 - Target word intelligibility by participant



In the case of participant 1, a process of overgeneralization seems to have occurred. In the pretest, P1 pronounced *beat* and *bit* as /bɪt/, while in the posttest *beat*, *bit*, *hit* and *heat* were all realized with the lax /ɪ/ sound. This suggests that the use of application affected her pronunciation of /i/, but was not enough for her to produce /i/ and /ɪ/ consistently and intelligibly.

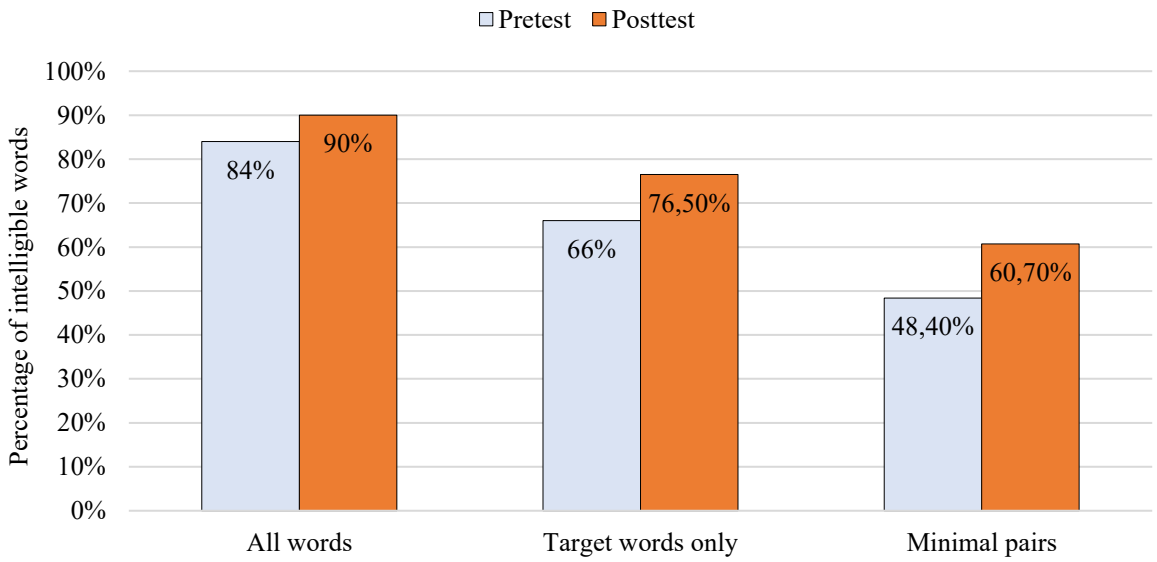
As Figure 16 shows, when only considering the 130 /i-ɪ/ minimal pairs, 63 (48.4%) of these words were intelligible, while in the posttest 79 (60.7%) were. Again, a paired samples T-test showed a significant difference between the pretest (M= 4.8 SE= 1.86) and the posttest scores (M=6, SE= 0.95); t(12)=3.41, p = .005, 95% CI [0.44, 2].

Figure 16 - Intelligibility of the minimal pairs



There were different levels of intelligibility for the words in each category. As can be seen in Figure 17, the participants had highest percentage of intelligibility when all words were accounted for, and had the lowest when only considering the /i-ɪ/ minimal pairs, with less than half of these words being correctly transcribed in the pretest.

Figure 17 - Percentage of intelligible words by category



Although the minimal pairs were the least intelligible words in the pretest (60.7%), they were the words that had the most gains percentagewise: 12.3%, compared to 10.5% when considering all target words and 6% for overall intelligibility. Still, the lower level of intelligibility of these words highlights how challenging they can be for Brazilian learners of English, who often produce them indistinctively (GONÇALVES, SIQUEIRA, 2014). For learners with lower levels of proficiency, as was the case of many participants of this study, appropriate production of these vowels is likely to be even more demanding. This difficulty can be noticed in some of the listeners' comments:

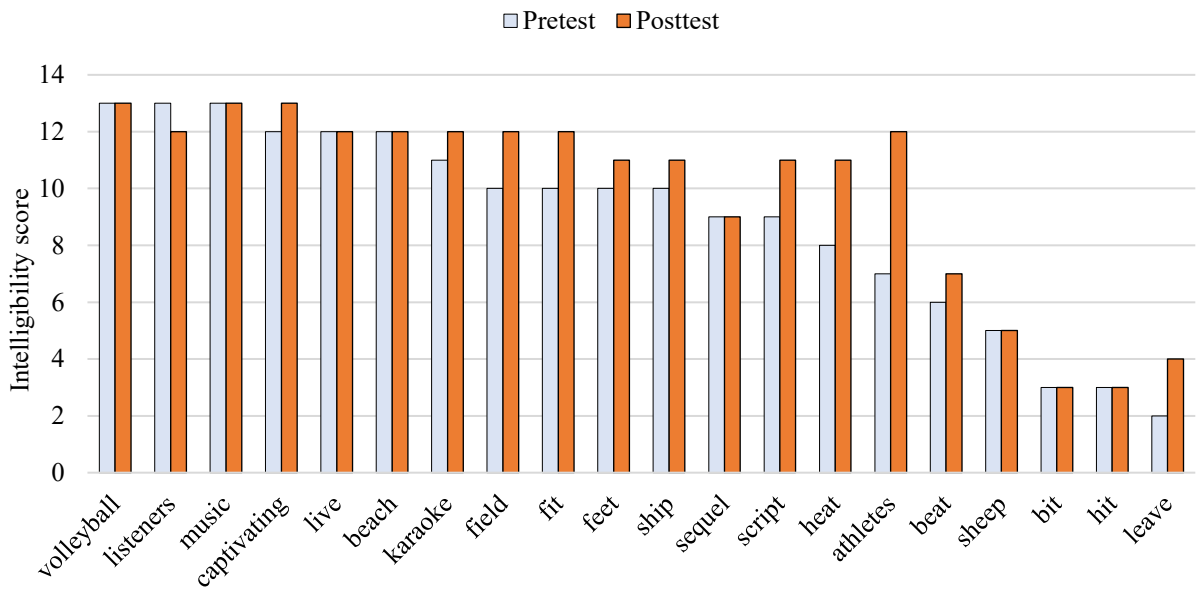
> L4: Some of the vowels had the wrong sound (for example, live rhymed with eve), (…) and the vowels kinda [sic] got mixed up.
> L5: I felt like I could understand most of what was being said (…) but the "I" and "ee" sounds took more listening skills to hear the difference (…) e.g., Sheep and Ship. They sound nearly the same however the second speaker [referring to the posttest recordings] was better at emphasizing which made it clearer for me to hear the difference as compared the first speaker.
> L7: It sounds like there were some homophones but that is difficult to understand for native speakers too.
> L8: (…) mostly vowels were hard to understand (e.g leave vs. live, field vs. felt).
> L9: The times it is more difficult are with particular vowel and consonant sounds (the most noticeable being "th" and "i/e") and when the words blend together a little.
> L10: Vowels are mispronounced in some cases. Some sounds are omitted by the speaker, too. I think there are many examples of negative transfer from Portuguese.
> L11: In my opinion, the participants had a thick accent, some problems to pronounce minimal pairs such as /i:/ /I/ and /u/ /u:/ (…).

The comment made by L5 reveals a possible effect of the use of application that is not evident in the intelligibility tests. She commented that the participant's pronunciation of 'sheep' and 'ship' was very similar, but that the 'second speaker', which in fact was the posttest of the same participant, was able to emphasize them differently. It is likely that this emphasis refers to the durational cues provided, that is, the participant produced the tense /i/ sound in words such as 'sheep' with more length, which allowed the listener to differentiate between them. Although native speakers of English have been found to weigh spectral cues over temporal ones (FLEGE; BOHN; JANG, 1997), temporal cues seem to have helped this listener distinguish the /i - ɪ/ pair. It is possible that, since she was somewhat familiar with the English spoken by Brazilians, she was aware that these speakers may pronounce /i/ with more emphasis and/or with increased duration so as to differentiate it from /ɪ/.

After using the application, the participants' productions of /i-ɪ/ minimal pairs became 12.3% more intelligible, which suggests that the application can assist them in the development

of these sounds. Nonetheless, given that they were still 40% unintelligible, the use of the application alone may not be enough for Brazilian speakers of English to produce them consistently and intelligibly. This was specially the case for the words ‘bit’ and ‘hit’, two words with the lax /ɪ/ sound. As Figure 18 shows, these words were correctly transcribed only three times in the pretest, with no intelligibility gains in the posttest.

Figure 18 – Intelligibility by target-word



In similar fashion to Gonçalves and Silveira (2014) and to Silveira and Silva (2018), the palatalization[60] of the final /t/ in words such as ‘bit’ resulted in unintelligible speech. The word ‘bit’, for example, was transcribed as ‘beach’ twice in the pretest. To improve the intelligibility of this word, these participants would have to alter the pronunciation of two distinct sounds, that is, they would have to change the quality of the vowel and also produce /t/ without palatalization. The application did not seem to help with this process, as there were no gains for the word ‘bit’ in the posttest, when it was transcribed as ‘beach’ four times.

A more positive effect of the use of the application took place with the words ‘leave’, transcribed as ‘live’ ten times in the pretest, and ‘heat’, transcribed as ‘hit’ once in the pretest. In the case of ‘leave’, as coda modification was not an issue, it is possible that the participants were able to focus on changing the quality of the vowel /i/. What is more, for the high front

[60] As when alveolar stops /t/ and /d/ are produced with the palate, becoming /tʃ/ and /dʃ/. This process is also referred to as affrication in Silveira and Silva (2018).

vowels, the application offered sound identification lessons and instructional videos on top of the ASR feedback, while the only way issues such as palatalization could be developed was via the feedback.

Because the minimal pairs were the last word of otherwise identical sentences (e.g., 'they will sell the ship' and 'they will sell the sheep'), the listeners could not rely on contextual clues to transcribe them, which meant that the participants' pronunciation of the minimal pairs was paramount for their intelligibility. This issue was noticed by one of the listeners, who commented that the "the [participant's] pronunciation is fine but maybe she spoke some words I'm not familiar with or I have to know more about the context to get the correct word because some words have the same pronunciation" (L6).

This was not the case for the words 'volleyball', 'listeners', 'music' and 'captivating', which are not minimal pairs and were embedded in sentences where other words facilitated comprehension by providing semantic context to the listeners. For example, in the sentence 'we listen to country music', the words 'listen' and 'country' probably facilitated the intelligibility of the word 'music'. In similar fashion, 'fit' in 'I don't think they fit' and 'feet' in 'she works with her feet' were largely intelligible even though they were minimal pairs, most likely because their sentences had contextual clues that guided listeners. In this regard, one listener commented that "(…) sometimes I could make up some of the words from the context (e.g., the sentence on Karaoke)". As Derwing and Munro (2005) point out, context plays an important role in intelligibility as "the gap between what is unintelligible and what is merely heavily accented but still understandable might be explained in part by the fact that listeners use context to interpret speech" (p. 386). This was also shown in Silveira and Silva (2018), in which words embedded in sentences with more semantic information were generally more intelligible than words that were in semantic limited contexts.

As can be seen in Figure 18, the minimal pair 'leave' and 'live' had very different intelligibility levels. While 'leave' was correctly transcribed only twice in the pretest and four times in the posttest, 'live' had 12 correct transcriptions in the pretest and 12 in the posttest. One explanation for this is that, in the lack of appropriate pronunciation of 'leave', the listeners assumed that the participants intended to say 'this is where I live' because it is a more frequent and probable sentence than 'this is where I leave'. As pointed out in Becker (2013), frequency plays a role in intelligibility, with more frequent tokens being more intelligible. Indeed, upon checking the Corpus of Contemporary American English, the sentence 'this is where I live' was

found to be more frequent, with 97 instances reported, versus 25 of ‘this is where I leave’ (DAVIES, 2008).

The word ‘athletes’ was the least intelligible target word that was not a minimal pair. In the pretest, it was transcribed as ‘ladies’, ‘adverts’, ‘at it’, and was completely unintelligible for 3 listeners. For this word, mispronunciation of the high front vowels may not have been the only issue, as some participants pronounced [ˈæθliːt] as [æˈtlit], [ɑˈtlit] or [ɑˈtlet]. In these cases, inappropriate stress placement may also have hindered intelligibility, as was the case in Cruz (2003), in which stress errors were the strongest source of unintelligibility. This issue was noted by one of the listeners, who commented that the participants “sometimes stressed the wrong syllables, e.g., captivating and athletes”. In the posttest, however, ‘athletes’ was unintelligible only once. As stated, ELSA Speak has lessons that focus on appropriate stress placement, which may have contributed to these positive effects. The improvement in terms of stress placement was noticed by L5, who commented that “the difference between the two speakers was their ability to give emphasis to certain parts of the words”. This is a positive finding considering that inappropriate stress placement was found to be one of the most common sources of unintelligibility of Brazilian learners of English (CRUZ, 2012).

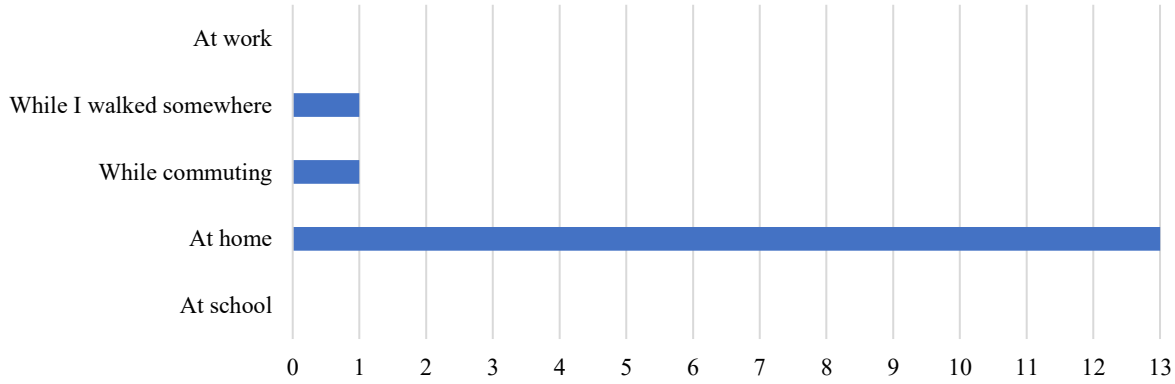
4.3 RQ3 – HOW DID THE PARTICIPANTS ENGAGE IN SELF-REGULATED LEARNING WITH THE MOBILE APPLICATION?

The participants of the current study engaged in self-regulated learning with a mobile application. Under SLR theory, this meant that the participants themselves were responsible for creating and establishing appropriate environments for learning, for planning and controlling the amount of instruction they received, and for regulating and maintaining their motivation. To investigate how they created and established appropriate environments for learning, the participants answered questions about: a) the places where the participants used the application and the role of mobility in their experience with it; and b) the activities they performed while using the application. To investigate how they planned and controlled the amount of instruction they received, data was collected with the application dashboard. Finally, they were asked questions about their motivation and organization motivation to use the application. The collected data in this regard is presented and discussed in the following four subsections, with an answer to RQ3 presented in subsection 4.3.5.

4.3.1 Places where the participants used the application.

Given that one of the main characteristics of mobile learning is the possibility of ubiquitous learning, the participants of this study were allowed to decide where they used the application. Figure 19 shows that all participants used the application at their homes, and two participants reported using the application outdoors.

Figure 19 - Where did the participants use the application?



At first sight, these results may be interpreted as a sign that the participants did not engage in mobile learning. However, as discussed in the review of literature, mobile learning has been defined as any learning that takes place when the learners are not in fixed, predetermined locations (O'MALLEY et al., 2005). In the case of this study, mobile learning took place inasmuch as the participants did not use the application in a predetermined location, that is, they had the freedom to use it anywhere, but still opted to use it at home. O'Malley et al. (2005) also argued that mobile learning happens when learners take advantage of the opportunities offered by mobile technologies. Thus, the participants were asked to provide details on the role of mobility in their experience with the application. In this regard, participants 9 and 10 responded that:

> P9[61]: To me it [the mobility] was important because you can find somewhere where the internet works well in your house. But to use it in other places, you need silence, you need earphones so you can hear, so it is hard to use the app in the bus or during the break. But sometimes it happens that you find some silent time, a quiet place. At

[61] Pra mim a mobilidade foi importante porque você pode encontrar um lugar onde a internet funciona bem na sua casa. Mas para usar em outros lugares, você precisa de silêncio, você precisa de fones de ouvido para poder ouvir, então é difícil usar o app no ônibus ou no intervalo. Mas às vezes acontece de você achar um momento silencioso, um lugar silencioso.

> home, I used the application in my bedroom, which was very quiet, but sometimes I had connection issues.
> P10[62]: I went to my grandfather's house, and I could take the phone and do the activities there, I didn't have to carry all my material there. And at home I felt more relaxed because I could use the phone whenever I wanted, because it was not like a class that I had to go to every day at a specific time.

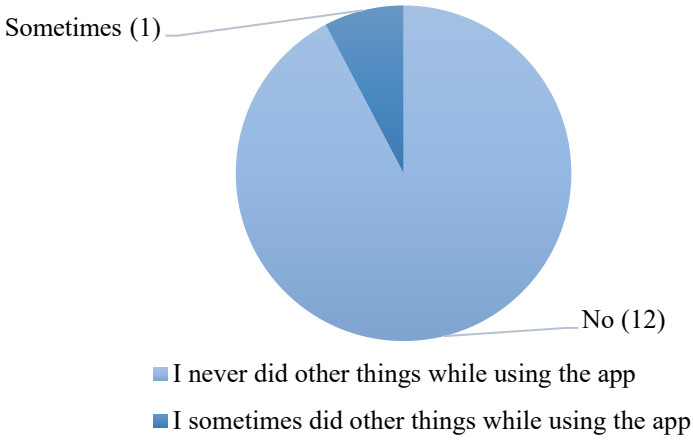
Taking these responses into account, even though the application was predominantly at their homes, the mobility of the devices seems to have helped the participants to find environments they deemed suitable for pronunciation practice, e.g., quiet places or places with good internet connection. What is more, it seems to have allowed pronunciation practice to happen in places where it was less likely to happen with more stationary devices, such as the house of P9's grandfather. In this regard, the possibility of learning in a relaxed and private manner was likely an affordance of the mobility of their smartphones because it allowed them to find places where they felt comfortable to practice pronunciation. As is further discussed in section 4.4.1, practicing pronunciation in a more relaxed manner, without the pressure of other learners, was considered the greatest advantage of learning with the application.

4.3.2 What did the participants do while using the application?

One of the issues with mobile learning is that mobile devices are not solely used for learning purposes (STOCKWELL, 2013; HAAG; BERKING, 2019), which may lead to interruptions that hinder the establishment of an adequate learning environment. In this regard, the participants were asked if they did other activities while using the application. As illustrated in Figure 20, most of them said they did not.

[62] Eu fui para a casa do meu avô, e eu pude levar o celular e fazer as atividades lá, eu não precisei levar todo meu material lá. E em casa eu me senti mais relaxada porque eu pude usar o celular quando eu queria, porque não era uma aula que eu tinha que ir todo dia, em um dia específico.

Figure 20 - Did the participants do other things while using the application?



The fact that none of the participants reported to have accessed social media or other entertainment applications while practicing ELSA Speak suggests that they were not distracted by the social and entertainment options available in their smartphones, and were able to establish an adequate learning environment. However, one participant (P12) reported to have done other activities while using the application. More specifically, P12 stated she did other domestic chores at home. However, in a subsequent question, she revealed that she appreciated the possibility of browsing the web while using the application, suggesting that the use of the application was not solely interrupted by domestic chores. As previously shown, the overall intelligibility of this participant did not improve in the posttests. Although it is not clear whether the task-switching involved in using the application and doing the chores influenced her learning outcomes, this is certainly a possibility, as performance reductions have been observed not only in situations when two tasks are carried out simultaneously, but also in situations that require the repeated shifting between two tasks (WORRINGER et al., 2019).

4.3.3 Participants' control over the instruction

Before receiving access to the application, participants were instructed to complete all the 38 lessons from unit 8 within the period of four weeks, and to complete around 10 lessons each week. These instructions were given to provide participants with a goal for the month (completing the 38 lessons), and a goal for each week (completing around 10 lessons). As stated, being able to control the form and amount of instruction is an important aspect of self-

regulated learning (ZIMMERMAN, 1990), and is also an integral part of mobile learning (SHA et al., 2012). Thus, in this study the participants were able to decide how much they would dedicate to the application on a given day or week, as well as the dates they preferred to use it.

The data from the dashboard reveals that the participants' control over the instruction varied, as the time they spent on each lesson varied, and as many did not complete the assigned units within the stipulated deadline, as can be seen in Table 5.

Table 5 - Use of the application by the participants

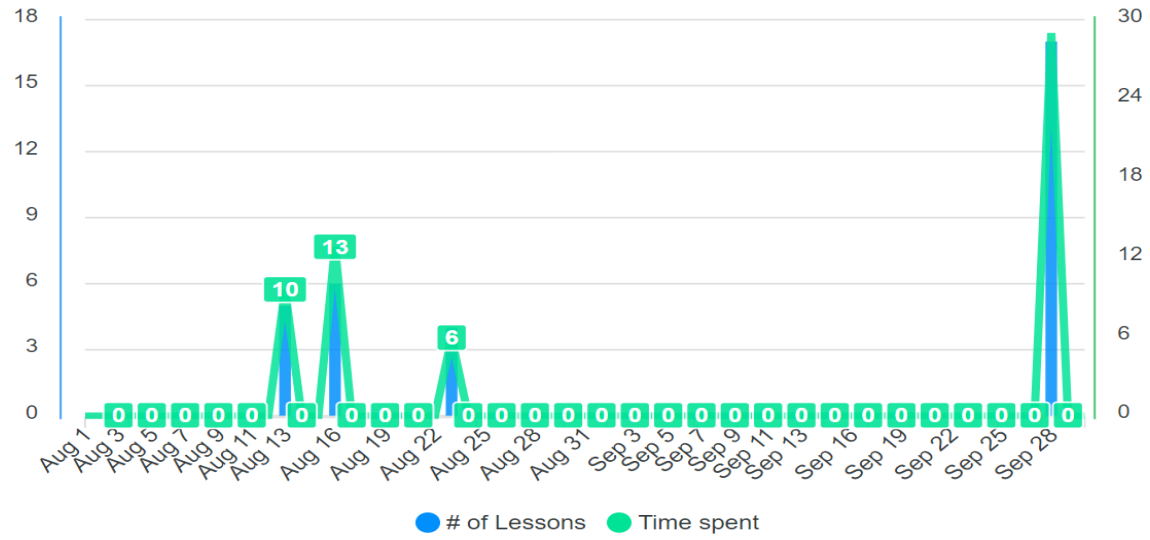
Participant	Minutes per lesson	Complete unit 8 in 1 month?	Completed 10 lessons a week?	Intelligibility gains in the posttest
P1	1.8	No	No	-4
P2	3.2	Yes	Yes	14
P3	2.1	Yes	No	3
P4	2	Yes	No	17
P5	1.6	Yes	No	0
P6	1.9	Yes	No	3
P7	1.6	Yes	No	7
P8	2.2	Yes	Yes	2
P9	1.2	Yes	No	10
P10	1.8	Yes	No	14
P11	3.4	Yes	Yes	10
P12	1.2	No	No	-1
P13	1.4	Yes	No	12

As Table 5 shows, the fastest participants (P9 and P12) took an average of 1.2 minutes to complete one lesson, while the slowest (P11) took 3.4 minutes to do so. Overall, they spent an average of 2 minutes to complete each lesson. A Pearson's R test showed no statistically significant relationship between the time spent on each lesson and intelligibility gains, $r(11) = .29$, $p = .324$. In other words, the time spent a lesson did not seem to be a predictor of intelligibility gains. For example, P9, P10 and P13 spent less than the average 2 minutes taken to complete a single lesson, and still improved their intelligibility (+10, +14 and +12, respectively). This result signals that, at least in the context of MALL, learning can take place within short time frames, which fits the notion that MALL lessons should be short and concise, with segmented sections that can be quickly completed independently as a single unit

(STOCKWELL, 2013; TRAXLER, 2007). However, this does not mean that MALL lessons must be completed in a rush. For instance, P2, who spent more time on each lesson than the other participants, also had improved intelligibility (+14). These findings signal that many participants were able to use the application at a pace that was suitable for them.

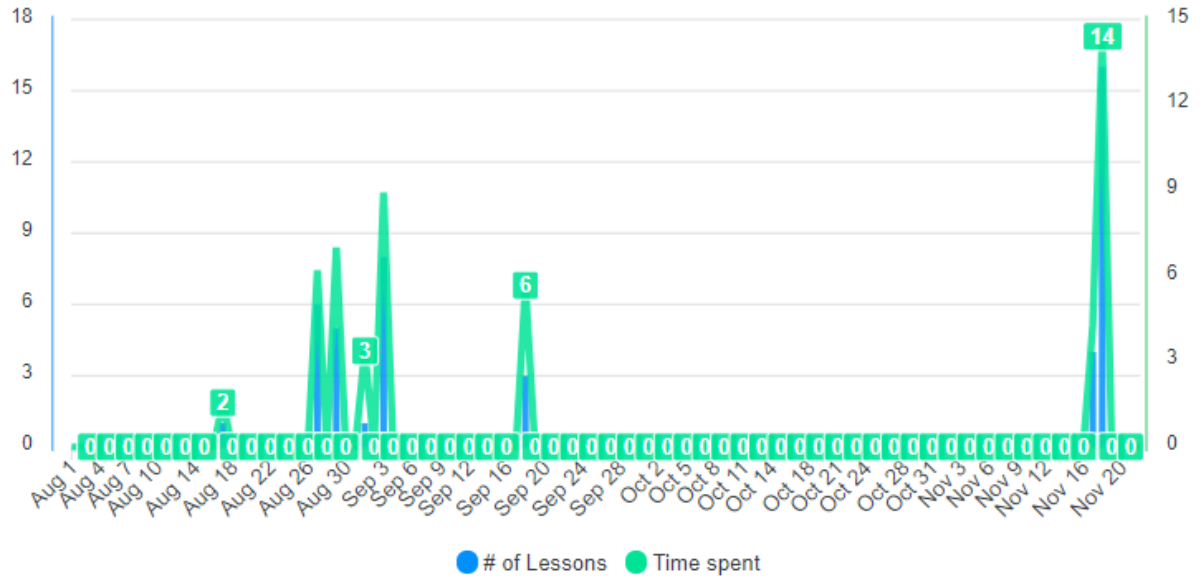
As was shown in Table 5, while 11 participants completed the 38 lessons from unit 8 in the stipulated period of a month, only three followed the instruction to complete around ten lessons weekly. Two (P1 and P12) did not follow either of these instructions. Figures 21 and 22 show the dates in which these participants accessed the application and the number of lessons they completed in these days.

Figure 21 - Use of the application by P1



Source: Dashboard from ELSA Speak

Figure 22 - Use of the application by P12



Source: Dashboard from ELSA Speak

As Figure 21 shows, P1 completed 18 lessons from August 13th to August 25th, but then spent more than a month without using the application. In similar fashion, P12 began using ELSA Speak on August 17th, completed 23 of the 38 lessons until September 18th and then spent two months without using the application. Having noticed the absence of these participants, I contacted them via Moodle and asked if they were still interested in participating in the study. They said they were, but they had been really busy with school assignments. They asked if they could still use the application, to which I answered that they could. I showed them the lessons they still needed to complete, and reminded them that, if they had any doubts, they could contact me. These participants then proceeded to complete the remaining lessons. However, it is likely that, by the time they took the intelligibility posttest, they no longer remembered what they had studied in the first lessons. As Elis (2012) points out, learning and memory are affected by recency, that is, we tend to remember recent experiences more easily and more fluently. Since many weeks had passed since the first week in which they used the application, the lack of recency may have affected their performance in the posttest.

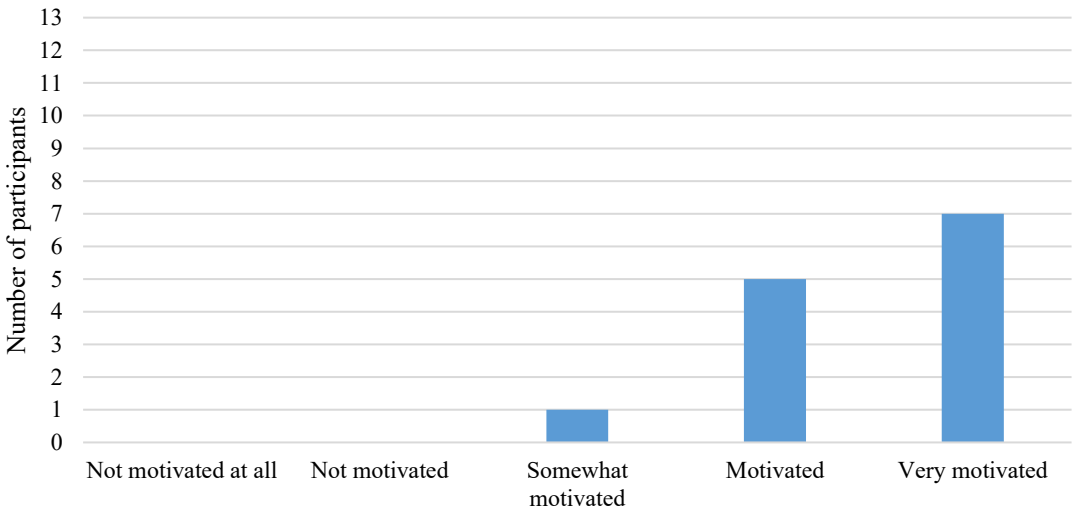
The usage pattern in Figures 21 and 22 shows that participants 1 and 12 were not able to finish the assigned lessons within the stipulated period of one month and suggests that they struggled to regulate their learning process and to maintain their motivation. In self-regulated learning, students should be able to self-monitor their learning process, which involves the

observation of one's progress on constant basis (SCHUNK, 1997). The fact that they needed to be reminded to finish the lessons signals that they were not monitoring their progress. More than the inability to self-monitor, this may be the result of insufficient motivation, as is discussed in the following subsection.

4.3.4 Participants' motivation to use the application

Using a 5-point Likert scale, the participants answered the question "how motivated were you to use the application at the beginning of the study". Their answers are shown in Figure 23.

Figure 23 - Motivation to use the application at the beginning of the study

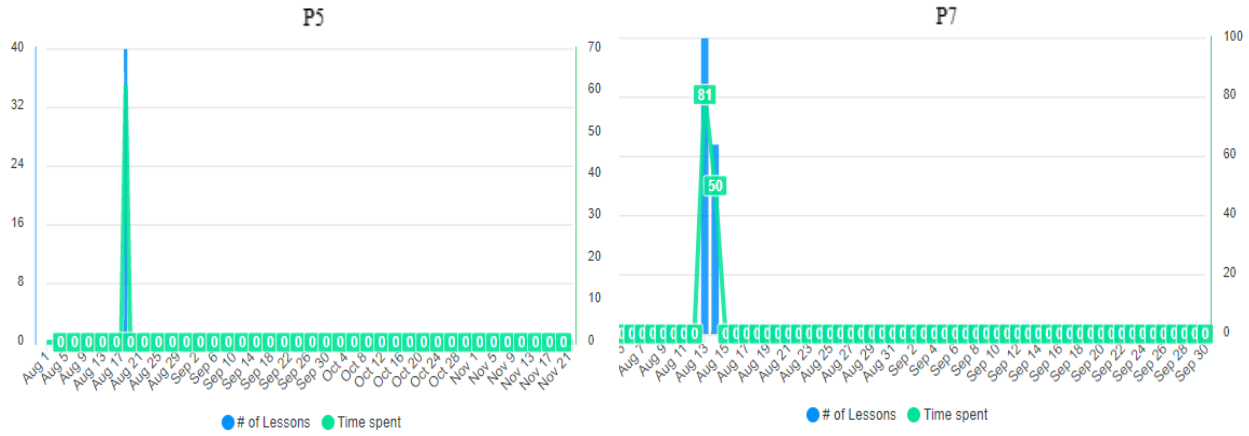


As can be seen in Figure 23, five participants felt motivated and six felt very motivated to use the application at the beginning of the study, that is, they had high entering motivation. Garrison (1997) argued that learners tend to have more entering motivation when they believe that the learning goals are of their interest, and when they think there is a good chance of achieving them. As discussed in section 3.4.1, this is likely the case of participants of the current study, who stated that they were very interested in studying and practicing pronunciation, and believed this is a very important skill to be developed in an L2.

As explained, the participants were asked to complete 10 lessons weekly and to finish all 38 lessons within a month, but had the liberty to complete the lessons at their own pace. As Figure 24 shows, P5 finished all assigned lessons on a single day, and P7 completed them in

only two days. The fact that these participants completed all assigned lessons in such short time highlights how motivated they initially were to use the application.

Figure 24 - Use of the application by P5 and P7.



Source: Dashboard from ELSA Speak

After completing the assigned lessons, the participants could continue using the application as they wished until their three-month ELSA Speak licenses expired. Six participants did this, with participants 3, 7, 8 and 13 completing each at least 30 extra lessons on topics of their choice.

Participants were also asked if they maintained their motivation during the use of the application, to which four participants (P1, P9, P12 and P13) did not. This indicates that the high levels of entering motivation were not sustained throughout the use of the application. Even those who completed more lessons than what was required did not maintain a regular use of ELSA Speak until their licenses expired. Garrisson (1997) points out that, in order to sustain motivation, students must be active learners, and that if they are not active in the learning process, motivation tends to decay. In the case of participants 1 and 12, it is likely that, having spent many days without accessing the application, they lost the motivation they initially had. Similar results were found in Botero et. al (2019) and Loewen et al. (2019), in which university students struggled to maintain their use of MALL in a context that was not directly linked to their formal learning obligations.

A different scenario was portrayed in Derwing et al. (2014), in which the participants, who were adult factory workers, spent more time on their homework assignments than was

required, and were highly motivated because of the demands of their jobs, which is deemed to be a type of extrinsic motivation. In the current study, there were no extrinsic factors, such as course assignments or extra points in their average grade, to motivate them to use the application appropriately. Rather, their motivation had to be mostly intrinsic, that is, they needed believe that the use of the application was personally engaging, even though it would not bring them any immediate, tangible benefits.

Studies such as Botero et al. (2019), Ushioda (2013) and Rodrigues et al. (2019) have observed that, when digital technologies are involved, learners are initially motivated by the possibility of exploring new technologies, rather than by their will to learn the instructed content. However, once this technology is no longer novel, motivation is likely to fade, which has been referred to as a novelty effect. In the case of the current study, it is likely that using their smartphones to practice pronunciation with an ASR-based application represented a novelty for them and the possibility of exploring a paid application without any costs may have been initial motivating factors. However, as discussed in Ushioda (2013), this initial technology-driven engagement tends to be superficial rather deep, and it still debatable whether mobile technologies can, on their own, promote and sustain deeper levels of engagement and motivation.

In 2021, all classes and projects from Colégio de Aplicação took place remotely, which meant that the students from the school spent several hours on their computers and/or smartphones. The fact that the participation in the current study involved additional time using a digital device may have led to a decline in their motivation, especially when many students already believed to be spending excessive time with their school assignments. In this regard, participants six and seven commented that:

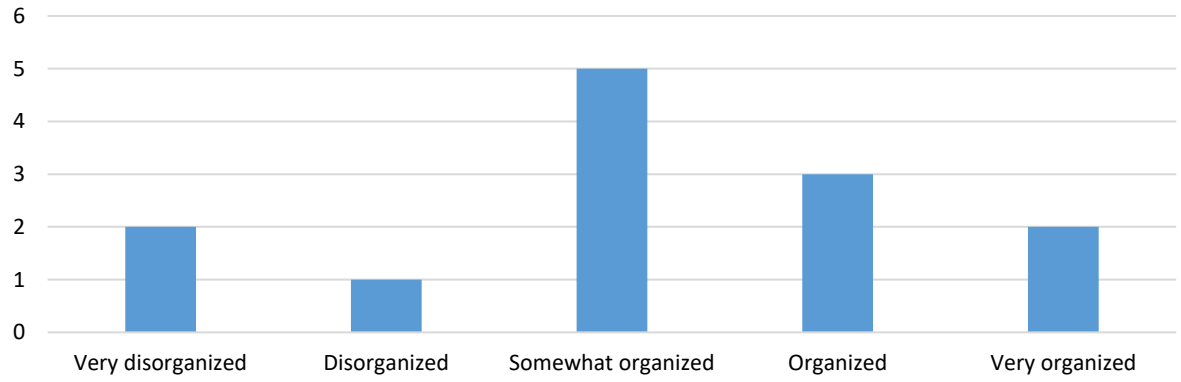
> P6: As I was exposed to a computer screen for hours and hours because of the school requirements, I had to stop using the application for a few days, because I was so tired. I think that affected my organization to participate in the research.[63]
> P7: I think we got so attached to the screens we ended up forgetting to do what we really had to do.[64]

As noted by P6, it is possible that the lack of motivation expressed by some participants resulted in a disorganized use of ELSA Speak. As Figure 25 shows, their answers ranged from

[63] Por ficar horas exposta a uma tela de computador para estudar deixei de lado o aplicativo por alguns dias, em virtude do cansaço. O que afetou minha organização nos meus estudos para este projeto.
[64] Acho que a gente acaba ficando tão preso as telas que acaba se esquecendo do que realmente devemos fazer.

very disorganized to very organized, with most of them believing to have had a somewhat organized use of the application.

Figure 25 – Participants perceived organization to use the application



As previously stated, participants were told that an organized user was someone who used the application in appropriate times and places, read and followed the provided instructions, and established weekly/daily goals to complete the lessons. Thus, the 'somewhat organized' users have likely followed some, but not all of the instructions provided, while the 'very organized' ones believed to have followed all of them. Three participants stated that their use of the application was disorganized, namely P1, P9 and P12. These responses show that they were aware that their use of the application was not optimal in terms of organization. This is further evidenced considering that these three participants had the perception[65] that their lack of organization and dedication was the greatest disadvantage of mobile learning.

Overall, a process seems to have occurred in which the majority of the participants were initially motivated to learn with the application, as it represented something novel to them. However, not all were able to sustain their motivation, and those with low task motivation struggled to complete the assigned lessons within the stipulated time frame, i.e., did not have adequate control over the amount of instruction they received. Having spent many days without practicing with the application, P1 and P12 had diminished results in the intelligibility posttests.

As will be discussed in section 4.4, the difficulties some of participants faced in their self-regulated use of the application did not seem to be related to the design of application, given that their perceptions of it were predominantly positive.

[65] These and other perceptions are presented and discussed in the following section (section 4.4).

4.3.5 Summary of findings for RQ3.

In light of the literature in SRL, self-regulated learners should be able to: a) establish optimal learning environments; b) plan and control the form and amount of instruction they need; and c) regulate and maintain their motivation (AN et al, 2021; SHA et al, 2012; ZIMMERMAN, 1990). Based on the results presented in subsections 4.3.1 and 4.3.2, the participants managed to establish adequate learning environments, and explored the mobility of their devices to practice pronunciation in places where they felt comfortable doing so. Moreover, the fact that only one participant reported using the application while doing other activities suggests that most of them focused on the use the application. Although many participants were able to plan and control the use of the application so as to conclude all of the assigned lessons within the stipulated time, two participants spent more than a month without accessing ELSA Speak, which indicates they did not manage to control the amount of instruction, as is necessary in SRL. Finally, some participants struggled to sustain their motivation throughout the use of the application, even though they initially reported to be highly motivated. All things considered, most of the participants of this study were able to engage in self-regulated learning with a mobile application, but faced difficulties to plan and control its use, and to maintain their motivation to use it.

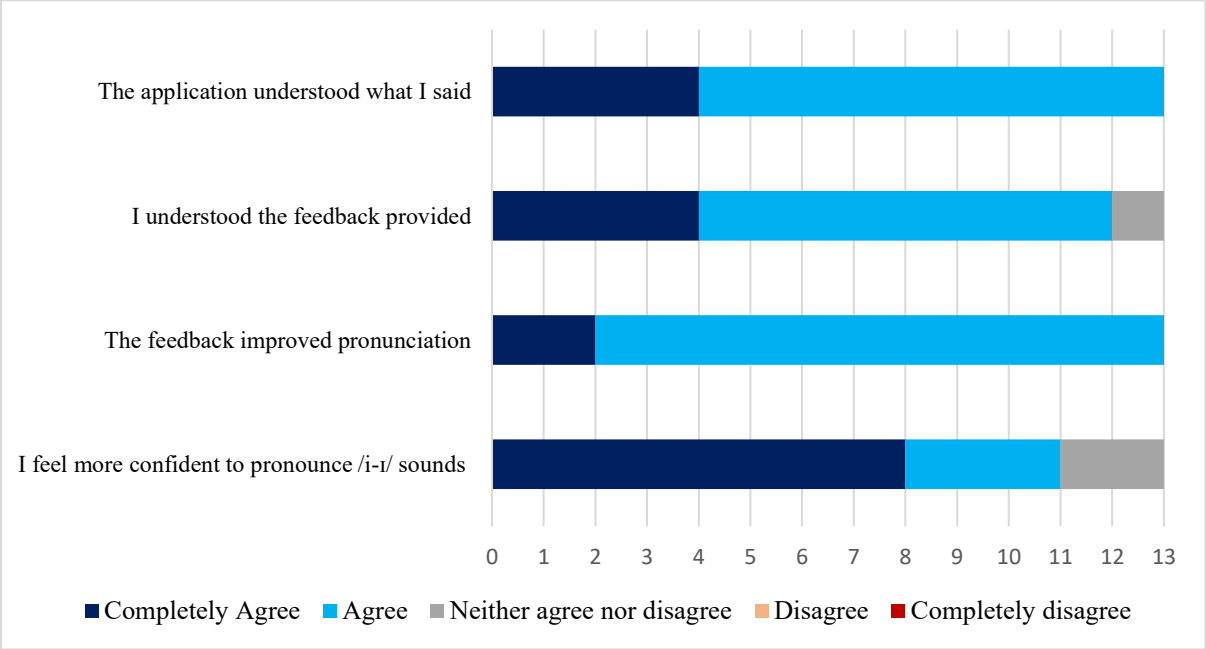
4.4 RQ4 - WHAT WERE THE PARTICIPANTS' PERCEPTIONS OF THE APPLICATION AND OF MOBILE LEARNING?

After completing unit 8 from ELSA Speak, the participants shared their perceptions of the application and mobile learning on a questionnaire with close and open-ended questions. More specifically, they shared their perceptions on the application's ASR system, on the feedback it provided, and on the advantages and disadvantages of learning pronunciation with an ASR-based mobile application.

4.4.1 Perceptions of the ELSA Speak's ASR and feedback

Using five-point Likert Scales ranging from 'completely agree' to 'completely disagree', the participants expressed whether they agreed or disagreed with four statements regarding the application's ASR system and the feedback. As Figure 26 shows, none of them chose the options 'disagree' or 'completely disagree', which in this case indicates that they had overall positive perceptions over the speech recognition and the feedback of ELSA Speak.

Figure 26 - Participants' perceptions of ELSA Speak



Four participants strongly agreed and seven agreed with the statement 'the application understood what I said', which reveals that the participants thought the ASR system was able to understand the words they pronounced. This is an indication that speech recognition technologies have improved and can now reliably recognize non-native speech, which was not the case in the late 1990s and early 2000s, as reported in Byrne et al. (1998) and Derwing, Munro, Carbonaro (2000). Ehsani and Knodt (1998) believed that for ASR systems to recognize non-native speech, they needed to collect appropriate speech data and train with non-native models. In this regard, the developers of ELSA Speak claim that their ASR system has users from more than a hundred countries, and uses artificial intelligence to train "with a diverse set of non-native accent data" (ELSA SPEAK, 2022). Thus, the improvements in ASR can be partially attributed to the use of artificial intelligences, which, with enough speech input, can

learn[66] the pronunciation patterns of non-native speech and feed the data to ASR systems. As a result, ELSA Speak may be more apt to recognize and provide feedback to L2 learners in comparison to applications that do not train with non-native speech. For example, although the participants of Botero et al. (2019) appreciated learning vocabulary with Duolingo, they reported issues with its ASR, which may have provided positive feedback to words that had been mispronounced.

Even though all participants believed the application was able to recognize their speech, six of them reported instances when they thought to have pronounced a given word or sentence correctly but were receiving negative feedback. These participants were asked to share details about these instances and what words caused them trouble. Their answers are transcribed below:

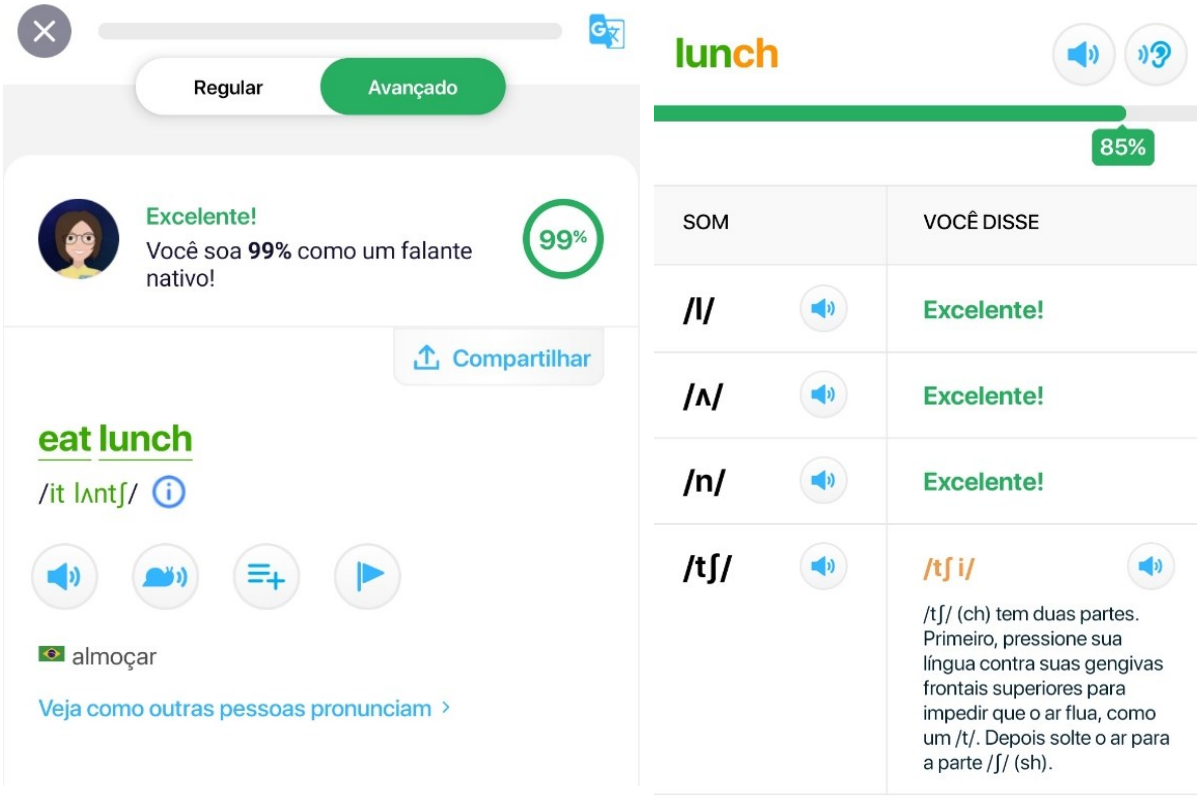
> P2: Sometimes the app asked me to say 'lunch'. I listened to the word and then I repeated exactly what I heard, but still it told me I was mispronouncing it.[67]
> P5: Sometimes I pronounced [a word] in a certain way, and it said it was wrong, and when I tried to correct, but ended up pronouncing the same thing, and it said it was correct.[68]
> P7: I tried to pronounce the word 'listeners' about 30 times, and the app still told me pronunciation was incorrect, so I tried to pronounce it in many different ways, but it still would not accept it. [69]
> P9: Sometimes I had the impression that I had to change my intonation, like, change my voice, so that it counted a word correctly.[70]
> P10: If I remember it, I pronounced the same sentence dozens of times and the app told me it was wrong, so I tried to repeat exactly as the app was pronouncing, but it didn't work. By then I was already angry and then I pronounced the sentence without caring much. And it worked…[71]
> P13: It happened so many times I forgot what the words were.[72]

Because it is not possible to hear what the participants actually pronounced in these instances, it is unclear whether their pronunciation was target-like and the application failed to

[66] In general terms, current artificial intelligences learn with help of algorithms that contain few rules and ingest training data to learn by trial and error (HUTSON, 2018). The exact way in which ELSA Speak employs AI is not part of the scope of this study.
[67] Algumas vezes o app me pedia para dizer "lunch", eu escutava a palavra e então repetia exatamente o que eu ouvia, mas mesmo assim o app dizia que eu estava pronunciando errado.
[68] Às vezes eu pronunciava [uma palavra] de um jeito e dava como errado, e quando ia corrigir acabava pronunciando do mesmo jeito mas ia como certo.
[69] Tentei falar a palavra "listeners" umas 30 vezes, mesmo acertando a pronúncia o app dizia que tava errado, aí comecei a tentar de várias maneiras diferentes e mesmo assim não aceitava
[70] As vezes dava a impressão que eu tinha que mudar a entonação, tipo mudar a voz, pra ele contar como correto
[71] Se eu estou lembrado, eu respondi uma mesma frase dezenas de vezes e o aplicativo dava como errado, dai eu tentei responder copiando exatamente o app e deu errado, naquele momento eu estava com raiva já e falei de uma vez a frase, sem pensar muito. E deu certo...
[72] Aconteceu tantas vezes que eu me esqueci quais palavras.

recognize it, or if they actually mispronounced the words, but believed they did not. In the case reported by P2, it is unlikely that he/she pronounced the word *lunch* as /lʌnʧ/, because the application did not give any negative feedback when I pronounced the word as such, as Figure 26 shows. However, when the word is mispronounced as /lʌnʧɪ/, that is, with the insertion of the vowel /ɪ/, the application quickly recognized this mistake and gave appropriate feedback, as also shown in Figure 27.

Figure 27 - Feedback on the word 'lunch'



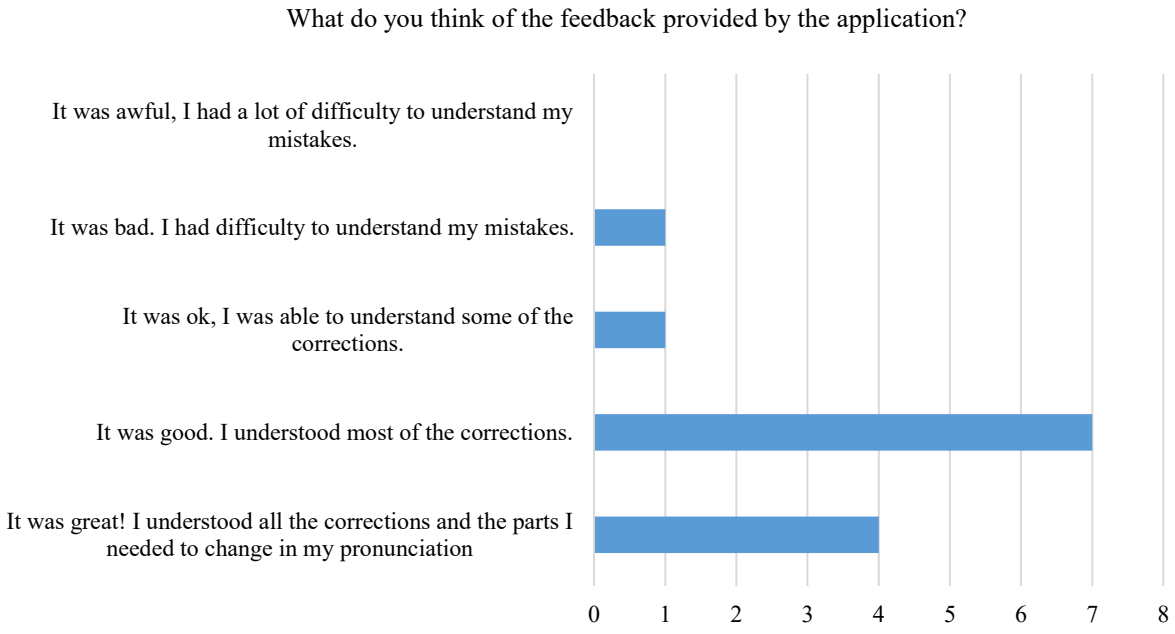
Although the application was able to recognize my pronunciation of the word 'lunch', there were instances when what I believed to be perfectly intelligible words had to be pronounced several times to receive a good score from the application. For example, the schwa sound (/ə/) in the words 'academy' and 'difficult' had to be pronounced very subtly, in an almost unhearable manner[73], while aspirated consonants such as /k/ in *country* required native-like aspiration, which may have led to some of the difficulties reported by the participants. Even

[73] This happened when I first analyzed the application in 2020. It is relevant to point out that, when using the application in 2022, I realized that the word 'difficult' was easily recognized, which signals that the training with non-native data and the use of artificial intelligences may have improved the recognition of non-native speech.

though the developers claim that ELSA Speak “accurately recognizes and gives feedback to non-native speakers” as its artificial intelligence “trains with a diverse set of non-native accent data” (ESLA API, 2022), there were instances in which intelligible speech received negative feedback. This does not mean that the ASR is faulty, as it was able to recognize and provide feedback appropriately. Rather, it signals that the application may not have been designed with the intelligibility principle in mind, something that is evidenced in the feedback that says ‘you sound 99 percent like a native speaker’, as can be seen on the top-right corner of Figure 24. In other words, ELSA Speak does not seek to tell learners if a word is intelligible or not, but attempts to recognize their speech, score it and provide feedback based on its speech models, which are claimed to be based on the speech of English speakers with various accents (ELSA SPEAK, 2022).

As was shown in Figure 27, four participants completely agreed and eight agreed with the statement “I understood the feedback provided by the application”. To learn more about their perceptions of the feedback, they were asked to choose one of five options to answer the question “what did you think of the feedback provided by the application?”. Figure 28 shows that their perceptions in this regard were overwhelmingly positive.

Figure 28 - Perceptions on the feedback



As Neri, Cucchiarini and Strik (2002) point out, feedback is an essential part of ASR-based pronunciation instruction, and it should be relevant and easy to interpret. As previously shown, ELSA Speak uses simple verbal and visual clues to tell users what aspects it deemed to be incorrectly pronounced, and how their pronunciation can be improved, which helped the participants of this study to understand the feedback received. This became evident in the answers some participants wrote:

> P5: I liked the different colors showed as I pronounced the words, sometimes the "t" in "bit" ended up disappearing in my pronunciation, and it showed these sounds in yellow or red colors for that letter. This made me more aware and helped me pay more attention.[74]
> P7: I really liked how the application explained exactly what my mistakes were in the pronunciation of a word, in an informative and comprehensible way.[75]

The participants also agreed that the feedback received helped improve their pronunciation, which means that not only did they understand the feedback, but they were able to use this information to enhance their pronunciation, as P11 highlighted.

> P11: I think what I enjoyed the most about the application was that I was able to see what the pronunciation of each letter and word should be, and with this I was able to improve my pronunciation, because I knew which part of the word was incorrect.[76]

In spite of predominantly positive perceptions, one participant (P13) believed the feedback was only 'OK', and another (P2) stated it was difficult to understand the mistakes she made. As previously stated, P2 reported issues with the pronunciation of the word 'lunch', which she believed to be pronouncing exactly as the application did. Moreover, in a subsequent question, this participant wrote that the application should be able to tell exactly where the mistakes are when it identifies incorrect pronunciation. If her pronunciation was in fact incorrect, the application may not have provided feedback accordingly. Another possibility is that feedback was accurate, but the participant was not able to alter her pronunciation based on it.

---

[74] Gostei das diferenças de cores conforme eu ia pronunciando, as vezes o "t" de "bit" acabava sumindo na pronúncia, e aí ele dava uma cor amarela ou vermelha para aquela letra. isso me fez ficar mais ciente e prestar mais atenção na pronúncia.

[75] Gostei muito de como o aplicativo explica exatamente o que estou errando na pronúncia da palavra, de uma forma didática e compreensível.

[76] Acho que o que mais gostei no aplicativo e que conseguia ver como era a pronúncia de cada letra, palavra e assim conseguir melhorar a pronúncia, pois sabia em que parte da palavra estava errando.

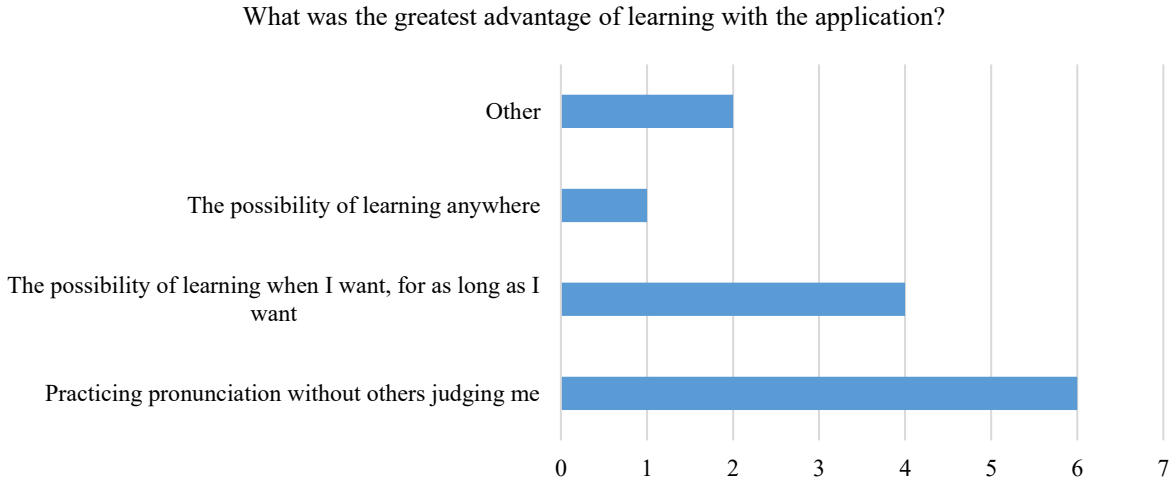
It is worth noting that even though P2 and P13 reported issues with the speech recognition and the feedback, they were among the participants with the highest intelligibility gains. On the other hand, P1 and P12 did not report any recognition problems and had positive perceptions on the feedback, but few gains in terms of intelligibility. This suggests that the reported issues with the application did not prevent learners from improving their pronunciation, and that intelligibility gains are unlikely to be a result of the design of the application or of the participants' perceptions of it, but were likely a consequence of their lack of motivation and of their inconsistent use of the application, as was discussed in section 4.3.3.

Finally, the participants were asked if they agree with the statement "now that I have completed unit 8, I feel more confident to pronounce words with the target /i - ɪ/ sounds". As can be seen in Figure 23, eight of them completely agreed, 3 agreed, and 2 participants did not agree nor disagree. These answers reveal that they believed the application helped them develop the pronunciation of this specific pair of vowels, which are often quite challenging for Brazilian English learners to pronounce.

4.4.2 Perceived advantages and disadvantages of learning pronunciation with an ASR-based mobile application.

Four questions were asked to gather data on what the participants perceived as advantages and disadvantages of learning with a mobile application such as ELSA Speak. As shown in Figure 29, for six of the 13 participants, the greatest advantage of learning pronunciation with a mobile application was the possibility of practicing it in a more relaxed manner and without feeling judged by others. Four participants stated the greatest advantage was the possibility of choosing when to learn, while only one said it was the possibility of learning anywhere.

Figure 29 - Advantages of learning with the application



As Figure 29 shows, two participants chose the option 'other', and wrote their own answers. One wrote that the greatest advantage were the feedbacks and the videos the application provided, while the other wrote that it was a combination of the three first options, namely, the possibility of practicing pronunciation without feeling judged, the possibility of learning of choosing when to learn, and the possibility of learning anywhere.

The perception that practicing without the feeling of being judged was an important advantage corroborates the findings of Aulia (2018) and Bashori et al. (2021), in which mobile technologies helped reduce the anxiety many students feel when speaking in front of other colleagues, and of Liakin, Cardoso, Liakina (2015), in which participants reported feeling more comfortable and less nervous when practicing pronunciation with no one else around. As stated, L2 courses in basic schools such as Colégio de Aplicação tend to have students with various[77] proficiency levels, and students who perceive their level as low frequently refrain from speaking in the target language as they fear they will be ridiculed or judged by more proficient learners. The high number of responses for this option may be a consequence of such feeling. Moreover, there are students who are inherently shy, and do not like to speak in front of others, even in their own L1.

The responses shown in Figure 28 indicate that being able to control when and how long to study was more valuable than being able to learn at any place, and seem to reflect their actual use of ELSA Speak, which varied immensely in terms of the time spent on each lesson

[77] This was also evidenced in the proficiency tests taken for this study, whose results ranged from starter (A0) to upper intermediate (B2) levels in the CEFR.

and the time taken to complete the assigned lessons, but predominantly took place at a single place i.e., their homes. It is possible that the context in which the study took place affected their use and perceptions of mobile learning. Because of pandemic, the participants studied from their homes and did not commute to school or went to a variety of places where short amounts of time and space could be exploited for learning (TRAXLER, 2007). Although only one participant believed that the possibility of learning anywhere as the greatest advantage, this does not mean that mobility did not play a role in their use of the application. As previously discussed, it seems that mobility allowed them to find suitable places to study inside their homes, even though it was not explored to learn in places other than their homes.

To gain more insight into the positive aspects of the application, the participants were asked to write the things they enjoyed the most about ELSA Speak in an open-ended question. Many of their answers showed appreciation for the application's ASR system and the quality of the feedback provided, as exemplified in the answers of six participants:

> P4: I really liked that it was able to recognize which words were wrong and which ones were right, and that really helps with the feedback.
> P5: I liked the different colors showed as I pronounced the words, sometimes the "t" in "bit" ended up disappearing in my pronunciation, and it showed these sounds in yellow or red colors for that letter. This made me more aware and helped me pay more attention
> P7: I liked how it explained exactly what my pronunciation mistakes were, in a didactic and comprehensible way.
> P9: I liked that it always gives you tips on how to pronounce the words, like "i" being a shorter vowel or something like that.
> P11: I think what I liked the most about the app was that I was able to see what the pronunciation of each letter and word should be, and with this I was able to improve my pronunciation, because I knew which part of the word was incorrect
> P12: The corrections, which were always detailed and spot on.

These answers corroborate the results presented earlier (Figures 23 and 25), in that the participants believed the application was able to understand their speech, and that they were able to understand the feedback it provided. As these five participants noticed, although ELSA Speak offered feedback in great detail, its corrections were objective and used a color scheme that made it easy to understand what sounds needed improvement. This can be considered an improvement over previous generations of ASR-based systems, which once required learners to compare their waveforms to those of native speakers (NERI, CUCCHIARINI; STRIK, 2003), or offered much simpler forms of feedback, as the yes or no responses from the system used in Burleson (2007). Another key difference is that ELSA Speak was able to provide feedback for target and non-target sounds, while systems such as the one in Burleson (2007)

could only do so for the target words, and were restricted to word level corrections. What is more, the application used in the current study not only signaled the sounds it recognized as incorrect, but also provided instructions on how these sounds could be improved.

Still concerning the aspects of the application perceived positively by the participants, there were comments that showed appreciation for its instructional design and content, as the following answers reveal:

> P1: I really liked the possibility of browsing while learning new words.[78]
> P3: I thought the application went straight to the point, and the materials are easy to understand, and it seems to have a lot of content. I also loved the mini videos it provided.[79]
> P6: I liked the progressive difficulty of the activities and the videos with people pronouncing the sentences on display.[80]
> P8: I enjoyed everything! I liked the types of exercises (…) the ones where I had to identify the changes in sound (…) I also loved the exercises that were like a dialogue. The videos also made a lot of things clear. I thought it is an incredible app, it seems like they thought about everything! It was really good to be able see the translation of the sentences.[81]
> P10: The practicality.[82]
> P13: The pronunciation test videos, I learned very fast how to pronounce with them.[83]

Three participants (P3, P6 and P13) stated they enjoyed learning with the videos that could be accessed by tapping on the instructed words. As stated, these are YouTube videos that are found using YouGlish.com, a tool that looks for specific parts of videos where the searched words are pronounced in various contexts. Because there are several videos that can be watched for each word, users get to hear pronunciations from different places around the world and can see the situations in which they are used, such as in movie scenes, documentaries, speeches, interviews and lectures. Encountering L2 items in multiple contexts is an important step to learn new words and their pronunciations (CARLO et al., 2008; TAKAČ, 2008). What is more, they can find videos about topics of their interest, which should help with their engagement and enjoyment while learning to pronounce the target words.

---

[78] Gostei muito de ter a possibilidade de navegar enquanto aprendia as palavras novas.
[79] Achei ele bem direto ao ponto, os materiais disponibilizados são de fácil entendimento e aparenta ter bastante conteúdo nele. Eu também amei os mini vídeos que ele disponibilizava.
[80] Eu gostei do aumento progressivo da dificuldade das atividades e do acervo de vídeos de pessoas pronunciando a sentença apresentada.
[81] Gostei de tudo! Gostei dos tipos de exercícios (...) aqueles em que eu tinha que identificar a mudança do som de acordo com a palavra (...) também gostei muito dos exercícios que vinha em forma de diálogo. Os vídeos também esclareceram bastante. Achei incrível o app, parece que eles pensam em tudo. Foi muito bom conseguir ver as traduções das frases.
[82] A praticidade.
[83] Os vídeos de pronounciation test, aprendi bem rápido como pronunciar com eles.

Three participants showed appreciation for the ‘can you hear the difference’ lessons. These were lessons in which the application pronounced one word from a /i - ɪ/ pair and then asked learners to identify the one that was pronounced.

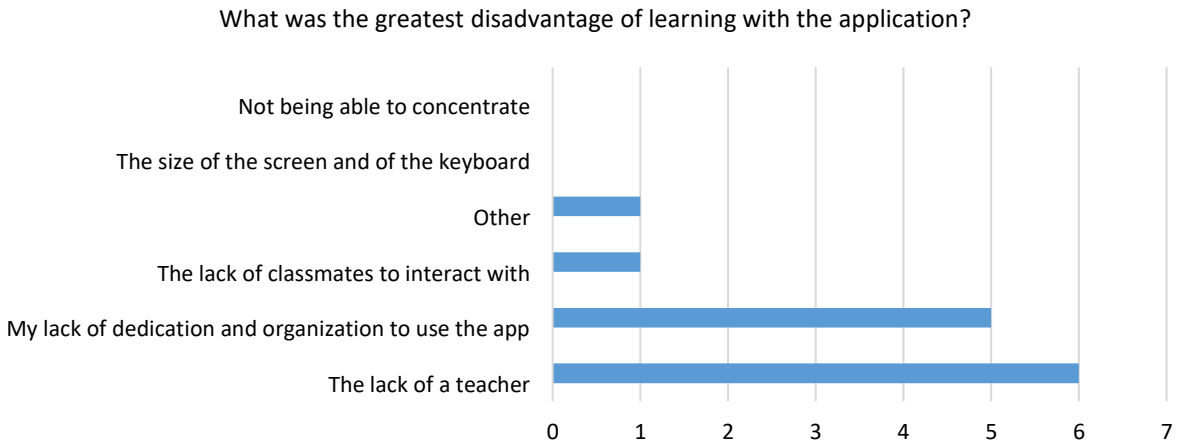
> P2: I liked that it showed you words with similar sounds for you to distinguish and choose the one that was pronounced (P2).
> P3: (…) I liked the types of exercises, especially those where I had to identify the sound changes according to the word, I felt that this type of exercise helped me a lot).
> P7: (…) I loved the activities of recognizing the word by their pronunciation, they were my favorite and they helped me a great deal in understanding the differences between the pronunciations[84]

These lessons were appropriately placed at the beginning of each level, that is, they were a source for the presentation of word pronunciation and meaning, and did not require learners to pronounce anything. Indeed, the instructional design of the application is organized so that learners first practice in a more controlled manner, then gradually move towards activities that are less controlled and require more active participation, such as the conversation activities in which learners participate in a dialogue with the application. This instructional design did not go unnoticed by P6, who welcomed the “progressive difficulty of the activities”. Although such progressivity may be a basic tenet in the instruction of second languages, its incorporation has not been a commonplace in MALL (BURSTON, 2014; PIRES, 2018). These findings indicate that the instructional design of ELSA Speak is pedagogically sound, and validate the high scores Baldissera (2020) attributed to it in her analysis of applications for developing English pronunciation.

When asked what the greatest disadvantage of learning pronunciation with a mobile application was, six participants said it was the lack of a teacher, while five believed it was their inability to use the application with more dedication and organization, as Figure 30 displays. One student mentioned missing classmates to interact with, and one used the ‘other’ option to write that she could not think of any disadvantage.

[84] Eu adorei as atividades de reconhecer a palavra pela sua pronúncia, foram as minhas preferidas e me ajudaram bastante a entender as diferenças entre as pronúncias.

Figure 30 - Disadvantages of learning with the application



The fact that nobody chose the option ‘the size of the screen and of the keyboard’ shows that this is not a major issue anymore, as was reported in earlier MALL studies (CHAE; KIM, 2004; CHINNERY, 2006) when screens and keyboards were smaller and often unreliable. Likewise, none of the participants chose the option ‘not being able to concentrate’. This is coherent with the finding that only one participant reported doing other activities while using the application, and reinforces that, at least in the current study, the lack of concentration was not an issue. The mobility of their smartphones may have been a contributing factor in this regard, as it allowed them to use the application in places that favored concentration inside their own homes.

Conversely, the lack of dedication and organization seems to have affected their mobile learning experience. This was revealed not only by their answers in the questionnaire, but also by the way they used the application. As previously discussed, some participants used the application with no regularity, spent many days without accessing it, and probably lacked the self-regulated skills necessary in this type of learning, such as maintaining motivation and regulating the amount of instruction. It is possible that these issues have also led many participants to choose the option ‘the lack of a teacher’ as the greatest disadvantage of learning with the application. In their regular classes, students from elementary and middle school maintain regular contact with teachers, who not only help with questions about the content, but also remind them of the assignments and deadlines, and motivate them to achieve their learning goals. In similar fashion, the participants’ from Botero et al. (2019) reported that they would be more motivated if the activities from the application they used were graded by a teacher, or if

teachers were checking their progress. Their participants also reported that because the activities from Duolingo were not assessed by a teacher and did add to the grades of their regular classes, using the application was simply not a priority. Likewise, in the current study, P6 said she had to stop using the application for a few days because of her formal school requirements. Thus, learners seem to miss the presence of teachers not only because of the instructional and organizational support they provide, but also because they represent a link to the requirements of their formal instruction.

When asked to write about the things they did not like about the application, four participants stated that sometimes it struggled to recognize the words they pronounced:

> P2: I didn't like that many times I pronounced the same word the application said, and it told me I said the word wrong.[85]
> P8: There were moments when I pronounced things and the app said it was right, but as I had not understood what I had changed in my pronunciation, I did it again. In some of these moments, when I repeated the word that I'd already gotten right, the application said it was incorrect, so I did them until I got them right once more. However, I did not always understand what I'd done right or wrong. I believe this is a matter of practicing, not of the app itself.[86]
> P10: Sometimes it (the application) did not understand I said.[87]
> P12: I think I liked almost everything in the application, the only problem was when it did not hear or understood what I was pronouncing, and I had to repeat a couple of times.[88]

There are a couple of factors that may have caused their reactions to the application. First, their pronunciation may actually have been incorrect, in which case the application did not fail to recognize their speech and provided appropriate feedback. Second, it is possible that its ASR system was not able to recognize intelligible speech because it did not perfectly match its models. However, this would contradict the claim that ELSA "was developed using voice data of people speaking English with various accents" which "allows ELSA to recognize the speech patterns of non-native speakers, setting it apart from most other voice recognition technologies" (ELSA, 2022). Finally, it might have been the case that the quality of the acoustic signals was impaired i.e., sounds that were not speech signals were captured and impeded

[85] Não gostei que na maioria das vezes que eu pronunciava a mesma palavra que o app dizia, ele colocava que eu falei a palavra errado.
[86] Às vezes eu pronunciava e o app dava como certo, mas como eu não tinha entendido o que tinha mudado na minha pronúncia eu fazia de novo. Algumas dessas vezes, quando eu repetia a palavra que eu já tinha acertado o app dava como errado, então eu fazia elas até acertar novamente. Porém nem sempre entendia o meu erro ou acerto. Mas acredito que isso seja uma questão de prática minha e não um erro do app.
[87] As vezes ele não entendia o que eu falava.
[88] Acho que gostei de quase todo o aplicativo, o único problema foi quando ele não ouvia ou entendia o que estava falando e tinha que repedir algumas vezes.

appropriate speech recognition. This is likely to happen if the application is used in noisy places or when there are other people communicating in the proximity. The issue reported by P8 seems to involve not only the recognition of speech, but also the users' comprehension of the feedback provided. It seems like the participant altered his speech to receive positive feedback, but was not aware of what speech aspects were altered or why positive feedback was given.

Although issues with the ASR-based system were the most prevailing in the responses, the participants also commented on other aspects of the application that they did not appreciate:

> P1: I didn't like that I did not find the option of Brazilian Portuguese and that many times I had to translate what it [the application] said.[89]
> P3: At the beginning I was having trouble to watch the videos because whenever I clicked on them, they paused soon after (I believe that is a bug). That happened when I tried to watch the videos in the application. But after a while I realized that if I clicked on the YouTube button and went to the YouTube application the video worked fine.[90]
> P9: I think that the only thing I did not like is that it is a little confusing at first, but then I got used to it.[91]
> P13: I didn't like that you can only see how many lessons you have completed after you complete one.[92]

The comments made by these four participants reveal some of the difficulties they faced while using ELSA Speak, as the bug reported by P3 and the initial confusion felt by P8. While P3 and P9 managed to overcome the issues they reported, P1 was not able to access the application in Brazilian Portuguese as she wanted, which likely affected her learning experience as she had to translate the sentences and instructions to her native language. However, as was shown in section 3.4.3, users can choose the display language of the application, with one of them being Brazilian Portuguese. With this option, translations of the target sentences are available in most of the activities, including the YouTube videos with people pronouncing the target words. This feature was complimented by P8, who commented that "it was really good to be able to see the translations of the sentences". Thus, it is likely that P1 missed the option to access the application in her native language when she first started using the application, and could not find the option to alter this on the settings menu, which can be done at any time.

[89] Eu não gostei que eu não encontrei a opção do português do Brasil e muitas vezes tive que traduzir o que dizia.
[90] No começo eu estava tendo problemas para assistir os vídeos, pois sempre que eu clicava neles eram pausados logo em seguida (acredito que isso seja um bug). Isso acontecia quando eu tentava ver os vídeos dentro do aplicativo. Mas depois de quebrar um pouco a cabeça eu percebi que se eu clicasse no botão do YouTube e fosse para o aplicativo do YouTube o vídeo rodava normal.
[91] Eu acho que a única coisa que eu não gostei foi que ele é um pouco confuso no início, mas depois eu me acostumei.
[92] Eu não gostei que você só consegue ver quantas lições você fez depois de terminar uma.

Considering the importance that the translations seem to have had, an ideal self-regulated learner would have attempted to fix the issue so that the form of instruction suited the learners' needs. The fact that P1 did not manage to do so highlights how challenging this type of learning was for her.

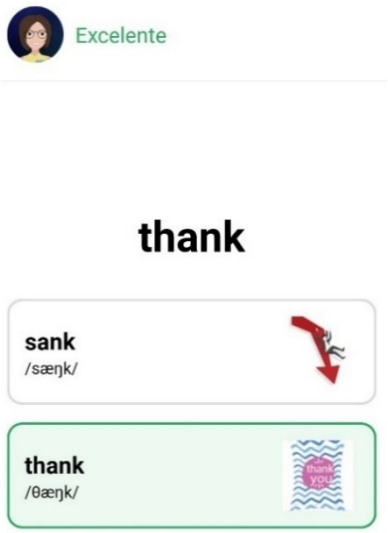
To learn more about the perceptions of the application, the participants were asked to write about one thing they would like to improve about it. Five of them did not have anything to say in this regard, but six participants made a few recommendations:

> P2: Improve its word comprehension, or, if it identifies a word that we mispronounced, it should show us exactly where we made that mistake.[93]
> P3: Allow us to watch the videos on the application itself, and maybe some more ways to show feedback[94]
> P9: It would be good to have a basic tutorial, that explained each 'section', as there are many different 'sections' with different phonetic questions in them (but I don't know, maybe I missed it)[95]
> P11: I don't remember any because my only issue was that the application sometimes did not hear me, and I don't know how to fix this.[96]
> P12: Add images along with the words and their meanings.[97]
> P13: I wish I could see how many lessons I've completed without having to complete a lesson first.[98]

In sum, two participants believed there should be improvements to the ASR-system (P2 and P11), and four mentioned design-related issues (P3, P9, P12 and P13). P3 stated he would like to be able to see the videos on the application itself. However, as shown in figure 2, this is already possible, and the issue is probably related to his smartphone software configuration. Likewise, P13 said he would like to be able to see the number of completed lessons without having to complete a lesson, which was already possible when the participants used the application in 2021. Participant 12 suggested adding images to represent word meanings, which was not a feature at the time when the students used the application in 2021. However, this feature was added in 2022, as can be seen in Figure 31.

Figure 30 - Images shown to illustrate word meanings

[93] Aumentar a compreensão de palavras dele, ou, se ele identifica uma palavra que falamos errado, mostre exatamente onde estamos errando.
[94] Poder assistir os vídeos dentro do aplicativo, e talvez mais algumas formas de feedback
[95] Seria bom ter um tutorial bem básico, que explicasse cada 'seção' porque tem muitas seções diferentes com questões fonéticas diferentes em cada (mas não sei, talvez eu não tenha visto).
[96] Agora pelo que me lembre nenhuma, pois meu único problema foi o do aplicativo não ouvir as vezes e não sei como poderia melhorar isso.
[97] Adicionar imagens juntamente das palavras e seus significados.
[98] Poder ver a quantidade de lições que você fez sem ter que completar uma lição para poder ver.



Source: ELSA Speak (2022)

Nonverbal representations of vocabulary, such as illustrations and videos, are relevant for vocabulary learning, as they provide learners with cues to identify words and help meaningful elaboration in the form of nonverbal context (SADOSKI, 2005). In the case of Figure 31, the verb ‘sank’ is illustrated with an arrow pointing downwards, which should help learners perceive the difference in meaning between the words ‘sank’ and ‘thank. The fact that these illustrations were implement shows that P12’s demand was relevant, and is a sign that the application receives updates that are in line with L2 learning theory.

Regarding the basic tutorial suggested by P9, it is true that the application does not provide an explicit tutorial on how to use it and on how to interpret its ASR feedback. P9’s response revealed that he missed more detailed explanations of the phonetic features that came up in the lessons. As previously shown, the application has introductory videos on how to produce /i/ and /ɪ/, but for other sounds the feedback shown in unit 8 was mostly text-based. For example, for mispronunciations of /b/, the feedback provided was ‘press your lips, stop the air and hold it for a second, then open your lips and make the sound /b/’. Although such feedback is relevant and is more accessible than visualizing spectrographic information, as was the case in earlier ASR-based systems (NERI; CUCCHIARINI; STRIK, 2003), it is probably not as clear and visual as the videos shown to introduce the high front vowels in unit 8.

Having presented and discussed the findings for the four research questions that guided this study, the following section presents the final remarks for this doctoral study.

5 FINAL REMARKS

The current study sought to investigate the effects of the use of an ASR-based mobile application the speech intelligibility of Brazilian learners of English. To do this, 13 elementary and high school students used a smartphone application named ELSA Speak and completed 38 lessons focused on the instruction of the English high-front vowels. This chapter presents the final remarks of this doctoral study. First, a summary of the findings of this study is presented, followed by a discussion of its pedagogical implications, limitations and suggestions for future research on related topics.

5.1 SUMMARY OF FINDINGS

The current study sought to investigate the effects of the use of an ASR-based mobile application on the speech intelligibility of Brazilian learners of English. To do this, 13 elementary and high school students used a smartphone application named ELSA Speak and completed 38 lessons focused on the instruction of the English high-front vowels. Significant differences were found for overall and target word intelligibility, which indicates that its use can improve L2 speech intelligibility. Although the minimal pairs were the least intelligible words in the pretest, they were the words that had the most gains percentagewise (12.3%). Participants with lower intelligibility scores in the pretests obtained more gains than those with higher scores, signaling that the application is more suitable for learners with lower intelligibility levels. In spite of the overall positive effects, the use of the application was not sufficient to improve the intelligibility of some the target words, such as ‘bit’, ‘hit’ and ‘sheep’, possibly because learners had to deal with multiple phonetic issues for the same word, such as the palatalization of the alveolar stops, as well as the /i - ɪ/ distinction.

Regarding the participants engagement in self-regulated learning with a mobile application, it was found that, although the application was used predominantly at home, the mobility of the devices appears to have helped them find suitable environments for pronunciation practice. Only one participant reported doing other activities while using ELSA Speak, which signals that the distractions typically associated with smartphones did not interfere in their learning establish, and indicates that they were able to establish adequate learning environments. Although the participants were asked to complete 10 lessons a week

and do all the assigned lessons within one month, their control of the application was not uniform, with some completing all the 38 assigned lessons in a single day or week, while two took more than two months to do so. It is likely that, by the time they took the intelligibility posttest, these two participants no longer remembered what they had studied in the first lessons, i.e., the content studied lacked recency, which may have hindered their intelligibility gains. Despite the high levels of entering motivation, a few participants lost motivation as they progressed through the lessons, which may have caused them to spend so many days without using the application. This attests to the novelty effect reported in studies such as Botero et al (2019) and Rodrigues et al. (2022). Completing all lessons in a short period of time, as in a day or in a week, did not seem to affect the posttest scores, nor did the average time spent on each lesson. These findings validate the flexibility of MALL, as learners could control the amount of instruction that fit their needs and learning styles, and still have good learning outcomes as long as they did not interrupt their learning process for extended periods e.g., more than a month without accessing the application. They also validate that the claims that MALL lessons can be short and succinct (BURSTON, 2014; STOCKWELL, 2013; TRAXLER, 2007), as learners took an average of two minutes to complete each lesson, which means they could squeeze them between other activities, as suggested in Burston (2014). All in all, the participants of the current study managed to engage in self-regulated learning with a mobile application, but struggled to control its use and to sustain their motivation.

In general, the participants had positive perceptions of the application. They believed they were able to understand the feedback it provided and appreciated that the feedback was presented in a comprehensible manner, which involved the use of colors to signal what needed of improvement and what had been appropriately pronounced. They believed that the feedback provided helped improve their pronunciation and stated they became more confident to pronounce words with the English high-front vowels. The participants thought the greatest advantage of learning with application was that they could practice pronunciation without feeling pressured or judged. This was argued to be an affordance of mobile learning, as it was the participants' mobility that allowed them to be in places where pronunciation practice could take place in a more relaxed manner. The possibility of controlling the amount and the time of instruction was perceived positively, and this perception that was corroborated by their actual use of ELSA Speak, which varied considerably in terms how long they took to complete the assigned lessons. There were also positive comments on the instructional design and resources

offered by ELSA Speak, such as the progressive difficulty of the lessons, the different types of activities it had, and the possibility of watching videos to check how people pronounced words in various contexts.

The participants perceived the lack of a teacher to ask questions, and their lack of dedication and organization as the greatest disadvantages involved in their mobile learning experience. These perceptions help explain the difficulties some of the participants had to complete the assigned lessons in due time, and indicate that the guidance of L2 teachers would be beneficial for their learning. The size of the screen and of the keyboards of their devices was not seen as a disadvantage, suggesting that this is no longer a serious issue for MALL. When asked to write about the things they did not appreciate about the application itself, the most mentioned issue was the difficulty it had to recognize their speech. Even though all participants agreed that the application was overall able to understand their speech, they reported instances in which they had to repeat certain words multiple times until they received positive feedback, which was something they believed should be improved. There were also instances when they thought to be pronouncing words appropriately, but the application repeatedly gave them negative feedback, which led some participants to develop strategies to circumvent this issue.

Having presented the summary of the findings, the following section discusses the pedagogical implications of the current study.

5.2 PEDAGOGICAL IMPLICATIONS

As previously discussed, time for pronunciation instruction in L2 classrooms is notoriously limited. While L2 learners may find other sources for developing and improving their pronunciation, they do not always involve pronunciation feedback. For instance, the participants of this study said they developed their English pronunciation by watching movies and TV shows, playing video games and listening to music in the target language. In spite of the value of such practices, pronunciation feedback is not part of any of them. ASR-based mobile applications can play an important role in this context because, unlike L2 teachers, who have time restraints and are only available at certain times and places, they are available virtually anywhere and at any time, and for as long as learners need. Cognizant of this, the current study investigated if one such application was capable of assisting L2 learners in developing their pronunciation and found that its use led to improvements in intelligibility. I

believe this finding has pedagogical implications for L2 teachers and learners. For one, they should feel confident that current ASR-based systems, such as ELSA Speak, are capable of reliably recognizing the L2 speech. What is more, they can provide feedback that can be comprehended by learners with little or no knowledge of phonetic symbols, and by young learners as the ones who participated in this study. With the belief that these systems can reliably recognize L2 speech and provide them with comprehensible feedback, I now present and discuss some pedagogical implications and possibilities for their use inside and outside of classrooms.

One of the challenges of L2 pronunciation instruction is that students often have difficulties with different phonetic features, even when they have similar proficiency levels or language backgrounds (MUNRO, 2021). In the case of Colégio de Aplicação, where students have highly heterogeneous backgrounds and levels of proficiency (FARIAS, 2020), even when they are in the same grade, the instruction of pronunciation under a one-size-fits-all approach runs the risk of being ineffective because students tend to have different difficulties and needs.

I believe ASR-based systems can help overcome this challenge in a couple of ways. First, they can help identify the common pronunciation difficulties for a particular group of students. Although well-trained L2 teachers are certainly capable of identifying common pronunciation issues among their students, doing so is time consuming. Thus, having the aid of an application in the process could save time and provide valuable data for future classes. For instance, the participants of this study, who had very low levels of intelligibility for the words ‘bit’ and ‘hit’, could all benefit from teacher-led instruction and practice on these and other phonetically similar words. Second, these systems can be used to learners develop the specific sounds they have difficulty with, without spending too much time on the ones they already know. In practical terms, teachers could ask students to use the application to improve the specific features they struggle with, and then promote classroom activities in which these features can be practiced in different ways. Mobile application such as ELSA Speak can be particularly useful in this regard because are run in devices that the students already own, which allows students to practice in their classrooms and also at their homes using with the same device.

ASR-based mobile applications can be particularly valuable for flipped[99] classroom strategies. This is because learners can practice the pronunciation of a set of words to get prepared for tasks that will happen in their classrooms. With this, the explicit pronunciation instruction that takes place with the use of the application allows classroom time to be focused on other types of practices, such as more communicative ones. For example, in one of the lessons of ELSA Speak, learners get to practice the pronunciation of sentences typically involved in a job interview. With this practice, they would likely feel more confident to enact this conversation in front of their classmates. As previously discussed, the possibility of practicing pronunciation without the pressure of peers was particularly important for the participants of this study.

Overall, the results of this study suggest that the use of an ASR-based mobile application can improve the intelligibility of English learners. Nevertheless, the use of the application alone did not seem to improve the intelligibility of words such as 'bit', 'hit' and 'sheep', and did not increase the overall intelligibility of three participants. These findings show limitations of learning independently with an application, and indicate that its use alone should not be expected to enhance all aspects of learners' pronunciation. In fact, I believe that closer assistance from L2 teachers can contribute to the learning process in two major ways. First, they could complement the pronunciation instruction of the application, especially for the features that are challenging for learners. Secondly, and perhaps most importantly, they can assist learners with their organization by helping establish learning routines and goals, while also taking part in the monitoring process. In the current study, there were participants who struggled to use the application within the stipulated time period, and who stated they lacked the organization necessary to use it appropriately i.e., they were not able to self-regulate their learning effectively. As discussed in Tumolo and Santibanes (2016), Brazilian students are used to learning in a heteronomous system, that is, they tend to rely on the decisions and instructions of others rather than taking responsibility for their own learning. Having the support of teachers can help learners transition from heteronomy towards autonomy, as they can initially work together to conceive and maintain goals and routines, while gradually promoting learners' independence up to a point when their assistance is no longer required.

[99] The flipped or inverted classroom is an instructional model in which activities that traditionally take place in the classroom (e.g., content presentation) become home activities, and activities that are typically homework become classroom activities (AKÇAYIR, G; AKÇAYIR, M, 2018.)

The fact that the participants perceived the lack of a teacher as the greatest disadvantage of their mobile learning experience indicates that they were aware of their difficulties to study in a more autonomous fashion. I believe that one pedagogical implication from this finding is that better outcomes can be achieved with MALL when the learners' ability to self-regulate are taken into account. For example, learners who are known to have difficulties with accomplishing tasks independently should not be expected to immediately perform well in an independent use of MALL. This does not mean that they cannot benefit from applications such as ELSA Speak. Rather, it means that more support, supervision and motivation are necessary. However, even if learners have the support of teachers, they may not be committed to MALL if it is not pedagogically integrated to schools' curricula. Souza Neto (2020) noticed that Brazilian teachers have explored DICTs in their educational processes, but points out that these technologies have not been pedagogically integrated to the curricula and to the culture of schools, and that their use tends to be fragmented and decontextualized. As a consequence, learners may perceive activities involving the use of technologies as less important, especially when they are not part of the curricula.

Souza Neto (2020) also mentions that teachers often feel insecure to use digital technologies because they lack the necessary training and knowledge, and fear they will lose authority and control over their classes. The author highlights how necessary the integration of DICTs, and that this should be done with the participation of the students so as to take into account their needs and their social contexts, but is aware of the challenges involved in this processes (SOUZA NETO, 2020). Considering the integral and indispensable role DICTs have played in schools around the world since 2020, I believe that their incorporation to the curriculum should not be seen as a thing of future, but as a necessity of the present.

5.3 LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

The results from the present study indicate that the use an ASR-based applications can have positive effects on the speech intelligibility of English learners. Nevertheless, these results should be interpreted with caution as no causal relationships between the use of such applications and its effects on intelligibility were established. This is due to a couple of reasons, which are presented in the following paragraphs, along with suggestions for future research based on the limitations of this study.

First, given that this study aimed at investigating the self-regulated use the mobile application, the participants' use of the application was far from controlled. As shown, they accessed it in different days, and took different amounts of time to complete the assigned lessons. What is more, it is possible that they developed their pronunciation in other ways that did not involve the application, e.g., by studying the target sounds on with other, even though they were asked not to do so. Finally, because the sentences for the intelligibility tests were recorded using their own devices, the quality of the recordings was not the same for all participants. Although none of the listeners stated any issues in this regard, it would be optimal to have their speech recorded with professional equipment, and with the same equipment for everyone. In light of this, future studies can investigate the effects of ASR-based applications such as ELSA Speak in more controlled ways. For instance, participants could use an application for a specific number of hours, at specific places and under the supervision of the researchers, thus guaranteeing that the only input they have received came from the application.

Second, the lack of a control group means that the intelligibility gains may have been an effect of time. In other words, it is possible that similar gains would have been found even if the application were not used, as a result of their input to the English language in their classes or elsewhere. As stated, their teachers did not plan to work with explicit pronunciation instruction during the months when data was collected for the study, but the language input received during the classes may still have led them to develop their pronunciation. Thus, future research should, whenever possible, include control groups and compare the results of those who received ASR-based instruction and those who did not. Ideally, this would be done under more controlled experiments, as previously suggested.

Third, the number of participants in this research was limited. Although 35 students initially agreed to participate, only 13 successfully completed their participation, which limits the generalizability of the results. The high dropout rate may have been due to a novelty effect, that is, students may have been initially enthusiastic about learning with a new technological device, but dropped out when they realized the commitment necessary to complete all the assignments. Moreover, the overwhelming number of hours the learners spent in front of computer screens during emergency remote learning is likely to have affected their willingness to participate in a piece of research that involved spending even more hours in front of a screen. Thus, future research can attempt to replicate the current study in more favorable conditions, that is, with more participants, but without the challenges that hampered learning throughout

the COVID-19 pandemic. All in all, future studies that replicate the method employed in the current research are necessary to assess generalizability of its findings.

Participation in this study required the ownership of a functional mobile device and adequate access to internet, which may have impeded those who had limited or no access to these digital technologies from taking part in the research. Due to the restraints posed by social distancing measures, it was not possible to access facilities such as the LIFE[100] laboratory, where the current piece of research was initially intended to take place. In light of this, future research should seek to provide the material conditions necessary to participate in studies involving mobile learning, especially in developing countries such as Brazil. This would make research more inclusive, and would allow the participation of more learners, which would strengthen the statistical power of the results. Moreover, with more participants, more experimental groups could be created. That would allow future research to compare a self-regulated use of ASR-based applications to a use that is always supervised and guided by language instructors, thus providing more insights into the issue of self-regulated learning.

One issue that emerged during this study was that some participants claimed their pronunciation of certain words was accurate but was not recognized or scored accordingly by the application. As discussed, it is not possible to determine the validity of their claim because we do not have access to their actual pronunciation of these words, so the question of whether appropriate feedback was provided is still unanswered. To shed light on this issue, studies that examine the recognition accuracy of ASR-based systems have to be carried out on a constant basis. Although such examinations have been done for systems based on native speaker models (DERWING; MUNRO; CARBONARO, 2000), ELSA Speak is said to be trained with non-native speech, which calls for studies focused on evaluating the ability of the application to recognize non-native speakers of English from different countries.

Finally, in the current study, learners with lower intelligibility levels benefitted more from the use of the application than those with higher intelligibility levels, with comparable results reported in studies such as Hincks (2005) and Yuan and Liu (2020). Although the data available so far seems to indicate that ASR-based pronunciation instruction is more suitable for learners with lower levels pronunciation levels, more research is necessary to understand the

[100] LIFE (Laboratório Interdisciplinar de Formação de Educadores) is a laboratory at Colégio de Aplicação – UFSC, which seeks to foment the creation of spaces and resources for the development of teachers and students, and provides access tablets and laptop computers for learners to use.

suitability of this type of instruction for learners with different levels of pronunciation and proficiency.

REFERENCES

AKÇAYIR, G.; AKÇAYIR, M. The flipped classroom: A review of its advantages and challenges, **Computers & Education,** v. 126, p. 334-345, 2018. https://doi.org/10.1016/j.compedu.2018.07.021.

AN, Z. et al. Technology-Assisted Self-Regulated English Language Learning: Association with English Language Self-Efficacy, English Enjoyment, and Learning Outcomes. **Frontiers in Psychology**, v. 11, n.1, p. 1-14, 2021.

ANDRADE, I. R. Aprendizagem de língua assistida por dispositivos móveis (ALADIM): uma proposta alternativa para o ensino da língua espanhola. Orientadora: Dra. Denise Bértoli Braga, 2017. 351 f. Tese (Doutorado). Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2017.

ANGUERA, X.; VAN, V. English Language Speech Assistant. In: Interspeech 2016: Show & Tell contribution, 2016, San Francisco, p. 1962 – 1963.

ASABERE, N. Benefits and Challenges of Mobile Learning Implementation: Story of Developing Nations. **International Journal of Computer Applications,** v. 73, n. 1, p. 23-27, 2013.

ATAŞ, A.; ÇELIK, B. Smartphone Use of University Students: Patterns, Purposes, and Situations. **Malaysian Online Journal of Educational Technology**, v. 7, n. 2, 2019. http://dx.doi.org/10.17220/mojet.2019.02.004.

AULIA, V. Reducing Students' Pronunciation Anxiety by Utilizing WhatsApp Group in Learning English Consonant Sounds. ELT Worldwide, v. 5, n.2, p. 165-178, 2018.

BALDISSERA, L. App resources for developing English pronunciation: a focus on mobile technology. Dissertação (mestrado) – Universidade Federal de Santa Catarina. Florianópolis, Brasil, 2020.

BAMGBOSE, A. Torn between the norms: innovations in world Englishes. **World Englishes**, v. 17, n. 1 p. 1–14, 1998.

BARAN-LUCARZ, M. The relationship between language anxiety and the actual and perceived levels of foreign language pronunciation. Studies in Second Language Learning and Teaching, v. 4, n. 1 – p. 491 – 514, 2011.

BASHORI, M. et al. Effects of ASR-based websites on EFL learners' vocabulary, speaking anxiety, and language enjoyment. **System**, v. 99, n1, p. 1-16, 2021.

BATAINEH, A.M.; AL-QADI, N. The Effect of Using Authentic Videos on English Major Students' Prosodic Competence. **Journal of Education and Practice**, v. 5, p. 157-172, 2014.

BECKER, M. R. Inteligibilidade da língua inglesa sob o paradigma da Língua Franca: Percepção de discursos de falantes de diferentes L1s por brasileiros. Tese (doutorado) - Universidade Federal do Paraná. Curitiba: Brasil, 2013.

BOHN, O.-S. Cross language speech perception in adults: First language transfer doesn't tell it all. *In*: W. STRANGE , W(Ed.). **Speech Perception and Linguistic Experience: Issues in Cross-language research**, Timonium, York Press, 1995, p.279-300. Disponível em: https://www.researchgate.net/publication/244456830_Cross_Language_Speech_Perception_in_Adults_First_Language_Transfer_Doesn"t_Tell_It_All. Acesso em novembro de 2021.

BOTO, C. A Liturgia escolar na idade moderna. Campinas: Papirus, 2017.

BROWN, J. Designing a Language Study. In: Classroom Teachers and Classroom Research. 1997, p. 55-70.

BROWN, J. Mixed Methods Research for TESOL. Edinburgh University Press, 2014.

BROWN, T. The Role of m-learning in the future of e-learning in Africa? In: ICDE World Conference, XXI, 2003, Hong Kong. Proceedings, 2003, p. 1-12.

BURLESON, D. Improving Intelligibility of Non-Native Speech with Computer-Assisted Phonological Training. The Journal of the Acoustical Society of America, v. 121, n.8, p. 1-1, 2007.

BURSTON, J. MALL: The pedagogical challenges. Computer Assisted Language Learning, v. 27, n. 4, p. 344-357, 2014.

BUSS, L. Pronunciation from the perspective of pre-service EFL teachers: An analysis of internship reports. In J. Levis & K. LeVelle (Eds.). Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference. Aug. 2012. (pp. 255-264). Ames, IA: Iowa State University, 2013.

BYRNE, W. et al. Is automatic speech recognition ready for non-native speech? A data collection effort and initial experiments in modeling conversational Hispanic English. Proceedings of Speech Technology in Language Learning, 1998.

CANHOTA, C. Qual a importância do estudo piloto? In: SILVA, E. E. (Org.). *Investigação passo a passo: perguntas e respostas para investigação clínica*. Lisboa: APMCG, 2008. p. 69-72.

CARLO, M. et al. Closing the Gap: Addressing the Vocabulary Needs of English Language Learners in Bilingual and Mainstream Classrooms. **Journal of Education**, 189, 57-76, 2008.

CASEY, C. Incorporating cognitive apprenticeship in multi-media. **Educational Technology Research and Development**, v. 44, n.1, p. 71–84, 1996.

CASPERS, C. The influence of erroneous stress position and segmental errors on intelligibility, comprehensibility and foreign accent in Dutch as a second language. **Linguistics in the Netherlands**, p. 17-29, 2010. DOI: 10.1075/avt.27.03cas.

CECI, L. Most popular Google Play app categories as of 1st quarter 2022, by share of available apps. 2022. Disponível em: https://www.statista.com/statistics/279286/google-play-android-app-categories/. Accesso em 22 de julho de 2022

CECI, L. Worldwide Mobile Education App Downloads from 1st Quarter 2017 to 1st Quarter 2020, by Platform. 2021. Disponível em: https://www.statista.com/statistics/1128262/mobile-education-app-downloads-worldwide-platforms-millions/. Accesso em 22 de julho de 2022.

CELCE-MURCIA, M.; BRINTON, D. M.; GOODWIN, J. M. **Teaching Pronunciation:** A Reference for Teachers of English to Speakers of Other Languages. 14. ed. New York: Cambridge University Press, 2010.

CHAE, M.; KIM, J. Do size and structure matter to mobile users? An empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones. **Behaviour & Information Technology**, 23(3), p. 165-181, 2004.

CHAU, T. et al. The Effects of L2 Pronunciation Instruction on EFL Learners' Intelligibility and Fluency in Spontaneous Speech. The Electronic Journal for English as a Second Language, v. 24, n. 24, 2022.

CHEN, C.; CHEN, L. YANG, S. An English vocabulary learning app with selfregulated learning mechanism to improve learning performance and motivation. **Computer Assisted Language Learning,** v. 32, n. 3, p. 237-260, 2019. DOI: 10.1080/09588221.2018.1485708

CHENG, H.; NIZIOLEK, C.; BUCHWALD, A.; MCALLISTER, T. Examining the Relationship Between Speech Perception, Production Distinctness, and Production Variability. Frontiers in human neuroscience, v. 15, 2021. https://doi.org/10.3389/fnhum.2021.660948.

CHINNERY, G. M. (2006). Going to the mall: mobile assisted language learning. Language Learning & Technology, 10, 9-16.

COHEN, J. Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers, 1998.

COSTA, B. Pronunciation Teaching Is Not a One-Size-Fits-All Endeavor: EFL Teachers' Beliefs and Classroom Practices. Dissertação (mestrado) – Universidade Federal de Santa Catarina, Centro de Comunicação e Expressão – Programa de Pós-Graduação em Inglês: Estudos Linguísticos e Literários. Florianópolis, SC, 2016.

COTTERALL, S. Readiness for autonomy: investigating learner beliefs. **System**, v. 23, n. 2, p. 195-205. 1995.

CROMPTON, H. A historical overview of mobile learning: Toward learner-centered education. In: BERGE, Z.L.; MUILENBURG, L.Y. (eds.) Handbook of mobile learning. Nova Iorque: Routledge, p.3-14, 2013a.

CROMPTON, H. Mobile learning: new approach, new theory. In: BERGE, Z.L.; MUILENBURG, L.Y. (eds.) Handbook of mobile learning. Nova Iorque: Routledge, p.47-57, 2013b.

CROMPTON, H. The Benefits and Challenges of Mobile Learning. Learning and Leading with Technology, v. 41, p. 38-39, 2013c.

CROMPTON, H.; et al. Learning with technology during emergencies: A systematic review of K-12 education. British Journal of Educational Technology, v. 54, n. 4, p. 1554-1575, 2021.

CRUZ, N. C. An exploratory study of pronunciation intelligibility in the Brazilian learners' English. **The Specialist**, v. 24, n. 2, p. 155-175, 2003.

CRUZ, N. C. Inteligibilidade de pronúncia de aprendizes brasileiros de inglês: um corpus de pequeno porte. **Revista Intercâmbio,** v. 41, p. 137-157, 2019.

DASHTESTANI, R. Moving bravely toward mobile learning: Iranian students' use of mobile devices for learning English as a foreign language. **Computer Assisted Language Learning**, v. 29, n. 4, p. 815–832, 2016.

DAUER, R. The Lingua Franca Core: A New Model for Pronunciation Instruction? **TESOL Quarterly**, v. 39, n. 3, p. 543-550, 2005.

DAVIES, M. Word frequency data from The Corpus of Contemporary American English (COCA). 2008. Disponível em: https://www.wordfrequency.info. Acesso em julho de 2022.

DELATORRE, F; SILVEIRA, R; GONÇALVES, A. The intelligibility of English verbs in the simple past tense: native and non-native speakers and Brazilian listeners. **Veredas Online**, v. 2, 2017, p. 57-77.

DERWING, T,; MUNRO, M.; WIEBE, G. Pronunciation instruction for fossilized learners: Can it help? **Applied Language Learning**, v. 8, p. 185–203, 1997.

DERWING, T. et al. Opening the window on comprehensible pronunciation after 19 years: a workplace training study. **Language Learning,** v. 64, p. 526–548, 2014.

DERWING, T.; MUNRO, M. Accent, intelligibility, and comprehensibility: evidence from four L1s. **Studies in Second Language Acquisition**, v.19, p. 1–16, 1997.

DERWING, T.; MUNRO, M. Pronunciation Fundamentals: Evidence-Based Perspectives for L2 Teaching and Research. **Language Teaching Series**, v. 42, n. 1, 2015.

DERWING, T.; MUNRO, M. Putting accent in its place: rethinking obstacles to communication. **Language teaching***,* v. 37, n. 1, p. 476-490, 2009.

DERWING, T.; MUNRO, M. Second Language Accent and Pronunciation Teaching: A Research-Based Approach. **TESOL Quarterly**, v. 39, n. 3, 2005.

DERWING, T.; MUNRO, M.; CARBONARO, M. Does popular speech recognition software work with ESL speech? **TESOL Quarterly**, v. 34, 592-603, 2000.

DERWING, T.; ROSSITER, M. The effects of Pronunciation Instruction on the Accuracy, Fluency and Complexity of L2 Accented Speech. **Applied Language Learning**, v. 13, n. 1, p. 1-17, 2003.

DETERDING, D. Segmentals. In REED, M., LEVIS, J (orgs). **The Handbook of English Pronunciation**. Wiley Blackwell: Oxford. p. 69 - 83, 2015. doi:10.1006/jpho.1997.0052.

DÖRNEY, Z. Research methods in Applied Linguistics: quantitative, qualitative and mixed methodologies. Oxford: Oxford University, 2007.

DUMAN, G.; ORHON, G.; GEDIK, N. Research trends in mobile assisted language learning from 2000 to 2012. ReCALL, v. 27, n. 2, p. 197-216, 2015.

DZUBAK, C. Multitasking: The Good, the Bad, and the Unknown**. The Online Journal of the Association for the Tutoring Profession**, v.1, 2008.

EGAN, K. Speech recognition application to language learning: Echos. Proceedings of CALICO, 1996.

EHSANI, F.; KNODT, E. Speech technology in Computer-Aided Language Learning: strengths and limitations of a new call paradigm. Language Learning & Technology, v. 2, n.1, p 54-73, 1998. Disponível em: https://www.lltjournal.org//

EISNER, E. Instruction, Teaching and Learning: an attempt at differentiation. The Elementary School Journal, v. 65, n. 3, p. 115-119, 1964.

ELIMAT, A.K.; ABUSEILEEK, A.F. Automatic Speech Recognition Technology as an Effective Means for Teaching Pronunciation. **JALT CALL Journal**, v.10, n.1, p. 21-46, 2014.

ELLIS, N. The Psychology of Foreign Language Vocabulary Acquisition: Implications for CALL. Computer Assisted Language Learning, v. 8, n. 2, 1995.

ELLIS, N. What can we count in language, and what counts in language acquisition, cognition, and use? In: MOUTON, D. G. Frequency Effects in Language Learning and Processing, Boston, De Gruyter, p. 7-34, 2012.

ELSA SPEAK. Science behind Elsa. Disponível em: https://elsaspeak.com/en/. Acesso em junho de 2022.

ELSA, Corp. ELSA Speak. Califórnia, EUA: ESLA Corp, 2021.

ELSA, Corp. ELSA Speak. Califórnia, EUA: ESLA Corp, 2022.

ESCUDERO, P.; BOERSMA, P. Bridging the Gap Between L2 Speech Perception Research and Phonological Theory. **Studies in Second Language Acquisition**, v. 26, n. 4, p. 551-585, 2004.

EVANS, J. Straightforward Statistics for the Behavioral Sciences. Brooks/Cole Publishing, Pacific Grove, California, 1996.

EXAMBOT. Why the ESL Market is set to boom in the next 10 years. **CEFR Language Exam Resource Centre Search.** Disponível em: https://cefrexambot.com/esl-market-due-boom-next-10-years-part-1. Acesso em fevereiro de 2020.

FARIAS, P, F. Critical EFL and Critical Literacy: **The Impacts of designing and implementing a cycle of tasks in a public school setting for critical language and critical**

literacy development. Tese (doutorado) – Universidade Federal de Santa Catarina, Centro de Comunicação e Expressão, 2020.

FIELD, J. Intelligibility and the Listener: The Role of Lexical Stress. TESOL Quarterly, v. 39, n. 3, p. 399-423, 2005.

FLEGE, E. Second language speech learning: theory, findings, and problems. *In:* W. STRANGE, W. (ed.). **Speech Perception and Linguistic Experience: Issues in Cross-language research**, p. 233-277, Timonium, York Press, 1995.

FLEGE, E.; BOHN, O.; JANG, S. Effects of experience on non-native speakers' production and perception of English vowels. **Journal of Phonetics**, v. 25, n. 4, 437-470, 1997.

FOUZ-GONZÁLEZ, J. Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app. **Language Learning & Technology**, v. 21, n.1, p. 62-85, 2020.

FRANCO, H. et al. Automatic pronunciation scoring for language instruction. Proceedings of ICASSP, April, 1471-1474, 1997.

FRASER, H. Teaching Pronunciation: A Handbook for Teachers and Trainers. New England, EUA: Department of Education Training and Youth Affair, 2001.

FRASER, H. **Teaching Pronunciation: A Handbook for Teachers and Trainers**. New England, EUA: Department of Education Training and Youth Affair, 2001.

GARRISSON, D. Self-directed learning: toward a comprehensive model. **Adult Education Quarterly**, v. 46, n. 1, p. 18-33, 1997.

GEORGE, S.; SERNA, A. Introducing Mobility in Serious Games: Enhancing Situated and Collaborative Learning. 14th International Conference, Orlando, 2011.

GIEGERICH, H. J.. English phonology: an introduction. 8th ed. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press, 2005

GILAKJANI, A. P.; SABOURI, B. N. Why Is English Pronunciation Ignored by EFL Teachers in Their Classes? **International Journal of English Linguistics**, v. 6, n. 6, p. 195-208, 2016.

GONÇALVES, A.; SILVEIRA, R. In search of speech intelligibility: the case of English high front vowels produced by Brazilian learners. **Linguagem & Ensino**, v.17, n.2, p. 361-390, 2014.

GONÇALVES, A.; SILVEIRA, R. Intelligibility research in Brazil: empirical findings and methodological issues. **Horizontes de Linguística Aplicada**, n.1, p. 51-81, 2015.

GOTTARDI, W.; ALMEIDA, J.; TUMOLO, C. Automatic speech recognition and text-to-speech technologies for L2 pronunciation improvement: reflections on their affordances. **Texto livre**, v. 15, p. 1-15, 2022.

HAAG, J.; BERKING, P. Design Considerations for Mobile Learning. In ZHANG, Y. A., CRISTOL, D. (eds.), Handbook of Mobile Teaching and Learning. Singapore: Springer Nature, 2019, p. 221-240.

HALUK, T. App attrition in computer-assisted language learning: **focus on Duolingo**. Tese (doutorado) - McGill University, Department of Integrated Studies in Education, 2020.

FARIAS, P, F. Critical EFL and Critical Literacy: **The Impacts of designing and implementing a cycle of tasks in a public school setting for critical language and critical literacy development**. Tese (doutorado) – Universidade Federal de Santa Catarina, Centro de Comunicação e Expressão, 2020.

HEIL, C. R.; WU, S. J.; LEE, J. J. (2016). A review of mobile language learning applications: trends, challenges and opportunities. **The Eurocall Review**, v. 24, n. 2, p. 32-50, 2016.

HENDERSON, A. et al. The English pronunciation teaching in Europe survey: Selected results. **Research in Language**, v. 10, n. 1, p. 5–27, 2012

HICE, S. Learning Styles and Field Dependency. Disponível em: https://www.houghton.edu/current-students/center-for-student-success/academic-support-and-accessibility-services/study-advisement/general-study-information/learning-styles-field-dependency/. Acesso em junho de 2022.

HIGGINS, J. (2013). Minimal pairs for English RP. Disponível em: http://minimal.marlodge.net/minimal.html

HINCKS, R. (2005). Computer support for learners of spoken English. Unpublished doctoral thesis, KTH, Stockholm, Sweden. Disponível em: https://www.yumpu.com/en/document/read/26490066/computer-support-for-learners-of-spoken-english-speech-music-.

HOLLIDAY, A. Native speakerism. **English Language Teaching Journal**, v. 60, p. 385–387, 2006.

HSU, T. C. Learning English with augmented reality: Do learning styles matter? Computers &Education, v. 106, p. 137–149, 2017. https://doi.org/10.1016/j.compedu.2016.12.007

HUANG, X. et al. Spoken Language Processing: A guide to Theory, Algorithm and System Development. Prentice Hall, Upper Saddle River, New Jersey, 2001.

HUTSON, M. How researchers are teaching AI to learn like a child. Science. Disponível em: https://bit.ly/3OkKjjF. Acesso em junho de 2022.

ISAACS, T. Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-Speaking graduate students. **The Canadian Modern Language Review**. V.64, n. 4, p. 555-580, 2008.

JENKINS, J. **The Phonology of English as an International Language.** Oxford, UK: Oxford University Press, 2000.

KAISER, D. Mobile-Assisted Pronunciation Training: The iPhone Pronunciation App Project. **IATEFL Pronunciation Special Interest Group Journal**, v. 58, n.1, p. 38–52, 2018.

KAMASAK, R. et al. The Effectiveness of Mobile-Assisted Language Learning (MALL): A Review of the Extant Literature. In: MOURA, A; REIS, P; CORDEIRO, M. **Determining the Reliability of Online Assessment and Distance Learning**. IGI Global Publishing, 2020.

KAUR, J. Intelligibility in global contexts. In: KANG, Okim; THOMSOM, Ron I; MURPHY, John M (orgs). **The Routledge Handbook of Contemporary English Pronunciation**. Nova Iorque: Routledge, p. 524-555, 2018.

KAUR, J. Intelligibility in global contexts. In: KANG, Okim; THOMSOM, Ron I; MURPHY, John M (orgs). **The Routledge Handbook of Contemporary English Pronunciation**. Nova Iorque: Routledge, 2018, p. 524-555.

KHADDAGE, F., MÜLLER, W., FLINTOFF, K. Advancing Mobile Learning in Formal and Informal Settings via Mobile App Technology: Where to From Here, and How?. Educational Technology & Society, v. 19, n. 3, v. 16–26, 2016.

KIVISTÖ-DE SOUZA, H. Effects of Speech Style on the Use of Temporal and Spectral Cues in the Production and Perception of a Non-native Vowel Contrast. Dissertação (mestrado) – Universitat de Barcelona - Department of English and German Philology - Applied Linguistics and Language Acquisition in Multilingual Contexts. Barcelona, 2011.

KIVISTÖ-DE SOUZA, H. et al. Vowel inventory size matters: Assessing cue-weighting in L2 vowel perception. **Ilha do Desterro**, v. 70, n. 3, p. 33-46, 2017.

KNOWLES, M. S. Self-directed learning. New York: Association Press, 1975.

L, O.; KHALIL, M.; BAARS, M. Self-Regulated Learning and Learning Analytics in Online Learning Environments: A Review of Empirical Research. In Proceedings of the 10th International Conference on Learning Analytics & Knowledge (LAK'20), New York, NY, USA, 2020.

language learning by children. **Computer Assisted Language Learning,** v. 21, n. 5, p. 393-408, 2008.

LEASK, A. **Student vs Learner. Enable Education**, 2013. Acesso em 25/05/2020. Disponível em https://enableeducation.com/student-vs-learner/.

LEVIS J. Changing contexts and shifting paradigms in pronunciation teaching. **TESOL Quarterly**, vol. 39/3: 369–77, 2005.

LIAKIN, D., CARDOSO, W., LIAKINA, N. Learning L2 pronunciation with a mobile speech recognizer: French /y/. **Computer Assisted Language Instruction Consortium**, v. 32, n.1, p. 1-25, 2015.

LIMA, J. A distintividade das vogais anteriores do inglês na comunicação de falantes de português. Dissertação (mestrado) – Universidade Federal da Bahia. Bahia, Brasil, 2005

LOEWEN, S. et al. Mobile-assisted language learning: A Duoligo case study. **ReCALL FirstView**, 1-19. https://doi.org/10.1017/S0958344019000065

LUSTIG, Robin. Can English remain the 'world's favourite' language?. **BBC World Service.** Disponível em https://www.bbc.com/news/world-44200901. Acesso em fevereiro de 2020.

MACRAE, I. Teaching Is Different From Training: How to Use Both Effectively. **Training Industry.** Disponível em: https://trainingindustry.com/articles/workforce-development/teaching-is-different-from-training-how-to-use-both-effectively/. Acesso em 21/07/2022.

MAHDI, S. H., KHATEEB, A. A. The effectiveness of computer-assisted pronunciation training: A meta-analysis. Review of Education, v. 7, n. 3, p. 733 – 753, 2019. DOI: 10.1002/rev3.3165

MARTINHÃO, M. S. Pesquisa sobre o uso das tecnologias de informação e comunicação nos domicílios brasileiros: TIC Domicílios 2018. São Paulo: Comitê Gestor da Internet no Brasil, 2019.

McARTHUR, T. (Ed.) The Oxford Companion to the English Language. Oxford: OUP, 1992.

MCCROCKLIN, S. Pronunciation learner autonomy: The potential of Automatic Speech Recognition. System, v. 57, p. 25-42, 2016.

MCCROCKLIN, S.; HUMAIDAN, A.; EDALATISHAMS, E. ASR dictation program accuracy: Have current programs improved? 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. 191-200). Ames, IA: Iowa State University.

MEIRELLES, F. 30ª Pesquisa Anual do uso de TI nas empresas, 2019.

MERTENS, D. M. *Research and Evaluation in Education and Psychology: Integrating Diversity with Quantitative, and Mixed Methods*. Thousand Oaks, California: Sage. 2005.

MIKANOWSKI, Jacob. Behemoth, bully, thief: how the English language is taking over the planet. **The Guardian**. Disponível em: https://www.theguardian.com/news/2018/jul/27/english-language-global-dominance. Acesso em fevereiro de 2020.

MIODUSER, D. et al. Web-based learning environments (WBLES): Current implementation and evolving trends. **Journal of Network and Computer Application**, v. 22, p.233-247, 1999.

MOURA, F.; LOPES, D. O reconhecimento de voz como alternativa para o ensino da pronúncia da língua inglesa em aulas remotas: estudo de caso com o aplicativo Elsa Speak. Hipertextus, v. 23, n.1, p. 76-97, 2021.

MOYER, A. Foreign Accent: The Phenomenon of Non-Native Speech. New York: Cambridge University Press, 2013.

MUNRO, M. On the Difficulty on Defining “Difficult” in Second-Language Vowel Acquisition. **Frontiers in Communication**, v. 6, n.1, p. 1 – 15, 2021.

MUNRO, M., DERWING, T. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. **Language Learning**, v. 45, n.1, p. 73–97, 1995.

MUNRO, M., DERWING, T. Intelligibility in Research and Practice: Teaching Priorities. In REED, M., LEVIS, J (orgs). **The Handbook of English Pronunciation**. Wiley Blackwell: Oxford. p. 377 - 396, 2015.

MUNRO, M., DERWING, T. Segmental acquisition in adult ESL learners: a longitudinal study of vowel production. Language Learning, v. 58, p. 479–502, 2008.

MUNRO, M., DERWING, T.; SAITO, K. 2013. English L2 vowel acquisition over seven years. In: Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference, J. Levis and K. LeVelle (eds.), Ames, IA: Iowa State University, 2012

MUNRO, M.; DERWING, T; MORTON, S. The mutual intelligibility of L2 Speech. **Studies in Second Language Acquisition**, v. 28, n. 1, p. 111-131, 2006.

NAIR, R., KRISHNASAMY, R., DE MELLO, G. Rethinking the teaching of pronunciation in the ESL classroom. **The English Teacher**, v.35, p. 27–40, 2017.

NAISMITH, L. Literature review in mobile technologies and learning. Bristol: Nesta Futurelab, 2004.

NARCISS, S.; PROSKE, A.; KOERNDLE, H. Promoting self-regulated learning in web-based learning environments. **Computers in Human Behavior**, v.23, p. 1126–1144, 2007.

NERI, A. et al. The effectiveness of computer assisted pronunciation training for foreign

NERI, A., CUCCHIARINI, C., STRIKM W. ASR-based corrective feedback on pronunciation: does it really work? In: Proceedings of Interspeech 2006- ICSLP, Ninth International Conference on Spoken Language Processing, Pittsburgh, PA, USA, September 17-21, 2006.

NERI, A., CUCCHIARINI, C., STRIKM W. Automatic Speech Recognition for second language learning: How and why it actually works. **Speech works**, p. 1157 – 1160, 2003.

NERI, A., CUCCHIARINI, C., STRIKM W. Feedback in Computer Assisted Pronunciation Training: When technology Meets Pedagogy. In: Proceedings of the 10th International CALL Conference on "CALL professionals and the future of CALL research", University of Antwerp, Belgium, pp. 179-188, 2002.

O’MALLEY, C.; VAVOULA, G.; GLEW, J.; TAYLOR, J.; SHARPLES, M. Guidelines for learning/teaching/tutoring in a mobile environment. Hal Archives-ouvertes, 2005. ffhal-00696244ff. Acesso em 05/2020 https://hal.archives-ouvertes.fr/hal-00696244/document.

OLIVEIRA, D. N. (2007). **The effect of perceptual training on the learning of English vowels by Brazilian Portuguese speakers**. Doctoral dissertation. Florianópolis: Universidade Federal de Santa Catarina.

OROZCO, V. Mobile-based solutions can strengthen human capital gains disrupted by COVID-19 in developing countries. **Digital Developmen**t, 2021. Disponível em https://blogs.worldbank.org/digital-development/mobile-based-solutions-can-strengthen-human-capital-gains-disrupted-covid-19. Acesso em maio de 2022.

PALALAS, A.; WARK, N. The relationship between mobile learning and self-regulated learning: a systematic review. **Australasian Journal of Educational Technology**, v. 36, n. 4, p. 151-172, 2020.

PALMAI, K., SMALE, W. **UK leading the way in use of language-learning apps.** BBC News. Publicado em 11 de janeiro de 2021. Acesso em 30 de novembro de 2021 Disponível em: https://www.bbc.com/news/business-55562267.

PARK, A. The Study on Automatic Speech Recognizer Utilizing Mobile Platform on Korean EFL Learners' Pronunciation Development. **Journal of Digital Contents Society**, v. 18, n. 6, p. 1101-1107, 2017.

PAULA, G. S. **5 vantagens da Computação em Nuvem (Cloud Computing) para sua empresa**. Acesso em 01/2020 https://www.treasy.com.br/blog/5-vantagens-da-computacao-em-nuvem-cloud-computing-para-sua-empresa/

PENG, H., CHIEN CHOU, Y., TSAI, C. Ubiquitous knowledge construction: mobile learning re-defined and a conceptual framework. **Innovations in Education and Teaching International**, v. 46, n. 2, p. 171-183, 2009.

PENNINGTON, M., ELLIS, N. Cantonese speakers' memory for English sentences with prosodic cues. **The Modern Language Journal**, v. 84, n.3, p. 372–389, 2000.

PERLMUTTER, M. Intelligibility rating of L2 speech pre- and postintervention. Perceptual and Motor Skills, v. 68, p, 515-521, 1989.

PERSSON, V.; NOURI, J. A Systematic Review of Second Language Learning with Mobile Technologies. **International Journal of Emerging Technologies in Learning**, v.13, n. 2, p. 188-210, 2018.

PFEIFFER, V., GEMBALLA, S., JARODZKA, H., SCHEITER, K., GERJETS, P. Situated learning in the mobile age: mobile devices on a field trip to the sea. **Research in Learning Technology**, v. 17, n. 3, p. 187-199, 2009. DOI: 10.1080/09687760903247666

PI-HUA, T. Bridging pedagogy and technology: User evaluation of pronunciation-oriented CALL software. Australasian Journal of Educational Technology, v. 2, n. 3, p. 375–397, 2006.

PI-HUA, T. Computer-assisted pronunciation learning in a collaborative context: A case

PINTRICH, P.; DE GROOT, E. Motivational and self-regulated learning components of classroom academic performance. **Journal of Educational Psychology**, v. 82, n. 1, p. 33–40, 1990.

PIRES, D.; TUMOLO, C. L2 vocabulary instruction: an analysis of smartphone applications for English learning. Diacrítica, v. 34, n. 1, p. 225-247, 2020. DOI: doi.org/10.21814/diacritica.303

PIRES, R. D. (2018). L2 Vocabulary Instruction an Analysis of Smartphone Applications for English Learning. Doctoral dissertation. Florianópolis: Universidade Federal de Santa Catarina.

PRICE, M. L. The subjective experience of foreign language anxiety: Interviews with highly anxious students. In: HORWITZ, E.; YOUNG, D. (eds.). Language anxiety: From theory and research to classroom implications. Upper Saddle River, NJ: Prentice Hall, 1991p. 101-108.

RACHEL'S ENGLISH. American English - EE [i] Vowel - How to make the EE Vowel. YouTube, 17 de março de 2016. Disponível em: https://youtu.be/C5WKc4R7mD4. Acesso em abril de 2021.

RAUBER, A. S. (2006). **Perception and production of English vowels by Brazilian EFL speakers**. Doctoral dissertation. Florianópolis: Universidade Federal de Santa Catarina.

ROACH, P. (2009). *English phonetics and phonology: a practical course*. 4 ed. Cambridge: CUP.

RODRIGUES, L. et al. Gamification suffers from the novelty effect but benefits from the familiarization effect: Findings from a longitudinal study. International Journal of Educational Technology in Higher Education, v. 19, n.13, p. 1 – 25, 2022. https://doi.org/10.1186/s41239-021-00314-6.

ROGERS, Y.; CONNELLY, K.; HAZLEWOOD, W.; TEDESCO, L. Enhancing learning: a study of how mobile devices can facilitate sensemaking. **Pervasive Ubiquitous Computing,** v.14, p. 111-124, 2010.

ROGERSON-REVELL, P. English vowels and consonants. In: KANG, Okim; THOMSOM, Ron I; MURPHY, John M (orgs). **The Routledge Handbook of Contemporary English Pronunciation**. Nova Iorque: Routledge, 2018, p. 93-121.

RYAN, R.; DECI, E. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. **Contemporary Educational Psychology**, v. 25, p. 54–67, 2000.

SADOSKI, M. A Dual Coding View of Vocabulary Learning. **Reading & Writing Quarterly: Overcoming Learning Difficulties**, v. 21, p. 221-238, 2005. Available in http://dx.doi.org/10.1080/10573560590949359

SAITO, K. Effects of Instruction on L2 Pronunciation Development: A Synthesis of 15 Quasi-Experimental Intervention Studies. **TESOL Quarterly**, v. 46, n. 4, p.807-819, 2012.

SAITO, KAZUYA. Effects of Instruction on L2 Pronunciation Development: A Synthesis of 15 Quasi-Experimental Intervention Studies. **TESOL Quarterly**, v. 46, n. 4, p.807-819, 2012.

SAKS, K; LEIJEN, A. Distinguishing Self-Directed and Self-Regulated Learning and Measuring them in the E-learning Context. **Social and Behavioral Sciences,** v. 112 p. 190 – 198, 2014.

SCHUNK, D. Self-monitoring as a Motivator during Instruction with Elementary School Students. In: Annual Meeting of the American Educational Research Association, 1997, Chicago, p. 1 – 24.

SHA, L., LOOI, C., CHEN, W., ZHANG, B.H. Understanding mobile learning from the perspective of self-regulated learning. **Journal of Computer Assisted Learning**, v. 28, p. 366-378, 2012.

SHADIEV, R., LIU, T., HWANG, W. Review of research on mobile-assisted language learning in familiar, authentic environments. **British Journal of Educational Technology**, v. 51, n. 3, p. 709-720, 2020.

SHARPLES, M. et al. A theory of learning for the mobile age. In: The Sage Handbook of E-Learning Research (eds R. Andrews & C. Haythornthwaite), pp. 221–245. Sage, London, 2007.

SILVA, E. Análise dos Dados Socioeconômicos do Colégio de Aplicação. Equipe do Serviço Social do Colégio de Aplicação, 2021.

SILVA, T. Pronúncia do Inglês para falantes do Português Brasileiro. São Paulo, SP. Editora Contexto, 2012.

SOUZA NETO, A. Escola, currículo e tecnologias: desafios da integração pedagógica. São Paulo: Pimenta Cultural, 2020.

SWEDBERG, R. Exploratory Research. In: ELMAN, C; GERRING, J; MAHONEY, J. (eds.). The Production of Knowledge: Enhancing Progress in Social Science, United Kingdom, Cambridge University Press, 2020, p. 17-42.

STAUFER, B. What's the Difference Between Online Learning and Distance Learning? **Applied Educational Systems**. Access on 29/04/2020. https://www.aeseducation.com/blog/online-learning-vs-distance-learning.

STOCKWELL, G. Mobile-assisted language learning. In: THOMAS, M., REINDERS, H., WARSCHAUER, M (eds). **Contemporary Computer-Assisted Language Learning**. London: Bloomsbury, 2013, p. 201-216. WARSCHAUER, M (eds). Contemporary Computer-Assisted Language Learning. London: Bloomsbury, 2013, p. 201-216.

STOCKWELL, G., HUBBARD, P. Some emerging principles for mobile-assisted language learning. **The International Research Foundation for English Language Education**. Retrieved from: http://www.tirfonline.org/english-in-the-workforce/mobile-assisted-language-learning.

THOMSON, R. I. ESL teachers' beliefs and practices in pronunciation teaching: Confidently right or confidently wrong? In J. Levis & K. LeVelle (Eds). **Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference**, August 2012. (pp. 224–233). Ames, IA: Iowa State University.

THOMSON, R.; DERWING, T. The Effectiveness of L2 Pronunciation Instruction: A Narrative Review. **Applied Linguistics**, v. 36, n. 3, p. 326-344, 2014.

TIC EDUCAÇÃO 2020. Pesquisa sobre o Uso das Tecnologias da Informação e da Comunicação no Brasil. São Paulo: Comitê Gestor da Internet no Brasil, 2020. Disponível em http://www.cetic.br/

TRAXLER, J. Defining, discussing and evaluating mobile learning: The moving finger writes. **International Review of Research in Open and Distance Learning**, v. 8, n. 1, p. 1-12, 2007.

TRIFONOVA, V. How Device Usage Changed in 2018 and What it Means for 2019. **Global Web Index**, 20 de novembro de 2018. Disponível em: https://blog.globalwebindex.com/trends/device-usage-2019/. Acesso em maio de 2020.

TUMOLO, C.; SANTIBANES; V. Autonomia em graduação na modalidade a distância: estratégias de organização e de aprendizagem. In: IX Congresso Brasileiro de Linguística Aplicada, 2011, Rio de Janeiro. Anais Eletrônicos, Rio de Janeiro, 2011: p. 1-25.

UNITED NATIONS. Policy Brief: Education during COVID-19 and beyond. 2020. Disponível em: https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/08/sg_policy_brief_covid-19_and_education_august_2020.

VAN LIESHOUT, C.; CARDOSO, W. Google Translate as a tool for self-directed language learning. **Language Learning & Technology**, v. 26, n. 1, p. 1–19, 2022. http://hdl.handle.net/10125/73460

WEST, M.; CHEW, H.E. (2014). Reading in the mobile era: a study of mobile reading in developing countries. Paris: Unesco.

WINTERS, N. What is mobile learning? In SHARPLES, M. Big Issues in Mobile Learning. p. 4-8). Kaleidoscope: University of Nottingham, 2006.

WORRINGER, B. et al. Common and distinct neural correlates of dual-tasking and task-switching: a meta-analytic review and a neuro-cognitive processing model of human multitasking. Brain Structure and Function, p. 1845-1869, 2019.

YANG, Z.; ASBURY, K.; GRIFFITHS, M. A Cancer in the Minds of Youth?" A Qualitative Study of Problematic Smartphone Use among Undergraduate Students. **International Journal of Mental Health and Addiction**, v. 19, p. 934–946, 2019.

YUAN, Y.; LIU, X. An Empirical Study of the Effect of ASR-Supported English Reading Aloud Practices on Pronunciation Accuracy. International Conference on Technology in Education, p. 75-87. https://doi.org/10.1007/978-981-33-4594-2

ZAMPINI, M. L. L2 speech production research: findings, issues, and advances. In Hansen Edwards, J. G. & Zampini, M. L. (Eds**.). Phonology and Second Language Acquisition** (pp. 219-249). Amsterdam: John Benjamins, 2008.

ZIELINSKI, B. The Segmental/Suprasegmental Debate. In REED, M., LEVIS, J (orgs). The Handbook of English Pronunciation. Wiley Blackwell: Oxford. p. 397 - 412, 2015.

ZIMMERMAN, B. Self-regulated Learning and Academic Achievement: **An Overview. Educational Psychologist**, v. 25, n. 1, p. 3-17, 1990.

APPRENDICES

APPENDIX A – Consent form for participants' parents

This consent form was sent for the parents and/or legal guardians of the participants who had less than 18 years of age. It was written in Portuguese as this is their L1.

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO – RESPONSÁVEL LEGAL

Prezada (o) Sra./Sr. pai/mãe ou responsável legal,

Meu nome é Daniel Reschke Pires, estudante de Doutorado da Universidade Federal de Santa Catarina (UFSC). Faço pesquisa na área de Língua Inglesa no Programa de Pós-Graduação em Inglês, sob a orientação do professor Celso Henrique Soufen Tumolo.

Convido seu filho(a) a participar na pesquisa: "Os efeitos do uso de um aplicativo baseado no Reconhecimento Automático de Fala na pronúncia da língua inglesa".

Por favor, leia este Termo de Consentimento Livre e Esclarecido para decidir se autoriza seu filho(a) a participar da pesquisa em questão.

Por que esta pesquisa está sendo realizada?

Esta pesquisa busca investigar se o uso de um aplicativo chamado "ELSA Speak" pode tornar a fala de aprendizes brasileiros da língua inglesa mais inteligível. Esta pesquisa visa contribuir à área de ensino e aprendizagem de línguas estrangeiras, uma vez que os dados coletados podem ajudar a compreender o processo de aprendizagem de língua estrangeira assistida por tecnologias móveis, as quais vem se tornando cada vez mais importantes, especialmente em modalidades de ensino não-presenciais.

O que vai acontecer?

Caso aceite participar desta pesquisa, seu filho(a) (i) responderá a um questionário de identificação e sobre sua trajetória de aprendizagem da língua inglesa, via Formulários Google; (ii) fará um teste online para avaliar sua proficiência em língua inglesa; (iii) antes de iniciar o uso do aplicativo, participará de uma sessão na qual a sua pronuncia de algumas palavras e frases em língua inglesa será gravada em áudio; (iv) caso seja alocado para o grupos experimental I, completará uma unidade do aplicativo ELSA Speak, tendo um mês para fazer isso, e caso seja alocado para o grupo experimental II utilizará este aplicativo em encontros semanais onde terá o auxílio do pesquisador para seu uso; (v) após o período de uso do aplicativo, participará de uma sessão na qual irá pronunciar algumas palavras e frases em língua inglesa; (vi) responderá um questionário via Google Forms sobre seu uso e suas percepções do aplicativo, e sobre sua experiência com a aprendizagem móvel durante a pesquisa.

A identidade dos participantes será revelada?

Durante todas as fases da pesquisa, todos os dados coletados serão absolutamente confidenciais. Não haverá divulgação de quaisquer informações que possam revelar sua identidade. Durante a coleta de dados, os dados serão armazenados em uma plataforma de armazenamento virtual segura, a qual poderá ser acessada somente pelo pesquisador e pelo professor orientador. Apesar da plataforma utilizada oferecer altíssimos níveis de proteção, existe o remoto risco, ainda que não intencional e involuntário, da quebra de sigilo. Para reduzir ainda mais este risco, os pesquisadores utilizarão uma plataforma de armazenamento paga, que fornece procedimentos de segurança adicionais.

Uma vez concluída a coleta de dados, os pesquisadores farão o download dos dados coletados para um dispositivo local e apagarão todo e qualquer registro da plataforma virtual.

Os resultados desta pesquisa poderão ser apresentados em apresentações ou revistas científicas. Todavia, serão apresentados somente os resultados obtidos como um todo, sem revelar qualquer informação pessoal sobre os participantes.

Haverá algum risco ao participar dessa pesquisa?

Embora esta pesquisa não ofereça riscos físicos, existem alguns riscos de natureza social e psicológica que os participantes poderão experienciar durante a pesquisa, entre eles: a) cansaço, aborrecimento e/ou desconforto ao responder o teste de proficiência e os questionários; b) desconforto ou constrangimento ao ter que pronunciar palavras e frases em língua inglesa; c) cansaço ou aborrecimento ao realizar as atividades do aplicativo. Salientamos que precisa seu filho(a) não precisa responder a nenhuma questão da pesquisa que lhe cause desconforto ou qualquer tipo de constrangimento.

A participação nesta pesquisa não oferece riscos de contágio por COVID-19, uma vez que todos os procedimentos serão realizados remotamente.

Haverá algum benefício ao participar dessa pesquisa?

Como benefício da participação nesta pesquisa, os participantes receberão, gratuitamente, uma assinatura do aplicativo Elsa Speak durante 3 meses. Os custos do acesso ao aplicativo serão pagos exclusivamente pelos pesquisadores. O uso do aplicativo pode auxiliar no desenvolvimento da sua pronúncia e vocabulário em língua inglesa. Além disso, após o encerramento da pesquisa, será enviado aos participantes um relatório com a análise da sua pronúncia antes e depois do uso do aplicativo, o qual poderá auxiliá-lo a compreender seus pontos fracos e fortes na pronúncia da língua inglesa.

Esta pesquisa também trará benefícios à área de ensino e aprendizagem de línguas estrangeiras, uma vez que os dados coletados podem ajudar a compreender: a) o processo de aprendizagem de língua estrangeira assistida por tecnologias móveis; b) o feedback automatizado provido por sistemas de reconhecimento de fala e c) as percepções dos alunos em relação a aprendizagem móvel. Por isso, e considerando o contexto de ensino remoto atual no qual as tecnologias móveis são amplamente utilizadas, as potenciais contribuições desta pesquisa são de grande valor para a comunidade escolar, em especial para os professores e pesquisadores de língua estrangeira.

A participação nessa pesquisa é obrigatória?

A participação na pesquisa é totalmente voluntária. Este documento se trata de um convite. A decisão de participar ou não será de cada um e será respeitada, e isso não afetará a sua relação com a instituição, tampouco as notas dos seus filhos(as).

É possível desistir de participar ou cancelar essa autorização?

Mesmo após ter aceitado participar da pesquisa, e por qualquer razão, qualquer participante poderá desistir da participação, a qualquer momento, bastando, para isto, comunicar-se com o pesquisador, por meio do e-mail daniel.reschke.pires@gmail.com ou pelo telefone (48) 99910-1405. A desistência cessará qualquer coleta de dados posteriormente e impedirá o uso de qualquer informação coletada anteriormente, e não acarretará nenhum prejuízo para o participante em termos de notas, no caso de alunos, e nem mesmo problemas de relação com a instituição.

Ao responder os questionários, os participantes têm o direito de não responder qualquer questão, sem necessidade de explicação ou justificativa para tal, podendo também se retirar da pesquisa a qualquer momento.

Haverá alguma despesa?

Não. Os custos com a assinatura do aplicativo serão pagos pelos pesquisadores. As coletas de dados irão acontecer no contraturno das aulas, de maneira remota, por isso não há gastos previstos com deslocamento.

Caso seja comprovada alguma despesa extraordinária associada à pesquisa, você será ressarcido pelos pesquisadores. Além disso, você tem o direito de indenização por danos comprovadamente decorrentes da participação nesta pesquisa.

Como faço o contato para esclarecer dúvidas?

Caso tenha alguma dúvida sobre os procedimentos, os participantes e/ou responsáveis poderão entrar em contato com os pesquisadores pelos e-mails daniel.reschke.pires@gmail.com ou celsotumolo@yahoo.com.br, ou pelo telefone (48) 99910-1405.

Se necessário contato com o Comitê de Ética em Pesquisas com Seres Humanos - CEPSH-UFSC, que é o departamento que aprova a realização desse tipo de pesquisa, você pode escolher uma das seguintes formas de contato:

Comitê de Ética em Pesquisas com Seres Humanos - CEPSH-UFSC
Prédio Reitoria II (Edifício Santa Clara),
R: Desembargador Vitor Lima, nº 222, sala 401, Trindade, Florianópolis/SC
CEP 88.040-400
Contato: (48) 3721-6094
cep.propesq@contato.ufsc.br

Sobre o CEPSH-UFSC: o CEPSH é um órgão colegiado interdisciplinar, deliberativo, consultivo e educativo, vinculado à Universidade Federal de Santa Catarina, mas independente na tomada de decisões, criado para defender os interesses dos participantes da pesquisa em sua integridade e dignidade e para contribuir no desenvolvimento da pesquisa dentro de padrões éticos.

O pesquisador responsável, Celso Henrique Soufen Tumolo, compromete-se a conduzir a pesquisa de acordo com o que preconizam as Resoluções nº 466 de 12/02/2012 e nº 510/16 de 07/04/2016, que trata dos preceitos éticos e da proteção aos participantes da pesquisa.

DECLARAÇÃO DE CONSENTIMENTO PÓS-INFORMAÇÃO

Agora que você foi esclarecido(a) sobre a pesquisa "Os efeitos do uso de um aplicativo baseado no Reconhecimento Automático de Fala na pronúncia da língua inglesa", responda a seguinte pergunta:

Você autoriza seu filho(a) a participar desse estudo?

( ) SIM

( ) NÃO

SEÇÃO 4 DO FORMULÁRIO – *Disponível apenas para aqueles que responderem SIM na seção 3.

Você receberá no seu e-mail uma cópia deste termo assinada digitalmente pelos pesquisadores responsáveis. Para isso, por favor preencha os campos abaixo. Sugerimos que você guarde salve este e-mail.

Nome do responsável

_____

Nome do filho(a):

_____

E-mail do responsável:

_____

Muito obrigado!

APPENDIX B – Assent form for participants

This was the assent form which the students from Colégio de Aplicação filled to participate in this study.

TERMO DE ASSENTIMENTO LIVRE E ESCLARECIDO – RESPONSÁVEL LEGAL

Prezado(a) estudante:

Meu nome é Daniel Reschke Pires, estudante de Doutorado da Universidade Federal de Santa Catarina (UFSC). Faço pesquisa na área de Língua Inglesa sob a orientação do professor Celso Henrique Soufen Tumolo.

Convido você̂ a participar na pesquisa: "Os efeitos do uso de um aplicativo baseado no Reconhecimento Automático de Fala na pronúncia da língua inglesa".

Por favor, leia este Termo de Consentimento Livre e Esclarecido para decidir se deseja participar da pesquisa em questão.

CLIQUE EM "PRÓXIMA" PARA PROSSEGUIR

Por que esta pesquisa está sendo realizada?

Esta pesquisa busca investigar se o uso de um aplicativo chamado "ELSA Speak" pode tornar a fala de aprendizes brasileiros da língua inglesa mais inteligível. Esta pesquisa visa contribuir à área de ensino e aprendizagem de línguas estrangeiras, uma vez que os dados coletados podem ajudar a compreender o processo de aprendizagem de língua estrangeira assistida por tecnologias móveis, as quais vem se tornando cada vez mais importantes, especialmente em modalidades de ensino não-presenciais.

O que vai acontecer?

Caso aceite participar desta pesquisa, você (i) responderá a um questionário via Google Forms sobre sua trajetória e seu perfil enquanto aprendiz de língua estrangeira; (ii) responderá um teste para avaliar sua proficiência em língua inglesa; (iii) antes de iniciar o uso do aplicativo, participará de uma sessão na qual a sua pronuncia de algumas palavras e frases em língua inglesa será gravada em áudio; (iv) caso seja alocado para o um grupo experimental, utilizará o aplicativo ELSA Speak para praticar pronúncia em língua inglesa; (v) após o período de uso aplicativo, participará de uma sessão na qual irá pronunciar algumas palavras e frases em língua inglesa; (vi) responderá um questionário via Google Forms sobre seu uso e suas percepções do aplicativo, e sobre sua experiência com a aprendizagem móvel durante a pesquisa.

A identidade dos participantes será revelada?

Durante todas as fases da pesquisa, todos os dados coletados serão absolutamente confidenciais. Não haverá identificação nominal dos participantes, nem divulgação de quaisquer informações que possam revelar sua identidade. Todos os dados serão armazenados em uma plataforma de armazenamento digital segura, a qual poderá ser acessada somente pelo pesquisador e pelo professor orientador. Apesar da plataforma utilizada oferecer altíssimos níveis de proteção, existe o remoto risco, ainda que não intencional e involuntário, da quebra de sigilo. Para reduzir ainda mais este risco, os pesquisadores utilizarão uma plataforma de armazenamento paga, que fornece procedimentos de segurança adicionais. Os resultados desta pesquisa poderão ser apresentados em apresentações ou revistas científicas. Todavia, serão apresentados somente os resultados obtidos como um todo, sem revelar qualquer informação pessoal sobre os participantes.

Haverá algum risco ao participar dessa pesquisa?

Embora esta pesquisa não ofereça riscos físicos, existem alguns riscos de natureza social e psicológica que você poderá experienciar durante a pesquisa, entre eles: a) cansaço, aborrecimento e/ou desconforto ao responder o teste de proficiência e os questionários; b) desconforto ou constrangimento ao ter que pronunciar palavras e frases em língua inglesa; c) cansaço ou aborrecimento ao realizar as atividades do aplicativo. Salientamos que você não precisa responder a nenhuma questão ao longo da pesquisa que lhe cause desconforto ou qualquer tipo de constrangimento.

A participação nesta pesquisa não oferece riscos de contágio por COVID-19, uma vez que todos os procedimentos serão realizados remotamente.

Haverá algum benefício ao participar dessa pesquisa?

Como benefício da participação nesta pesquisa, você receberá, gratuitamente, acesso ilimitado ao aplicativo Elsa Speak durante 3 meses. Os custos do acesso ao aplicativo serão pagos exclusivamente pelos pesquisadores. O uso do aplicativo pode auxiliar no desenvolvimento da sua pronúncia e vocabulário em língua inglesa. Além disso, após o encerramento da pesquisa, você receberá um relatório com a análise da sua pronúncia antes e depois do uso do aplicativo, o qual poderá auxiliá-lo a compreender seus pontos fracos e fortes na pronúncia da língua inglesa.

A participação nessa pesquisa é obrigatória?

A participação na pesquisa é totalmente voluntária. Este documento se trata de um convite. A decisão de participar ou não será de cada um e será respeitada, e isso não afetará a sua relação com a instituição, nem tampouco as suas notas.

É possível desistir de participar ou cancelar essa autorização?

Mesmo após ter aceitado participar da pesquisa, e por qualquer razão, qualquer participante poderá desistir da participação, a qualquer momento, bastando, para isto, comunicar-se com o pesquisador, por meio do e-mail daniel.reschke.pires@gmail.com ou pelo telefone (48) 99910-1405. A desistência cessará qualquer coleta de dados posteriormente e impedirá o uso de qualquer informação coletada anteriormente, e não acarretará nenhum

prejuízo para o participante em termos de notas, no caso de alunos, e nem mesmo problemas de relação com a instituição.

Haverá alguma despesa?

Não. Os custos com a assinatura do aplicativo serão pagos pelos pesquisadores. As coletas de dados irão acontecer no contra turno das aulas, de maneira remota. Caso alguma despesa extraordinária associada à pesquisa venha a ocorrer, você será ressarcido.

Como faço o contato para esclarecer dúvidas?

Caso tenha alguma dúvida sobre os procedimentos, os participantes poderão entrar em contato com os pesquisadores pelos e-mails daniel.reschke.pires@gmail.com ou celsotumolo@yahoo.com.br, ou pelo telefone (48) 99910-1405.

Se necessário contato com o Comitê de Ética em Pesquisas com Seres Humanos - CEPSH-UFSC, que é o departamento que aprova a realização desse tipo de pesquisa, você pode escolher uma das seguintes formas de contato:

Comitê de Ética em Pesquisas com Seres Humanos - CEPSH-UFSC
Prédio Reitoria II (Edifício Santa Clara),
R: Desembargador Vitor Lima, nº 222, sala 401, Trindade, Florianópolis/SC
CEP 88.040-400
Contato: (48) 3721-6094
cep.propesq@contato.ufsc.br

O pesquisador responsável, Celso Henrique Soufen Tumolo, que também rubrica e assina esse documento, compromete-se a conduzir a pesquisa de acordo com o que preconiza a Resolução 510/16 de 07/04/2016, que trata dos preceitos éticos e da proteção aos participantes da pesquisa.

Duas vias deste documento estão sendo rubricadas e assinadas por você, participante, e pelo pesquisador responsável. Guarde cuidadosamente a sua via, pois é um documento que traz importantes informações de contato e garante os seus direitos como participante da pesquisa.

Se você estiver de acordo em participar desta pesquisa, assine no espaço indicado na página seguinte.

DECLARAÇÃO DE ASSENTIMENTO PÓS-INFORMAÇÃO

Eu, _____(nome completo), declaro que fui esclarecido(a) sobre a pesquisa "Os efeitos do uso de um aplicativo baseado no Reconhecimento Automático de Fala na pronúncia da língua inglesa" e concordo em participar desse estudo e que os dados obtidos sejam utilizados para a realização da mesma.

Nome completo do responsável legal _____

_____

CPF_____

Florianópolis – SC, _____ de _____ de 2021.

_____

Assinatura do participante

_____	_____
Daniel Reschke Pires	Celso Henrique Soufen Tumolo
Doutorando-pesquisador	Orientador e pesquisador responsável

APPENDIX C – Consent forms for listeners

This consent form read and signed by the English speakers who participated as listeners in this study. This form is available on: https://forms.gle/6LdaWPqiAuDmxZPj9.

Hi!

My name is Daniel Reschke Pires. I am a PhD candidate at Universidade Federal de Santa Catarina (UFSC). I carry out research on the field of Second Language Acquisition at The Graduate Program in English (PPGI-UFSC) and my advisor is professor Celso Henrique Soufen Tumolo.

This document invites you to participate as a listener in the research project named "The effects of the use of an ASR-based application on the speech intelligibility of English".

Please read this Consent Form to understand the role of listeners in this research and to decide if you want to take part in it.

- Why is this research being conducted?

This research seeks to investigate if the use of an ASR-based mobile application can improve the speech intelligibility of Brazilian learners of English. This research aims to contribute to the field of Mobile Assisted Language Learning (MALL) as the data acquired can help understand the educational uses of technology and how mobile devices can support language learning, especially in times of remote learning.

- What will be your role in this research?

In this research, high school students from Colégio de Aplicação (UFSC) will be invited to use an application called "ELSA Speak" for a month. They will take pronunciation tests to measure how intelligible their speech is before and after the use of the application. In

this research, the intelligibility of the participants will be measured by the ortograpic transcription done by listeners.

Should you accept this invitation, your role will be to listen and orthographically transcribe sentences recorded by the participants of the research. This means you will receive audio files recorded by the students, listen to them and type what you hear in a text file. You will also answer a questionnaire with questions regarding your experience with the English and the Portuguese languages.

- Will my identity be revealed?

Throughout all the phases of the research, all collected data will be strictly confidential. There will be no sharing of any information that may reveal your identity. During the data collection, all data will be stored in a safe, virtual platform, which will be accessed only by the researcher and his advisor. Although this platform offers high levels of protection, there is a risk, even if remote and involuntary, of data disclosure. To reduce this risk even further, the researchers shall use a paid storage cloud, which provides additional safety measures.

Once the data collection is concluded, the researchers will download all collected data to a local device and will erase any data on the cloud.

- Is there any risk involved in participating in this research?

Even though this research has no physical risks involved, there are some risks of social and psychological nature that the participants may feel during the research. Listeners may feel bored, tired and/or frustrated as a result of transcribing the sentences.

Your participation in this project does not pose any COVID-19 risks, as all the procedures will be conducted remotely.

- Are there any benefits to participating in this research?

Even though there are no direct benefits for those who decide to participate as listeners, this research will be beneficial to teachers and learners of English as a second/foreign/additional

language, as it will allow us to understand: a) the process of teaching and learning a second language with the aid of mobile technologies; b) the feedback given by Automatic Speech Recognition (ASR) technologies; and c) the perceptions of the students regarding mobile learning. Considering that mobile technologies are widely used for remote learning these days, the potential contributions of this research are of great value for the school community, especially for teachers and researchers of foreign languages.

- Am I required to participate in this research?

No. Participation is completely voluntary. The decision to participate or not will be respected, and it will not affect your relationship with the researcher or the institution.

- Can I quit after accepting to participate?

Yes. Even after accepting to participate, you can leave at any point. To do so, please email me at daniel.reschke.pires@gmail.com or at the phone (48) 99910-1405. By doing so, no further data will be collected, and any data previously collected will not be used. This will not affect your relationship with the researcher or the institution.

- Will there be any expense?

No. However, if there are any proven extraordinary expenditures associated with the research, you will be reimbursed. Also, you have the right to compensation in case there is any damage resulting directly from your participation in this research.

- How do I get in touch if have any questions?

If you have any doubts or questions, you can be in touch with one of the researchers via the emails daniel.reschke.pires@gmail.com or celsotumolo@yahoo.com.br.

This research has been approved by the Ethics Committee from UFSC under the number 44097421.0.0000.0121. This committee is responsible for approving this type of research. You can reach them to verify the integrity of this research with following information:

Comitê de Ética em Pesquisas com Seres Humanos - CEPSH-UFSC
Prédio Reitoria II (Edifício Santa Clara),
R: Desembargador Vitor Lima, nº 222, sala 401, Trindade, Florianópolis/SC
CEP 88.040-400
Contato: (48) 3721-6094
cep.propesq@contato.ufsc.br

About CEPSH-UFSC: this committee is an interdisciplinary, deliberative, advisory, and educational organization, which is attached to UFSC, but independent in its decision taking. It was created to defend the interests of research participants and their integrity and dignity, and to contribute to the development of research within ethical standards.

The responsible researcher, Celso Henrique Soufen Tumolo, is committed to conducting this research according to resolutions nº 466 de 12/02/2012 and nº 510/16 de 07/04/2016, which provide the ethical principles and the protections to research participants.

Having read and understood your role and the terms of this research, please answer the following question:

Do you want to participate in this research as a listener?

( )Yes

( ) No

Thank you very much!

We will get in touch soon with further instructions.

You will receive a copy of this term digitally signed by the researchers. For this, please write your name and e-mail below. Make sure your email is typed correctly. We suggest you keep a copy of this email.

APPENDIX D – Questionnaire on the participants' English learning background

The participants answered a questionnaire on Google Forms to gather data on their background as learners of English. This questionnaire is available on: https://forms.gle/zn8YAuUY38NUUefh7

APPENDIX E – English proficiency test

This is the English proficiency test which the participants took before starting the use of the application: https://www.esl-languages.com/en/online-language-tests/english-test

APPENDIX F – Intelligibility tests

The intelligibility pre-test can be accessed at https://form.jotform.com/211473775259666.

The intelligibility post-test can be accessed at https://form.jotform.com/212366160101642.

APPENDIX G – Sentences of the intelligibility tests

These are the sentences the participants recorded in the pre and post intelligibility tests, presented here in the order in which they read them.

	Sentence
1	Fiction fans are detailed planners
2	I really like this beat
3	These women are famous athletes
4	A lot of good books
5	I really like this bit
6	That movie was so captivating
7	The tickets sold out fast
8	I didn't feel the hit
9	They will make a sequel
10	They released the new album
11	I didn't feel the heat
12	I recommend playing beach volleyball
13	Fiction readers are good listeners
14	This is where I leave
15	He works in the field
16	He wrote the best script
17	I don't think they fit
18	She sings at the karaoke
19	This is where I live
20	She works with her feet
21	We listen to country music
22	They will sell the sheep
23	I live on the beach
24	They will sell the ship

APPENDIX H – Sentences of in the intelligibility tests, sorted by category

These are the sentences the participants recorded in the pre and post intelligibility tests, presented here according to three sentence categories.

Type of sentence	Sentence
Sentences with minimal pairs	I really like this **beat**
	I really like this **bit**
	I didn't feel the **heat**
	I didn't feel the **hit**
	This is where I **leave**
	This is where I **live**
	They will sell the **sheep**
	They will sell the **ship**
	I don't think they **feet**
	She works with her **fit**
Sentences with target words	These women are famous **athletes**
	That movie was so **captivating**
	They will make a **sequel**
	They released the new **album**
	I recommend playing beach **volleyball**
	Fiction readers are good **listeners**
	He works in the **field**
	He wrote the best **script**
	She sings at the **karaoke**
	These women are famous **athletes**
Distractors	Fiction fans are detailed *planners*
	A lot of good *books*
	The tickets sold out *fast*
	The released the new *album*

APPENDIX I – Sentence Transcription Task

An example of the sentence transcription task carried out by the listeners is available on: https://forms.gle/Duk3Y3t5tkYaNWfu9.

APPENDIX J - Instructions for participants on how to use ELSA Speak.

In this file, the participants could access the instructions and suggestions on how to use ELSA Speak appropriately in this study. It is available on: https://drive.google.com/file/d/1rF9yFkvLOyZE9OcrK9kN4lT25kfveREa/view?usp=sharing

APPENDIX K – Questionnaire on the participants' use and perceptions of the application.

After completing the assigned lessons, the participants answered a Google Forms questionnaire with close and open-ended questions about their use of ELSA Speak, and their perceptions of the application and mobile learning. It is available on: https://docs.google.com/forms/d/e/1FAIpQLSdeqf_BshWsy4p4ntP0-wrz4yW8WKzObrzV49vx_45IET_Q3g/viewform?fbzx=3946557328700795657

APPENDIX L – Consent Form for teachers

This is the consent form sent to the English teachers from Colégio de Aplicação, whose syllabi was read in order to verify if there were plans to work with the target sounds of this study.

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO PARA OS PROFESSORES DOS PARTICIPANTES

Prezado(a) professor(a)

Meu nome é Daniel Reschke Pires, estudante de Doutorado da Universidade Federal de Santa Catarina (UFSC). Faço pesquisa na área de Língua Inglesa no Programa de Pós-Graduação em Inglês, sob a orientação do professor Celso Henrique Soufen Tumolo.

Convido você̂ a participar na pesquisa: "Os efeitos do uso de um aplicativo baseado no Reconhecimento Automático de Fala na pronúncia da língua inglesa".

Por favor, leia este Termo de Consentimento Livre e Esclarecido para decidir se deseja participar da pesquisa em questão.

Por que esta pesquisa está sendo realizada?

Esta pesquisa busca investigar se o uso de um aplicativo baseado no Reconhecimento Automático de Fala pode tornar a fala de aprendizes brasileiros da língua inglesa mais inteligível. Esta pesquisa visa contribuir à área de ensino e aprendizagem de línguas estrangeiras, uma vez que os dados coletados podem ajudar a compreender o processo de aprendizagem de língua estrangeira assistida por tecnologias móveis, as quais vem se tornando cada vez mais importantes, especialmente em modalidades de ensino não-presenciais.

O que vai acontecer?

Nesta pesquisa, os estudantes de inglês do ensino médio do Colégio de Aplicação da UFSC serão convidados a utilizar o aplicativo "ELSA Speak", e farão testes para verificar os níveis de inteligibilidade antes e depois do uso do aplicativo.

Uma vez que a pesquisa busca compreender os efeitos do uso do aplicativo na pronúncia, e considerando que os participantes estarão tendo aulas regulares de inglês durante o período de uso do aplicativo, será necessário verificar, antes do início do uso do aplicativo, os planos de ensino dos(as) professores(as) da língua para saber se existem momentos planejados para o ensino explícito de pronúncia. Este será o único intuito da verificação dos planos de ensino, não sendo observados ou analisados quaisquer outros fatores. Também não será requisitada qualquer alteração nos planos, mesmo que existam planos para o ensino de pronúncia.

A identidade dos participantes será revelada?

Durante todas as fases da pesquisa, todos os dados coletados serão absolutamente confidenciais. Não haverá divulgação de quaisquer informações que possam revelar sua identidade. Durante a coleta de dados, os dados serão armazenados em uma plataforma de armazenamento virtual segura, a qual poderá ser acessada somente pelo pesquisador e pelo professor orientador. Apesar da plataforma utilizada oferecer altíssimos níveis de proteção, existe o remoto risco, ainda que não intencional e involuntário, da quebra de sigilo. Para reduzir ainda mais este risco, os pesquisadores utilizarão uma plataforma de armazenamento paga, que fornece procedimentos de segurança adicionais.

Uma vez concluída a coleta de dados, os pesquisadores farão o download dos dados coletados para um dispositivo local e apagarão todo e qualquer registro da plataforma virtual.

Os resultados desta pesquisa poderão ser apresentados em apresentações ou revistas científicas. Todavia, serão apresentados somente os resultados obtidos como um todo, sem revelar qualquer informação pessoal sobre os participantes.

Haverá algum risco ao participar dessa pesquisa?

Embora esta pesquisa não ofereça riscos físicos, existem alguns riscos de natureza social e psicológica que os participantes poderão ter durante a pesquisa. No caso dos(as) professores(as) participantes, existe o risco de constrangimento e ou aborrecimento por conta da observação dos planos de ensino.

A participação nesta pesquisa não oferece riscos de contágio por COVID-19, uma vez que todos os procedimentos serão realizados remotamente.

Haverá algum benefício ao participar dessa pesquisa?

Embora não haja benefícios diretos para os professores-participantes, esta pesquisa trará contribuições à área de ensino e aprendizagem de línguas estrangeiras, uma vez que os dados coletados podem ajudar a compreender: a) o processo de aprendizagem de língua estrangeira assistida por tecnologias móveis; b) o feedback automatizado provido por sistemas de reconhecimento de fala e c) as percepções dos alunos em relação a aprendizagem móvel. Por isso, e considerando o contexto de ensino remoto atual no qual as tecnologias móveis são amplamente utilizadas, as potenciais contribuições desta pesquisa são de grande valor para a comunidade escolar, em especial para os professores e pesquisadores de língua estrangeira.

A participação nessa pesquisa é obrigatória?

A participação na pesquisa é totalmente voluntária. Este documento se trata de um convite. A decisão de participar ou não será de cada um e será respeitada, e isso não afetará a sua relação com a instituição ou com o pesquisador.

É possível desistir de participar ou cancelar essa autorização?

Mesmo após ter aceitado participar da pesquisa, e por qualquer razão, qualquer participante poderá desistir da participação, a qualquer momento, bastando, para isto, comunicar-se com o pesquisador, por meio do e-mail daniel.reschke.pires@gmail.com ou pelo telefone (48) 99910-1405. A desistência cessará qualquer coleta de dados posteriormente e impedirá o uso de qualquer informação coletada anteriormente, e não acarretará nenhum prejuízo para o participante.

Haverá alguma despesa?

Não. Todavia, caso seja comprovada alguma despesa extraordinária associada à pesquisa, você será ressarcido pelos pesquisadores. Além disso, você tem o direito de indenização por danos comprovadamente decorrentes da participação nesta pesquisa.

Como faço o contato para esclarecer dúvidas?

Caso tenha alguma dúvida sobre os procedimentos, os participantes e/ou responsáveis poderão entrar em contato com os pesquisadores pelos e-mails daniel.reschke.pires@gmail.com ou celsotumolo@yahoo.com.br, ou pelo telefone (48) 99910-1405.

Se necessário contato com o Comitê de Ética em Pesquisas com Seres Humanos - CEPSH-UFSC, que é o departamento que aprova a realização desse tipo de pesquisa, você pode escolher uma das seguintes formas de contato:

Comitê de Ética em Pesquisas com Seres Humanos - CEPSH-UFSC
Prédio Reitoria II (Edifício Santa Clara),
R: Desembargador Vitor Lima, nº 222, sala 401, Trindade, Florianópolis/SC
CEP 88.040-400
Contato: (48) 3721-6094
cep.propesq@contato.ufsc.br

Sobre o CEPSH-UFSC: o CEPSH é um órgão colegiado interdisciplinar, deliberativo, consultivo e educativo, vinculado à Universidade Federal de Santa Catarina, mas independente na tomada de decisões, criado para defender os interesses dos participantes da pesquisa em sua integridade e dignidade e para contribuir

no desenvolvimento da pesquisa dentro de padrões éticos.

O pesquisador responsável, Celso Henrique Soufen Tumolo, compromete-se a conduzir a pesquisa de acordo com o que preconizam as Resoluções nº 466 de 12/02/2012 e nº 510/16 de 07/04/2016, que trata dos preceitos éticos e da proteção aos participantes da pesquisa.

DECLARAÇÃO DE CONSENTIMENTO PÓS-INFORMAÇÃO

Agora que você foi esclarecido(a) sobre seu papel na pesquisa "Os efeitos do uso de um aplicativo baseado no Reconhecimento Automático de Fala na pronúncia da língua inglesa", responda a seguinte pergunta:

Você deseja participar desse estudo?

( ) SIM

( ) NÃO

Você receberá no seu e-mail uma cópia deste termo assinada digitalmente pelos pesquisadores responsáveis. Para isso, por favor preencha os campos abaixo. Sugerimos que você guarde salve este e-mail.

Nome do professor:

_____

Escreva seu e-mail abaixo:

_____

Muito obrigado!

FIM DO TCLE