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

FHELIPPE WALTELON SOUZA DOS SANTOS

UNDERSTANDING THE DIMENSIONS OF AUTOMATIC SPEECH RECOGNITION (ASR): AN ASSESSMENT OF THE RELATIONSHIP BETWEEN FOREIGN LANGUAGE VOWEL PRODUCTION, AFFECTIVE FACTORS AND TECHNOLOGY USE

FLORIANÓPOLIS
2024

Fhelippe Waltelon Souza dos Santos

Understanding the dimensions of Automatic Speech Recognition (ASR): an assessment of the relationship between foreign language vowel production, affective factors and technology use

Tese submetida ao Programa de Pós-graduação em Inglês: Estudos Linguísticos e Literários da Universidade Federal de Santa Catarina para a obtenção do título de Mestre em Inglês: Estudos Linguísticos e Literários
Orientador: Prof. Dr. Hanna Kivistö-de Souza

Florianópolis
2024

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Santos, Fhelippe Waltelon Souza dos
Understanding the dimensions of Automatic Speech Recognition (ASR): an assessment of the relationship between foreign language vowel production, affective factors and technology use / Fhelippe Waltelon Souza dos Santos ; orientadora, Hanna Kivistö-de Souza, 2024.
110 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro de Comunicação e Expressão, Programa de Pós-Graduação em Letras/Inglês e Literatura Correspondente, Florianópolis, 2024.

Inclui referências.

1. Inglês. 2. Prazer em Aprendizagem de uma Língua Estrangeira,. 3. Ansiedade de Pronúncia. 4. Inteligibilidade. 5. Reconhecimento Automático de Fala,. I. Souza, Hanna Kivistö-de . II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Letras/Inglês e Literatura Correspondente. III. Título.

Fhelippe Waltelon Souza dos Santos

Understanding the dimensions of Automatic Speech Recognition (ASR): an assessment of the relationship between foreign language vowel production, affective factors and technology use

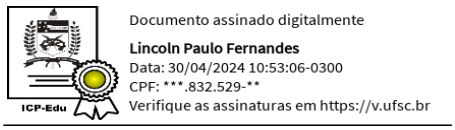
O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Profª. Drª. Hanna Kivistö-de Souza
Universidade Federal de Santa Catarina (UFSC)

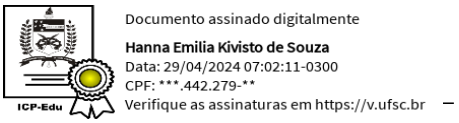
Prof. Dr. Ronaldo Mangueira Lima Jr
Universidade de Brasília (UnB)

Prof. Dr. Walcir Cardoso
Concordia University (Montreal, Canada)

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título Mestre em Inglês: Estudos Linguísticos e Literários, na área de concentração Estudos da Linguagem



Prof. Dr.(a)
Coordenador(a) do Programa



Prof. Dr.(a)
Orientador(a)

Florianópolis, 22 de Abril de 2024

ACKNOWLEDGEMENTS

This MA thesis is first and foremost a dedication to all teachers and researchers who built the foundations of my knowledge in the field of phonetics and phonology. Before these long two years, I was only peripherally aware of the field, but today I feel closer to an expert. And if only not an expert yet, it is because those who taught me have shown endless wisdom to which I still aspire to have. For that in particular, and for the opportunity given, thank you Prof. Dra. Isabel Seara for your knowledge in acoustic and articulatory phonetics. Thank you Prof. Dra. Rosane Silveira for assisting me during the MA process and in providing me with insights on the design of my research. Thank you Dr. Ronaldo Lima Jr. for participating in my qualifying session and giving me invaluable feedback on my thesis.

Most importantly, thank you Prof. Dra. Hanna Kivistö-de Souza for believing in me during these 2 years and in enabling me to carry out my first quantitative study and my first study in the field of phonetics and phonology. All of the assistance you provided me during each step of the study is and was central to the success of this present thesis. It is because of the example of brilliance that I saw in your work that I felt the necessity to give all of my energy to craft something you would at least be proud of participating in.

I'd like to extend my gratitude to PPGI and to FAPESC for making this opportunity exist in the first place, as without the structures and support behind me, I would not have been able to conclude and present my ideas. The stability I had during my MA process aided me in dedicating all of my time to the process of writing, collecting data, and applying my experiment.

When it comes to support, none other than my family comes first in this. Without the support I received from my mother, Claudia Sinara de Souza, and from my father, Luciano dos Santos, I wouldn't have created the habit, routine, and opportunities necessary to further explore my dreams of becoming an aspiring researcher. Had there been any doubts about my success many years ago, I hope that I can make my family prouder with each step I take, until I can give back to my parents for the time they invested so intensely in me.

My dearest friends, Vitor Marques and Elis Regina, are also pillars which hold my work above ground today. The very few who stuck around to see the results of this long road are those who I cherish lovingly, and I hope that we'll continue together as a close bond the

next time I have to write future acknowledgements. Thanks for all the moments we shared together in the last two years.

Lastly, thank you Patricia Pimenta. Because of you, today I can stand on two feet. I love you, *thank you for the memories.*

ABSTRACT

Pronunciation is argued to be one of the most anxiety-provoking skills for FL language learners. In particular, overcoming the hurdle that is anxiety in pronunciation classes may have to tackle intrinsic aspects to the FL learning context, which involves learners' self perceptions, and fear of negative evaluation from their peers and professors. Furthermore, EFL pronunciation instruction may also greatly benefit from the feeling of joy; fostering enjoyment in learners of FL English has been a widely discussed individual difference in language studies. As an attempt to aid learners with their affective factors, beliefs, and overall intelligibility, this study implements automatic speech recognition (ASR) through consciousness-raising activities of two target English vowel pairs in order to determine the effect of an out-of-classroom tool by which learners can foster autonomous practices in isolated contexts. Thus, this study applied the tool *Microsoft Word Dictate* on L1 Brazilian Portuguese speakers of FL English (n= 30) through a longitudinal quasi-experimental study, where participants' FL English vowel intelligibility, pronunciation anxiety, and foreign language enjoyment were measured before and after remote ASR instruction. This present study posits that higher levels of foreign language enjoyment and lower degrees of pronunciation anxiety are linked to higher FL English vowel intelligibility scores, and that while some change can be observed from pre-test to post-test, ASR tool MW Dictate did not result in significant change in participants' anxiety, enjoyment, or overall intelligibility. While participants' overall intelligibility scores averaged 50%, and despite seeing little change after instruction, it is believed that approaches to learning FL English pronunciation in such manner may allow participants to develop awareness regarding certain FL phonological aspects in a controlled environment that does not pose significant threats to participants' individual differences.

Keywords: Pronunciation Anxiety, Foreign Language Enjoyment, Intelligibility, Automatic Speech Recognition, Consciousness-raising Activities

RESUMO

A pronúncia é considerada uma das habilidades que mais provoca ansiedade nos alunos de línguas estrangeiras. Em particular, a superação do obstáculo que é a ansiedade em aulas voltadas para a pronúncia pode ter de lidar com aspectos intrínsecos ao contexto de aprendizado de inglês como língua estrangeira, que envolve a autopercepção dos alunos sobre si mesmos e o medo da avaliação negativa de seus colegas e professores. Além disso, o ensino da pronúncia do inglês como língua estrangeira também pode se beneficiar muito da sensação de alegria, e promover o prazer nos alunos de inglês fluente no idioma tem sido uma diferença individual amplamente discutida nas pesquisas de estudos linguísticos. Na tentativa de ajudar os alunos com seus fatores afetivos, crenças e inteligibilidade, este estudo implementa o reconhecimento automático de fala (RAF) por meio de atividades de conscientização de dois pares alvos de vogais em inglês para determinar o efeito de uma ferramenta utilizada fora da sala de aula por meio da qual os alunos podem promover práticas autônomas em contextos isolados. Assim, este estudo aplicou a ferramenta Microsoft Word Dictate em falantes do português brasileiros não-nativos de inglês como língua estrangeira (n=30) por meio de um estudo quase experimental longitudinal, no qual a inteligibilidade das vogais do inglês como língua estrangeira, a ansiedade de pronúncia e o prazer com a língua estrangeira dos participantes foram medidos antes e depois da instrução remota via RAF. O presente estudo postula que níveis mais altos de satisfação com a língua estrangeira e graus mais baixos de ansiedade de pronúncia estão ligados a escores mais altos de inteligibilidade de vogais do inglês, e que embora algumas mudanças possam ser observadas do pré-teste para o pós-teste, a ferramenta de RAF Microsoft Word Dictate não resultou em mudanças significativas nas crenças dos participantes ou na inteligibilidade. Embora a média das pontuações de inteligibilidade dos participantes tenha sido apenas 50% e, apesar de ter havido pouca mudança após a instrução, acredita-se que estas formas de abordagens para o aprendizado da pronúncia do inglês como língua estrangeira podem permitir que os participantes desenvolvam consciência em relação a determinados aspectos fonológicos do inglês como língua estrangeira em um ambiente controlado que não apresente ameaças às diferenças individuais dos participantes.

Palavras-Chave: Prazer em Aprendizagem de uma Língua Estrangeira, Ansiedade de Pronúncia, Inteligibilidade, Reconhecimento Automático de Fala, Atividades de Conscientização

LIST OF APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASR - Automatic Speech Recognition

CALL - Computer-Assisted Language Learning

CAPT - Computer-Assisted Pronunciation Teaching

EFL - English as a Foreign Language

ESL - English as a Second Language

FL - Foreign Language

FLCAS- Foreign Language Classroom Anxiety Scale

FLE - Foreign Language Enjoyment

IPA - Intelligent Personal Assistant

TSA - Time Spent Abroad

MPA - Measure of Pronunciation Anxiety

PA - Pronunciation Anxiety

SLM-r - Revised Speech Learning Model

TABLE OF CONTENTS

1. INTRODUCTION

Some of the biggest hurdles in foreign language learning are related to the negative effects of anxiety, as it can hinder learners' willingness to participate in class activities, and cloud their self-perception regarding their own linguistic skills (Dewaele; MacIntyre, 2014; Horwitz; Horwitz; Cope, 1986; MacIntyre et al, 1998). As Horwitz and colleagues first discussed, matters such as communication apprehension and fear of negative evaluation can trigger the feeling of foreign language anxiety in students, and are both elements that can occur due to the classroom environment due to peer and teacher judgment, and learners' self-perceptions of their own FL proficiency (Horwitz et al., 1986). For the past decades, research on foreign language anxiety has been frequently discussed in the area of linguistic studies (reading, writing, speaking, and listening). In specific, an advancement made in analyzing pronunciation anxiety (Baran-Łucarz, 2014, 2014b, 2016, 2017), which had earlier been measured by generic tools such as different tools such as the Foreign Language Classroom Anxiety Scale (Shams, 2006), or measurements on speaking anxiety. Research has also shifted from studies on FLA towards foreign language enjoyment (Dewaele; Macintyre, 2014, Dewaele et al., 2017), emphasizing the role of positive emotions in foreign language (FL) classrooms.

One tool that has been proven to reduce anxiety is ASR, a technology that converts speech into text, and used by many applications such as word processors (Microsoft Word, Google Docs), and language learning apps (Duolingo, ELSA). Many studies on ASR done in the past have reported a number of limitations which hindered the ability of the tool in being desirable for non-native speakers of English. In particular, the biggest challenge present in the past decades were in how ASR technologies in the past had an inability to carry out human-like transcription of speech recordings (Coniam, 1999, Derwing et al. 2000). Since this technology was often trained to understand and/or transcribe speech produced by standard varieties of English produced by native speakers, non-native speakers often struggled to make themselves understood by the algorithm (Edalatishams, 2017). Considering the linguistic focus that has placed intelligibility above measures of accentedness and accuracy (see chapter 2 of this thesis) (Chun, 2019; Derwing, 2010; Mroz,

2018; Thompson; Derwing, 2014), there has been more attention being paid to ASR as they are now developed to include and transcribe varieties of non-native English. As McCrocklin and Edalatishams (2020) affirm: ASR technologies today do appear to significantly close the gap between human and machine transcription ability. Consequently, this current study believes that in line with recent studies which have considered ASR as a tool for learning FL pronunciation and speaking, more consideration should be given to this technology since they now may provide for novel didactic designs that contribute with learners' intelligibility by providing for learning opportunities that extend beyond regular classroom settings. In addition, the field of Computer-Assisted Pronunciation Teaching (CAPT) has seen enhancements in the area of technologies pertaining to ASR (Tejedor-Garcia et al, 2020), which may provide benefits for addressing affective factors in FL learners (Bashori et al, 2021).

There has been a significant increase in studies analyzing pronunciation tools, apps, and ASR (Baldissera, 2020; Mroz, 2018; Tejedor-Garcia et al. 2020). However, it could be interesting to place more focus on ASR tools' potentiality as a didactic approach for EFL learners who do not enjoy practicing English pronunciation, either due to anxiety or boredom. Furthermore, pronunciation instruction is often avoided by teachers due to their lack of confidence and/or adequate didactic materials (Silveira, 2006). This study proposes that such challenges could be amended with digital technologies that provide for an out-of-classroom approach to instruction, permitting EFL students to practice their pronunciation from home with the use of cheap and accessible devices.

The objective of this study is to observe the effects of an Automatic Speech Recognition tool on EFL learners' individual differences regarding pronunciation anxiety (PA) and foreign language enjoyment (FLE), while also assessing its applicability for English vowel instruction. Concerned with the parallels of intelligibility, anxiety, enjoyment, and speech technologies such as ASR, this study aims to analyze the relationship between these variables through a quasi-experimental and quantitative approach that observes intermediate-level EFL students in environments tailored for pronunciation learning through ASR. By using tools such as the *Measurement of Pronunciation Anxiety* (MPA) (Baran-Łucarz, 2017), together with the measure of foreign language enjoyment (FLE) (Dewaele; Macintyre, 2014), pronunciation assessment tasks, and remote ASR instructions

for English vowel intelligibility, this study hopes to provide future researchers and instructors in the field with a clearer perspective on the applicability of this digital technology for pronunciation instruction. The results provided in this thesis may be used by EFL teachers to consider the specificities of the ASR tool *Word Dictate* as to understand the qualities and limitations linked to the design and features linked to this digital technology, further cementing considerations for how and when professionals should rely on these tools.

This thesis is organized into four different sections. Section 2 provides a literature review of contents related to the research providing an overview of intelligibility and foreign language speech learning, linking aspects such as autonomous learning and technology with benefits for FL speech instruction. Then, learners' individual differences for EFL/ESL pronunciation are discussed, followed by an overview of the field of ASR throughout the decades. These two subsections hope to argue why pronunciation anxiety and foreign language enjoyment could be aided with the use of digital technologies such as ASR.

Section 3 is dedicated to presenting the method of the study. Here, participants, study duration, assessment tools, treatment, data collection and procedures are described. The study's conceptual framework, instruments and design are presented to clarify each step of the data collection.

Section 4 presents this project's results and discussion. Here, the research questions are answered and discussed in order to argue whether ASR can be a pedagogically viable tool to aid participants with their FL pronunciation skills..

Finally, Section 5 summarizes and analyzes the findings obtained by pointing out the pedagogical implications of using ASR for FL speech learning. It also reports the limitations of the study, and the future directions for further research.

2. REVIEW OF LITERATURE

2.1 INTELLIGIBILITY AND FOREIGN LANGUAGE SPEECH LEARNING

This section aims to elaborate on the differences between accuracy and intelligibility, and on which aspects of FL communication are to be prioritized when conducting research in FL pronunciation. Then, FL pronunciation instruction is discussed by observing how FL pronunciation is taught in classrooms, and how certain consciousness-raising activities can aid learners in developing FL phonological awareness through the development of proceduralized explicit knowledge of FL pronunciation features and sounds (Kivisto-De Souza, 2015). Finally, the pronunciation difficulties of Brazilian Portuguese learners of English are addressed.

For this present study, it is important to understand what is considered intelligible speech. In specific, Munro and Derwing (1995) defined three dimensions to assess pronunciation- *intelligibility*, *comprehensibility,* and *accentedness*. *Intelligibility*, according to the authors, is "(...)the extent to which a speaker's message is actually understood by a listener" (Munro; Derwing, 1999, p. 289), which differs from the dimension of *comprehensibility*, as the latter is concerned with the listeners' perceived ease of comprehension and judgment of ease of a certain utterance. These dimensions, according to Munro and Derwing (1999), are separate from a speakers' degree of *accentedness*, which while related to intelligible speech, does not equate to a direct measure of level of intelligibility. Thus, the authors suggested that language instructors should not exclusively judge students' accent as the problematic phonological factor if their objective is to teach and assess their degrees of intelligibility and comprehensibility, apart from specific situations where the accented speech directly interferes with the other dimensions of FL speech.

Levis (2005), building upon Munro and Derwing's concept of intelligibility, further developed the differences between the *nativeness principle* and the *intelligibility principle*. The author claims that most pronunciation materials leaned towards the reduction of speaker accent and promotion of accuracy, as is in line with the *nativeness principle*. The *nativeness principle* reinforces that achieving near-native pronunciation is the ultimate goal, thus

claiming that it is desirable to reach a state of native-like pronunciation. This principle in particular guided and still guides many pronunciation instruction practices aimed at the idea of reducing or eliminating a foreign accent, often in an attempt to lead instructors into teaching language learners a way to sound like a native speaker of a specific FL (Levis, 2005). As will be discussed in the following sections, this phenomenon is reflected in the development of ASR tools, which despite not being developed specifically for language learning purposes, also posed significant challenges in acknowledging non-native speakers of English as intelligible speakers in their transcriptions (Coniam, 1999; Derwing, 2000; Mccrocklin; Edalatishams, 2020).

The *intelligibility principle*, on the other hand, is founded upon the idea that "speakers simply need to be understandable" (Levis, 2005, P. 370). It returns to the proposals by Munro and Derwing (1999) that actively separated the dimensions of accentedness and intelligibility, reinforcing that accented speech can still be well understood and intelligible. Levis' claims that within the *intelligibility principle*, teaching requires using didactic materials and instruction methods that give learners opportunities to learn and prioritize aspects of pronunciation that affect overall intelligibility, both at the word and prosodic level (Levis; Silpachai, 2022). Often, measures of intelligibility are analyzed by giving select speech samples to a group of listeners tasked with transcribing the spoken input as accurately as possible (Derwing, Munro, 1999). Other forms of measuring intelligibility can include assessing intelligibility at the level of discourse by listener recall of main ideas and details (Hahn, 2004) assessing intelligibility at the segment level by listener perception of minimal pair words varying on the vowel segment (Lee; Lyster, 2016), and utilizing ASR transcriptions to measure intelligibility of accented speech (Gottardi; Kivistö-de Souza, 2022; McCrocklin; Edalatishams, 2020; Mroz, 2018). It is important to remember that intelligibility is not entirely distinct from a measurement of accuracy, and that assessing intelligibility at the segmental level requires understanding what sounds are problematic for intelligible communication, thus requiring higher degrees of accurate production.

Turning attention to the teaching of pronunciation, it is important to emphasize that pronunciation teaching often struggles with having a limited presence in the classroom, and that the amount of time dedicated for actual pronunciation teaching to occur is often little to

none (Kivistö-de Souza, 2015; Silveira, 2004). Historically, FL pronunciation was often relegated to being a proficiency that learners would "pick up" on naturally by being exposed to input (Celce-Murcia Brinton; Goodwin, 2010), but the idea that implicit learning of any linguistic feature can occur has been contested (Schmidt, 1990, 2010). Specifically Schmidt (1990), when presenting the *noticing hypothesis* argues that learning can only occur when learners are aware at the time of learning. Unlike L1 speech, FL speech often requires speakers to "consciously notice the form in the L2 phonology and consciously notice the gap between their interlanguage perception and production, and the target language." (Kivistö-de Souza, 2015, p. 91). That is, at some point during their FL learning process, learners must develop awareness of the FL phonological process in question, and then notice differences between their production as opposed to what has been noted in the retrieved intake. This, combined with the arduous task of re-organizing their L1 articulatory and phonetic systems in order to accommodate new FL sounds (Flege; Bohn, 2021), indicate that teachers should at least promote awareness of salient FL features so that FL learners can begin bridging the gap between their production and the expected goals within the TL.

In order to promote FL phonological awareness, it is important to aid learners by providing them with consciousness-raising activities regarding the desirable FL target sounds. Consciousness-raising is defined by Sharwood Smith (1981) as the act of drawing attention to a linguistic feature by providing explicit knowledge of said feature in order to facilitate language learning. The author claims that this ultimately creates or provides opportunities for learners to access shortcuts to explicit knowledge that can aid them in the learning process. These consciousness-raising activities can often be both highly explicit and elaborate, and or simple and indirect, as long as awareness is placed upon the desired TL feature.

Some empirical studies evidence that language awareness does correlate with higher FL pronunciation ratings (Kennedy; Trofimovich, 2010), that high-variability perceptual training results in FL perceptual improvement (Carlet; Cebrian, 2014; Rato, 2014), and that consciousness-raising through practice with automatic speech recognition software leads to higher awareness by providing opportunities for learners to notice the gap between actual and desirable TL productions (Mroz, 2018). The latter study in particular analyzed how learners of French as a foreign language could become more aware of their own degree of

intelligibility through the use of the automatic speech recognition built in Gmail. This study tested two student groups enrolled in different FL speech course dynamics throughout four months by giving participants tasks related to spontaneous, semi-structured interactions with the automatic speech recognition tool. The results bundle together all of the aforementioned aspects deemed so far relevant for FL pronunciation learning to occur (autonomous learning, awareness and FL phonological awareness, additional out-of-classroom solutions to limited classroom time) through the application of a consciousness-raising activity. Learners became more aware of how intelligible they were by noticing the gap between themselves and the automatic speech recognition tool, they attempted to bridge the gap between their productions and the resulting text transcriptions of the tool, and the gap between their declarative and procedural knowledge. They were also provided with opportunities for activation of explicit knowledge of the features of the TL phonological system through autonomous learning, and how ASR use allowed them to reflect on the difficulties of the TL phonological system.

One of the tasks pronunciation instructors need to tackle is to decide which aspects of pronunciation are relevant for intelligibility, in other words: which aspects should be taught. For example, when teaching segmental features, it is possible to select appropriate target sounds concerning aspects of pronunciation that present a higher functional load (Munro; Derwing, 2006). Brown (1988) devised the measure of functional load for pronunciation teaching, indicating that many aspects of a sound's functional load can determine its relevance for learners' speech. For example, vowels that constitute minimal pairs such as /æ-ɛ/ may be more relevant for FL pronunciation instruction than vowel pairs with low functional load. That is because /æ-ɛ/ could be considered core sounds of English, indicating two distinct long and short vowels (Jenkins, 1988). While some studies may indicate that vowels bring greater contribution to speech and word intelligibility (Fogerty; Humes, 2012), studies conducted on the functional load of English consonants indicate that high functional load consonants may also pose a significant problem for listeners' perception and comprehension of speech (Munro; Derwing, 2006). Lastly, it is important to address that much of FL research on pronunciation instruction regarding intelligibility first started emphasizing suprasegmental aspects of FL speech (Celce-Murcia; Brinton; Goodwin, 2010). Levis (2005), for example criticizes the overemphasis on the research

regarding suprasegmental instruction for intelligible communication, suggesting that researchers and instructors should consider more carefully which aspects of the language to teach. This is important for communication between non-native FL speaker as while native speakers of a language may better understand speech which uses suprasegmental features of their language appropriately, interlanguage interactions may require a more balanced approach that pays attention to both segmental and suprasegmental features of the spoken language (Jenkins, 1998).

2.1.1 Target Vowels

For this study, the vocalic contrasts /æ-ɛ u-ʊ/ were selected based on Flege and Bohn's (2021) phonetic category subcomponents regarding their Revised Speech Learning Model (SLM-r). This study in particular adopts the authors' ideas in how speakers acquiring a second language will acquire certain sounds depending on their L1 phonetic categories and on how precisely these categories are previously defined. According to Flege and Bohn (2021) L2 phonetic contrasts that exist in the speakers' L1 are less likely to be categorized within the same phonetic categories, and more likely to be categorically assimilated if speaker' judge the L2 sound to be too similar to a sound present in their L1. In these situations, the authors claim that speakers often form composite L1-L2 phonetic categories, which is shaped by the phonetic input of both languages at play. Furthermore, studies have shown that BP speakers assimilate vowels /æ-ɛ u-ʊ/ into one pre-existing phonetic category for each pair of vowels(LIMA JR, 2015, 2017), and that even L1 BP highly proficient English speakers may not have yet formed distinct L2 phonetic categories for these sounds (Rauber et al., 2005). Thus, it can be assumed that L1 speakers of BP are likely to develop composite L1-L2 categories in which the production of the selected vowels will depend on the participants' L1 category precision. Thus, it can be argued that L1 BP speakers may find difficulties discerning between the English high-back /u-ʊ/ vowels and mid-front vowels /æ-ɛ/, as they are both vowel pairs that can be discerned by spectral differences, and are likely to cause categorical assimilation problems for L1 BP speakers.

According to Thomson and Derwing (2014, p. 10), "(...)the first step of any L2 pronunciation study should be to establish whether the form under consideration is

problematic for intelligibility/comprehensibility". Taken into account previous studies and the postulations of SLM-r and the existence of many minimal pairs, it is believed that the English vowels /æ-ɛ u-ʊ/ chosen for analysis are particularly problematic for intermediate to advanced level Brazilian learners of English. Furthermore, the selected sounds are some of the sounds deemed to be appropriate for ASR instruction in this experiment through the use of minimal-pair or near-minimal-pair tasks.

2.2 ANXIETY, ENJOYMENT, AND FL PRONUNCIATION

This section aims to analyze the dimension of emotion in the foreign language learning process by discussing affective factors such as anxiety and enjoyment. It will attempt to summarize the most noteworthy studies done in the field of individual differences pertaining to aspects such as foreign language classroom anxiety, foreign language enjoyment, and pronunciation anxiety (Horwitz et al., 1984; Dewaele; Macintyre, 2014; Baran-Łucarz, 2014, 2016), while also attempting to exemplify their effects in the FL learning process. Furthermore, it will attempt to discuss how to reduce or promote these emotions through the use of technology, computer-assisted language learning, and the internet (Bashori et al., 2020; Mroz, 2018; Lee; Lee, 2020).

2.2.1 General and Specific Anxiety in a Foreign Language

Anxiety can be characterized as a construct constituted by physiological, behavioral and cognitive manifestations, often linked to attentional and memory biases towards threatening stimuli or situations (Vasa; Pine, 2004). Besides the physiological representations of anxiety (such as tense body positions, dry mouth or restlessness), anxiety can be of various types, such as in behavioral and cognitive acts of avoidance and concern, or in ones' fear of negative evaluation and overall social anxiety. The feeling of concern, for example, is part of what constitutes the dimension of *Worry* in Liebert and Larry's seminal study which first defined Test Anxiety (1967), which is in itself a construct related to the threat of specific achievement failures. It encompasses the anxiety-specific state of physiological arousal and cognitions related to failure (Pekrun, 2000). For example, the

more worried a FL learner is about their performance on a language exam, the lower their performance expectancy should be, which can hinder their overall performance by limiting their cognitive working capacity (Pekrum, 2000). On the other hand, phobias such as social anxiety can be better evidenced by one's fear of negative evaluation, social status and identity (Hofmann; Asnaani; Hinton, 2010). Both dimensions of test and social anxiety are representations of anxiety that often constitute the foundations of what can be considered language anxiety[1] (Baran-Łucarz, 2022).

The topic of foreign language anxiety, which can be defined as "a distinct complex of self-perceptions, beliefs, feelings, and behaviors related to classroom language learning arising from the uniqueness of the language learning process" (Horwitz et al., 1986, p. 128) is argued to be a highly problematic psycho-physiological factor for FL learners, as it can directly affect their participation and self-perceptions in FL language interactions (Dewaele; Macintyre, 2014; Dewaele et al., 2017, Horwitz et al., 1986; Macintyre et al., 1998). It can inhibit particularly anxious learners from practicing and engaging in oral communication, both in and outside of the classroom (Baran-Łucarz, 2014a). MacIntyre et al. (1998), for example, has affirmed how state anxiety[2] may arise in FL communicative situations depending on the factors surrounding the ongoing interaction, such as the number of listeners present during social interactions. Similarly, the authors also emphasize how *state communicative self-confidence*, which characterizes one's own judgment of their capabilities of communicating at any particular moment, may hinder FL performance, especially in novel situations. These factors are part of the heuristic model that composes the concept of FL Willingness to Communicate (FL-WTC), defined by MacIntyre et al. (1998, p. 547) as the "readiness to enter a discourse at a particular time with a specific person or persons, using an L2[3]". As state anxiety can affect one's *state communicative self-confidence*, it may be a significant cause of hindrance of FL use (Macintyre et al., 1998).

---

[1]Similarly to how it was used in the chapter written by Baran-Łucarz (2022), Language Anxiety, Foreign Language Anxiety, and Foreign Language Classroom Anxiety, will be used interchangeably in this study.

[2]State anxiety is defined by Spielberger (1983) as "the transient emotional reaction defined by feelings of tension and apprehension, accompanied by autonomic nervous system arousal" (as cited by MacIntyre et al. 1998)

[3]In this study, L2 and FL are used interchangeably, but most definitions were replaced with FL for ease of reading. Quotes remain unchanged.

One of the first studies that addresses anxiety in the FL classroom was conducted by Horwitz and colleagues (1986), which take into account the dimensions of communicative apprehension, test anxiety, and fear of negative evaluation. Communicative apprehension, according to the authors, is "(...) a type of shyness characterized by fear of or anxiety about communicating with people." (Horwitz et al., 1986, p. 127). Regarding test anxiety, it refers to a speaker's fear of failure of not achieving desirable scores on tests and exams, which directly affects foreign language learning in classes as the environment often evaluates its students on their linguistic performance. Lastly, fear of negative evaluation constitutes a speaker's apprehension of being evaluated by other speakers of the same foreign language in any context. Horwitz and colleagues' study (1986) on foreign language classroom anxiety is still a relevant way of measuring students' language anxiety, which is done by using the Foreign Language Classroom Anxiety Scale (FLCAS). This scale consists of a 33 item questionnaire that evaluates learners' language anxiety in all three aforementioned dimensions in a 5-point Likert scale ranging from strongly agreeing to strongly disagreeing with each item. The FLCAS has been used in a number of studies since its conception as a way to measure language anxiety, such as in the study conducted by Awan and colleagues (2010), where foreign language classroom anxiety was compared with language achievement of 149 undergraduate students in order to determine the influence of language anxiety. According to authors, language anxiety was negatively correlated with learners' language achievement (r= .-273) at a weak level, which suggests that language anxiety is responsible for a small amount of variance of participant's GPA scores.

Furthermore, a meta-analysis by Botes, Dewaele and Greiff (2020) evaluated the correlation of foreign language classroom anxiety with general, reading, writing, listening and speaking linguistic achievement, assessing 67 studies conducted utilizing the FLCAS as a measure of language anxiety. For all dimensions of linguistic achievement, FLCA had either moderately or largely negative correlation, with writing and listening being the most impacted by language anxiety. The FLCA, however, is not without fault, as some studies indicate that its measurement may be actually measuring participants' self-knowledge about their language skills (Sparks; Ganschow, 2007), and others claim that ultimately, FLCA may not necessarily be a problem that can be directly addressed in the classroom environment (Dewaele et al., 2017). The criticism posed by the aforementioned authors suggest that

FLCA is not a clear indicator of actual language anxiety, and that unlike affective factors such as enjoyment, foreign language learners' anxiety may not be a variable which teachers can directly aid in reducing or controlling inside the classroom context.

As just previously mentioned, affective factors can directly affect general FL learning, but they can also play a role in FL learners' listening and pronunciation proficiency. A study conducted by Baran-Łucarz (2013) aimed to analyze FL listening anxiety, that is, one's predisposition to feel anxiety when listening and interpreting FL speech. It also observed how FL learners' perceived and actual levels of FL pronunciation aimed were related with these dimensions of anxiety on FL listeners. In the study, a negative correlation between the participants' listening anxiety and both levels of FL pronunciation can be observed, with perceived FL pronunciation playing a higher role in determining the degree of listening anxiety of the subjects in the study. This is relevant as it highlights the impact FL learners' self-perceptions and beliefs about their FL pronunciation may have on their overall FL comprehensibility, while also clarifying how anxiety may be tied to FL pronunciation.

FL pronunciation is one of the aspects that generates the most anxiety amongst FL learners (Phillips 1992; Price 1991; Young, 1992 apud Baran-Łucarz, 2014c), which prompted Baran-Łucarz to further expand on the concept of FL listening anxiety by developing the concept of Pronunciation Anxiety (PA) and its respective measurement scale. The author designed a multidimensional construct that categorizes FL pronunciation anxiety as being formed of "(...)negative FL pronunciation self‑perceptions, fear of negative evaluation, and beliefs about the importance of pronunciation, difficulty of learning and the sound of the FL pronunciation" (Baran-Łucarz, 2014a). Furthermore, Baran-Łucarz (2014a, 2014c) inspired by the FLCAS proposed by Horwitz et al. (1986), developed the *Measurement of Pronunciation Anxiety in Foreign Language Classroom*s (MPA-FLC), which consists of a questionnaire asking participants to agree/disagree with statements pertaining to fear of negative evaluation related to pronunciation self‑image, pronunciation self‑assessment, and pronunciation self-efficacy. Pronunciation self-image refers to a speaker's self-perception regarding their appearance when speaking a foreign language, often encompassing aspects of identity and overall self-evaluation of that identity. Pronunciation self-efficacy, on the other hand, is related to one's personal beliefs about their

own ability to acquire target sounds of a foreign language, tied to aspects of linguistic motivation and self evaluations of perceived competence. Lastly, pronunciation self-assessment represents a speaker's self-evaluation of their own foreign language pronunciation level, often done by comparing themselves to other speakers of the same language (Baran-Łucarz, 2014a).

The MPA-FLC was first used in an quasi-experimental study (Baran-Łucarz, 2014a, 2014b, 2016) analyzing the relationship between Pronunciation Anxiety and FL Willingness to Communicate in FL classrooms. FL-WTC was measured through a questionnaire that assessed both the level of acquaintance and group size between participants in specific scenarios. Baran-Łucarz reports a significant negative correlation of .60 between FL WTC and PA for the 151 participating students, meaning that those who feel more anxious about their pronunciation, were significantly less willing to communicate in a foreign language. The author stated that the only insignificant correlation was seen when taking into account the level of pronunciation anxiety when speaking to people they did not know. This result, according to the author, indicates "(...)that for many learners FL classroom is, among others or maybe first and foremost, a platform or scene at which they feel they are constantly evaluated" (Baran-Łucarz, 2014a, P. 45). It is important to mention that to date, very few studies have attempted to measure how PA may affect learners' degree of intelligibility in a FL, which puts into question the degree to which PA may hinder learners FL pronunciation learning and overall intelligibility. It can be assumed, however, that learners with low FL WTC are likely to not participate in learning activities, resulting in less FL use (Hashimoto, 2002; MacIntyre; Charos, 1996).

2.2.2 Foreign Language Enjoyment and FL Pronunciation

Foreign Language Enjoyment (FLE) may also be considered as a relevant affective factor for FL pronunciation learning due to how it can be controlled in FL classrooms (Dewaele et al., 2017). FLE can be defined as a positive emotion derived from joy present in positive classroom experiences, often including aspects such as laughter and motivation, and influenced by positive teacher dynamics (Dewaele; MacIntyre, 2014). The authors thus state that "the process of language learning will implicate the two key sources of enjoyment:

developing interpersonal relationships and making progress toward a goal" (Dewaele; MacIntyre, 2014, p. 242), both of which can be nurtured within the classroom context. With the emergence of positive psychology and the work of Dawaele and MacIntyre (2014), researchers claim that FLE may be more prevalent than FLCA in the classroom environment (Dewaele et al., 2017; Saito et al., 2018), despite both being separate psycho-physiological factors. Thus, they are not affective factors that exist on opposite ends of a scale, but rather coexisting, independent emotions that occur in parallel depending on the context and situation of the learning environment (Dewaele; Macintyre, 2014). The authors argue that positive emotions such as enjoyment may be beneficial for the achievement of goals and tasks, and that joy is an essential component of achieving the state of *flow*[4] (Csikszentmihalyi, 1999 apud Dewaele; MacIntyre, 2014, p. 242). Thus, the role of foreign language enjoyment is that by promoting joy in the foreign language learning environment, students may be less hindered by the negative emotions surrounding the activity, while also enabling new experiences by promoting exploration and engagement, which could result in positive effects for learning (Fredrickson, 2003), and may even aid in lowering participants' Pronunciation Anxiety in situations where FLE is high (Baran-Łucarz; Lee, 2021).

As previously mentioned, the goals set out by enjoyment in the foreign language classroom are related to both interpersonal and personal aspects in the learning context. This, according to Dewaele and MacIntyre (2016) indicates that foreign language enjoyment can be divided into two categories, those being social and private dimensions of FLE. The authors state that FLE-social is constituted of positive classroom dynamics, including aspects such as laughter, positive environment, and sociable and kind peers and teachers. Extracted from the questionnaires employed by the researchers, it was possible to notice in the learners a need to preserve their identity in front of others through improved linguistic performance (Dewaele; MacIntyre, 2016). Furthermore, the role of teachers in this dimension of FLE is quintessential, since their role in managing the emotional atmosphere of the classroom plays a significant part in transforming potentially anxiety-provoking situations into enjoyment inducing ones instead. In contrast, FLE-private deals with learners' internal enjoyment of the foreign language learning process. This dimension

[4]Flow is described here as "(...)a positive state where challenges and skills to meet them are aligned well" (Csikszentmihalyi, 1999 apud Dewaele; MacIntyre, 2014, p. 242)

includes the feeling of accomplishment in learning a foreign language, attached to keywords such as "*I enjoy it, it's fun, don't get bored, learnt interesting things* and *feel proud*" (Dewaele; MacIntyre, 2016, p. 223). Thus, it can be assumed that in particular, private foreign language enjoyment may be related to overall learner attitudes about a foreign language.

Returning to the role of the teacher in promoting enjoyment in classrooms, it is important to understand how FLE may be more directly controlled by teachers than FLA. In two studies by Dewaele et al. (2017) and Dewaele and MacIntyre (2019), learner and context variables were analyzed in order to find which variables were more responsible for each affective factor in classrooms. Specifically, Dewaele et al. (2017) found that the participants' attitudes towards the teacher were unrelated to FLCA, and that the degree of anxiety of a learner in a FL classroom may be just as low or high regardless of teacher behavior. However, attitude towards the teacher was positively correlated with FLE, indicating that there is much a teacher can do to promote an enthusiastic, joyful environment for their students. Furthermore, Dewaele and MacIntyre (2019) further investigated the differences between internal and context variables by analyzing learner personality traits. In this study, the authors sought to determine how much personality traits such as cultural empathy, open-mindedness, social initiative, emotional stability and flexibility could predict the degree of variance of FLCA and FLE. Similar results were found, with personality traits predicting about 30% of variance in FLA, but only 10% of variance in FLE. Furthermore, in the reported feedback from the participants, the authors also found that the teacher as a context variable was significantly more mentioned as the cause of FLE for the participants involved, whereas this was not the case for FLCA. It is possible to conclude, then, that the role of the teacher should not necessarily be concerning themselves with the dimension of anxiety, as it may occur independently of context, but rather in focusing on promoting pleasant experiences in the classroom.

Despite the awareness on how enjoyment may promote positive outcomes for the foreign language learning process, it is also important to understand how it correlates to academic achievement. In specific, Botes et al., (2022) meta-analysis of the effects of foreign language enjoyment reports significant, moderate positive correlations (r= .30) between FLE and academic achievement in 28 different studies. The authors also found a

similar positive correlation (r=.27, p<0.001) for self-perceived achievement, taken from the 9 effect sizes found in the included data. In a study by Dewaele and Li (2022) they analyzed the correlations between FLE, FLCA and general and domain-specific (actual and self-perceived) language achievement by testing 1415 Chinese EFL students. The authors utilized a combination of an English proficiency test and three questionnaires pertaining to self-perceived competence, foreign language enjoyment, and foreign language classroom anxiety. In this study, it was found that not only perceived language achievement predicted more variance of FLE and FLCA, but that certain perceptions of language specific domains were predictors of both affective factors (speaking, grammar), while the perception of other domains (listening, writing, vocabulary) predicted neither. This, the authors claim, is relevant since learners' self-perceptions about their academic achievement may be a reliable predictor of their language competence (Li, 2020 apud Dewaele; Li, 2022). Despite this, however, very little research has been conducted considering the dimensions of FLE and speaking (Bashori et al., 2021; Saito et al., 2018), and to date, very few studies directly address FL pronunciation as a separate proficiency in relation to FLE.

Furthermore, there appears to exist a link between learners' FLE and their ability to learn FL speech. While most studies in the field assess aspects such as motivation and identity (Nagle, 2018), or Foreign Language Speaking Anxiety (Bashori et al., 2020; Han; Tanriöver; Şahan, 2016), some studies directly address enjoyment in relation to intelligibility (Wilang; Singhasiri, 2016; Matsuura, 2007, Mroz, 2018). Saito and colleagues (2018), in specific, observe a vast number of affective factors in relation to FL speech learning and comprehensibility. The author assessed the dimensions of experience (FL background, FL use over time), motivation (integrativeness, instrumentality, family influence, attitudes to FL community and culture), and emotion (foreign language anxiety, foreign language enjoyment), and how these dimensions related to the development of the participants' FL speech learning. This study applied both longitudinal and cross-sectional analyses by testing participants at two separate times (T1 and T2) with the use of questionnaires and speech tests. Surprisingly, the data acquired by the authors indicated that during that at the initial time of testing, anxiety was the biggest predictor of overall FL oral performance (10.6%), while their FL longitudinal development (T2) was primarily explained by private enjoyment (7.5%), an emotion derived from foreign language

enjoyment. These results indicate that FLE, together with foreign language anxiety are intrinsic to the FL speech learning process.

Another aspect of FLE to be considered is in how it relates to technology use. In particular, some studies point to how out-of-classroom learning practices, together with the use of CALL tools or internet use (Lee; Lee, 2020; Lee, 2022), affect the level of enjoyment of foreign language learners. For example, Lee and Lee (2020) has observed the relationship between informal digital learning of English and dimensions of motivation and emotion of 661 Korean EFL learners ranging from middle school learners to university students. In particular, informal digital learning of English involves practices such as using everyday technological practices as a form of incidental language learning practice, that is "self-directed, informal English learning using a range of different digital devices (e.g., smartphones, desktop computers) and resources (e.g., web apps, social media) independent of formal contexts" (Lee; 2017, p. 2). Practices such as watching Youtube videos, interacting and reading in social media and video-games, and talking to other speakers of English may be considered informal digital learning of English. In Lee and Lee (2020), it is possible to observe that those who practiced these activities with more frequency had higher levels of enjoyment in learning a foreign language, indicating that there may be a positive aspect to be derived from out-of-classroom FL learning experiences. To date, however, only one study has directly addressed the relationship between enjoyment, technology and FL speaking (Bashori et al., 2021), which suggests further investigation should be considered.

Ultimately, The role of FLE in this study is twofold, since enjoyment appears to be a variable which can be directly impacted in the classroom context, as creating an enjoyable learning environment may facilitate learning (Dewaele; MacIntyre, 2014, 2017). On the other hand, FLCA appears to not be as easily controlled by teacher attitudes (Dewaele et al. 2017), which indicates that students' anxiety may also be related to other elements that compose the classroom environment. Taking into account both aspects aforementioned, managing FLCA (specifically, pronunciation anxiety) and promoting FLE has been addressed and may be manageable through the use of technologies and individualized instruction, such as with the use of Automatic Speech Recognition, as studies that observe anxiety, enjoyment and ASR instruction (Bashori et al., 2021; Mroz, 2018) have already shown promising results in favor of extending the instruction format beyond the classroom.

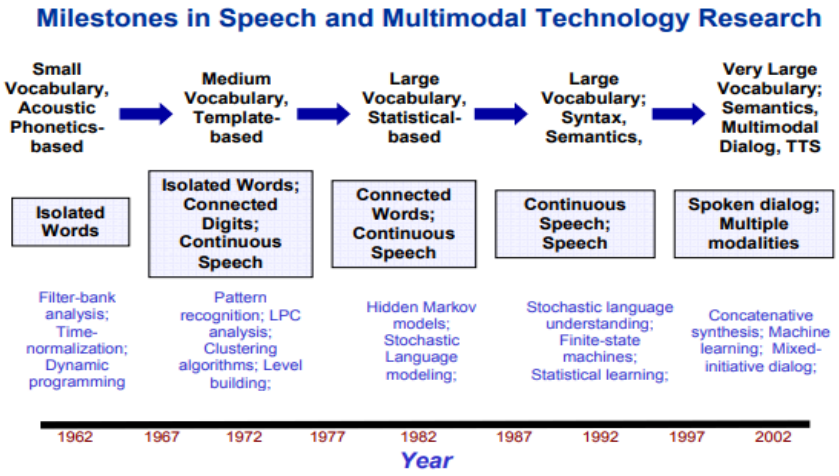
2.3 AUTOMATIC SPEECH RECOGNITION (ASR) AND FL LEARNING

This section is concerned with defining, contextualizing and exemplifying automatic speech recognition, from both its earliest implementations to modern technologies, and in both general and specific uses for foreign language learning. It will address the historical evolution of automatic speech recognition as a technology, how these tools managed the transcription of non-native speech throughout the years, the benefits and negatives of current dictation ASR technology, how computer assisted pronunciation training apps and tools were developed using it (and their efficacy in training EFL learners), and address the limited number of studies tackling learners' individual differences through ASR use. It will also briefly cover how learner autonomy (Holec, 1981; Ellis, 2002) can be a beneficial factor to be promoted in FL pronunciation instruction, and how it can be promoted through technology and ASR use.

Automatic speech recognition (ASR) is a technology made with the purpose of accurately translating spoken language into digital text. According to Levis and Suvorov (2012, p. 1), the technology of ASR "is an independent, machine-based process of decoding and transcribing oral speech". This is done by transcribing user input (audio) into words, sentences, and paragraphs, and is present in many different apps and services such as YouTube's automatic subtitling system, Google and Microsoft's word processing tools, mobile foreign language learning apps, and virtual assistants such as Alexa and Cortana. Even in the field of FL pronunciation instruction, ASR technologies have been implemented dating back to 1982 with the "German-by-Satellite" program conducted by Wohlert (1984, apud Wachowicz; Scott, 1999), where limited ASR technologies (consisting of only the transcription of small sets of single words) were applied via satellite to German as a foreign language students. These tools have seen significant development since its earliest implementations, which were first developed in the 1960s and often could only transcribe either single digits, syllables, or a small selection of words (Juang; Rabiner, 2005). Nowadays, however, most ASR tools are capable of transcribing entire sentences, paragraphs, and videos. As seen in Figure 1, much of the foundation of ASR technologies that still see use today were quickly developed from the 1980s up until the 1990s, where the

implementation of hidden markov models (HMMs)[5] greatly enhanced the capabilities of continuous speech recognition (for a more detailed summary, see Levis; Suvorov, 2012). This is where most ASR technologies would begin being commercially developed. *Dragon NaturallySpeaking,* for example, was one of the technologies that was then developed as speech recognition software aimed as an alternative to traditional note-taking practices done by a number of professionals such as lawyers and doctors. As a tool, it required participants to train the software to recognize a single voice through a number of tasks done in setting up the program for the first time, often by uttering select sentences prompted by the program itself. As we will later see, this tool was eventually applied for FL pronunciation research, with limited success (Coniam, 1999; Derwing, 2000).

Figure 1 - Milestones in Speech Recognition and Understanding Technology over the Past 40 Years.



Source: Juang; Rabiner (2005)

As seen in Figure 1, another technology that had been applied to ASR was machine learning, or Artificial Neural Networks (O'shaughnessy, 2008). However, it was only up until deep learning techniques became feasible to conduct within these network models that the technology started being widely applied to speech recognition (Padmanabhan, 2015). The accessibility to these new techniques due to better hardware led to significant advancements in how ASR tools were trained to transcribe speech, resulting in significantly more accurate transcription of input, specially of non-native speech of a language

[5] Hidden Markov models are probabilistic frameworks where the observed data are modeled as a series of outputs generated by one of several (hidden) internal states" (CHRISTOPHER, 2020)

(Mccrocklin; Edatalishams, 2020). As such, many of the modern ASR tools today such as *Google Voice* and Microsoft's *Azure AI* have greatly benefited from the technologies and provide for much more capable machine transcription of human speech. These tools may often use a combination of L1 (native) speech training with a combination of either conversion techniques (voice conversion, accent conversion) for non-native speech, or FL speech data taken from corpuses such as L2-ARCTIC (Shibano et al., 2021). It is, however, important to note that FL speech corpuses appear to be scarce (Shibano et al., 2021), requiring further additional procedures to generalize existing FL data onto and from previously trained L1 speech.

It is also important to consider that ASR technologies can be divided into distinct systems that serve different overall purposes. In their literature review, Wachowicz and Scott (1999) classified ASR tools into four categories, being them speaker-dependent, speaker-independent, discrete word, and continuous systems. This taxonomical approach to ASR technologies is still applied today to categorize distinct ASR systems. Relevant to this research are the differences between speaker-dependent and independent systems, as speaker-independent ASR tools are often made to accurately understand any non-specific speaker of a language, while speaker-dependent systems require the ASR technology to be trained on a specific speaker only. It is important for ASR tools to be capable of transcribing users' speech accurately despite differences in their voice features (gender, age, dialect), and despite their degree of accentedness, as a large percentage of users of these technologies today are non-native speakers of English.

ASR is widely present in *Computer-Assisted Pronunciation Training* (CAPT), often used for FL pronunciation teaching of segments, usually through the practice of minimal pair vocabulary, enhancing FL learner intelligibility (Chun, 2019; Liakin; Cardoso; Liakina, 2017). It may also be used for practices of FL pronunciation learning, often through autonomous learning practices (Mroz, 2018). In particular, the use of dictation software for FL pronunciation training (ASR software not made specifically for FL instruction) has been on the rise, with researchers Evers and Chen (2020) claiming its large speech database, free cost, and ease of use to be efficient for both explicit instruction and autonomous learning practices. Dictation software does come at a price, as its lack of phonetic descriptions and perception features may make noticing the gap between learners' own pronunciation and the

desired pronunciation more difficult (Evers; Chen, 2020). Despite this, the experiment by Evers and Chen (2020) appears to indicate that the use of dictation software significantly improved participants FL comprehensibility, both in the experiment group (paired with peer corrective feedback) and in the control group (practicing on their own). The 31 participants involved in the study also gained improvements in their abilities in performing FL spontaneous speech and had positive attitudes about the tools and its ease of use.

Furthermore, ASR-based CAPT tools such as *Talk To Me - English* were developed as the practice of utilizing speech recognition software for pronunciation instruction started becoming more frequent. This program by *Auralog* used to provide FL English learners with the opportunity to practice dialog sequences where they would be given questions to respond to orally, and choose between three alternatives to sound out correctly in order to progress the conversation. It provided for articulatory and acoustic features as learners could observe vocal tract animations, waveforms and pitch contour comparisons, while also applying a multimodal approach to learning as it contained visual stimuli present in the programs' user interface (Hincks, 2003). Furthermore, many literature reviews conducted on pronunciation instruction, CAPT and automatic speech recognition for FL pronunciation instruction indicate the contributions these tools could bring to FL learners, such as individualized instruction, large practice load, promotion of autonomy in FL pronunciation learning, and significant enhancements on FL pronunciation intelligibility (Chun, 2019; Thomson; Derwing, 2014; Wachowicz; Scott, 1999).

As previously mentioned, the time allocated in the classrooms for FL speech learning is often very limited, which could lead to the use of digital technologies and out-of-classroom approaches to further extend and improve the FL speech learning process. Chun (2019) provides a review on CAPT tools and how they have been used in previous studies, indicating that they can provide for "unlimited input, practice, and repetition, for individualized and instantaneous feedback, for both listening and speaking activities, and for providing a visualization of segmentals and suprasegmental" (p. 1). Thomson and Derwing (2014) in their literature review also claim similar benefits to approaching FL speech learning through the lenses of CAPT, citing learner autonomy and individualized instruction as appealing aspects of using these tools, programs and apps.

Previous research has examined several computer-assisted language learning (CALL) tools for FL pronunciation learning, such as using computer-mediated communication tools such as Skype (Alastuey, 2011), acoustic tools that provide waveform feedback (Motohashi-Saigo; Hardison, 2009), ultrasound devices for biofeedback (Li et al., 2019), and ASR programs (Tejedor-Garcia, 2020), all with the purpose of developing FL oral and pronunciation skills. Many mobile teaching applications also include FL oral and pronunciation training (Baldissera, 2020; Ahn; Lee, 2016). However, not many studies have considered how to adapt CAPT tools within an intelligibility-focused framework. In specific, Rogerson-Revell (2021) has noticed how these tools are often not concerned with modern pedagogical sensibilities, and instead prioritize developing novel technological advancements. Furthermore, most pronunciation apps and tools are not individualized enough and still consider native-like pronunciation as the ultimate FL speech learning goal, and thus, assess learners and users through an often unattainable scale by using "descriptive adverbs (...) such as 'poor pronunciation' or 'non-nativelike intonation'" (Rogerson-Revell, 2021), rather than assessing intelligibility and fluency. Still, the few studies that address intelligibility through the use of CAPT technologies have shown promising results through the use of widely available automatic speech recognition software (Mroz, 2018; Kasrani et al., 2018; Chung, Bong, 2022), both by promoting consciousness-raising activities and through explicit pronunciation teaching.

Over the years ASR technology has greatly enhanced its ability to transcribe FL English speakers, but this has not always been the case. As mentioned earlier in this section, the ASR software *Dragon NaturallySpeaking* has been previously used for FL pronunciation instruction and learning, despite being designed for note-taking purposes. However, this tool when previously used in quasi-experimental studies was criticized for its inability to accurately recognize FL speech (Coniam, 1999; Derwing, 2000), casting doubt about its implementation of said tool for FL pronunciation purposes. In particular, both studies claim that the ASR tool used (*Dragon NaturallySpeaking)*, despite being speaker-dependent, and thus built around each individual speakers' voice recognition training, failed to achieve recognition of FL non-native speech to the same degree as it did with English native speech, and falling significantly below (~20% difference) to humanlike

comprehension of the listener raters involved. This is exacerbated by the fact that both studies included exclusively highly proficient EFL speakers.

Recently, however, studies using ASR technologies such as *Google Voice Typing* and intelligent personal assistants such as *Alexa* have closed the gap between human and automated intelligibility (Mccrocklin; Edalatishams, 2020; Moussalli; Cardoso, 2019). In particular, the study by McCrocklin and Edalatishams (2020) aimed at revisiting ASR technologies and current popular dictation software *Google Voice Typing* in order to compare its ability to transcribe non-native FL speech to previous data by Coniam (1999) and Derwing (2000). The study sought to assess the level of accuracy of the dictation software in transcribing both native and non-native FL speech, while also observing whether the results correlate to human listener recognition and ratings of comprehensibility and accentedness of the speech data of 30 participants (10 L1 English, 10 L1 Mandarin Chinese, 10 L1 Spanish speakers). As opposed to the less than 20% accurate transcriptions of non-native speech (compared to human raters) found in the study by Coniam (1999) using the tool *Dragon NaturallySpeaking,* McCrocklin and Edalatishams (2020) found that currently, *Google Voice Typing* fell just under 5% less accurate than human raters, accounting for an impressive >90% rate of accuracy for both Spanish and Chinese L1 non-native EFL speech. While the reason behind this leap in transcription accuracy is not explained in either study, it can be speculated that the modern dictation software technologies that apply a combination of artificial neural networks, deep learning techniques, large speech corpuses trained on non-native speakers, and voice and accent conversion techniques may have an impact on the ability to develop more inclusive and accurate ASR software.

Many studies that have applied ASR tools designed exclusively for FL pronunciation instruction have resulted in positive gains towards FL learner's intelligibility (Burlerson 2007; Hincks, 2003; Neri et al., 2008; Tejedor-Gárcia et al, 2020), and oral proficiency (Chiu et al., 2007). However, due to the difficulty of finding or creating CAPT programs that adapt modern pedagogical approaches that allow for more inclusive non-native FL speech recognition, a dictation tool ASR tool may still be preferable. However, as Baldissera (2020) claims in her review of current mobile assisted language learning apps for FL pronunciation, many current ASR pronunciation apps (*English Pronunciation Tutor, Elsa,*

EnglishPronunciation, Juna) do not allow for flexibilities in transcribing and providing non-native EFL speech. Furthermore, most ASR apps prioritize accuracy as a measurement of speech and heavily favor American English as the standard variety of English. Due to these issues that can be found in modern pronunciation apps that contain ASR, consideration should be given to dictation ASR technologies such as *Google Voice Typing* and intelligent personal assistants such as *Alexa*, as it is believed that they can be implemented in aiding FL speakers with their intelligibility (Mroz, 2018), due to their ability to efficiently transcribe accented FL speech (Mccrocklin; Edalatishams, 2020; Inceoglu et al., 2020; Moussalli; Cardoso, 2019).

Promising results regarding ASR have also been found by researchers using ASR tools built into intelligent personal assistants (Dizon; Tang, 2020; Moussalli, Cardoso; 2019). In particular, the ability to accurately transcribe accented speech of non-native EFL language speakers has been reported by Moussalli and Cardoso in their mixed-methods analysis of *Amazon Echo*. The authors applied a battery of tests consisting of surveys, interviews, and listener judgements in order to assess whether not only would *Amazon Echo*‘s synthesized speech be comprehensible and intelligible to the participants, but also how well would the transcription of the intelligent personal assistant be when compared to listener judgment of two native English speakers. Eleven non-native English speakers seven different linguistic backgrounds had to spend approximately 30-45 minutes interacting with *Alexa* (the “voice” of Amazon Echo) by asking it a number of pre-established questions (e.g. “Alexa, what is the definition of convoluted?”) and were incentivized to also make-up their own questions after getting accustomed with the tool. Participants also had to fill-out a survey and transcribe a sentence uttered by *Alexa* in response to the question “Alexa, are you in love?” in order to judge Echo's comprehensibility and intelligibility. Based on 5-point Likert scale survey answered by the speakers, the authors were able to confirm that not only *Alexa* was comprehensible to them (M= 4.18), but that they were understood by the IPA (M= 3.55), fact which was later supported by the relatively small 23% word error rate found in the transcriptions of participants’ speech by *Amazon Echo*. Lastly, similarly to other studies such as McCrocklin and Edalatishams (2020), intelligibility values across native English listeners and the built-in ASR tool within *Amazon Echo* only deviated about 12%

(95% for the listeners, 83% for *Echo*). This further reinforces the notion that ASR tools as a whole have greatly improved their ability to understand non-native, accented EFL speech.

Speech recognition tools have also been analyzed in regards to FL pronunciation and speaking gains and positive changes in learners' emotions towards a FL. For example, Bashori and colleagues (2021) conducted a study that suggests significant benefits in FL learners' emotions towards a FL by using ASR for FL learning. The study applied the use of two CAPT tools with ASR included, those being *NovoLearning* and *ILoveIndonesia*, in order to assess the impact of the web-based learning tools on 232 EFL learners. Thus, participants were divided into two groups (ILI and NOVO). Furthermore, their vocabulary gains were also assessed based on the treatment consisting of 4 sessions of 90 minutes practicing their speaking and pronunciation of select FL target items, including pronunciation feedback. The participants were tested before and after treatment with a number of general and vocabulary tests and two questionnaires pertaining to foreign language speaking anxiety and foreign language enjoyment, and their data was compared to a control group which was included. The authors reported higher negative gain scores for foreign language speaking anxiety (FLSA) in the post-test (*M*= -0.642 for ILI, M= -0.704 for NOVO, and r= -0.053 for Control) and increased gain scores foreign language enjoyment (FLE) for participants through ASR-based website instruction (*M*= 0.321 for ILI, *M*= 0.203 for NOVO, and *M*= 0.052 for Control)

However, a study by Shams (2006) reports a non-significant relationship between language anxiety and ASR-based CAPT pronunciation instruction, measured through the use of the Foreign Language Classroom Anxiety Scale (FLCAS). The study utilized the CAPT tool *Talk to me- English*, where learners of French as a foreign language were separated into two groups: a comparison group (practice via cassette recordings in a listening lab), and an experimental group (ASR-based CAPT practice in a computer lab). The study also aimed to measure learners' overall pronunciation gains by assessing them before and after the treatment of seven sessions (one per week) of 50 minutes each. The 65 subjects in the study appeared to have negative gains regarding their foreign language classroom anxiety (*M* = -8.89) for the comparison group, *M*= -9.73 for the experimental group), but the differences between groups did not appear to be statistically significant. Besides the possible limitations imposed by the treatment approach, such as the lack of

autonomous learning and out-of-classroom learning implementations (treatment was done specifically in a computer lab), the author specifically states that the FLCAS may not be sensible enough to address the specificities related to pronunciation anxiety. Unfortunately, 17 years since the conclusion of that study, very few authors have yet attempted to implement the *Measure of Pronunciation Anxiety (*MPA) (Baran-Łucarz, 2014a, 2014c, 2016) with the use of ASR tools to assess the effect on pronunciation anxiety.

If we are to consider ASR as a potential technology for FL learning, it is also relevant to discuss the role of autonomous learning (Holec, 1981), especially when taking into account the aforementioned difficulties of conducting extensive FL pronunciation teaching in classrooms, specifically due to the commonplace time constraints encountered in educational contexts (Silveira, 2004; Carlet, Kivistö de Souza, 2018). Here, autonomous learning is not the idea of learning without guidance or aid (Little, 1991) but rather as the practice of giving learners the capacity to reflect, act independently, and to define their own goals, strategies, and contents of their own learning process (Holec, 1981 apud Dickinson, 1995; Little, 1991). Thus, autonomous learning in this context is not wholly independent, but instead includes an aspect of interdependence that sets the boundaries of which independent actions and autonomous learning occurs (Benson, 2013; Kohonen, 1992). Amongst benefits that can be cultivated by promoting autonomous learning practices for FL learners are matters such as increased motivation (Dickinson, 1995; Koçak, 2003; Kormos, Csizér, 2014), bigger exposure to input (Lai, 2017; Tudor, 1993), and the promotion and development of learners' identity as active participants in the FL learning process (Benson, 2013; Kormos, Csizér, 2014).

These aspects of autonomous learning can also be achieved through the use of CALL tools, ranging from the broad use of the Internet to specific programs made for language learning (Blin, 2004). The link between the use of digital technologies and autonomous learning seems evident, as the online digital space provides for opportunities to develop 1) learner motivation (Peterson, 2006; Reinders; Wattana, 2011), 2) individualized pace, schedule, content and strategies (Mutlu, Eroz-Tuga, 2013) and 3) rich quantity and quality of FL language input, either by other speakers of the language or through linguistic content found in programs, videos, websites and video-games (Chun, 2019; Thompson; Von Gillern, 2020). In specific for the practice of FL pronunciation learning, since aspects such

as identity should be considered important in order to address learners' anxiety (Baran-Łucarz, 2016), autonomous learning practices again may facilitate the development of an active and motivated identity.

Regarding FL pronunciation learning, studies have also supported ASR use for autonomous practice and overall consciousness raising activities (Mccrocklin, 2016; Dizon, 2020; Liakin et al., 2017; Spring; Tabuchi, 2022). In specific, the study done by McCrocklin (2016) evidenced how automatic speech recognition use can significantly boost learners' autonomous FL pronunciation learning and autonomous behavior. This study sought to answer whether utilizing ASR as a learning strategy led to higher degrees of self-reported autonomous behavior and autonomous beliefs by testing 48 ESL students via surveys, language learning logs, and interviews. The participants were divided into three groups: 1)CONV= conventional face-to-face training; 2)STRAT= minimal ASR strategy training; 3)HYBRID= half of the workshop consisting of ASR strategy training. One group required ASR practice as part of the pronunciation workshop (HYBRID). Post-workshop results indicate that the two groups that received strategy training reported increased beliefs about autonomous learning due to the feedback they received from the dictation ASR tools used. Furthermore, the HYBRID group, which had required ASR practice, was the group to have the most significant autonomous pronunciation learning and ASR use after the workshop.

In this section, literature on ASR and foreign language learning was reviewed. Qualities, peculiarities, and challenges regarding this novel technology for FL speech learning and for aiding FL learners with their affective factors were discussed in order to achieve the objectives set out at the beginning of this study. In particular, this study is concerned with observing the extent of the effects of an automatic speech recognition tool (ASR) on participants' degree of non-native English intelligibility, pronunciation anxiety, and foreign language enjoyment. In order to observe the relationship and the effects of ASR on participants' PA, FLE and intelligibility, the following research questions were devised:

RQ1: What is the relationship between vowel production intelligibility and pronunciation anxiety and foreign language enjoyment in intermediate-to-advanced level EFL learners?

RQ2: What are the effects of instruction with ASR tool *Word Dictate* on intermediate-to-advanced level EFL learners' production of high-back /u-ʊ/ and low-front /æ-ɛ/ vowel intelligibility?

RQ3: What are the effects of ASR tool *Word Dictate* on intermediate-to-advanced level EFL learners' pronunciation anxiety (PA)?

RQ4: What are the effects of ASR tool *Word Dictate* on intermediate-to-advanced level EFL learners' foreign language enjoyment (FLE)?

H1: It is expected that participants' English vowel intelligibility scores will share a negative correlation with pronunciation anxiety, and a positive correlation with foreign language enjoyment

H2: It is expected that the ASR tool *Microsoft Word Dictate*, used in this study, will aid participants in developing higher intelligibility scores for vowels /u-ʊ/ /æ-ɛ/, enhance (lower) their pronunciation anxiety levels, and heighten their foreign language enjoyment emotions in the FL learning process.

The following section presents the method that was designed in order to answer these research questions. In the following section, all of the framework, information about the participants, stimuli, and procedures, as well as the steps taken to analyze the collected data, will be discussed.

3. METHOD

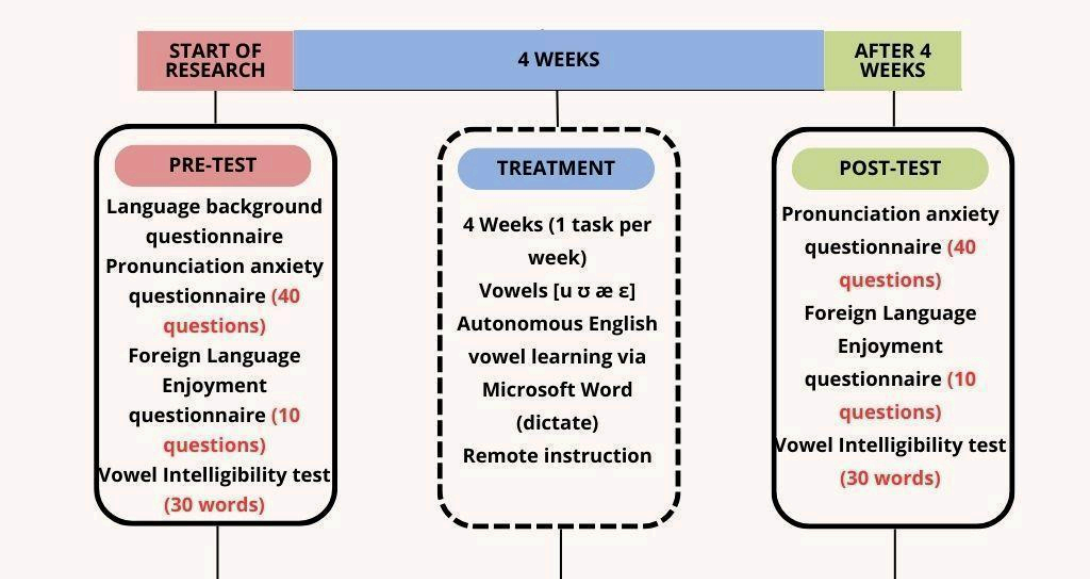
The objective of this quasi-experimental study is to assess the effects of an automatic speech recognition tool, *Microsoft Word Dictate* (Microsoft, 2022) on pronunciation anxiety (Baran-Łucarz 2014, 2014b, 2016) and foreign language enjoyment (Dewaele; Macintyre, 2014; Dewaele et al., 2018), and to evaluate its abilities in developing English vowel intelligibility in native Brazilian Portuguese speakers. This section will first present the

framework of the study, and participants. Then, the instruments of the research will be detailed, followed by an explanation of the applied treatment and the procedures for data analysis, followed by a description of the method's ethical considerations.

3.1 FRAMEWORK

The design of this study follows a pre-test post-test design, as seen in Figure 2. The pre-test consists of measures of pronunciation assessment and questionnaires on background information and individual differences. In the post-test, the same instruments were applied to evaluate the effect of the treatment. Participants' English vowel pronunciation intelligibility was assessed through a pronunciation assessment test, where their FL pronunciation anxiety was collected through the pronunciation anxiety (MPA) questionnaire (Baran-Łucarz, 2017), and participants' foreign language enjoyment (FLE) was retrieved using the foreign language enjoyment scale (Dewaele; Macintyre, 2014). In order to assess the effectiveness of ASR in aiding participants with their English vowel intelligibility, PA, and FLE, this study included a treatment period consisting of four weeks where participants were given weekly pronunciation tasks to complete remotely and at home with the use of the ASR technology *Microsoft Word Dictate*.

Figure 2: Study design



Source: Santos (2023)

3.2 PARTICIPANTS

The participants were intermediate-level English-speaking Brazilian college students at the Federal University of Santa Catarina (UFSC), invited through class interventions which were organized by the researcher and the respective class instructors. Thus, inclusion criteria were students who were currently enrolled in oral English classes, whose proficiency levels ranged from intermediate to advanced. The exclusion criteria included participants whose L1 was not Brazilian Portuguese and/or participants that had a language disorder. These participants were chosen specifically for fitting the criteria of having intermediate-to-advanced English proficiency, since the vowel pairs chosen in this study appear to be problematic for FL language learners at similar proficiency levels. (Rauber et al, 2005; Lima Jr, 2015, 2017). Overall, participants from three classes at UFSC were recruited, those being from the first year, second year, and third year of the course *Letras - Inglês*.

In total, 35 participants agreed to participate in this study, where one was excluded for not being an L1 BP speaker, and four were excluded as their vowel intelligibility test data did not appropriately save due to connectivity issues. The participants' background information can be observed in Table 1, with factors such as age of onset of learning English, time spent abroad, and FL proficiency and use. The latter two variables were assessed using two different Likert scales, one ranging from 1-5 (1= low proficiency, 5= very high proficiency) in five different FL skills (writing, reading, speaking, listening, and pronunciation), and another ranging from 0 to 100 in total percentage, measuring FL use in four different contexts (educational, social, professional, and entertainment).

Table 1: Descriptive statistics of participant's background information

	Age	**AOL**	**Time Spent Abroad**	**Perceived FL proficiency**	**FL use**
L1 BP	23 (6.6)	13.37 (6.06)	0.18 (0.83)	3.77 (0.66)	54 (15.83)
N=30	18-51	0-32	0-5	3-5	20-90

Source: Santos (2023)

Age, AOL, and Time Spent Abroad expressed in years. Standard deviation between brackets. Range of responses on the second line. Perceived FL experience was measured by calculating the mean of participants' responses to five proficiency categories (writing, reading, listening, speaking, and pronunciation). AOL refers to the first contact with English in an instructed setting.

3.3 INSTRUMENTS

The following sections will present each individual instrument used to measure the dimensions of English vowel intelligibility, pronunciation anxiety, and foreign language enjoyment. Furthermore, a questionnaire on language background was also designed with means to measure overall linguistic proficiency.

3.3.1 Language Background Questionnaire

Participants in this experiment were requested to fill in a form to report on their linguistic background. This questionnaire, developed by the researcher, was designed to gather information about participants' years studying EFL, time studying abroad, age, self-assessment of linguistic proficiency, and FL use frequency and context. The focus of this questionnaire was to allow for the selection of adequate participants for the research. Self-assessed linguistic proficiency was used for qualitative comparative analysis to the participants' scores on the tests. Due to how questions are directed to participants' self-assessment, this instrument cannot accurately assess overall FL English proficiency, and no other instrument attempted to measure such variable. This study believes that it is difficult to control for FL pronunciation proficiency, especially considering the dynamicity of the learning process regarding FL pronunciation (Lima Jr; Alves, 2019), and it is expected that even participants in the same oral English class may have vastly different FL pronunciation skill levels.

3.3.2 Vowel Intelligibility Measure

In order to assess the participants' English vowel intelligibility, the following steps were taken: a) minimal vowel pairs /æ-ɛ u-ʊ/ were selected according to category assimilation found between L1 Brazilian Portuguese (BP) and L2 English; b) the ASR instrument *Word Dictate* was selected according to prior pilot-testing of its transcription of non-native EFL speech; c) an ASR-based English vowel intelligibility assessment test was developed to evaluate the participants' English vowel intelligibility and; d) An English vowel intelligibility assessment test was applied prior to treatment (pre-test), and after the treatment (post-test). This vowel intelligibility test uses an ASR tool (*Word Dictate*) with the objective of testing participants' intelligibility on the target vowels /æ-ɛ u-ʊ/.

The tool Word Dictate used in the vowel intelligibility test is a feature present in Microsoft Word's text editor program that allows users to directly speak into a recording device (e.g,. a microphone) and have their input be transcribed into text by an automatic speech recognition algorithm. Here, the machine uses contextual clues found in whole sentences to make up for intelligibility issues found in the recorded speech samples. In order to avoid the influence of context in the transcriptions, this test's aim is at transcribing speech at the word-level.

3.3.2.1 Stimuli

The stimuli chosen for this study includes the vowel pairs /æ-ɛ u-ʊ/, which are believed to be problematic for the participants enrolled in this research. The test included 20 CVC words combined with 10 distractor words, as can be seen in Table 2. Out of the 20 target words, 12 contain voiceless stops and 6 are voiced stops on coda position. The remaining target word contains the structure CVC ending in the voiced alveolar lateral approximant /l/. For the distractor words, 6 contained voiceless stops, and 4 have voiced stops. These decisions were taken in order to create a controlled phonetic context for the production of the target vowels, as to limit the variation of transcriptions that could be computed by the ASR technology. Furthermore, this prevents the process of coarticulation from interfering with the participants' productions. This study, in order to completely

eliminate any influence of context from the ASR tool's judgment of intelligibility, chose to present these words to the listener without a carrier sentence.

Table 2 - Carrier words included in the English vowel pronunciation task and during treatment

Target Vowels	/æ/	/ɛ/	/ʊ/	/u/		
Distractors					/i/	/ɑ/
Pre-test/ Post-test Carrier Words	Bat	Pet	Foot	Boot	Beat	Lot
	Rack	Peck	Look	Luke	Peak	Dock
	Rap	Pep	Took	Loot	Keep	Top
	Sad	Red	Full	Pool	Feed	Rod
	Bad	Head	Hood	Dude	Need	God
Treatment Carrier Words	Pat	Bet	Nook	Nuke		
	Pack	Deck	Look	Loop		
	Cat	Tech	Hook	Root		
	Hat	Let	Put	Toot		
	Rat	Neck	Book	Poop		
	Hack	Heck	Cook	Coop		
	Lap	Rep	Good	Food		
	Tap	Net	Wood	Mood		
	Rad	Bed	Could	Rude		
	Had	Said	Would	Who		
	Mad	Dead	Pull	Cool		
	Lad	Led	Bull	Fool		

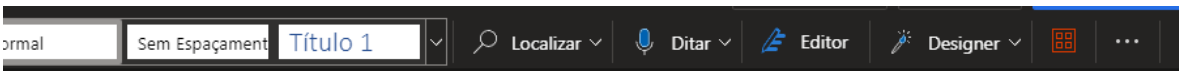
Source - Santos (2023)

3.3.2.2 Microsoft Word Dictate

For this study, Microsoft Word Dictate was chosen as the ASR technology, as it has proven to be a viable alternative to competing services such as Google Voice Typing (Gottardi; Kivistö-De Souza, 2022; Nelson; Cardoso, 2023). As argued in the section above, it is believed that many instruction-focused ASR technologies still bring significant issues regarding their approach to instruction, opting to grade pronunciation in a "native/non-native" scale. This, in combination with the notion that Microsoft's Azure AI technology can be argued to be just as efficient as Google's in regards to ASR and recognizing non-native English speech, as they appear to share similar intelligibility ratings in transcription of non-native EFL speakers (Gottardi; Kivistö-De Souza, 2022), resulted in giving preference to Microsoft's word processing program due to its ease of use and accessibility. *Microsoft Word* is a program available for both mobile and home devices such as computers and laptops, readily available online or locally, and often free with most versions of *Windows* and on app stores such as *Google Play*.

For both the Vowel Intelligibility test and for the instruction period, participants were directly interacting with this ASR technology found within Microsoft Word's display, seen in Figure 1.

Figure 3 - Microsoft Word display menu bar, featuring Dictate feature



Source: MICROSOFT (2023)

3.3.3 Measure of Affective Factors

In order to assess the engagement of ASR technologies with learners' individual differences, this study applied two questionnaires that consists of the original 50 questions included in the Measure of Pronunciation Anxiety (Baran-Łucarz, 2017), and ten additional questions regarding foreign language enjoyment present in the FLE Scale (Dewaele; Macintyre, 2014), as shown in Appendix - A and B. The questions here were designed to

ask participants about their feelings regarding certain scenarios, social interactions, beliefs and preferences regarding the FL learning process. In particular, 50 questions are specifically focused in assessing participants' beliefs about FL pronunciation, while the remaining 10 questions are concerned with general language beliefs within the classroom context.

These instruments have been consistently used to measure affective factors that involve the dimensions included in this experiment (Baran-Łucarz, 2014a, 2014c, 2016; Dewaele JM.; Macintyre, 2016, 2019; Dewaele et al.; 2017; Li; Jiang; Dewaele JM., 2018; Saito et al., 2018), and other studies have previously used this research design for assessing the dimensions of PA and FLE together (Baran-Łucarz; Lee 2021). In this dissertation, composite scores will be used for presenting the results, following previous studies which have not separated the dimensions of PA and FLE into categories (Baran-Łucarz, 2014; Dewaele et al.; 2017). Lastly, the researcher responsible for this present study also translated these questionnaires to Brazilian Portuguese in order to avoid comprehension issues during the assessment.

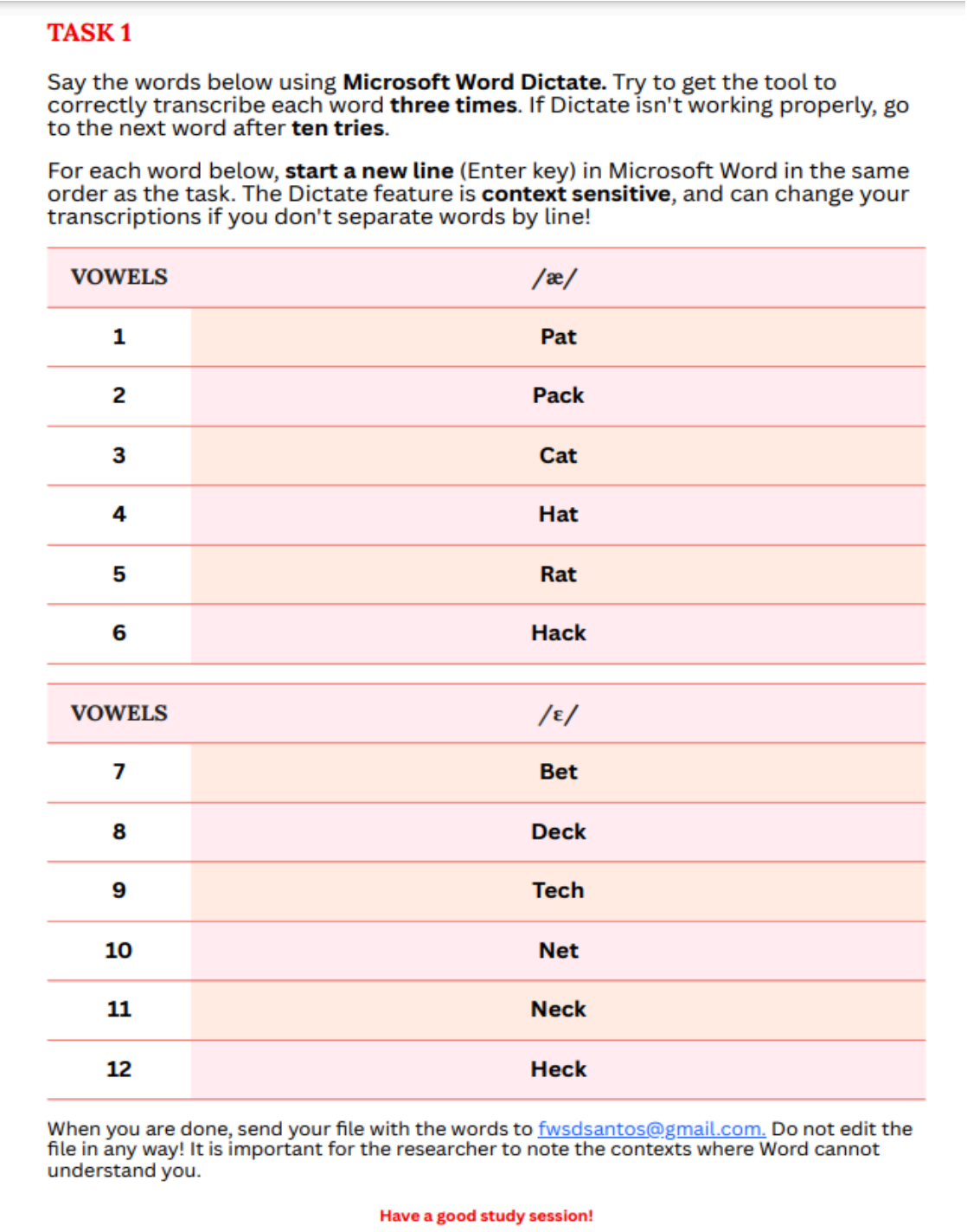
3.4 TREATMENT

In this study, a treatment phase was also included, where participants interacted with an ASR technology that according to previous studies, should result in reduced negative affective factors (anxiety) (Bashori et al., 2021), increased positive affective factors (enjoyment) (Bodnar et al., 2017) and improved overall intelligibility (Burleson, 2007; Neri; Cucchiarini; Strik, 2008; Mroz, 2018). Participants were given weekly printed instructions that provide additional guidance on how to operate the ASR tool *Word Dictate* used in this study for transcription and instruction. Weekly printed tasks were included alongside the printed instructions guiding participants on how to operate the tool. Participants completed four tasks during the period of four weeks. These tasks have the objective of assisting the participants in autonomously learning the English vowels /æ-ɛ u-ʊ/. Participants did these tasks at home or at another silent location either through the use of a mobile phone or a personal computer or laptop.

The treatment lasted four weeks and was organized as follows: a) the first and third weeks consisted of guiding participants in engaging with the minimal pair /æ-ɛ/; b) the second and fourth weeks were dedicated to training the production of the /u-ʊ/ vowels. Participants practiced a total of 12 CV(C) words per task. Additionally, they were given example sentences to facilitate comprehension for less frequently used words. Out of the 44 words selected, all of the 20 minimal or near minimal pair words equivalent to the words present in the pre-test were selected. The remaining words were chosen according to the word structures present in the pre-test to ensure a controlled and consistent phonetic context throughout the experiment.

In the instruction material given to participants each week were instructions on how to use the ASR tool *Word Dictate*, tips on the differences between words that are more infrequent, instructions on how to perform the task, and the carrier words containing the target vowels themselves. On the next page, in Figure 4, an example of these instruction materials can be observed. All of the instruction materials for each week is also attached to this present study, as Appendices E, F, G and H.

Figure 4 - Instruction material for week 1



TASK 1

Say the words below using **Microsoft Word Dictate.** Try to get the tool to correctly transcribe each word **three times**. If Dictate isn't working properly, go to the next word after **ten tries**.

For each word below, **start a new line** (Enter key) in Microsoft Word in the same order as the task. The Dictate feature is **context sensitive**, and can change your transcriptions if you don't separate words by line!

VOWELS	/æ/
1	Pat
2	Pack
3	Cat
4	Hat
5	Rat
6	Hack

VOWELS	/ɛ/
7	Bet
8	Deck
9	Tech
10	Net
11	Neck
12	Heck

When you are done, send your file with the words to fwsdsantos@gmail.com. Do not edit the file in any way! It is important for the researcher to note the contexts where Word cannot understand you.

Have a good study session!

Source - Santos (2023)

Words chosen for the treatment were chosen to give participants opportunities to practice the target vowels in a varied number of contexts. The first week of the treatment

provided for words where the vowels /æ-ɛ/ are preceded by a consonant and followed by a voiceless stop. The third week contained /æ-ɛ/ minimal pair words occurring in less controlled phonetic contexts, where the first consonant included rhotic, lateral, fricative, nasal, plosive consonants, and the last consonant included voiced and voiceless stops. Controlling for the phonetic context in the treatment is done in order to maintain consistency with the vowel intelligibility test target words.

Table 3 - Words included in the first and third weekly printed papers for pronunciation training

Task	Task 1		Task 3	
Target vowels	/æ/	/ɛ/	/æ/	/ɛ/
	Pat	Bet	Lap	Rep
	Pack	Deck	Tap	Net
	Cat	Tech	Rad	Bed
	Hat	Let	Had	Said
	Rat	Neck	Mad	Dead
	Hack	Heck	Lad	Led

Source - Santos (2023)

The second week was dedicated to the first task about /u-ʊ/ English vowel training, which consists of CVC words with voiceless stops /p/, /t/, /k/. The fourth week and last activity of the treatment involved the remaining CV(C) structures with word-final voiced stops, and laterals, as seen in Table 4. Again, consideration was given to the words present in the vowel intelligibility test, and the phonetic context in the treatment aimed to train participants in these vowel-consonant clusters.

Table 4 - Words included in the second and last weekly printed papers for pronunciation training

Task	**Task 2**		**Task 4**	
Target vowels	**/u/**	**/ʊ/**	**/u/**	**/ʊ/**
	Nuke	Nook	Food	Good
	Loop	Look	Mood	Wood
	Root	Hook	Rude	Could
	Toot	Put	Who	Would
	Poop	Book	Cool	Pull
	Coop	Cook	Fool	Bull

Source - Santos (2023)

3.5 PROCEDURES

The tests and questionnaires were conducted individually and in-person with each participant in a room equipped with adequate sound quality and with the supervision of the researcher. The order in which participants concluded the tasks in both pre-test and post-test were as follows: 1) they answered both PA and FLE questionnaires, and 2) participants then were invited to a room to conduct the vowel intelligibility test, and 3) participants answered the linguistic background questionnaire. Regarding the vowel intelligibility test, participants were instructed to speak the aforementioned words into the recording device one at a time, where they were asked to wait a minimum of five seconds per production. The ASR tool Word Dictate was used during this period to automatically transcribe the 30 productions per participant, as in this study it is assumed that MW Dictate is capable of performing as a human listener (see section 3.6). The supervising researcher recorded the computer screen during the test and separated each recorded transcription per line to avoid contextually influenced transcriptions. Each individual participant took around 5-10 minutes to complete the test. Participants were not able to see their productions in order to not impact their affective factors throughout the test.

Pilot-testing was conducted by testing the ASR tool using secondary data of L1 BP speakers of FL English available at Speech Accent Archive. Both paragraph and isolated words were assessed individually by presenting either the entire passage our individual carrier words containing the target vowels /æ-ɛ u-ʊ/ to the tool in order to address its ability in transcribing this group of non-native FL English speakers. Furthermore, the words chosen for the test are different from the words that were presented and were selected for the treatment to avoid training effects during the post-test and delayed-test assessments. Lastly, most of the words in the treatment phase are the opposite minimal or near-minimal pairs of the words in the vowel intelligibility test to minimize variability between the contents present in each section of the experiment, further controlling the phonetic environment and the measurement accuracy of the study, and ensuring the treatment and the test are measuring the same variable. For example, in the vowel intelligibility test, participants will spoke the word *bat* (/æ/), and for the first weekly printed paper they practiced the word *bet* (/ɛ/). In order to confirm whether all of the words in both the pre-test and treatment could be correctly and incorrectly transcribed, the researcher tested each word individually by speaking it with the proper vowel, and then with both incorrect spectral and temporal

qualities in order to observe whether mistranscriptions could occur due to either error in pronunciation.

During the treatment phase, in order to operate the *Dictate* feature, users clicked the microphone icon located in the program's display and waited until they heard a sound indicating the tool was ready to transcribe. After that, users were able to speak into a microphone to have *Microsoft Word* transcribe their audio input into text. The tool can be disabled by clicking the microphone button again. *Dictate* will continue to transcribe any audio input into text until it is disabled. For the pre-test and post-test procedures in particular, participants waited until the new word was shown in the assisting computer screen before speaking each carrier word. Furthermore, the mobile phone screen which was used to run the transcriptions through MW Dictate was dimmed in order to prevent participants from checking their answers during the test. If the tool had any issues transcribing the participants, the researcher stopped transcriptions briefly and asked participants to repeat the previous word in the test.

Participants also completed a pronunciation anxiety and a foreign language enjoyment questionnaire. These self-report Likert scale questionnaires were handed out during the first week of the experiment alongside the vowel intelligibility test, which were reapplied at the post-test to observe possible changes in affective behavior towards pronunciation anxiety and foreign language enjoyment post treatment. At all periods of data collection, these questionnaires took about 15-20 minutes collectively for completion by the participants involved.

During the treatment, participants were instructed to practice speaking the carrier words present in the instruction materials using *Word Dictate*. The participants were expected to create a new document file in Microsoft Word for each session, where they are to click the button "dictate" present in the program's interface, and speak into the ASR the aforementioned words. Participants were instructed via printed paper to speak the task words until they achieve a total of three correct transcriptions. They were also instructed to separate each word per line, to avoid contextually biased ASR transcriptions. After the end of each week, participants were also expected to send their text document containing all transcriptions to the researcher via email for monitoring.

Lastly, this research is within the parameters of the CNS resolution 510/16, and the guidelines of the OFÍCIO CIRCULAR N° 2/2021/CONEP/SECNS/MS for research involving virtual environments, thus being committed to meet all the necessary ethical and legal issues. Therefore, for dealing with personal data, the researchers involved with this study are also committed to respect the *Lei Geral de Proteção de Dados* (N°. 13,709, of August 14, 2018), thus valuing the protection and anonymization of the data of the enrolled participants.

3.6 DATA ANALYSIS

The data collected during the pre test and post test was analyzed using the software *SPSS*. First, the data was coded by classifying each variable from the background questionnaire, PA and FLE questionnaires, and vowel intelligibility test. For all of the instruments besides the background questionnaire data, questions were labeled numerically in order to obtain individual values for each question and carrier word. Data was then separated by each participant, labeled from S01 to S30 (Speaker). Then, mean values were calculated for all the coded variables.

Descriptive statistics were run on each individual variable, as well as normality tests. After observing the Shapiro-Wilk values, the distribution of the values in the histograms, and due to the low number of participants (N= 30), it was decided that the use of non-parametric tests would be more adequate. This is further exacerbated by the fact that in the following post-test data, the number of participants dropped significantly (N=10), which resulted in even more unlikely normally distributed data. Furthermore, additional attention was paid to missing data and outliers, which resulted in the removal of 4 participants from the final data analysis process, as they did not have adequate vowel intelligibility test data. This occurred due to software error, as *Microsoft Word Online* uses cloud data to save text documents, and it can be assumed that connectivity issues resulted in the loss of the respective data files.

Regarding the calculation of mean scores for each variable, the following steps were taken. For the vowel intelligibility score, values for front vowels, back vowels, total vowels, and individual vowels were analyzed. This was done by calculating the sum of the total

numbers of correctly transcribed words and subtracting by the total number of words, multiplied by 100 to acquire scores in percentages. This was done for each of the intelligibility score categories for each vowel and vowel pair. For PA and FLE scores, questions were totaled up and divided by total number of questions (50 for PA, 10 for FLE) in order to acquire the mean score of participants, creating thus one composite score for each measure. In specific, for PA, some questions were reverse coded (Baran-Lucarz, 2014) as they were worded positively instead of negatively. For PA, lower scores indicate less anxious, while for FLE, higher scores indicate more enjoyment. Composite scores were used for both scales, and no categorical grouping of questions was conducted, as was previously stated in the method. This means that in this study, we will not be assessing PA by different subcategories attached to the three different types of self-perceptions, or dividing FLE in FLE-social or FLE-private (Dewaele; MacIntyre, 2016).

In order to run nonparametric partial correlation tests to answer RQ1, a syntax script was created in order to allow for the computing of Pearson partial correlations of Spearman's correlation rho values of the desired variables (PA, FLE, and vowel intelligibility scores). Lastly, in order to answer RQ2-4, related samples Wilcoxon signed rank tests were run on vowel intelligibility, PA and FLE values at pre and post test. This was done in order to observe the degree of instruction effects after four weeks of pronunciation learning via ASR use. Again, nonparametric tests were chosen due to the low number of participants (N=10).

The following section aims to present the descriptive statistics for all variables, as well as the results from the aforementioned tests.

4. RESULTS AND DISCUSSION

This section presents and discusses the results of the vowel intelligibility tests, pronunciation anxiety and foreign language enjoyment questionnaires, and the respective tests that were applied in order to answer research questions 1-4. As aforementioned, non-parametric statistics were used as the data was not normally distributed.

First, descriptives of the vowel intelligibility tests, PA and FLE questionnaires taken at T1 will be presented. In the following paragraphs, vowel intelligibility scores will be

reported, and then, PA and FLE will be discussed. In the table below, participants' levels of PA, FLE, and overall and vowel-specific intelligibility:

Table 5 - PA, FLE, and Intelligibility scores averaged across participants in the pre-test

	Intelligibility score %								
			Overall intelligibility	Vowel intelligibility					
	PA	**FLE**		Front	Back	æ	e	u	ʊ
N=30	3.17 (0.73) 2-5	4.35 (0.51) 3-5	50,6 (15.1) 25-85	46.3 (14.9) 20-80	55.6 (19.0) 30-100	52,6 (24,9) 0-100	40.6 (21.3) 20-80	46 (22.9) 20-100	63.3 (22.9) 20-100

Source - Santos (2023)

PA measured in a 6-point Likert scale. FLE measured in a 5-point Likert scale. Vowel Intelligibility means measured in percentages. Standard deviation between brackets. Range of responses on the third line.

Before continuing with further analyses, it is important to note that participants were recruited from three different classes ministered by two different instructors, where class 1, 2, and 3 came from the first year, second year, and third year of the course *Letras - Inglês* at UFSC. This could have resulted in skewed data regarding FLE and PA, as participants had different classroom environments. In order to ensure that participants could be grouped together, descriptive statistics were run after separating each participant according to their class, as seen in Table 6

Table 6 - Participant descriptive values separated by class

N=10	Self-assessed Proficiency	FL use (%)	T1 PA	T1 FLE	Overall Intelligibility (%)
CLASS 1 (*N*=6)	3.8	62.9	3.7	4.0	45.8
CLASS 2 (*N*=11)	3.6	49.9	2.8	4.5	51.3
CLASS 3 (*N*=13)	3.8	55.0	3.1	4.3	52.3

Source - Santos (2023)

A Kruskal-Wallis test showed that there was no significant difference of means (p>0.50) for self-assessed proficiency, FL use, PA, FLE and vowel intelligibility across all three groups. Thus, for all the following analyses present in this study, participants will be treated together as one group.

In Table 5, it can be observed that participants' degree of intelligibility was reportedly low (*M*=50.66%), indicating that on average, participants had 10 out of 20 words transcribed incorrectly by *Microsoft Word Dictate*, considering that any word transcribed outside of the exact target word is considered an error. This result appears to not align with previous studies done with non-native speakers of English (Mccrocklin, Edalatishams; 2020), and studies done with L1 BP speakers' samples of FL English speech (Gottardi, Kivistö- de Souza; 2022), which have found higher intelligibility rates for L1 BP accented English speech.

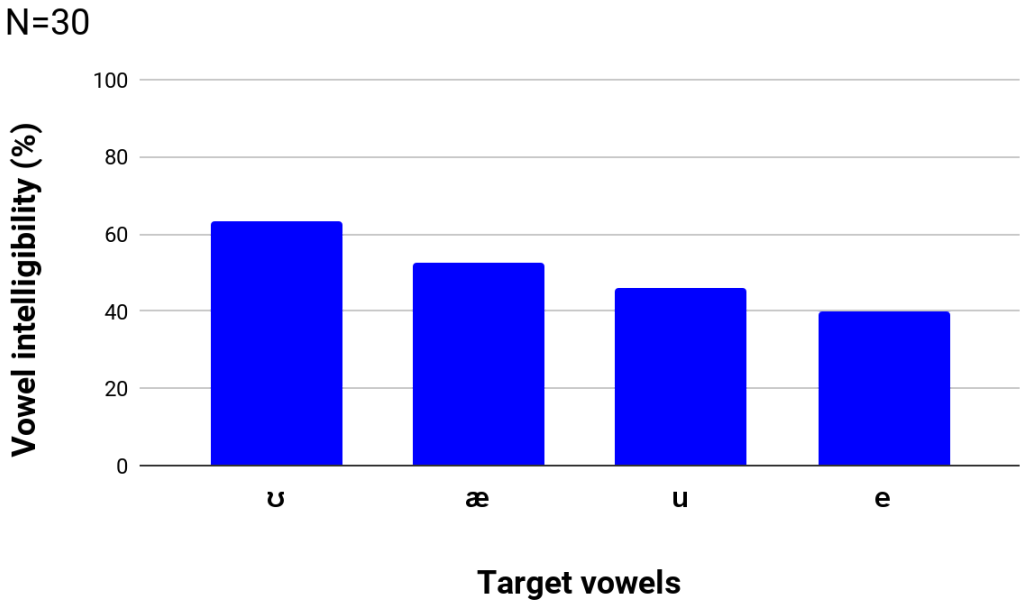
Furthermore, in order to answer RQ1, descriptive statistics of both PA and FLE variables were also run. The values on Table 5 above show that on average, the participants were moderately anxious about their pronunciation (*M*= 3.17, *N*=30) and present high FLE. (*M*= 4.35, *N*=30). As aforementioned, it is assumed that participants who score lower in the PA questionnaire (range 1-6) are less anxious, and participants who score higher in the FLE questionnaire (range 1-5) experience a higher degree of enjoyment in the FL classroom.

After presenting the descriptive data collected on vowel intelligibility, PA and FLE, it is important to further elaborate on participants' low vowel intelligibility scores. Specifically, the values obtained for each English vowel could be a result from the words chosen for testing, as word familiarity and frequency were not considered for this study due to the relatively low number of minimal pair words for the high-back pair /ʊ/ and /u/. The vowel /e/ also posed issues in finding frequently used carrier words, which thus resulted in using less familiar words in specific phonetic contexts. Furthermore, difficulties with pronouncing the target vowel /e/ could have been due to the measurement used in this paper. Even if participants had successfully gotten their vowel production /e/ to be recognized by MW Dictate, if the word wasn't the carrier word indicated in the test, it was regarded as an incorrect transcription. This was done in order to avoid addressing matters of coarticulation as incorrect words with correct vowel sounds (e.g; *Bag* instead of *Bat*) would result in different vowel quality for the expected production, leading the test design to analyze each mistranscription in a case-by-case manner, which was not possible with the limited time available for this present study.

Both /u/ and /e/ had carrier words that were often frequently mistranscribed by MW Dictate during testing, those words being *Pep* and *Luke*, which greatly swayed the scores of

participants as they posed great difficulty in correct pronunciation due to their overall unfamiliarity with the present words. When looking at the intelligibility values for each individual vowel, /ʊ/ appears to be the most intelligible sound (*M*=63.33%), while /e/ and /u/ make up for a significant portion of the overall low intelligibility scores across all participants (*M*=40.67% and *M*=46%, respectively). Due to the higher intelligibility mean score for the vowel /ʊ/, back vowels had overall higher intelligibility scores (*M*=55.67%), almost 10 points higher than the average mean for the front vowels (*M*=46.33%) The values for the degree of intelligibility of each individual vowel can be seen in bar graph below:

Graph 1 - Intelligibility vowel values in the pre-test



Source: Santos (2023)

Participants' low intelligibility scores for vowels /u/ and /e/ meant that further careful observation of the data had to be conducted in order to obtain vowel intelligibility data after controlling certain problematic words. Below are intelligibility values obtained for all participants in the pre-test after removing words which had less than 10 correct transcriptions out of 30 total productions transcribed by MW Dictate. Thus, the words *Pep,*

Luke, and *Peck,* were removed from all participants' scores, (*N*= 17 for total word count compared to the previous 20 words included in the test).

Table 7 - Intelligibility scores of averaged across speakers (with the exclusion of outlier carrier words) in the pre-test

	Intelligibility score %						
N=30	Overall intelligibility	Vowel intelligibility					
		Front	Back	æ	e	u	ʊ
	57.0 (15.2) 29-88	55.4 (16.9) 25-100	58.5 (19.9) 22-100	52.6 (24.9) 0-100	60.0 (26.8) 33-100	52.5 (25.7) 0-100	63.3 (22.9) 20-100

Source - Santos (2023)

As it appears, overall intelligibility scores increased a total of 4 points after controlling for outlier words. Participants' intelligibility of vowels /e u/ appear to differ from the original values considering all 20 carrier words, where vowels /e/ and /u/ reported a large 10-20 points increase in vowel intelligibility scores. However, participants' overall intelligibility still reports low-to-intermediate scores (*M* =57.0 *N*=30). This is important to consider for the following analyses as it can be assumed to some degree that participants' were unfamiliar with the differences between the vowel pairs presented in this research, and therefore were susceptible to targeted instruction on these specific vowels. Furthermore, after conducting a related samples Wilcoxon signed-rank test on the vowel intelligibility scores before and after controlling for outliers report a large significant increase in vowel intelligibility (z=4.769, N=30 p<.0.05), with an effect size of r= .87. Thus, for further analysis, the intelligibility score values after controlling for outlier words will be considered.

4.1 RELATIONSHIP BETWEEN VOWEL INTELLIGIBILITY, PA, AND FLE

This subsection will include analysis and discussion on research question 1, which involves assessing the relationship between participants' vowel intelligibility, pronunciation anxiety, and foreign language enjoyment. In particular, this study is also concerned with students' vowel intelligibility scores, PA and FLE at the pre-test stage. In order to answer

RQ1, correlation analyses were run. Initially, a bivariate Spearman correlation was run on PA and FLE in order to consider whether either variable should be controlled for when observing their effects on vowel intelligibility values. In particular, it is assumed that FLE and PA share a weak negative correlation, despite both being separate dimensions (Dewaele; Macintyre, 2014; Baran-Łucarz; Lee, 2021). As expected, Spearman correlation of PA and FLE variables in this study (N=30) showed a weak negative correlation (r_s=-.371, N=30, p<0.05). Afterwards, two additional bivariate Spearman correlations were run on PA, FLE and vowel intelligibility scores to assess the relationship between the variables. Here, PA and vowel intelligibility scores posit a moderate negative correlation (r_s= .-583, N=30, p<0.01) and a moderate positive correlation between FLE and vowel intelligibility was also observed (r_s=.431, N=30, p>0.05). However, correlation between FLE and vowel intelligibility scores do not appear to be statistically significant.

After confirming the weak correlation between both variables, partial Spearman correlations were considered for further analysis, in order to answer research question 1. When controlling for one variable at a time (e.g., controlling for FLE when analyzing the correlation between PA and vowel intelligibility scores), correlations between PA and vowel intelligibility were moderately negatively significant (r=-.505, N=30, p=0.04), while FLE and vowel intelligibility test only showed a weak correlation that was not statistically significant (r=.285, N=30, p=0.180). These results indicate that in 50% of the cases, participants with lower foreign language anxiety produced more intelligible vowels. On the other hand, FLE only posited a weak, non-significant positive correlation with learners' vowel intelligibility scores. This indicates that after controlling for each variable, FLE appears to not impact participants' vowel intelligibility scores as much as PA.

In order to answer RQ1, it is possible to further elaborate on the findings obtained through the prior statistical tests run in this study. First, it was noticed that participants' overall FL English vowel intelligibility scores are linked to pronunciation anxiety, but not to FLE. In particular, the participants that were less anxious about their pronunciation were more likely to score higher in the vowel intelligibility test. FLE, on the other hand, only revealed a weak, positive, and non-significant correlation with vowel intelligibility scores. This posits the argument that despite studies done on learners' affective factors in the FL classroom shifting towards the notion of enhancing learners enjoyment in classroom

contexts (Dewaele; MacIntyre, 2014; 2016), anxiety may still play a significant role in hindering learners FL pronunciation skills. This suggests that elaborating approaches that reduce learners' anxiety, such as using out-of-classroom instruction, and diminishes participants self-image concerns regarding their peers and teachers, may aid EFL learners in becoming more communicative, since pronunciation anxiety is linked to speakers' FL willingness to communicate (Baran-Lucarz, 2014). Similarly, while still being separate dimensions, it may be possible that despite linguistic anxiety often being more dynamic and inconsistent over FLE (Dewaele, J.M; Dewaele, L; 2017), the development of less anxiety-inducing contexts for pronunciation may aid learners in having more fun in the FL learning process.

Second, it is also important to understand participants' overall low FL English vowel intelligibility scores, as this was an unexpected outcome of the data retrieved. It is important to reinstate MW Dictate as a capable tool for assessing the intelligibility for non-native EFL speakers. Overall participants' FL English intelligibility appears to be significantly lower (50.66%) when compared to previous studies done with L1 Spanish, Chinese, and L1 BP speakers' pronunciation of FL English assessed via ASR tools (Mccrocklin, Edalatishams; 2020; Kivistö-De Souza, Gottardi, 2022). In this study, it is believed this could have been a result of a couple of factors. Initially, it was assumed that the use of unfamiliar and infrequent words may have affected a significant portion of the intelligibility score values for select vowels, both at pre-test and post-test. Due to the low amount of minimal-pair words available for vowels /ʊ/ and /u/, and the focus given to phonetic context in order to prevent coarticulation during the development of the testing materials, some unfamiliar and infrequent words were chosen for the stimuli. The phonetic context here was chosen with two specific criteria in mind: 1) match neighboring consonants for each vowel and vowel pair for all twenty words, and 2) choose minimal-pair or near-minimal pair words for the testing phase, but only present one word of the pair during testing, and the corresponding pair word in the treatment. In particular the presence of problematic words such as *Pep* and *Luke* which were almost mistranscribed by all participants in the study were assumed to have impacted participants' overall vowel intelligibility scores. As there was a significant change in participants' vowel intelligibility when controlling for outlier words, 3 total words were removed, but overall intelligibility did not increase past 60%.

An issue that could have affected the intelligibility scores is the sound quality of the input provided by the participants. Due to using a cellphone as the method of recording, it is possible that sound quality may have been hindered during the testing process. It is important to reinstate that the mobile phone was used as the primary recording device, and while a microphone was also used, it was secondary to the overall research design. This was done for a number of reasons: 1) all prior testing done with the target vowels and carrier words was done by the researcher using the same mobile device, being capable of replicating all 30 words (including distractors) accurately with no mistranscriptions. 2) microphone recorded data was on average, far less intelligible and reported some issues such as not picking-up any sound uttered by the speakers, and difficulties with maintaining the ASR tool activated (MW Dictate would often "turn off" as it did not pick up any sound for a long period of time). The room in which participants were tested also remained silent and the use of the mobile phone also allowed the researcher to always have internet connectivity, as the combination of Wi-Fi and mobile data could further prevent loss of audio input mid-transcription if one of the connectivity types were to not be available.

Due to these problems, it is important to consider that while participants' overall intelligibility was low compared to previous studies assessing intelligibility with ASR, it does not necessarily reflect negatively on the ASR tool's transcription ability. The only remaining question that could be asked regarding the low intelligibility scores of participants is regarding carrier word and sentence choice. In this study, it was chosen not to present words with a carrier sentence "*I say xxxx this time*", which was a conscious decision to avoid contextual clues from interfering with the transcription provided by MW Dictate. Further reflection on this decision poses a question whether a carrier sentence that doesn't include contextual clues could have been included in this study design. Nonetheless, as controlling for unfamiliar outlier words still reflected similar intelligibility scores for all participants, it can be argued that Microsoft Word Dictate is still capable of transcribing non-native EFL speakers, despite the low intelligibility scores.

This section aimed to answer research question 1 and the specificities related to the vowel intelligibility scores. It has discussed how PA is partially correlated to vowel intelligibility scores, while FLE does not pose a significant link to participants' vowel intelligibility scores. That is, participants with lower anxiety are more intelligible in

producing English vowels. This suggests that FL speakers who appear to have difficulties in FL pronunciation classes, or participants who have poor FL intelligibility may be highly anxious about their FL pronunciation self-image, assessment and efficacy. It may be particularly important for FL teachers to consider how pronunciation anxiety may directly hinder participants' willingness to participate in FL classroom contexts, and how PA may reflect on participants' English vowel intelligibility, as well on other aspects of FL pronunciation. Despite being an emotion reportedly less difficult to directly control in the classroom context (Dewaele et al., 2017), out-of-classroom approaches to teaching may aid in removing some of the threatening aspects related to speakers' pronunciation self-image. Furthermore, participants' FLE, while showing a weaker link to overall English vowel intelligibility, may still play a supportive role as an affective factor in the foreign language learning context. Considering the malleability of the emotion and influence of the teacher on participants' sense of joy in the classroom context, it is suggested that creating FL pronunciation classes that consider learners' preferences and subjectivities, while providing them with fun and joyful classroom experiences may boost participants' desire in communicating in FL pronunciation activities and becoming more intelligible. This is especially important if we are to consider that even highly anxious students may report high degrees of enjoyment, and that it may not always necessarily characterize anxiety as a "negative" emotion (Dewaele; MacIntyre, 2019). For the next section, the results obtained from analyzing the treatment effects will be presented and discussed.

4.2 THE EFFECTS OF ASR INSTRUCTION ON PARTICIPANTS' AFFECTIVE FACTORS AND VOWEL INTELLIGIBILITY

In this section, non-parametric tests were run in order to assess instruction effects on post-test scores (participants' vowel intelligibility, pronunciation anxiety, and foreign language enjoyment) and to answer research questions 2-4. Here it is important to note that a sharp drop in participants can be evidenced from pre-test to post-test, as only a total of ten participants finished all of the steps in this present research. This phenomenon will be later discussed when we observe patterns and behaviors of the remaining ten participants who did the post-test.

In order to observe the effects of the treatment with ASR instruction included in this study, related samples Wilcoxon signed-rank tests were run from T1 to T2 values on vowel intelligibility score, pronunciation anxiety composite scores, and foreign language enjoyment composite score. Table 10 shows the scores for all the variables before and after the time of testing for the participants who completed the instruction period (*N*=10).

Table 8 - PA, FLE and vowel intelligibility scores at pre-test and post-test (instruction effects)

	T1 (pre-test)			T2 (post-test)		
N=10	Mean	Sd	Range	Mean	Sd	Range
Back vowel Intelligibility (%)	62.2	19.0	33-100	67.7	16.1	44-88
Front Vowel Intelligibility (%)	61.2	17.1	38-100	66.2	13.2	50-87
Pronunciation Anxiety	3.0	0.7	1.8-4.0	3.0	0.7	2.1-3.9
Foreign Language Enjoyment	4.3	0.6	3.3-4.9	4.3	0.6	3.0-5.0

Source - Santos (2023)

As seen in table 8, the intelligibility scores appear to increase for front vowels and back vowel intelligibility at post-test. In order to test the influence of outlier words at T2, the same infrequent carrier words were controlled for in order to observe whether this could sway participants' overall intelligibility scores at post-test, which reflects an overall 5% increase over T2 values obtained prior to this analysis. However, participants' front vowel intelligibility (*M*=66.2, *N*=10), and back vowel intelligibility (*M*=67.7, *N*=10) remained relatively low-to-moderate.

Furthermore, when running related samples Wilcoxon signed-rank tests, no statistical significance for back vowel (z=.509, N=10, p=0.670), and front vowel intelligibility (z=.957, N=10, p=0.389) can be observed. This is also recurrent for the values of PA (z=1.173, N=10, p=.0.241) and FLE (z=.564, N=10, p=0.573) indicating that there is no statistically significant difference between pre-test and post-test scores. Thus, it can be assumed that instruction effects on participants were not enough to result in significant differences in affective factors or vowel intelligibility scores. The results in this study corroborates with Shams' (2006) findings regarding ASR and foreign language classroom anxiety, which suggests that the tool by itself may not be enough in aiding learners in becoming less anxious about their pronunciation. However, as will be later discussed, a

slight increase in vowel intelligibility scores may indicate that participants' awareness of the target vowels could have increased through testing, since learners could have become more conscious of the vowel pairs at T2, but still not have the appropriate motor skills to produce the target vowels distinctively. Again, observations made on participants' scores and results at T2 should be analyzed carefully considering the low number of participants in this part of the study. The lack of statistical power may be indicative of why little to no change was observed from T1 to T2 for PA, FLE and vowel intelligibility scores.

When analyzing the data, the behavior participants had in speaking certain carrier words were different. In particular, the word *Bat*, which was not uttered intelligibly by 26 participants at pre-test, was deemed to be intelligible by 7 out of the 10 participants at T2. This finding suggests that the treatment may have aided participants' with the production of the word. On the other hand, some carrier words containing the vowel /e/ still remained highly problematic both at T1 and T2, such as *Peck* and *Pep*, which were only uttered intelligibly by one participant for each individual word, and were removed from analysis. Furthermore, the vowel /æ/ became more intelligible in the production of the participants during T2, which may indicate that the familiarity with the English vowel may have increased during treatment, leading to bigger interference of the FL vowel on the L1 vowel (Flege, Bohn, 2021).

The results present in this study will be discussed by exploring possible reasons and causes for not seeing significant change in participants' vowel intelligibility from T1 to T2. First, since the instruction period only lasted a total of four weeks, it can be assumed that such a short amount of time may prove insufficient for the development of the required motor skills. Since participants' perception was not tested in this study, it is not possible to know whether they are more likely to discern between the vowel pairs included in the testing period. As perception can also be trained in short time periods (Rato, 2014) and that explicit awareness about these FL English vowels may have been acquired during the treatment period, it may be possible that the respective participants are in the middle of developing the necessary motor skills to produce these vowel pairs intelligibly, as the acquisition of new FL sounds often presents itself as a gradual process (Lima Junior; Alves, 2019). Furthermore, since participants did not get explicit feedback from the ASR tool, and due to the fact that MW Dictate was not designed specifically for FL pronunciation

teaching, participants could have struggled with the lack of appropriate feedback given by MW Dictate in order for participants to train their FL English production.

Since ASR instruction lacks appropriate feedback in the form of phonetic transcriptions, audio-visual cues (articulatory training), or immediate corrective feedback, participants in this study had to resort to the incorrect mistranscriptions in order to find what is deemed intelligible by MW Dictate. While this may be beneficial for developing FL phonological awareness, or for practicing learners' perception of English vowels, it may not transfer to participants' ability to produce English vowels intelligibly. Furthermore, the testing format (isolated words) may have made measuring intelligibility more difficult, as some particular words remained issues even after instruction, indicating that participants' overall scores both at T1 and T2 may have been brought lower due to the presence of specific unfamiliar words.

This section analyzed data obtained during the post-test in order to better understand the instruction effects of the teaching materials used in this study. According to the data presented here, the teaching material, while providing for a small increase in FL English vowel intelligibility, was not sufficient in providing for a statistically significant change in learners' overall intelligibility scores, pronunciation anxiety, and foreign language enjoyment. However, as the low number of participants during T2 could have impacted post-test scores, individual analyses for each participant should be considered.

After presenting the quantitative analysis resulting from the tests run on the collected data, this study now shifts towards discussing and observing the results from a qualitative point of view in order to better understand the results obtained and the possible effects that could have impacted the participants during testing. In order to better answer research questions 2-4 present in this study, the next subsection will discuss patterns and differences between the 10 participants who did all the steps. RQ2-4 will be discussed collectively, as the main focus will be to assess the instruction effects and the impact of the teaching materials on participant's FL English vowel intelligibility, and affective factors. By observing patterns, behaviors and difficulties encountered by individual participants, this study hopes to uncover to what extent ASR instruction could have directly impacted participants' vowel intelligibility, pronunciation anxiety, and foreign language enjoyment.

4.2.1 Individual Participant Analysis of ASR Instruction Effects

In this study, an out-of-classroom approach to reducing learners' PA and increasing FLE, while promoting vowel intelligibility gains was supported and tested, and the data reflected above still appear inconclusive regarding the use of ASR instruction in an out-of-classroom context. In order to further investigate the effects of the treatment in this study, the following subsection hopes to further analyze the present data in order to answer research questions 2-4, by providing further information on participants' behaviors, tendencies and emotions. The previously presented data was analyzed in order to observe the effects of ASR out-of-classroom instruction on participants' post-test vowel intelligibility scores, pronunciation anxiety, and foreign language enjoyment measures.

When observing each participant individually, some new findings can be discussed. Table 9 presents relevant information about the 10 participants who underwent the treatment.

Table 9 - Individual participant descriptive values at T1 and T2

N=10	Back vowel gain	Front vowel gain	Proficiency	FL use (%)	T1 PA	T2 PA	T1 FLE	T2 FLE	CLASS	Completed tests in time
S26	33.3	25.5	3	75.0	3.6	2.1	4.9	4.9	CLASS 3	No
S05	55.6	0.0	3	40.0	2.3	2.2	4.6	4.9	CLASS 2	Yes
S28	22.2	12.5	3	75.0	3.3	3.6	3.7	4.1	CLASS 3	Yes
S31	0.0	25.0	3	23.0	3.1	3.3	4.5	4.6	CLASS 3	No
S11	11.1	12.5	4	35.0	3.6	3.6	4.6	4.6	CLASS 2	No
S24	0.0	0.0	5	48.0	2.7	3.3	3.3	3.6	CLASS 3	No
S22	0.0	0.0	3	58.0	3.5	3.8	3.4	3.0	CLASS 3	No
S03	-22.2	12.5	4	50.0	1.8	2.2	4.8	4.7	CLASS 2	No
S17	-33.3	0.0	4	55.0	4.0	3.9	4.5	4.2	CLASS 1	No
S20	-11.1	-37.5	4	48.0	2.3	2.4	4.8	5.0	CLASS 1	No

Source - Santos (2023)

PA measured in a 6-point Likert scale. FLE measured in a 5-point Likert scale. Back and front vowel intelligibility gain scores measured in %. Completed tests in time indicate completion during the allotted 6 weeks.

First, out of the 10 participants who participated in all the steps in this study, only two participants finished all of the treatment activities without any delays. The remaining participants not only were often a week late to handing in their tasks, but also reported having a number of issues with using Microsoft Word. This was mainly reported to be due to two primary reasons: 1) older Microsoft Word versions often found in home computers are likely to not have the *Dictate* feature, which prompted participants to either switch to a different word processing tool (Google Docs), or to send emails to the researcher asking for guidance, and 2) some participants reported having issues setting the *Dictate* feature to American English in order to get appropriate transcriptions. Overall, most participants who completed all the tasks had to be emailed twice per week in order to be reminded of completing the steps. Unsurprisingly, this is reflected in participants' affective factors, which

most report marginally increased degrees of anxiety, and no changes to their enjoyment. Furthermore, participants' who reported experiencing technical difficulties (or were the most late to handing in all the tasks) were often the participants who performed worse at T2 than at T1 (S03, S17). In particular, both of these participants reported not finding MW Dictate early on through e-mail exchanges, and they report much lower intelligibility values for back vowels (30% less intelligible at T2). In fact, both participants who did the post test and who directly reported not knowing how to find and use MW Dictate, or had issues with the transcriptions, later showed increased anxiety and decreased enjoyment at T2 (S03, S22). Many of these technological frustrations may have been the reason as to why very few participants completed all 4 weekly tasks, and may have been a source of frustration or anxiety.

The only outlier in this analysis was S26, who not only had minor delays in completing all of the tasks, but also used Google Docs as opposed to MW Dictate. When observing the data in Table 9, it can be seen that S26 is not only the participant who had the most motivation at T1, but also demonstrated the largest positive change in pronunciation anxiety from T1 to T2. Furthermore, S26 was the most engaged participant during the treatment phase, maintaining a consistent and frequent email chain with the researcher by reporting issues and doubts about the procedures. Another observation that can be extracted from the values above is that on average, participants who reported the largest changes in their FLE (S28, S05) are amongst the participants that had the highest improvements to their English back vowel intelligibility scores at T2 (15-50%). The exception for this being participant S24, which evidenced only a 10% increase in back vowel intelligibility scores at T2. Furthermore, all three participants who observed the largest increases in vowel intelligibility scores at T2 (S31, S28, S26, S05) self-assessed their own FL proficiency as intermediate, suggesting that it is possible that at their degree of FL proficiency, the differences between these vowel pairs may still be a novel feature for these speakers.

Regarding participants with the lowest gain scores or decreased affective factor values at T2, it appears as if technological frustrations played a significant role. Participant S22, for example, had reported to the researcher personally during T2 that they felt very frustrated with their performance during the treatment, and that they felt they were worse than they thought they were at FL pronunciation. It can be argued that the way MW Dictate

provides feedback could be causing some degree of frustration to participants, as they do not get clear, thorough feedback with instruction on how to fix their productions. Since participants' have to "trial-and-error" their way into correct transcriptions, it may have made participants who performed worse feel more anxious about their FL pronunciation. Furthermore, those who self-reported higher degrees of FL proficiency may have felt as if the tool did not work properly, and that their productions were intelligible and should have been understood by MW Dictate. This points to another possible reason as to why PA and FLE values at T2 saw no significant change, and in particular, speaker S17 (who demonstrated a loss in FLE scores at T2 and assessed themselves as a intermediate-to-advanced EFL speaker), reported through email that:

> I was a little unsure about the question of attempts, but I tried to do what I understood: make the tool transcribe the word correctly three times within 10 attempts. **It was difficult to get it right**. I don't know if it appears in the document, but in the word "Net", there was a suggestion to correct it with "next" (My emphasis).

Thus, while MW Dictate may not directly affect participants' self-assessment related to pronunciation self-image (fear of being judged by peers or of not being able to speak in social settings), it could have had an impact in their beliefs regarding their pronunciation self-efficacy and self-assessment. This could have made the participants assume that they were not only less skilled than they believed to be, but also that English vowel sounds were perhaps more difficult to acquire than they believed them to be. As for FLE, it could be argued that some of these technological difficulties made the experience more tedious for the participants, since they had to troubleshoot the tool before being able to access them to their full potential. It can be assumed that participants' frustration with the technologies in use may have caused them to be indifferent about the act of learning. In specific, since FLE can be divided into FLE-social and FLE-private, it was assumed that part of their private enjoyment would have stemmed from their exploration with the ASR tool, due to the process of autonomous learning. However, it may be that the technology was not particularly enjoyment-inducing in this study, as presenting learners with isolated words may have not shown the full capabilities of MW Dictate. In this study, it was possible for

participants' to explore the tool on their own beyond using it for the weekly tasks (as a practice of autonomous learning), but it is unknown whether participants explored MW Dictate for their own curiosity, which would also partially explain why FLE did not see any significant change at T2.

It is important to note that FLE values at pre-test and post-test are nearing ceiling effect (M=4.35). This means that all participants, including those who did not complete all tasks, had positive to very positive degrees of enjoyment in the FL classroom. This may be due to the demographic chosen, as they were all students enrolled in an undergraduate program of English. Due to this, it may have been difficult to see an increase in FLE over the remaining 10 participants who completed all the steps, as they all already had very positive beliefs regarding FL learning. As previously stated, participants came from three different classes ministered by two different teachers, however, there was no statistical difference between groups at T1 in their degrees of PA, and FLE.

Finally, as this study was exclusively opt-in, with no monetary or academic benefit, students may have not been drawn to completing the tasks or fully interested in the experiment. Since this research was done during participants' academic semester, it had to compete with other obligatory tasks and activities that they were expected to resolve during the weeks of testing. Thus, this could have made participants overall less enthusiastic about participating, and it could have also had an impact on participants' anxiety levels, as they had to leverage doing an optional task as opposed to their obligatory course activities. The lack of direct feedback may have caused some participants to remain anxious or frustrated about their own FL English pronunciation skills, despite noticing differences in the vowel pairs. This is exacerbated by the fact that participants only had four weeks of practice, and that the short instruction period may have not been sufficient to train participants' motor skills on the target vowels. However, according to the PA and FLE values at T2, MW Dictate did not cause participants' beliefs and emotions about FL English intelligibility to be worse than prior to treatment, as participants on average did not report feeling any more anxious about their pronunciation, nor did they report feeling any less joy learning a FL. Overall, the effects of ASR on participants' affective factors appear inconclusive, and there is a chance that some of the benefits provided by ASR instruction could be affected by

technological difficulties, limited period of instruction, or frustrations with the method of feedback.

In this subsection, research questions 2-4 were discussed. It was argued that despite seeing a slight increase in overall vowel intelligibility scores, participants' FL English pronunciation of the high-back and low-front vowel pairs /æ-ɛ u-ʊ/ were not significantly higher at post-test. This indicates while participants could have become more conscious about these vowel pairs, they may not have developed the motor skills to produce the differences accordingly. The main takeaway from this discussion is that participants' recruited for this study could have very likely developed some degree of awareness over the vowel pair, as observed by the fact that speakers had difficulties with different words than the ones considered problematic at T1. The ability participants developed in making themselves intelligible when speaking words such as *Bat* indicate that to some degree, participants began noticing the peculiarities of the vowel pairs included in this study. It is likely that treatment using ASR instruction such as the one used in this study may serve a purpose in aiding participants with the development of their FL phonological awareness, as participants have also reported noticing the form (S17), and noticing the gap (S22) between their production and the desired target language pronunciation. Learners who are seeking to uncover problems with their English vowel sounds may benefit from testing their degree of intelligibility on tools such as MW Dictate, by utilizing a similar approach to instruction to the one used in this study (e,g., isolated words instead of full sentences), even if they do not ultimately acquire the appropriate motor skills required to pronounce the correct English vowels through testing.

Despite the aforementioned benefits, it is necessary to understand that MW Dictate poses some difficulties for intermediate-to-advanced level EFL speakers. As discussed, the lack of thorough feedback that includes instruction on how to pronounce target words or sounds, phonetic transcriptions, or visual articulatory videos and images, may not be sufficient for learners who are in the early stages of noticing a new FL sound. The use of additional feedback resources may contribute greatly to the functionality of ASR tools not made for instruction, such as MW Dictate and Google Docs. Another challenge instructors and teachers may face when integrating out-of-classroom ASR instruction to their course plan comes from fostering their students' autonomous learning as a practice. If learners are

to only use ASR tools for a limited amount of time, without attempting to explore the capabilities of what can be done with automatic speech recognition, it is likely that they will not extract all of the learning that could be acquired from the interaction.

When it comes to research questions 3 and 4, a similar pattern was observed. Again, no significant change in participants' PA and FLE was achieved at T2, which indicates that on average, ASR instruction did not aid participants' with the affective factors regarding FL pronunciation and FL learning. It is only when the individual learner was assessed that reasoning could be given for this specific occurrence. As previously mentioned, out of the 10 participants, many of them reported a number of difficulties and challenges related to the use of MW Dictate. Frustrations ranging from operating the tool, improper feedback, or their feelings of FL pronunciation self-efficacy and self-assessment were expressed from the reports given by the participants. This indicates that while ASR tools such as MW Dictate may provide benefits for participants' beliefs on FL pronunciation self-image, as it excludes their fear of negative evaluation from their peers, it still poses threats to participants by making them feel like their English vowel pronunciation isn't intelligible enough for the tool to understand. The technological difficulties may also make the process more tedious, so it is important for learners engaging with this tool to develop a sense of familiarity with the peculiarities of MW Dictate and how to utilize it. Language instructors should also emphasize that the harsh judgment given by MW Dictate to some of the utterances spoken by their students is not entirely indicative of their overall FL pronunciation skills, so that they don't feel as anxious with the feedback these learners receive.

Overall, for the fostering of FL affective factors, out-of-classroom ASR instruction proves at least to be as adequate as in-classroom teaching, as engagement with MW Dictate did not make participants feel more anxious or more bored. It is likely that some of the external issues aforementioned may have also played a role in their beliefs regarding the treatment and the tests, as participants likely had to prioritize other class activities during the period of testing. Furthermore, the nature of the treatment (1 task every week), consisting of strictly isolated words, may not have instilled participants' curiosity in further testing and exploring the *Dictate* feature, which may have hindered their practices of autonomous learning, and limited their ability to further develop a sense of joy when learning a foreign language. Just as mentioned above regarding vowel intelligibility, developing learners'

autonomous learning practices may pose essential for the development of more joy in FL learners engaging with ASR tools such as MW Dictate. Despite this, as participants' FLE scores were nearing ceiling effect, possibly due to the demographic chosen (undergraduate students), which indicates not much change could have occurred from T1 to T2.

5. CONCLUSIONS

This section aims to restate the objectives, procedures, results, and main findings of the study. It also hopes to add concluding remarks in order to answer the research questions that guided the steps of the research. It also hopes to shed light on some of the limitations present in this paper, and what future research in the field of ASR instruction for FL pronunciation learning could adopt.

In this study, the main objective was to assess the ASR tool *Microsoft Word Dictate* and the extent of how it relates to non-native speakers' English vowel intelligibility, pronunciation anxiety, and foreign language enjoyment. Specifically, it observed the effects that ASR instruction in an out-of-classroom setting could have in the development of more intelligible production of the target vowels /æ-ɛ u-ʊ/ by L1 BP non-native speakers of FL English. It also analyzed the effects of ASR instruction on participants' affective factors, as the leading hypothesis behind this study was that out-of-classroom instruction could aid anxious participants in practicing their FL pronunciation in a less anxiety-inducing environment. Furthermore, it was also assumed that the practice with MW Dictate would appear as something novel and fun for the participants, which would have resulted in a change in their beliefs on enjoyment in the foreign language learning context. In order to achieve this goal, a number of procedures were developed with the intent of testing intermediate-to-advanced FL English speaking participants recruited for this study before and after instruction with MW Dictate. First, the target vowels /æ-ɛ u-ʊ/ were chosen due to their difficult nature for L1 BP speakers of FL English. Then, a vowel intelligibility test was devised in order to assess participants using the ASR tool MW Dictate, combined with three questionnaires pertaining to linguistic background information, pronunciation anxiety, and foreign language enjoyment. Then, an instruction period was crafted considering the target

vowels chosen for this study, utilizing MW Dictate as a tool for consciousness-raising activities that fostered participants' autonomous learning practices, totaling four tasks over four individual weeks. Lastly, participants then underwent the procedures designed for this longitudinal study, totaling six weeks of testing and instruction.

In order to answer research question 1 - *What is the relationship between vowel production intelligibility and pronunciation anxiety and foreign language enjoyment in intermediate level EFL learners?* - Some considerations had to be first addressed. First, while not central to this study, it was observed that participants on average had low intelligibility score values obtained across the board. This was discussed in order to guarantee that MW Dictate was still suited for the transcription of intelligible non-native English speech. After controlling for outlier words and stating external problems that could have hindered participants' intelligibility scores, it is believed that MW Dictate still showed adequacy in understanding L1 BP non-native speakers of English, indicating that there could be a gap in participants' pronunciation and the desired TL production of the target vowels. Again, attention is given to the aspect that this is in part, addressing accurate productions of target vowels, but it is important that the production of these target vowels demonstrate a certain degree of accuracy should they not cause communication breakdowns and thus directly affect speakers' intelligibility.

Afterwards, a number of non-parametric correlation tests were run between PA, FLE and vowel intelligibility scores for all 30 participants enrolled in this research in order to answer research question 1. The data collected in this study suggests that there is a moderate negative correlation between PA and vowel intelligibility, and a weak positive non-significant correlation between FLE and vowel intelligibility. This implies that participants' pronunciation anxiety was to a degree predictive of their intelligibility scores in this study, and that aiding participants in feeling less anxious may result in better English vowel intelligibility. FLE, while indicating a much weaker relation to vowel intelligibility scores, may still play a secondary role in boosting participants' autonomous learning practices in order to more directly engage with the ASR tool on their own.

In particular, it was argued that speakers who have more pronunciation anxiety may not engage with the opportunities to practice their FL pronunciation as it was stated that PA and FL-WTC are linked (Baran-Lucarz, 2014). Furthermore, it was emphasized that despite

PA, and language anxiety of FL learners as a whole being a difficult individual difference to control by language teachers and instructors, that it still plays a significant role in predicting participants' degree of intelligibility. Thus, finding approaches that address learners' pronunciation self-image, self-assessment, and self-efficacy, may aid FL learners' ability to practice their intelligibility. To a degree, a similar argument can be said for fostering learners' enjoyment in the foreign language learning environment, as participants who enjoy learning a foreign language may be more likely to participate in FL pronunciation activities and opportunities. Despite this, FLE appears to not necessarily indicate that those who do participate in FL pronunciation activities are likely to be more intelligible.

After assessing the values at T1, it was necessary to observe the treatment effects using ASR instruction in order to answer research question 2. After explaining the dropout effect and the low number of participants who completed all the steps of the study, it was observed that on average, the use of out-of-classroom ASR instruction of English vowels /æ-ɛ u-ʊ/ did not show statistically significant differences to participants' vowel intelligibility scores at T2. This was argued to be a result of a number of possible causes, linked to the type of feedback given by MW Dictate, or the difficulty in acquiring the adequate motor skills to produce the target English vowels. Furthermore, limited time of instruction may have also contributed to the lack of immediate gains in vowel intelligibility.

However, after investigating the patterns of mistranscriptions encountered in participants' tests at T2, it was found that participants had overall fewer problematic words at T2, and that in particular, *Bat*, which was the most commonly mistranscribed word at T1, was uttered appropriately by more than half of the participants at post-test. This may indicate that participants were able to develop some degree of awareness over the vowel pairs /æ-ɛ u-ʊ/ and that while they may still not have fully developed the motor skills to produce these words intelligibly to MW Dictate, they may have noticed the form and the gap between their production and the desired TL sound. This was further evidenced by participants' reports of frustration with not getting the words to transcribe properly during treatment. It is possible that ASR instruction with MW Dictate may have had some benefit to participants' FL phonological awareness, but the present data in this study is yet inconclusive. Furthermore, when looking thoroughly into the individual differences of each participant, it was noticed that the participants which had less issues throughout the study,

and those who had more motivation at T1, were likely to have received greater improvements to their vowel intelligibility scores at T2, in particular for the back vowels /u-ʊ/. This suggests that participants who are more familiar with the technology, or have more enjoyment in the practice of FL learning, may benefit more from ASR instruction with MW Dictate. Language instructors should be careful about implementing ASR instruction without direct guidance or additional feedback for participants who do not appear to find joy in the FL learning context, since MW Dictate may not be fully appreciated by students who do not find it enjoyable to explore the tool for learning purposes.

In regards to research questions 3 and 4, related to participants' degrees of pronunciation anxiety and foreign language enjoyment, no statistically significant difference appears to have happened from pre-test to post-test. On average, participants retained similar levels of PA and FLE after instruction via ASR. It was noticed that some participants reported more anxiety and less enjoyment after post-test, while some reported the direct opposite effect. When observing each individual learner, some patterns emerged. It was discussed that the technological difficulties and type of feedback received may have obfuscated some of the possible benefits that could have been inherited from the use of ASR instruction in an out-of-classroom environment. That is, since participants felt frustrated about their experience at some points during the treatment, it may have caused them to feel more anxious or less enthusiastic about the practice, despite initially finding MW Dictate a tool that fosters overall less pronunciation anxiety and provides joy.

In particular, since it was assumed CALL tools used from a remote setting could benefit learners who are particularly anxious about their pronunciation self-image, it was expected that MW Dictate would remove some of those participants' fears, as they did not have to worry about being judged by their peers. However, since the tool does not directly address what was specifically wrong about their pronunciation, it may have made them anxious about their pronunciation self-assessment and self-efficacy. Nonetheless, despite this, overall scores for both affective factors still remained similar at T1 and T2, which suggests that despite frustrations encountered by participants, it does not mean MW Dictate is a less favorable option than in-classroom FL pronunciation instruction. As previously suggested above, giving FL learners additional feedback (phonetic transcriptions, articulatory tips with the use of images and videos, sound feedback in the form of

text-to-speech), as well more guidance in order to develop familiarity with the tool, may allow MW Dictate to perform more adequately and boost participants' affective factors. It may also be important to promote participants' autonomous learning practices by devising activities with MW Dictate that allow the participant to go beyond the task itself in order for them to foster FLE in interacting with ASR technology.

In this section, concluding remarks were made on the overall results of the study. While this study attempted to implement a method of ASR instruction that would aid participants' English vowel intelligibility, PA, and FLE, some limitations may have influenced the results of the research. In order to better understand the technology and the boundaries of how it can be used for FL learning, further research must be conducted, which will be discussed in the following subsection.

5.1 PEDAGOGICAL IMPLICATIONS, LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

Throughout the development of the study design, and during the testing and treatment phases, this present research encountered a number of limitations that may have contributed to the aforementioned results and effects. First, and most evident, is the number of participants included in this study. With such a low number of participants (N=30), which dropped to a total of only ten at post-test, many problems arose. The weak statistical power of this study means that further observations were required in order to complement the findings achieved through the analyses run in this study. Thus, the inclusion of observations of each participants' individual differences were also addressed in order to better understand the data collected in this study. Furthermore, the lack of participants also meant that a control group was not designed for this study, despite this being something which was desired in this paper in order to compare typical classroom instruction methods against ASR instruction. Furthermore, a control group could have also been used to test participants who did not receive treatment in order to observe if there were any differences between those who practiced and those who didn't.

Another limitation present in this study was the lack of an additional way of assessing participants' intelligibility. If this study had either included acoustic analysis or the

use of human listeners, it could have allowed the researcher to observe more directly whether MW Dictate was not transcribing intelligible non-native English speech, instead of creating an assessment with the removal of outlier words. This is also related to the question of whether the use of isolated words was the adequate choice for this type of study. While it is believed it was not a direct limitation of this study, it is possible results could have been different with the use of a carrier sentence.

Specifically, another issue that could have hindered the results present in this paper is the limited time of instruction. As previously discussed, the fact the instruction period was so short could have been the reason no change occurred to participants' vowel intelligibility scores, as developing the necessary motor skills may take longer to develop. An extended period of instruction could also give participants more opportunity to become more familiar with and explore MW Dictate, which could have impacted their affective factors further.

Finally, the last limitation pertaining to the study comes from the recording device used. While the mobile phone used by the researcher reported great results during pilot testing, and performed better than the microphones which were utilized, it is likely there could exist a better suited option for the recording of participants' voices. There is a chance that a better microphone could have allowed for better interaction between the participants and the ASR tool.

These limitations lead to a couple of suggestions for future studies, and the pedagogical considerations that will be (re)instated in this subsection. I believe that future studies should consider some of the limitations of the ASR tool and how it may reflect on what aspect of FL pronunciation can be improved. What this implies is that studies should consider whether tools such as MW Dictate may not be better suited for assessing and training perception instead of production. Furthermore, as this study used a consciousness-raising activity, future studies should also test participants' FL phonological awareness as opposed to their production.

Furthermore, studies should also consider comparing this method of instruction to other traditional means of instruction of hybridized instruction methods that include both ASR instruction and in-class guidance. It is likely that some of the limitations of the tool itself could be remediated by teacher interference, and this approach should also be tested for future research. Additionally, other technologies that include ASR should also be

considered for FL pronunciation instruction, such as the use of AI language models and intelligent personal assistants. Researchers should consider placing ASR as a mediator between learner and tools that go beyond direct transcriptions of speech, as it is likely that the potential of ASR as a tool for aiding participants' FL intelligibility and individual differences has not yet been fully observed.

Additionally, studies done with a larger number of participants, larger FL sound inventory, and a longer instruction period, may provide for a better insight in the specificities of the relationship between ASR instruction and FL learner. This present experiment serves as a platform for future research to explore many facets of ASR as a technology for FL learning, and it is necessary to further explore all of the yet inconclusive aspects of this novel technology applied into the educational context.

REFERENCES

AHN, T. Y.; LEE, S.-M. User experience of a mobile speaking application with automatic speech recognition for EFL learning. **British Journal of Educational Technology,** 47, n. 4, p. 778-786, 2016.

BALDISSERA, LG. **App resources for developing English pronunciation: a focus on mobile technology**. 2020. 103 f. Dissertação (Mestrado) - Curso de Programa de Pós-Graduação em Letras/Inglês e Literatura Correspondente, Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, Florianópolis, 2020. Disponível em: https://repositorio.ufsc.br/handle/123456789/215901. Acesso em: 25 out. 2022.

BARAN-LUCARZ, M. Pronunciation anxiety and willingness to communicate. **Proceedings of the International Symposium on the Acquisition of Second Language Speech (New Sounds 2013)**, v.5, p. 36-49, 2014. Disponível em: http://doe.concordia.ca/copal/documents/4_Baran-Lucarz_Vol5.pdf.

BARAN-LUCARZ, M. researching phonetics learning anxiety: A qualitative approach. **Second Language Learning and Teaching**, 27, p. 161-173, 2014.

BARAN-LUCARZ, M. The link between pronunciation anxiety and willingness to communicate in the foreign-language classroom: The Polish EFL context. **Canadian Modern Language Review-Revue Canadienne Des Langues Vivantes**, 70, n. 4, p. 445-473, 2014.

BARAN-LUCARZ, M. FL pronunciation anxiety and motivation: results of a mixed-method study. *In*: EWA PIECHURSKA-KUCIEL, E. S.-C., MAGDALENA SZYSZKA (Ed.). **At the Crossroads: Challenges of Foreign Language Learning**. 1 ed. Second Language Learning and Teaching: Springer Cham. cap. 2, p. 107-133. 2017

BARAN-LUCARZ, M.; LEE, J. H. Selected determinants of pronunciation anxiety. **International Journal of English Studies**, 21, n. 1, p. 93-113, 2021.

BARAN-ŁUCARZ, M. Foreign language pronunciation and listening anxiety: A preliminary study. *In*: PIECHURSKA-KUCIEL, E. e SZYMAŃSKA-CZAPLAK, E. (Ed.). **Language in Cognition and Affect**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013a. p. 255-274.

BARAN-ŁUCARZ, M. Conceptualizing and measuring the construct of pronunciation anxiety: results of a pilot study. *In*: PAWLAK, M. (Ed.). **Classroom-Oriented Research: Reconciling Theory and Practice**. Cham: Springer International Publishing, p. 39-56. 2016.

BASHORI, M.; VAN HOUT, R.; STRIK, H.; CUCCHIARINI, C. Effects of ASR-based websites on EFL learners' vocabulary, speaking anxiety, and language enjoyment. **System (Linköping),** 99, p. 102496, 2021.

BENSON, P. **Teaching and Researching: Autonomy in Language Learning**. 2. ed. London: Routledge, 2013. 296 p.

BLIN, F. CALL and the Development of Learner Autonomy: Towards an Activity-Theoretical Perspective. **ReCALL**, 16, n. 2, p. 377-395, 2004.

BODNAR, S.; CUCCHIARINI, C.; PENNING DE VRIES, B.; STRIK, H. Learner affect in computerised L2 oral grammar practice with corrective feedback. **Computer Assisted Language Learning**, 30, n. 3-4, p. 223-246, 2017.

BOTES, E.; DEWAELE, J.-M.; GREIFF, S. The Power to Improve: Effects of Multilingualism and Perceived Proficiency on Enjoyment and Anxiety in Foreign Language Learning. **European Journal of Applied Linguistics**, 8, p. 1-28, 09/12 2020.

BOTES, E.; DEWAELE, J.-M.; GREIFF, S. Taking stock: A meta-analysis of the effects of foreign language enjoyment. **Studies in Second Language Learning and Teaching**, 12, p. 205-232, 06/21 2022.

BROWN, A. Functional Load and the Teaching of Pronunciation. **TESOL Quarterly**, 22, n. 4, p. 593-606, 1988.

BUENO-ALASTUEY, C. Perceived benefits and drawbacks of synchronous voice-based computer-mediated communication in the foreign language classroom. **Computer Assisted Language Learning**, 24, p. 419-432, 12/01 2011.

BURLESON, D. Improving intelligibility of non-native speech with computer-assisted phonological training. **IULC Working Papers**. 2007.

CARLET, A.; CEBRIAN, J. Training Catalan speakers to identify L2 consonants and vowels: A short-term high variability training study. *In*, 2013. p. Volume 5, pp. 85-98.

CARLET, A.; KIVISTÖ-DE SOUZA, H. Improving L2 Pronunciation Inside and Outside the Classroom: Perception, Production and Autonomous Learning of L2 Vowels. **Ilha do Desterro**, 71, p. 99-123, 09/01 2018.

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.

CHIU, T.-L.; LIOU, H.-C.; YEH, Y. A Study of web-based oral activities enhanced by Automatic Speech Recognition for EFL college learning. **Computer Assisted Language Learning**, 20, n. 3, p. 209-233, 2007.

CHUN, D. M. Computer-assisted pronunciation teaching. *In*: **The Encyclopedia of Applied Linguistics**, p. 1-11, 2012.

CHUNG, B.; BONG, H.K.M. A study on the intelligibility of Korean-Accented English: Possibilities of implementing AI applications in English education. **J. Asia TEFL** 19, 19, 2022.

CONIAM, D. Voice recognition software accuracy with second language speakers of English. **System**, 27, n. 1, p. 49-64, 1999.

DERWING, T. Utopian goals for pronunciation teaching. **Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference,** Iowa State University, 2010.

DERWING, T. M.; MUNRO, M. J.; CARBONARO, M. Does popular speech recognition software work with ESL speech? **TESOL Quarterly**, 34, n. 3, p. 592-603, 2000

DEWAELE, J.-M.; MACINTYRE, P. The two faces of Janus? Anxiety and Enjoyment in the Foreign Language Classroom. **Studies in Second Language Learning and Teaching**, 4, p. 237-274, 2014.

DEWAELE, J.-M.; MACINTYRE, P. Foreign language enjoyment and foreign language classroom anxiety. The right and left feet of FL learning? **Positive Psychology in SLA.** p. 215-236, 2016.

DEWAELE, J.-M.; MACINTYRE, P. The predictive power of multicultural personality traits, learner and teacher variables on foreign language enjoyment and anxiety **Evidence-Based Second Language Pedagogy**. p. 263-286, 2019.

DEWAELE, J.-M.; WITNEY, J.; SAITO, K.; DEWAELE, L. Foreign language enjoyment and anxiety: The effect of teacher and learner variables. **Language Teaching Research,** 22, n. 6, p. 676-697, 2017.

DEWAELE, J.-M.; LI, C. Foreign Language Enjoyment and Anxiety: Associations With General and Domain- Specific English Achievement. **Chinese Journal of Applied Linguistics**, 45, p. 32-48, 02/10 2022.

DERWING, Tracey M.; MUNRO, Murray J.. Pronunciation Learning and Teaching. In: DERWING, Tracey M.; MUNRO, Murray J.; THOMSON, Ron I. (ed.). **The Routledge Handbook of Second Language Acquisition and Speaking**. New York: Routledge, 2022. p. 1-490.

DICKINSON, L. Autonomy and motivation a literature review. **System**, 23, n. 2, p. 165-174, 1995/05/01/ 1995.

DIZON, G. Evaluating intelligent personal assistants for L2 listening and speaking development. **Language, Learning and Technology**, 24, p. 16-26, 02/01 2020.

DIZON, G.; TANG, D. Intelligent personal assistants for autonomous second language learning: An investigation of Alexa. **The JALT CALL Journal**, 16, 10/12 2020.

ELLIS, R. The Place of Grammar Instruction in the Second/Foreign Language Curriculum. In: HINKEL, Eli; FOTOS, Sandra (ed.). **New Perspectives on Grammar Teaching in Second Language Classrooms**. London: Routledge, 2002.

EVERS, K.; CHEN, S. Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. **Computer Assisted Language Learning**, 35, p. 1-21, 11/03 2020.

FLEGE, J.; Second language speech learning: Theory, findings and problems. In: STRANGE, Winifred (ed.). **Speech perception and linguistic experience: Issues in cross-language research**. Baltimore: York Press, 1995. Cap. 8. p. 229-273.

FLEGE, J.; BOHN, O.-S. The Revised Speech Learning Model (SLM-r).**Second Language Speech Learning: Theoretical and Empirical Progress** p. 3-83, 2021.

FREDRICKSON, B.L. The value of positive emotions - The emerging science of positive psychology is coming to understand why it's good to feel good. **American Scientist**, **91:**330-335, 2003.

FOGERTY, D.; HUMES, L. The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences. **The Journal of the Acoustical Society of America**, 131, p. 1490-1501, 02/01 2012.

HAHN, L. D. Primary Stress and Intelligibility: Research to Motivate the Teaching of Suprasegmentals. **TESOL Quarterly**, 38, n. 2, p. 201-223, 2004.

HAN, T.; TANRIÖVER, A.; ŞAHAN, Ö. EFL students' and teachers' attitudes towards foreign language speaking anxiety: A look at NESTs and Non-NESTs. **International Education Studies**, 9, p. 1, 02/21 2016.

HASHIMOTO, Y. Motivation and willingness to communicate as predictors of reported L2 use: The Japanese ESL context. **Second Language Studies**, 20, 01/01 2002.

HOFMANN, S.; ASNAANI, M. A.; HINTON, D. Cultural Aspects in Social Anxiety and Social Anxiety Disorder. **Depression and Anxiety**, 27, p. 1117-1127, 12/01 2010.

HOLEC, H. Autonomy and Foreign Language Learning. **Oxford: Pergamon.** 9, 1981

HORWITZ, E. K.; HORWITZ, M. B.; COPE, J. Foreign Language Classroom Anxiety. **The Modern Language Journal**, 70, n. 2, p. 125-132, 1986.

INCEOGLU, S.; LIM, H.; CHEN, W.-H. ASR for EFL Pronunciation Practice: Segmental Development and Learners' Beliefs. **The Journal of AsiaTEFL**, 17, p. 824-840, 09/30 2020.

JUANG, B.; RABINER, L. Automatic Speech Recognition - a brief history of the technology development. **Elsevier Encyclopedia of Language and Linguistics.** 2005.

JENKINS, J. Which pronunciation norms and models for English as an International Language? **ELT Journal**, 52, 04/01 1998.

KASRANI, I.; VASOYA, M.; SHIVAKUMAR, A.; PEI, Y. A mobile cloud computing-based independent language-learning system with automatic intelligibility assessment and instant feedback. **4th International Conference on Mobile Learning**. 2018

KENNEDY, S.; TROFIMOVICH, P. Language awareness and second language pronunciation: A classroom study. **Language Awareness**, 19, p. 171-185, 08/01 2010

KING, R. D. Functional Load and Sound Change. **Language**, 43, n. 4, p. 831-852, 1967.

KIVISTÖ-DE SOUZA, H. (2015). **Phonological awareness and pronunciation in a second language.** Unpublished doctoral dissertation. University of Barcelona, Barcelona, Spain.

KIVISTÖ-DE SOUZA, H.; CARLET, A.; JULKOWSKA, I.; RATO, A. Vowel inventory size matters: Assessing cue-weighting in L2 vowel perception. **Ilha do Desterro**, 70(3), 33-46., 2017.

KOHONEN, V. Experiential language learning: second language learning as cooperative learner education. **Collaborative language learning and teaching**, *1439*. 1992

KORMOS, J.; CSIZÉR, K. The Interaction of Motivation, Self-Regulatory Strategies, and Autonomous Learning Behavior in Different Learner Groups. **TESOL Quarterly**, 48, 06/01 2014.

LAI, C. Learning beliefs and autonomous language learning with technology beyond the classroom. **Language Awareness**, 28, p. 1-19, 10/23 2019.

LEE, J. Informal digital learning of English and second language vocabulary outcomes: Can quantity conquer quality?: Informal digital learning of English. **British Journal of Educational Technology**, 50, 12/10 2017.

LEE, J. S.; LEE, K. The role of informal digital learning of English and L2 motivational self system in foreign language enjoyment. **British Journal of Educational Technology**, 52, n. 1, p. 358-373, 2021.

LEVIS, J. Changing contexts and shifting paradigms in pronunciation teaching. **TESOL Quarterly**, 39/3: 369–77, 2005.0

LEVIS, J.; SUVOROV, R. Automatic Speech Recognition. *In*, 2012.

LEVIS, John M.; SILPACHAI, Atif. Speech Intelligibility. In: DERWING, Tracey M.; MUNRO, Murray J.; THOMSON, Ron I. (ed.). **The Routledge Handbook of Second Language Acquisition and Speaking**. New York: Routledge, 2022. p. 1-490.

LI, J. J.; AYALA, S.; HAREL, D.; SHILLER, D. *et al.* Individual predictors of response to biofeedback training for second-language production. **The Journal of the Acoustical Society of America**, 146, p. 4625-4643, 12/01 2019.

LI, C.; JIANG, G.; DEWAELE, J.-M. Understanding Chinese high school students' Foreign Language Enjoyment: Validation of the Chinese version of the Foreign Language Enjoyment scale. **System,** 76, p. 183-196, 2018.

LIAKIN, D.; CARDOSO, W.; LIAKINA, N. Mobilizing Instruction in a Second-Language Context: Learners' Perceptions of Two Speech Technologies. **Languages**, 2, p. 11, 07/17 2017.

LIEBERT, R. M.; MORRIS, L. W. Cognitive and Emotional Components of Test Anxiety: A Distinction and Some Initial Data. **Psychological Reports**, 20, n. 3, p. 975-978, 1967.

LIMA JUNIOR, RM. The influence of metalinguistic knowledge of segmental phonology on the production of English vowels by Brazilian undergraduate students. **Ilha do Desterro,** 70, p. 117-130, 09/01 2017.

LIMA JÚNIOR, RM. A influência da idade na aquisição de seis vogais do inglês por alunos brasileiros. **Organon,** 30, n. 58, 07/06 2015.

MACINTYRE, P.; CHAROS, C. Personality, Attitudes, and Affect as Predictors of Second Language Communication. **Journal of Language and Social Psychology - J LANG SOC PSYCHOL**, 15, p. 3-26, 03/01 1996.

MACINTYRE, P.; CLÉMENT, R.; DÖRNYEI, Z.; NOELS, K. Conceptualizing Willingness to Communicate in a L2: a situational model of L2 confidence and affiliation. **The Modern Language Journal**, 82, p. 545-562, 1998.

MATSUURA, H. Intelligibility and individual learner differences in the EIL context. **System**, 35, n. 3, p. 293-304, 2007/09/01/ 2007.

MCCROCKLIN, S. Pronunciation learner autonomy: The potential of Automatic Speech Recognition. **System**, 57, p. 25-42, 04/01 2016.

MCCROCKLIN, S.; EDALATISHAMS, I. Revisiting popular speech recognition software for ESL speech. **Tesol Quarterly**, 54, n. 4, p. 1086-1097, Dec 2020

MROZ, A. Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. **Foreign Language Annals,** 51, 2018.

MOTOHASHI-SAIGO, M.; HARDISON, D. Acquisition of L2 Japanese Geminates: Training with Waveform Displays. **Language Learning and Technology**, 13, 06/01 2009.

MOUSSALLI, S.; CARDOSO, W. Intelligent personal assistants: can they understand and be understood by accented L2 learners? **Computer Assisted Language Learning**, 33, p. 1-26, 04/15 2019.

MUNRO, M. J.; DERWING, T. M. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. **Language Learning**, 49, p. 285-310, 1999.

MUNRO, M.; DERWING, T. The foundations of accent and intelligibility in pronunciation research. **Language Teaching**, 44, p. 316-327, 07/01 2011.

MUTLU, A.; EROZ, B. The Role of Computer-Assisted Language Learning (CALL) in Promoting Learner Autonomy. **Egitim Arastirmalari - Eurasian Journal of Educational Research**, 13, p. 107-122, 03/01 2013.

NAGLE, C. Motivation, Comprehensibility, and Accentedness in L2 Spanish: Investigating Motivation as a Time-Varying Predictor of Pronunciation Development. **The Modern Language Journal**, 102, n. 1, p. 199-217, 2018.

NERI, A.; CUCCHIARINI, C.; STRIK, H. The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch. **ReCALL.** European Association for Computer Assisted Language Learning, 20, p. 225-243, 2008.

NOBRE-OLIVEIRA, D. **The effect of perceptual training on the learning of english vowels by brazilian portuguese speakers**. 2007. 198 f. Tese (Doutorado) - Curso de Programa de Pós-Graduação em Letras/Inglês e Literatura Correspondente, Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, Florianópolis, 2007. Disponível em: https://repositorio.ufsc.br/bitstream/handle/123456789/90286/242580.pdf?sequence=1&isAllowed=y. Acesso em: 23 out. 2022.

O'SHAUGHNESSY, D. Invited paper: Automatic speech recognition: History, methods and challenges. **Pattern Recognition**, 41, p. 2965-2979, 10/01 2008.

PADMANABHAN, J.; PREMKUMAR, M. Machine Learning in Automatic Speech Recognition: A Survey. **IETE Technical Review**, 32, p. 1-12, 02/23 2015.

PEKRUN, R. A social-cognitive, control-value theory of achievement emotions. Motivational psychology of human development: Developing motivation and motivating development.: Elsevier Science**:** 143-163 p. 2000

PETERSON, M. Learner interaction management in an avatar and chat-based virtual world. **Computer Assisted Language Learning**, 19(1), 79-103, 2006.

RATO, A. **Cross-language perception and production of English vowels by Portuguese learners: The effects of perceptual training** (Unpublished Doctoral dissertation). University of Minho, Minho. 2013

RATO, A.; RAUBER, A. S., The effects of perceptual training on the production of English vowel contrasts by Portuguese learners. **ICPhS**. 2015.

RAUBER, A.; ESCUDERO, P.; BION, R.; BAPTISTA, B. The interrelation between the perception and production of English vowels by native speakers of Brazilian Portuguese. **Interspeech.** p. 2913-2916, 2005.

REINDERS, H.; WATTANA, S. Learn English or die: The effects of digital games on interaction and willingness to communicate in a foreign language. **Digital Culture and Education**. 3. 4-28, 2011.

ROGERSON-REVELL, P. M. Computer-Assisted Pronunciation Training (CAPT): Current Issues and Future Directions. **RELC Journal**, 52, n. 1, p. 189-205, 2021.

SCHMIDT, R. The role of consciousness in second language learning. **Applied Linguistics**, 11, 129-158, 1990

SCHMIDT, R. Attention, awareness, and individual differences in language learning. In: CHAN, Wai Meng et al (ed.). **Proceedings of CLaSIC 2010**. Singapore: National University Of Singapore, p. 721-737, 2010.

SAITO, K.; DEWAELE, J.-M.; ABE, M.; IN'NAMI, Y. Motivation, emotion, learning experience, and second language comprehensibility development in classroom settings: a cross-sectional and longitudinal study. **Language Learning,** 68, n. 3, p. 709-743, 2018.

SHAMS, A. N. **The use of computerized pronunciation practice in the reduction of foreign language classroom anxiety**. Orientador: LEBLANC, L. 2006. 114 f. Ph.D. -, The Florida State University, Ann Arbor. Disponível em:

https://www.proquest.com/dissertations-theses/use-computerized-pronunciation-practice-reduction/docview/305331776/se-2?accountid=26642

SHARWOOD SMITH, M. A. Consciousness-raising and the second language learner. **Applied Linguistics**, 2, 159-168, 1981.

SHIBANO, T.; ZHANG, X.; LI, M.; CHO, H. *et al.* Speech technology for everyone: automatic speech recognition for non-native english. In: International Conference on Natural Language and Speech Processing 2021 (ICNLSP 2021). Trento. **Anais…** Trento: Association for Computational Linguistics. 2021.

SILVEIRA, R. **The influence of pronunciation instruction on the perception and the production of English word-final consonants**. 2004. Tese (Doutorado em Letras). – Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, Florianópolis, 2004.

SPARKS, R.; GANSCHOW, L. Is the foreign language classroom anxiety scale measuring anxiety or language skills? (vol 40, pg 260, 2007). **Foreign Language Annals**, 40, p. 556-557, 09/01 2007.

SPRING, R.; TABUCHI, R. The Role of ASR Training in EFL Pronunciation Improvement: An In-depth Look at the Impact of Treatment Length and Guided Practice on Specific Pronunciation Points. 23, p. 163-185, 08/14 2022.

SUZUKIDA, Y. The Contribution of Individual Differences to L2 Pronunciation Learning: Insights from Research and Pedagogical Implications. **RELC Journal**, 52, n. 1, p. 48-61, 2021.

TEJEDOR-GARCÍA, C.; ESCUDERO-MANCEBO, D.; CÁMARA-ARENAS, E.; GONZÁLEZ-FERRERAS, C. *et al.* Assessing pronunciation improvement in students of english using a controlled computer-assisted pronunciation tool. **IEEE Transactions on Learning Technologies**, 13, n. 2, p. 269-282, 2020.

THOMSON, R.; DERWING, T. The effectiveness of L2 pronunciation instruction: a narrative review. **Applied Linguistics**, 2014, p. 1-20, 2014

Thompson, C. G.;von Gillern, S. Video-game based instruction for vocabulary acquisition with English language learners: A Bayesian meta-analysis. **Educational Research Review***,* 30, 100332, 2020.

TUDOR, I. Teacher roles in the learner-centered classroom. **ELT Journal,** 47 (1), 23-24, 1993.

VASA, R. A.; PINE, D. S. Anxiety disorders. *In*: **Child and adolescent psychopathology: Theoretical and clinical implications.** New York, NY, US: Routledge/Taylor & Francis Group, 2006. p. 78-112.

WACHOWICZ, K. A.; SCOTT, B. Software that listens: it's not a question of whether, it's a question of how. **CALICO Journal**, 16, n. 3, p. 253-276, 1999.

WILANG, J.; SINGHASIRI, W. Out-of-class anxiety in a non-English speaking context and its effects on intelligibility and comprehensibility. **Issues in Educational Research**, 27, p. 620-638, 08/02 2017.

APPENDIXES

APPENDIX A - Measure of Pronunciation Anxiety

Please indicate the extent to which you agree/disagree with the statements below by writing a digit next to each statement.

6 – strongly agree (completely true about me)
5 – agree
4 – slightly agree
3 – rather disagree
2 – disagree
1 – strongly disagree (definitely not true about me)

0. Example: I get nervous every time I am asked to answer a question in the foreign language.	2
1. During oral tasks in the foreign language classroom, I tend to have difficulties with concentration.	
2. When I speak English during the lesson, my performance is usually at a lower level than when I try (rehearse) speaking at home.	
3. I can feel my heart pounding, have a dry mouth, or clammy hands (or have other symptoms of being stressed) when I am asked to respond in English at the whole class forum.	
4. I frequently volunteer to answer questions in English.	
5. I feel shy when I am asked to read aloud in English.	
6. I feel more comfortable during classes that involve less talking and more writing (e.g. grammar or lexical exercises).	
7. Usually I feel embarrassed when asked to repeat after the teacher.	
8. I avoid eye contact with the teacher when he/she is looking for a learner to answer his/her question in the foreign language.	
9. I find it more difficult to improve my pronunciation than grammar or vocabulary.	
10. I remember the pronunciation of new words easily.	
11. My pronunciation is at a lower level than that of my classmates.	

12. I believe that after a 2- or 3-year course of English with a native foreign language speaker, my accent could become target language nativelike.	
13. I am satisfied (happy) with my present level of English pronunciation.	
14. I have a talent to pick up the pronunciation of foreign languages.	
15. My pronunciation of English is far from that of native speakers.	
16. I look funny pronouncing the 'th' sound.	
17. I like singing and/or speaking to myself in English.	
18. I do (would) not mind pronouncing English sounds and/or words with my native language accent.	
19. I like imitating English actors/singers.	
20. I look natural speaking English.	
21. The comprehensibility of a speaker depends on his/her level of pronunciation.	
22. I (would) feel uneasy pronouncing English sounds and/or words as they should be pronounced.	
23. I do not like listening to myself reading in English aloud.	
24. I think I sound unnatural speaking English.	
25. I would rather my classmates did not hear me making pronunciation mistakes.	
26. Some words in English sound awkward and/or funny.	
27. I feel stressed when the teacher corrects my pronunciation mistakes at the class forum.	
28. I fear my classmates might find my pronunciation of English strange or funny.	
29. The pronunciation of English is difficult for speakers of my first language.	
30. Some sounds of English seem silly and/or strange.	
31. English sounds like music to me.	

32. The level of pronunciation affects the ability to understand spoken language.	
33. I am worried what others might think of me when they hear my English pronunciation.	
34. Usually it bothers me when I mispronounce a word in English during a lesson.	
35. A speaker that mispronounces many sounds can still be understood by his interlocutor quite easily.	
36. There are several aspects of English pronunciation that are difficult for speakers of my mother tongue.	
37. I get nervous and feel shy of the teacher when making a pronunciation mistake.	
38. I feel stressed knowing that other students are listening to me.	
39. I feel more embarrassed making a pronunciation mistake than any other type of mistake (grammatical or lexical).	
40. Mastering correct word stress of English is not particularly difficult for speakers of my native language.	
41. I (would) worry about what other non-native speakers of English could think of me hearing my pronunciation of English.	
42. I can feel my heart pounding, have a dry mouth, or clammy hands (or have other symptoms of being stressed) when I have to join a conversation In English with other non-native speakers of English.	
43. Talking to another non-native speaker of English, I would fear that he could consider my English pronunciation funny or awkward.	
44. Usually I am embarrassed when talking to other non-native speakers of English.	
45. When talking to a non-native speaker of English, I worry that I might not be understood.	
46. I (would) feel comfortable and relaxed talking in English to native speakers .	
47. I (would) worry about what my native speaking interlocutors could think of me on the basis of my pronunciation of English.	
48. When I have to join a conversation In English with native speakers of English, I can feel my heart pounding, have a dry mouth, or clammy hands (or have other symptoms of being stressed).	

49. When talking to a native speaker of English, I worry that I might not be understood.	
50. Talking to a native speaker of English, I would fear that he could consider my English pronunciation funny or awkward.	

Source: BARAN-LUCARZ, M. FL Pronunciation Anxiety and Motivation: Results of a Mixed-Method Study. Tradução por: SOUZA DOS SANTOS. F. W. *In*: EWA PIECHURSKA-KUCIEL, E. S.-C., MAGDALENA SZYSZKA (Ed.). **At the Crossroads: Challenges of Foreign Language Learning**. 1 ed. Second Language Learning and Teaching: Springer Cham, 2017. cap. 2, p. 107-133. (Second Language Learning and Teaching).

APPENDIX B - Foreign Language Enjoyment Scale

As you respond to the following items, think about your current Foreign language class
To what extent do you agree with the following statements?
Strongly disagree/ Disagree /Undecided/ Agree /Strongly agree

1. I don't get bored
2. I enjoy it
3. I'm a worthy member of the Foreign language class
4. In class, I feel proud of my accomplishments
5. It's a positive environment
6. It's cool to know a Foreign language
7. It's fun
8. The peers are nice
9. There is a good atmosphere
10. We laugh a lot

Source: DEWAELE, J.-M.; MACINTYRE, P. The two faces of Janus? Anxiety and Enjoyment in the Foreign Language Classroom. Tradução por: SOUZA DOS SANTOS. F. W. **Studies in Second Language Learning and Teaching**, 4, p. 237-274, 09/27 2014.

APPENDIX C - Measure of Pronunciation Anxiety

Por favor indique até que ponto você concorda/discorda com as declarações abaixo, marcando a alternativa de acordo ao lado de cada declaração.

6 - concordo fortemente (completamente verdadeiro a meu respeito)
5 - concordo
4 - concordo um pouco
3 - discordo um pouco
2 - discordo
1 - discordo fortemente (definitivamente não é verdade sobre mim)

	1	2	3	4	5	6
0. Exemplo: Fico nervoso/a toda vez que me pedem para responder uma pergunta no idioma estrangeiro.		*				
1. Durante as tarefas orais na sala de aula de língua estrangeira, tenho a tendência de ter dificuldades de concentração						
2. Quando falo inglês durante a aula, meu desempenho geralmente é de nível mais baixo do que quando tento falar (ensaiar) em casa.						
3. Eu consigo sentir meu coração batendo, a minha boca seca, ou as mãos úmidas (ou outros sintomas de estresse) quando me pedem para responder em inglês durante as aulas.						
4. Frequentemente me ofereço como voluntário para responder perguntas em inglês.						
5. Sinto vergonha quando me pedem para ler em voz alta em inglês.						

						
6. Eu me sinto mais confortável durante as aulas que envolvem menos conversa e mais escrita (por exemplo, exercícios gramaticais ou lexicais).						
7. Normalmente me sinto envergonhado/a quando me pedem para repetir depois do professor.						
8. Evito contato visual com o professor quando ele está procurando um/a aluno/a para responder sua pergunta na língua estrangeira.						
9. Acho mais difícil melhorar minha pronúncia do que a gramática ou o vocabulário.						
10. Lembro-me facilmente da pronúncia de novas palavras.						
11. Minha pronúncia está em um nível mais baixo do que a de meus colegas da turma.						
12. Acredito que após um curso de 2 ou 3 anos de inglês com um **falante nativo** da língua estrangeira, minha pronúncia pode se tornar nativa						
13. Estou satisfeito (feliz) com meu nível atual de pronúncia do inglês.						
14. Tenho um talento para aprender a pronúncia de línguas estrangeiras.						
15. Minha pronúncia do inglês está longe de ser a de **falantes nativos**.						
16. Eu pareço engraçado pronunciando o som "th".						
17. Eu gosto de cantar e/ou falar sozinho em inglês.						

						
18. Não me importo (ou não me importaria) de pronunciar sons e/ou palavras em inglês com meu sotaque da **língua nativa**.						
19. Eu gosto de imitar atores/cantores ingleses.						
20. Pareço natural falando em Inglês.						
21. A compreensibilidade de um falante depende de seu nível de pronúncia.						
22. Sinto-me (me sentiria) desconfortável ao pronunciar sons e/ou palavras em inglês do jeito que deveriam ser pronunciadas.						
23. Eu não gosto de me ouvir lendo em inglês em voz alta.						
24. Acho que falo inglês de forma não natural.						
25. Preferiria que meus colegas de turma não me ouvissem cometendo erros de pronúncia.						
26. Algumas palavras em inglês soam estranhas e/ou engraçadas.						
27. Eu me sinto estressado/a quando o professor corrige meus erros de pronúncia na sala de aula						
28. Temo que meus colegas de turma possam achar minha pronúncia do inglês estranha ou engraçada.						
29. A pronúncia do inglês é difícil para os falantes da minha **língua nativa**.						
30. Alguns sons de inglês parecem bobos e/ou estranhos.						
31. O inglês soa como música para mim.						

						
32. O nível de pronúncia de um falante afeta a capacidade de compreender a língua falada.						
33. Estou preocupado/a com o que os outros possam pensar de mim quando ouvem minha pronúncia em inglês.						
34. Normalmente me incomoda quando pronuncio errado uma palavra em inglês durante uma aula.						
35. Um falante que pronuncia errado muitos sons ainda pode ser compreendido pelo seu interlocutor com bastante facilidade.						
36. Há vários aspectos da pronúncia em inglês que são difíceis para os falantes da minha **língua nativa**.						
37. Eu fico nervoso e me sinto tímido em relação ao professor quando cometo um erro de pronúncia.						
38. Sinto-me estressado sabendo que outros alunos estão me ouvindo.						
39. Sinto-me mais constrangido ao cometer um erro de pronúncia do que qualquer outro tipo de erro (gramatical ou lexical).						
40. Dominar a tonicidade correta das palavras em inglês não é particularmente difícil para os falantes da minha **língua nativa**.						
41. Eu (me preocuparia) me preocupo com o que outros **falantes não-nativos** de inglês podem pensar de mim ouvindo minha pronúncia de inglês.						
42. Eu consigo sentir meu coração batendo, a minha boca seca, ou as mãos úmidas (ou outros sintomas de estresse) quando tenho que participar de uma conversa em inglês com outros **falantes não-nativos** de inglês.						

						
43. Quando falo com outro f**alante não-nativo** de inglês, eu temo que ele pudesse considerar minha pronúncia em inglês engraçada ou embaraçosa.						
44. Normalmente fico envergonhado/a quando falo com outros **falantes não-nativos** de inglês.						
45. Quando falo com um **falante não-nativo** de inglês, temo que eu possa não ser compreendido/a.						
46. Eu (me sentiria) me sinto confortável e relaxado ao falar em inglês com **falantes nativos**.						
47. Eu (me preocuparia) me preocupo com o que meus **falantes nativos** de inglês podem pensar de mim com base na minha pronúncia do inglês.						
48. Quando tenho que participar de uma conversa em inglês com **falantes nativos** de inglês, eu consigo sentir meu coração batendo, a minha boca seca, ou as mãos úmidas (ou outros sintomas de estresse).						
49. Quando falo com um **falante nativo** de inglês, me preocupo que eu possa não ser compreendido/a.						
50. Quando falo com um **falante nativo** de inglês, eu temo que ele/a pudesse considerar minha pronúncia em inglês engraçada ou estranha.						

Source: BARAN-LUCARZ, M. FL Pronunciation Anxiety and Motivation: Results of a Mixed-Method Study. Tradução por: SOUZA DOS SANTOS. F. W. *In*: EWA PIECHURSKA-KUCIEL, E. S.-C., MAGDALENA SZYSZKA (Ed.). **At the Crossroads: Challenges of Foreign Language Learning**. 1 ed. Second Language Learning and Teaching: Springer Cham, 2017. cap. 2, p. 107-133. (Second Language Learning and Teaching).

APPENDIX D - Foreign Language Enjoyment Scale

Ao responder aos seguintes itens, pense em sua aula atual de língua estrangeira
Até que ponto você concorda com as seguintes declarações?
1- Discordo fortemente
2- Discordo
3- Indeciso/a
4-Concordo
5-Concordo fortemente

	Discordo fortemente				*Concordo fortemente*
	1	**2**	**3**	**4**	**5**
1. Eu não me aborreço					
2. Eu gosto					
3. Sou um membro digno da turma de língua estrangeira					
4. Nas aulas, sinto orgulho de minhas conquistas					
5. É um ambiente positivo					
6. É legal conhecer uma língua estrangeira					

	Discordo fortemente				*Concordo fortemente*
7. É divertido					
8. Os colegas são simpáticos					
9. Há uma boa atmosfera					
10. Nós rimos muito					

Source: DEWAELE, J.-M.; MACINTYRE, P. The two faces of Janus? Anxiety and Enjoyment in the Foreign Language Classroom. Tradução por: SOUZA DOS SANTOS. F. W. **Studies in Second Language Learning and Teaching**, 4, p. 237-274, 09/27 2014.

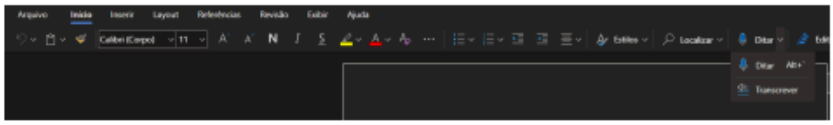
APPENDIX E - INSTRUCTION MATERIAL (WEEK 1)

Week 1 **Vowels /æ/ - /ɛ/**

Instruções

Esta semana, vamos praticar com os sons **/æ/** e **/ɛ/**, há muitas palavras nas quais o som errado destas vogais para certas palavras pode levar a significados completamente diferentes!

- Antes de tudo, vamos aprender a usar o recurso de **Ditar** do Microsoft Word!



- Em seu telefone ou dispositivo de computador, abra sua versão do Microsoft Word e procure o botão que diz **"Ditar"**. Em seguida, clique na roda dentada encontrada na janela que abrir e selecione o idioma **"Inglês (EUA)".** Clique em **salvar** depois de mudar o idioma.
- Em seguida, clique no microfone azul e fale a vontade!
- Se você ver o texto aparecendo em seu arquivo Word enquanto fala, isso significa que o recurso está funcionando corretamente! Sinta-se livre para usar o modo Ditar por diversão antes de começar a tarefa, para se familiarizar.

Antes de iniciar a tarefa, busque um lugar silencioso o suficiente para não afetar sua qualidade de áudio. Caso tenha dúvidas sobre o procedimento, entre em contato com o pesquisador via o e-mail disponibilizado no fim da página 2.

TASK 1

Say the words below using **Microsoft Word Dictate.** Try to get the tool to correctly transcribe each word **three times**. If Dictate isn't working properly, go to the next word after **ten tries**.

For each word below, **start a new line** (Enter key) in Microsoft Word in the same order as the task. The Dictate feature is **context sensitive**, and can change your transcriptions if you don't separate words by line!

VOWELS	/æ/
1	Pat
2	Pack
3	Cat
4	Hat
5	Rat
6	Hack

VOWELS	/ɛ/
7	Bet
8	Deck
9	Tech
10	Net
11	Neck
12	Heck

When you are done, send your file with the words to fwsdsantos@gmail.com. Do not edit the file in any way! It is important for the researcher to note the contexts where Word cannot understand you.

Have a good study session!

APPENDIX F - INSTRUCTION MATERIAL (WEEK 2)

Week 2 **Vowels /u/ - /ʊ/**

Instruções

Para esta semana, você estará praticando os sons **/u/** e **/ʊ/**. Estas vogais são particularmente difíceis em perceber as diferenças na hora da produção, mas podem levar palavras diferentes (e.g. **Pool /u/ vs. Pull /ʊ/**).

Algumas palavras nesta semana podem ser menos familiares:

- **Toot:** Verbo que indica apertar ou ativar uma buzina ou corneta. Normalmente usado como a seguinte expressão idiomática:

"Not to **toot my own horn**, but I did very good in the exam." (Não **quero me gabar**, mas eu fui muito bem na prova.)

- **Coop:** Uma gaiola ou galinheiro, normalmente para aves (especificamente galinhas).

"This is a chicken **coop**." (Isto aqui é um **galinheiro**.)

Antes de iniciar a tarefa, busque um lugar silencioso o suficiente para não afetar sua qualidade de áudio. Caso tenha dúvidas sobre o procedimento, entre em contato com o pesquisador via o e-mail disponibilizado no fim da página 2

TASK 2

Say the words below using **Microsoft Word Dictate.** Try to get the tool to correctly transcribe each word **three times**. If Dictate isn't working properly, go to the next word after **ten tries**.

For each word below, **start a new line** (Enter key) in Microsoft Word in the same order as the task. The Dictate feature is **context sensitive**, and can change your transcriptions if you don't separate words by line!

VOWELS	/u/
1	Nuke
2	Loop
3	Root
4	Toot
5	Poop
6	Coop

VOWELS	/ʊ/
7	Nook
8	Look
9	Hook
10	Put
11	Book
12	Cook

When you are done, send your file with the words to fwsdsantos@gmail.com. Do not edit the file in any way! It is important for the researcher to note the contexts where Word cannot understand you.

Have a good study session!

APPENDIX G - INSTRUCTION MATERIAL (WEEK 3)

Week 3 **Vowels /æ/ - /ɛ/**

Instruções

Esta semana retorna à produção e prática dos sons **/æ/** e **/ɛ/**. Desta vez, foram escolhidas palavras que terminam com o final **/p/** e **/d/:**

- **Red /ɛ/ vs. Rad/æ/**: "Wow! This is so _____". (Red/Rad - Isto é muito **vermelho/maneiro**)
- **Rep /ɛ/ vs. Rap /æ/:** "Having _____ is essential for our community." (Rep/Rap - Ter **a Rep/reputação** é essencial para a nossa comunidade)

Uma palavra nesta semana pode ser menos familiar:

- **Lad:** Um menino ou homem jovem.

"This **lad** is very well mannered." (**Este jovem** é muito bem educado)

Antes de iniciar a tarefa, busque um lugar silencioso o suficiente para não afetar sua qualidade de áudio. Caso tenha dúvidas sobre o procedimento, entre em contato com o pesquisador via o e-mail disponibilizado no fim da página 2

TASK 3

Say the words below using **Microsoft Word Dictate.** Try to get the tool to correctly transcribe each word **three times**. If Dictate isn't working properly, go to the next word after **ten tries**.

For each word below, **start a new line** (Enter key) in Microsoft Word in the same order as the task. The Dictate feature is **context sensitive**, and can change your transcriptions if you don't separate words by line!

VOWELS	/æ/
1	Tap
2	Lap
3	Rad
4	Had
5	Mad
6	Lad

VOWELS	/ɛ/
7	Rep
8	Let
9	Bed
10	Said
11	Dead
12	Led

When you are done, send your file with the words to fwsdsantos@gmail.com. Do not edit the file in any way! It is important for the researcher to note the contexts where Word cannot understand you.

Have a good study session!

APPENDIX H - INSTRUCTION MATERIAL (WEEK 4)

Week 4 **Vowels /u/ - /ʊ/**

Instruções

Esta semana retorna à produção e prática dos sons **/u/** e **/ʊ/**. Desta vez, foram escolhidas palavras que terminam com o final **/d/ e /l/**:

- **Food(/u/) and Good(/ʊ/)**: "This **food** is really **good**! (Esta **comida** é muito **boa**!)
- **Fool /u/ vs. Full /ʊ/**: "You have to be a **fool** to be this **full** of yourself." (Você tem que ser um **tolo** para ser tão **cheio** de si mesmo)

Uma palavra nesta semana pode ser menos familiar:

- **Bull:** Um touro.

"I've never seen a **bull** in person." (Eu nunca vi um **touro** pessoalmente)

Antes de iniciar a tarefa, busque um lugar silencioso o suficiente para não afetar sua qualidade de áudio. Caso tenha dúvidas sobre o procedimento, entre em contato com o pesquisador via o e-mail disponibilizado no fim da página 2

TASK 4

Say the words below using **Microsoft Word Dictate.** Try to get the tool to correctly transcribe each word **three times**. If Dictate isn't working properly, go to the next word after **ten tries**.

For each word below, **start a new line** (Enter key) in Microsoft Word in the same order as the task. The Dictate feature is **context sensitive**, and can change your transcriptions if you don't separate words by line!

VOWELS	/u/
1	Food
2	Mood
3	Rude
4	Who
5	Cool
6	Fool

VOWELS	/ʊ/
7	Good
8	Wood
9	Could
10	Should
11	Pull
12	Bull

When you are done, send your file with the words to fwsdsantos@gmail.com. Do not edit the file in any way! It is important for the researcher to note the contexts where Word cannot understand you.

Have a good study session!