UNIVERSIDADE FEDERAL DE MINAS GERAIS

Faculdade de Letras

Programa de Pós-Graduação em Estudos Linguísticos

Guilherme de Oliveira Lucena

Effects of vocabulary size on the speech perception of L2 phonological categories

Belo Horizonte

2024

Guilherme de Oliveira Lucena

Effects of vocabulary size on the speech perception of L2 phonological categories

Dissertação apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais como requisito parcial à obtenção do título de Mestre em Estudos Linguísticos.

Área de concentração: Linguística Teórica e Descritiva.

Linha de pesquisa: Processamento de Linguagem

Orientador: Prof. Dr. Rui Rothe-Neves

Coorientadora: Profa. Dra. Mara Passos Guimarães

Belo Horizonte

2024

L935e

Lucena, Guilherme de Oliveira.

Effects of vocabulary size on the speech perception of L2 phonological categories [manuscrito] / Guilherme de Oliveira Lucena. – 2024.

1 recurso online (41 f. : il., tabs., grafs. (algumas color.)) : pdf.

Orientador: Rui Rothe-Neves.

Coorientadora: Mara Passos Guimarães.

Área de concentração: Linguística Teórica e Descritiva.

Linha de Pesquisa: Processamento de Linguagem.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Faculdade de Letras.

Bibliografia: f. 38-41.

Exigências do sistema: Adobe Acrobat Reader.

1. Aquisição da segunda linguagem – Teses. 2. Bilinguismo – Teses. 3. Percepção da fala – Teses. I. Rothe-Neves, Rui, 1970-. II. Guimarães, Mara Passos. III. Universidade Federal de Minas Gerais. Faculdade de Letras. IV. Título.

CDD : 401.9

Ficha catalográfica elaborada pelo Bibliotecário Israel José da Silva – CRB/6-2128
Biblioteca Professor Rubens Costa Romanelli - FALE/UFMG

UNIVERSIDADE FEDERAL DE MINAS GERAIS

FOLHA DE APROVAÇÃO

Effects of vocabulary size on the speech perception of L2 phonological categories

GUILHERME DE OLIVEIRA LUCENA

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de PósGraduação em ESTUDOS LINGUÍSTICOS, como requisito para obtenção do grau de Mestre em ESTUDOS LINGUÍSTICOS, área de concentração LINGUÍSTICA TEÓRICA E DESCRITIVA, linha de pesquisa Processamento da Linguagem.

Aprovada em 16 de setembro de 2024, pela banca constituída pelos membros:

Prof(a). Rui Rothe-Neves - Orientador

UFMG

Prof(a). Mara Passos Guimarães - Coorientadora

UFMG

Prof(a). Ricardo Augusto de Souza

UFMG

Prof(a). Daniel Márcio Rodrigues Silva

UEMG

Belo Horizonte, 16 de setembro de 2024.

Documento assinado eletronicamente por **Mara Passos Guimaraes**, **Professora do Magistério Superior**, em 01/10/2024, às 14:05, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Daniel Márcio Rodrigues Silva**, **Usuário Externo**, em 01/10/2024, às 17:15, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Ricardo Augusto de Souza**, **Professor do Magistério Superior**, em 19/11/2024, às 09:42, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3493043** e o código CRC **14CFBDB5**.

À minha família e, especialmente, à minha mãe, Teresinha, que não me deixou desistir.

AGRADECIMENTOS

A Cristo, por não ter me esquecido mesmo nos vários momentos em que dele me esqueci.

Agradeço à minha mãe, por todo o amor e por não ter me deixado desistir. Ao meu pai, por todo o amor e por ser minha maior referência do que é ser um homem bom e justo. Aos dois, por todo o apoio emocional e financeiro ao longo dos anos.

Agradeço ao meu irmão, Gabriel, por ter sido um amigo sincero e leal, o melhor deles, durante todos os anos da minha vida. Agradeço também por ter aberto os caminhos acadêmicos na nossa família, sendo um matemático genial. Você é muito curioso!

Agradeço à minha noiva, Joyce, por topar construir comigo um amor sincero, tranquilo e ao mesmo tempo intenso. Minha vida é muito melhor com você, eu te amo.

Agradeço ao Alan, meu primo (e ao mesmo tempo também meu irmão), por todo o companheirismo, por cada cerveja e por ser exemplo do que é ser um pesquisador ético e incansável.

Agradeço ao Rui, pela orientação objetiva e respeitosa, e à Mara, pela orientação e por não me deixar entrar em pânico. Aos dois, por compartilharem um pouco de sua inteligência comigo. Vocês são fantásticos!

Agradeço ao Peter, meu gato companheiro de todas as horas, apesar de imaginar que ele terá um pouco de dificuldade em ler estes agradecimentos.

Agradeço a professora Valdeni Reis, por ter me dado a oportunidade de ingressar na pesquisa científica, e por todo o carinho durante os anos em que me orientou na iniciação científica.

Agradeço aos professores Dr. Daniel Silva e Dr. Ricardo Souza, por terem aceito participar da minha banca.

Agradeço a cada amigo que me ajudou a chegar até aqui. Ao Gustavo, por me mostrar que há alegria em ensinar uma segunda língua a alguém, e ao Nicolas, pelo companheirismo durante todos esses anos. Ao Victor Hugo, Eduardo, Pablo e tantos outros que se dispuseram a dar um pitaco nas minhas ideias para a dissertação. À tia Nicinha, por ter me recebido em sua casa e à Dani, ao Ed e a tantos outros amigos maravilhosos que me acolheram em Belo Horizonte. São muitos, então não ousarei tentar nominar todos.

Agradeço à Cultura Inglesa de Belo Horizonte por todas as oportunidades profissionais.

Agradeço muito especialmente aos voluntários, que disponibilizaram seu precioso tempo para participar deste estudo.

Agradeço ao Santos Futebol Clube, por me providenciar sempre um motivo para passar raiva fora do ambiente profissional e acadêmico. Nascer, viver e no Santos morrer! Agradeço também à cervejaria Heineken, por me fornecer meios de suportar ser torcedor do Santos.

Finalmente, agradeço à Universidade Federal de Minas Gerais, por ter me dado a oportunidade que é o ensino público, gratuito e de qualidade.

Então o “th” tem som de “d”?

-Algum aluno, em algum lugar

ABSTRACT

For decades, bilingualism has been a trending topic in speech perception studies. Among others, Flege (1995), Best (1995) and Van Lessen and Escudero (2015) have dedicated a great deal of work to explain whether and how one can acquire second-language phonological categories. Along with those authors, Best and Tyler (2007) believe second language vocabulary size contributes to effective L2 phonological category acquisition. However, few studies have investigated this matter regarding foreign language learning. The present study investigates how vocabulary size influences foreign language speech perception. Results of an AX categorial task and a vocabulary levels test (NATION, 1990) indicate that participants with more extensive vocabulary in the second language contrast L1 and L2 sounds more effectively than those with a smaller vocabulary.

Keywords: speech perception; vocabulary size; phonological acquisition; first language; second language.

RESUMO

Por décadas, o bilinguismo tem sido um assunto em alta nos estudos de percepção da fala. Entre outros, Flege (1995), Best (1995) e Van Lessen e Escudero (2015) dedicaram muito trabalho para explicar se e como alguém pode adquirir categorias fonológicas de segunda língua. Junto com esses autores, Best e Tyler (2007) acreditam que o tamanho do vocabulário na segunda língua contribui para a aquisição eficaz de categorias fonológicas na L2. No entanto, poucos estudos investigaram esse assunto em relação ao aprendizado de línguas estrangeiras. O presente estudo investiga como o tamanho do vocabulário influencia a percepção da fala em línguas estrangeiras. Os resultados de uma tarefa de percepção do tipo AX e um teste de níveis de vocabulário (NATION, 1990) indicam que os participantes com vocabulário mais extenso na segunda língua contrastam os sons L1 e L2 de forma mais eficaz do que aqueles com um vocabulário menor.

Palavras-chave: percepção da fala; tamanho do vocabulário; aquisição fonológica; primeira língua; segunda língua.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
AX	or AX-task, a speech-perception task that requires discriminating A from X
CEFR	Common European Framework of Reference for Languages
CG	Category-Goodness
CVC	Consonant-vowel-consonant sequence
EFL	English as a Foreign Language
ESL	English as a second language
F1	First formant
F2	Second formant
FL	Foreign Language
GLM	Generalized linear model
GLMM	Generalized linear mixed model
L1	First Language
L2LP	Second Language Speech Perception model
L2	Second Language
NA	Non-Assimilable
PAM	Perceptual Assimilation Model
SC	Single-Category
SLM-r	Speech Learning Model Revised
SLM	Speech Learning Model
TCLE	in Portuguese, Free and Informed Consent Form
TC	Two-Category
UFMG	in Portuguese, Federal University of Minas Gerais
UU	Both Uncategorizable
VLT	Vocabulary Levels Test

SUMMARY

1. Introduction

Beginning with the work by Miller and Nicely (1955), Lieberman, Hoffman and Griffith (1957) and others, speech perception studies date from the middle of the twentieth century. Until the latter part of the century, the field focused mainly on native language speech perception and acquisition. However, since the 1990s, three models of second language (from now on, L2) speech perception have arisen and become prominent in psycholinguistic studies: the Speech Learning Model (from now on, SLM) (FLEGE, 1995), the Perceptual Assimilation Model (from now on, PAM) (BEST and TYLER, 2007) and, finally, the Second Language Speech Perception model (from now on, L2LP) (VAN LEUSSEN and ESCUDERO, 2015). Despite their many differences, the three models share two similarities: they focus on the acquisition of a language in an L2 learning context (as opposed to a foreign language learning context), and they somehow, with very different degrees of emphasis, uphold the idea that vocabulary enlargement may have a positive effect on L2 speech perception; Bundgaard-Nielsen, Best, & Tyler (2011) even propose a model to explain this.

Both Krashen (1985) and Gass and Selinker (2008) stress the differences between second and foreign language (from now on, FL) learning. Krashen defines language acquisition as the subconscious process a child goes through when naturally learning a language and language learning as a conscious process of 'knowing about language', while Gass and Selinker refer to foreign language as "the learning of a nonnative language in the environment of one's native language' and L2 acquisition as "the learning of a nonnative language in the environment in which that language is spoken." One of the main differences to be expected from the two settings is the amount of L2 auditory input received by the learner, as one geographically inserted in a context where the target language (the one to be learned) is widely spoken is expected to have much more contact with it.

Due to the notable differences in second and FL learning settings, it is crucial to test whether the three models' postulates and hypotheses are valid for individuals who have learned (or have been learning) an L2 in an FL context. Perozzo and Alves (2016) have already made pertinent suggestions on how to adapt PAM to this context. Also, Santos and Alves (2021) have suggested attention components should be considered in SLM, which is relevant since attention is a key aspect of in-class language learning. This dissertation reports a study that tested the relationship between vocabulary size and L2 speech perception among learners who have Brazilian Portuguese (from now on, BP) as their native language and have never lived outside of Brazil. After that, we discuss the data according to the SLM, PAM and L2LP. By doing so, we expect to collaborate with speech perception studies specifically in an FL context and to provide fellow researchers with useful empirical data.

2. Theoretical Construct

2.1 The Bilingual Mind

Until the late twentieth century, it was believed that bilinguals have an independent language system for each language they speak. The idea that bilinguals have two separate and independent switchable lexicons that work similarly to the ones of monolinguals (MACNAMARA; KUSHNIR, 1971) was prominent then. Also prominent was the idea that a "true" bilingual speaks more than one language with the same mastery (BLOOMFIELD, 1933). Mackey (1962) even stood for the interference of one language in another, yet still committed to the assumption that the two belong to different representational systems inside the brain. Despite being long-established, such views on bilingualism became obsolete due to the work of, among others, Grosjean (1985), who argues that bilingual individuals do not have a separate lexicon for each of the languages they speak. According to Grosjean, they have a "unique and specific linguistic configuration" that considers all the languages they speak. Apart from Grosjean's remarks, some empirical studies attest to the influence of not only the first language (from now on, L1) on the L2 (SOUZA, 2017) but also of the L2 on the L1 (COOK, 2003).

Among the many instances in which aspects of the L1 influence the way one perceives and produces the L2, there is evidence (FLEGE, 1995; BEST and TYLER, 2007; ZIMMER; BITTENCOURT, 2014; SOUZA, 2017) that bilinguals tend to substitute approximate phonemes, which are at times exclusive to the L2, for the most similar phones which are licensed in the L1, as if they were allophones of those. Examples from Brazilian Portuguese-English bilinguals are /æ/ being pronounced as /ɛ/, /r/ being pronounced as /h/ and /x/, /ɪ/ being pronounced as /i/, /ð/ being pronounced as /d/ and /θ/ being pronounced as /f/, /s/ or /t/. Such substitutions happen not only when uttering an L2 but also when perceiving L2 speech. Ramires (2016) observed, for instance, how complex the acquisition of the discrimination between English /æ/ and /ɛ/ in both production and perception is for speakers of L1 PB and L2 English by conducting a production task (reading of minimal pairs in a carrier sentence) and both an identification and a discrimination task. Perozzo and Alves (2017) concluded that BP native speakers, independently of their proficiency in English, find it difficult to discern [t] with audible aspiration, which is very characteristic in English but not common in Portuguese. Also, Bernardino (2023) identified that a Portuguese-English bilingual struggles more to identify [ɪ] than its tense counterpart [i].

2.2 L2 Speech Perception in L2 context

Several approaches aim to explain how one deals with perceiving and processing speech. Among many differences and similarities, we can highlight the fact that some approaches state that linguistically relevant sounds are stored in the brain as an abstract phonetic category (FLEGE, 1995), while others (BEST, 1995) sustain that sounds are perceived according to one's perception of how they are articulated in the vocal tract.

Flege's Speech Learning Model (FLEGE, 1995) aims to explain age-related effects on the ability to produce L2 sounds, focusing on experienced rather than beginner L2 speakers. It is a mentalist approach to storing phonological categories in long-term memory. It claims speech perception becomes attuned to the consonant and vowel sounds of the L1, making it difficult for bilinguals to later pronounce L2 sounds natively. That difficulty is due to the possibility of an L2 phoneme being either assimilated to an L1 phoneme, for instance, a L1 BP speaker pronouncing /æ/ as /ɛ/ when speaking English, or to the L1 phonology "filtering out" features of the L2 sounds, as L1 BP speakers often do when they pronounce plosives without their characteristic aspiration in English. Also, the model caters to the possibility of motor-driven production inaccuracies, such as 'school' being pronounced as [ɪskuːl] by Brazilian Portuguese-English bilinguals.

According to the SLM, the elements which enable one to learn one's native language phonotactics remain functional throughout the lifespan and can be implemented in the learning of an L2 sound system. Also, Flege proposes that phonetic categories evolve to cater to the characteristics of foreign and native sounds, which are identified as realizations. Although native and L2 phonetic categories coexist in the same phonological space, bilinguals can differentiate them.

Flege hypothesizes that the relation between native and L2 sounds is perceptual; they work as positional allophones of each other. If a L2 sound differs from an L1 sound to at least some degree, a new category is formed, equal or different from the actual L2 one. The greater the dissimilarity, the higher the chance of forming a new category. For individuals to discern noncontrastive L2 sounds, the chance decreases according to how old they are when they learn the language. If they are not discerned, they belong to the same category.

Flege's SLM has been updated since 1995, and the current SLM is the Speech Learning Model Revised (from now on, SLM-r) (FLEGE and BOHN, 2021). SLM-r aims to explain how phonetic systems reorganize over the lifespan. According to the new version, novel phonological categories can be formed anytime during an individual's life. If it differs from the closest L1 category, a category based on phonetic input from both L1 and L2 is created. This formation depends on how phonetically dissimilar a sound is to the L1 sound closest to it, how much L2 sound input is received through meaningful conversations, and how well-defined the L1 category is when L2 learning starts (FLEGE and BOHN, 2021). Also, SLM-r defends the *Full access hypothesis*: a bilingual can learn to utter sounds shared by L1 and L2 with L2 features (e.g., a L1 BP speaker uttering /p/ with English-like aspiration). For the SLM-r, the L2 input one receives is key for successful phonological category formation.

Best (1995) proposes a perception model based on a direct realist (GIBSON, 1991) approach, which refers to the notion of human perception (in a whole, being it visual, tactile or related to any sense, not necessarily hearing) of things is not mediated by a neural component, but is direct. According to the direct realist approach, we perceive things as they are. Best's model, the Perceptual Assimilation Model (PAM), is direct realist in how it considers speech perception to be based not on abstract phonemic categories but on how we perceive a given sound to be articulated in the vocal tract. Although it is one of the most cited models in L2 speech perception studies, PAM has been criticized because of its assumptions of how speech perception works. Perozzo and Kickhöfel (2013) highlight the fact Gibson's direct-realist theory was conceived to explain sight (as opposed to hearing). While they criticize PAM's direct-realist philosophical support, they support perception as both mediated by abstract representation and continuously constructed through articulatory routines. Also, they suggest Albano's articulatory gesture (ALBANO, 2001) is adopted, as opposed to being linked only to motor procedures; it is also related to other linguistic levels of the language to be perceived. Ohala (1995) argues against the idea that speech sounds are categorized according to how one assumes them to be articulated. Among their main arguments are the fact that babies and animals can distinguish speech sounds and that humans can distinguish non-speech sounds, which are as complex as speech.

Regarding L2, PAM states that non-native sounds are assimilated in terms of how similar or distinct they are compared to native sounds, which are most similar to them in terms of articulation. Through that process, three are the possible outcomes for that new sound to be categorized:

- Assimilated to a native category: the L2 sound articulatory similarity to an existent L1 category leads it to be perceived as an instance of that L1 category

- Assimilated as an uncategorizable speech sound: perceived as speech but not as any instance of an existent native category.
- Not assimilated to speech: perceived as a nonspeech sound.

PAM predicts six scenarios during the acquisition of an L2 sound, each leading to a different degree of phonetic discrimination. Firstly, Two-Category assimilation (TC Type) occurs when two different L2 sounds are assimilated to two different L1 sounds, such as /æ/ and /a/ being perceived as /ɛ/ and /a/ by L1 BP English learners. In this scenario, discrimination is expected to be good. Then there is Category-Goodness Difference (CG Type) when two different L2 sounds are identified as realizations of a single native phoneme with different appropriateness. In this case, discrimination is expected to be mild to very good, depending on how appropriate the realization of the native sound each sound is considered. Next, when Single-Category Assimilation (SC Type) is the case, two non-native sounds are perceived as equally appropriate or inappropriate instances of a native sound, such as /æ/ and /ɛ/ being perceived as /ɛ/ by L1 BP English learners. In this case, discrimination is expected to be very weak.

Both Uncategorizable (UU Type) takes place when there is a pair of non-native sounds, and both are recognized as speech, but neither of them is linked to a specific native sound by the speaker. Depending on how similar or dissimilar these sounds are from each other and native categories, discrimination from poor to excellent can be expected. However, significant discrimination is expected in a pair where one of the non-native sounds is categorizable, and the other is not (Uncategorized versus Categorized, UC Type). Finally, if both non-native sounds in a pair are not Non Assimilable (NA Type) and not perceived as speech, then discrimination is expected to be very good. PAM's assumptions were tested in many opportunities, with Best et al. (1988) among them. In this study, native English speakers perceived many Zulu contrasts following PAM's predictions.

Apart from the SLM and PAM, there is also the Linguistic Perception Model (L2LP; VAN LEUSSEN and ESCUDERO, 2015), a connectionist computational model that embraces four levels of representation: acoustic, phonetic, phonemic, and lexical (see Figure 1, below).

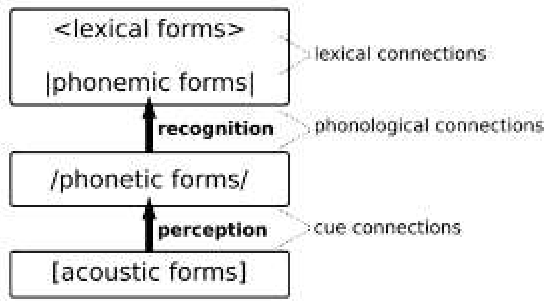

FIGURE 1 | The levels of representation and connection types in the L2LP model.

Figure 1: (VAN LEUSSEN and ESCUDERO, 2015, p.4)

According to the L2LP, the learner starts with an L1 perception grammar and the L2 grammar is attuned to it (L2 sounds are perceived as instances of L1 sounds). With time, experience and contact with a new language, the speaker forms a more accurate L2 perception grammar. This happens through the continuous act of choosing an optimal path. By an optimal path, Van Leussen and Escudero mean that a speaker is constantly comparing production instances of a given category in the L2 and, by trial and error, adjust that category to one adequately understood by the community of speakers of that language. This process is described as *meaning-driven* or *message-driven*, which means that L2 category attunement is improved along with how well what a speaker means is understood by others. It is also described as *error-driven*, meaning speakers adjust their L2 categories every time their production leads to misunderstandings. The process of continuously construing an optimal path is illustrated in Figure 2, below.

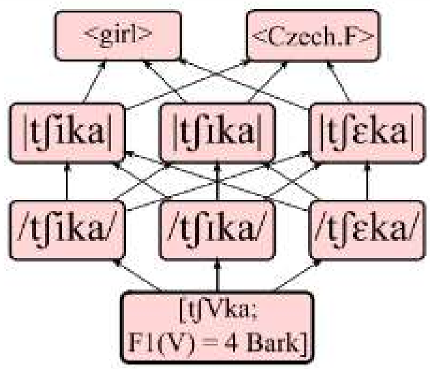

FIGURE 2 | Possible mappings for the input [tʃVka; F1(V) = 4 Bark], via phonetic and phonemic representations, to a lexical form. Each bottom-to-top path through the graph represents a possible pathway of perception and recognition.

Figure 2: (VAN LEUSSEN and ESCUDERO, 2015, p. 4)

Like PAM, the L2LP makes predictions for the perceptual development of sound contrasts. According to the L2LP, when two different L2 sounds are close to a single L1 category, learners face a *NEW* scenario (which is similar to PAM's Single Category Assimilation). When two L2 sounds are assimilated as two different L1 sounds, there is a *SIMILAR* scenario (similar to PAM's Two-Category Assimilation). Finally, the SUBSET scenario (similar to PAM's Uncategorized) happens when a single L2 sound is perceived as more than 1 L2 category. Unlike the other mentioned models, which focus on learners at the beginning of L2 acquisition (PAM) or experienced L2 speakers (SLM), the L2LP proposes a model to explain the whole L2 speech sound acquisition process.

The reason why these three models were selected and described in this study is that although they might present dissonant accounts and conclusions for the acquisition of L2 sounds in both speech perception and production, the three of them lead to a common outcome: Vocabulary acquisition might lead to an improvement in L2 speech perception.

2.3 L2 Vocabulary Size and Speech Perception

Vocabulary size is a relevant factor for successful speech perception in all three models above. Flege (2003) mentions that "the filtering of L2 speech input will not persist as learners acquire a dense network of L2 lexical items that need to be differentiated phonetically", which matches closely to L2LP's notion that the track to adjusting L2 categories (selecting an optimal path) is error-driven and message-driven. Regarding the L2LP, the link between vocabulary acquisition and L2 perception is described throughout the model. Since attunement to L2 sounds occurs due to adjustments made in response to misunderstandings and/or adequate production of words, opportunities for attunements to be successfully made increase along with L2 vocabulary acquisition since there are more opportunities for the link between phonetic and lexical categories to be made.

Regarding PAM, Best and Tyler (2007) believe vocabulary acquisition has a crucial role in L2 speech perception, but only in the early learning steps. The authors believe attunement to accurate perception and production of L2 sounds happens primarily with lexical and grammatical development. When these are already more established, L2 perception becomes stable as well. Later, together with Bundgaard-Nielsen and Tyler (2012), Best introduced the Vocabulary-tuning Model, which states that in the early stages of L2 acquisition, there is "a facilitating effect of adults' L2 vocabulary expansion on L2 perception and production" (p. 643), demonstrated by the correlation between the increase of the intelligibility of Japanese learners' Australian English and the growth of their English vocabulary during the early steps of L2 learning. The increase in intelligibility could be justified by the need to understand and acquire novel L2 phonemes to learn new vocabulary and properly speak a FL. Another study (Bundgaard-Nielsen, Best, & Tyler, 2011) also identified and tested such an effect: Japanese learners of English with larger vocabularies have more consistently identified L2 vowels than those with smaller vocabularies. The comparison between the performance of participants with high and low vocabulary in their study is illustrated below, in figure 3.

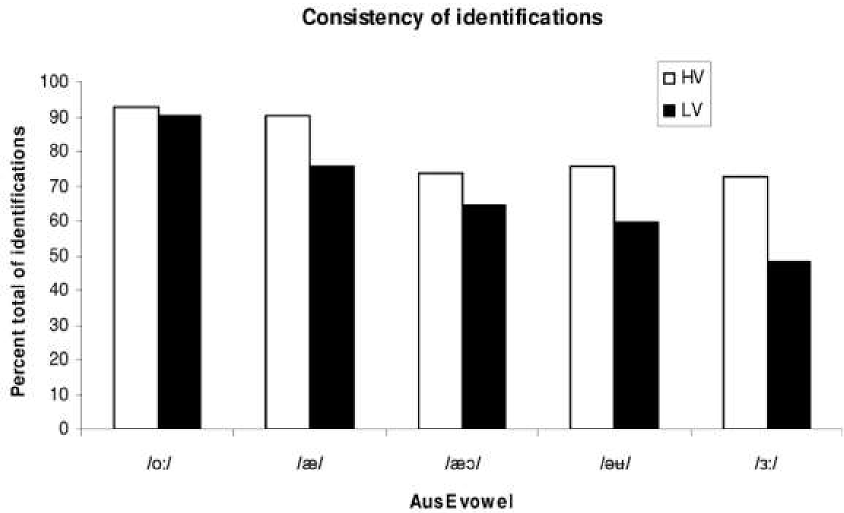

Figure 3. The percentages of the three most selected Japanese categories for the five uncategorized Western Sydney Australian English vowels for the high vocabulary group (HV) and the low vocabulary group (LV). (Bundgaard-Nielsen, Best, & Tyler, 2011, p. 62)

2.4 The Present Study Objectives and Hypotheses

Among the critical differences between the models mentioned earlier, SLM explains how L2 speech perception works for experienced learners. At the same time, PAM focuses on inexperienced/naïve learners, and L2LP proposes to explain the whole process of L2 phonetic/phonological acquisition. Also, while the SLM only mentions, without further development, vocabulary size as a factor, PAM (the vocabulary-tuning model, more precisely) and L2LP highlight vocabulary as a critical factor in L2 perception development. As for their similarities, all models focus on learners inserted in an L2, and learning an L2 as a FL is not the target of any of them.

Considering those facts, the main aim of this study is to investigate whether and how vocabulary size influences L2 speech perception development, specifically when learners are in an FL learning context. We expect that when discriminating between pairs of sounds that compare phonological categories from the L2, learners with more expanded vocabulary will perform better than those with lesser vocabulary.

A study was conducted to check this hypothesis. The methods will be described in the following section.

3. Materials and Methodology

3.1 Participants

The participants were 22 L1 BP speakers who had studied or had contact with English at some point in their lives, being that period at language schools, regular schools, private lessons, or self-study. All the participants were between 20 and 48 years-old at the time of the experiment and all of them had already completed high-school. 14 of them were women and 8 of them were men. Considering the study focused on learning English as a foreign language, none of the participants had lived in an English-speaking country. Since Bundgaard-Nielsen et al. (2011) suggest that vocabulary growth influences FL speech perception in the early levels of learning (the vocab model), the initial aim of the present study was to analyze only individuals currently or recently in the initial steps of learning English as an L2 (CEFR A1-A2 levels). However, this endeavor proved unfruitful, as many efforts were made to assemble enough participants who fulfilled these specific characteristics, and none were successful.

Considering that the objective of this study is to investigate vocabulary size influence on speech perception in an English as an FL (from now on referred to as EFL) learning context, eventual data from participants who have lived in countries where English is widely spoken was not considered. The study was previously analyzed and authorized by the UFMG Ethics Committee (CAAE 73202623.0.0000.5149), and all participants signed informed consent forms (TCLE, "Termo de Consentimento Livre e Esclarecido").

3.3 The AX Categorial Task

The AX task was selected for this study to reduce the load on the auditory memory (GERRITS & SCHOUTEN, 2004), since in this task the time one has to remember a sound is shorter, and each sound is compared to one other sound only. It is not an issue that the task might be highly dominated by a phoneme-based criterion since that is the very aspect we wished to investigate. The AX task consists of the participant listening to a pair of two sounds, either different or similar. Then, the participant points out if the sounds are equal (pressing a green key on a keyboard) or different (pressing a red key on a keyboard). AX is the least complex of discrimination tasks and is not very cognition-demanding

(COLANTONI et al., 2015), which is positive when it comes to low proficiency participants (PEROZZO, 2017).

Figure 4: A representation of the AX auditory discrimination task

The target phonemes were /i/, /ɪ/, /æ/ and /ɛ/, all presented in a single-syllable (CVC) word. These were chosen due to the abundance of minimal pairs they construe. Also, to the fact that /ɪ/ is never part of the stressed syllable in BP, so L1 BP learners of English as an L2 (ESL) tend to mispronounce this phoneme (SOUZA, 2017). The task was administered in PRAAT (BOERSMA, 2001) on the same laptop for all participants and all of them used the same headphones. The task contains 12 stimuli paired up in 24 different pair settings. Each possible pair was played in a random order 8 times, totalling 288 repetitions and approximately 20 minutes. The AX task demands that A = X in 50% of occurrences for control.

The list of pairs used as stimuli in the experiment was the following: "bat, bat", "sat, sat", "sack, sack", "bat, bet", "sat, set", "sack, seck", "bet, bet", "set, set", "seck, seck", "bet, beat", "set, seat", "seck, seek", "bet, bit", "set, sit", "seck, seek", "beat, beat", "seat, seat", "seek, seek", "bit, beat", "sit, seat", "sick, seek", "bit, bit", "sit, sit" and "sick, sick". Unfortunately, due to a technical problem, all answers for pairs containing the stimuli "beat" were discarded.

3.4 Stimuli

The words present in the AX task were recorded using the Google Translate text-to-speech tool. To make sure the synthesized words were adequate, vowel formants were analyzed in PRAAT to ensure sounds correspond to the average contrasts of such sounds (as in PETERSON & BARNEY, 1952), figures 5 and 6 demonstrate their distribution according to F1 and F2. F1 formants are similar to those in Peterson and Barney's study, while F2 followed a similar pattern but differed considerably in frequency. To ensure words would compare to real-life stimuli, a phonologically trained, experienced English speaker (an English teacher who has BP as his L1 and is a certified English C2 user), was asked to identify the words on the stimuli and write them. All words were correctly written except for "seck", which is very infrequent in English but effectively completes the design of the experiment.

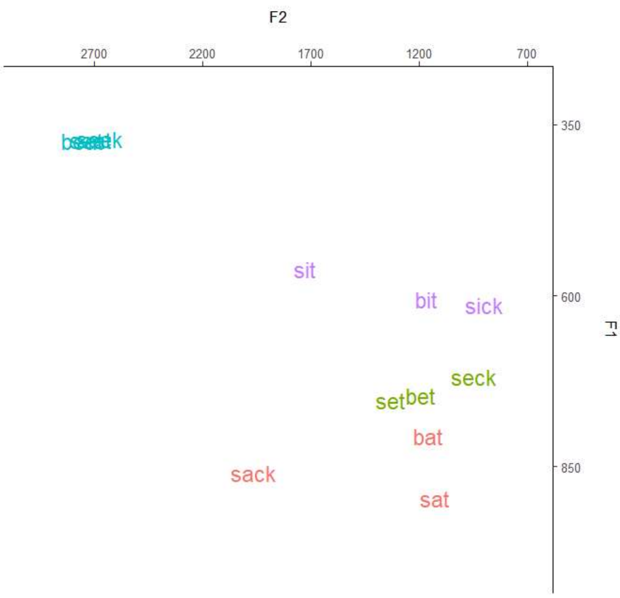

Figure 5: F1 and F2 of the stimuli used in the experiment. The three words which appear together in blue are "beat", "seat" and "seek".

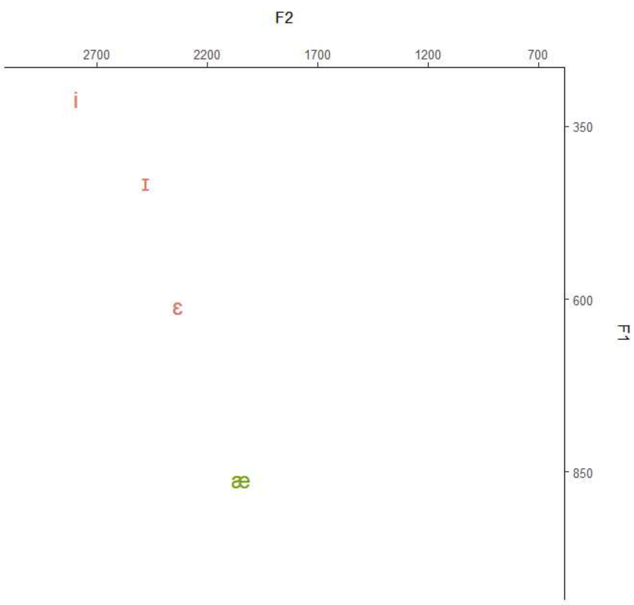

Figure 6: Average production of English Vowels (PETERSON & BARNEY, 1952)

The use frequency of the words in the experiment was registered. Words with higher frequency were preferred. Since frequency plays a crucial role in language processing and acquisition (ELLIS, 2002), beginner learners are more likely to know them. However, it was impossible to select only words with high use frequency. Word frequency was measured based on its number of occurrences in the English Web TenTen Corpus, and the word frequency rank appeared next to the word. Words are represented in the table below according to their position among the 5 thousand most frequent words in English, the ones with an x next to them are less frequent than the first 5 thousand words and were not available in the corpus.

word	**frequency**	**word**	**frequency**
bit	438	beat	2085
bet	3522	bat	x
sit	2077	seat	1981
set	198	sat	2354
sick	3027	seek	2072
seck	x	sack	x

Table 1: Frequency of the words used as stimuli, according to the English TenTen Corpus

3.5 The Vocabulary Levels Test

The Vocabulary Levels Test has been widely used by researchers as a diagnostic tool of L2 vocabulary breadth (NATION & BEGLAR, 2007); it consists of a five-level test, each level containing six items. Each item is composed of a column of three descriptions and a row of 6 words, with the objective of the test-taker to match the descriptions aligned vertically to their corresponding words, which are aligned horizontally among three distractors. The VLT version used in this experiment (WEBB, SASAO & BALLANCE, 2017) estimates vocabulary size by corresponding levels 1, 2, 3, 4 and 5 to the 1, 2, 3, 4 and 10 thousand most frequently used word families in English. Each Item of the VLT corresponds to 100 words of the correspondent frequency band (KREMMEL & SCHMITT, 2017). Therefore, we can assess vocabulary size through it.

The original VLT does not regard time as one of its parameters. However, LAUFER & NATION (2001) conducted an experiment in which a vocabulary test based on the VLT was administered while time latency was being registered. Their results highlighted a correlation between vocabulary breadth and meaning and word matching. So, in order to stimulate automaticity, a maximum time of 20 seconds was implemented for each item of the VLT administered to the participants.

	bar	conversation	neighbor	rain	rubbish	shirt
person who lives nearby						
things that are thrown away						
type of clothing						

Figure 7: An item from the Vocabulary Levels Test (WEBB, SASAO & BALLANCE, 2017)

A thorough description of the VLT can be found in SILVA (2016), which was consulted during the writing of this section. The VLT was opted for instead of other tests, such as the Vocabulary Size Test (NATION, 2007), because it served the research objective while making the study feasible since other options would have made it too long. The VLT was administered in Psytoolkit (STOET, 2010, 2017), one of its items can be seen in figure 7.

Participants were then divided in two groups according to their resulting score in the vocabulary test: group A (lower vocabulary) and group B (higher vocabulary). Group A consisted of 10 participants and group B of 10 participants. Participants in group A scored between bands 1 and 3 in the VLT, while participants in group B scored band 5. No participants scored band 4, surprisingly.

4. Data Analysis

4.1 Results

The table below displays the proportion of correct answers among all of the items presented by group, divided by category (comparison between vowels), language (whether the pair contains an L2-exclusive sound), and pair (whether the stimuli presented were the same or different from each other).

Group	Category	Language	Pair	Hit	Total	% correct
A	æ-æ	L2	Equal	328	336	0.97619
	ɛ-æ	L2	Dif.	437	672	0.650298
	ɛ-ɛ	L1	Equal	651	672	0.96875
	ɛ-i	L1	Dif.	216	224	0.964286
	ɛ-ɪ	L2	Dif.	441	448	0.984375
	i-i	L1	Equal	313	336	0.931548
	ɪ-i	L2	Dif.	412	448	0.919643
	ɪ-ɪ	L2	Equal	553	560	0.9875
B	æ-æ	L2	Equal	235	240	0.979167
	ɛ-æ	L2	Dif.	392	480	0.816667
	ɛ-ɛ	L1	Equal	471	480	0.98125
	ɛ-i	L1	Dif.	156	160	0.975
	ɛ-ɪ	L2	Dif.	313	320	0.978125
	i-i	L1	Equal	235	240	0.979167
	ɪ-i	L2	Dif.	295	320	0.921875
	ɪ-ɪ	L2	Equal	396	400	0.99

Table 2: Accuracy in the AX task according to group and vowel comparison

For statistical analysis, correct answers were defined as “equal” for pairs with the same vowels and “different” for pairs with different vowels. Below, Figure 8 shows that the first hypothesis is likely true: participants with higher vocabulary (group B) have more correct answers on L2-exclusive phonological categories compared to those with lesser vocabulary (group A). The same applies to pairs which contain L1 phonological categories.

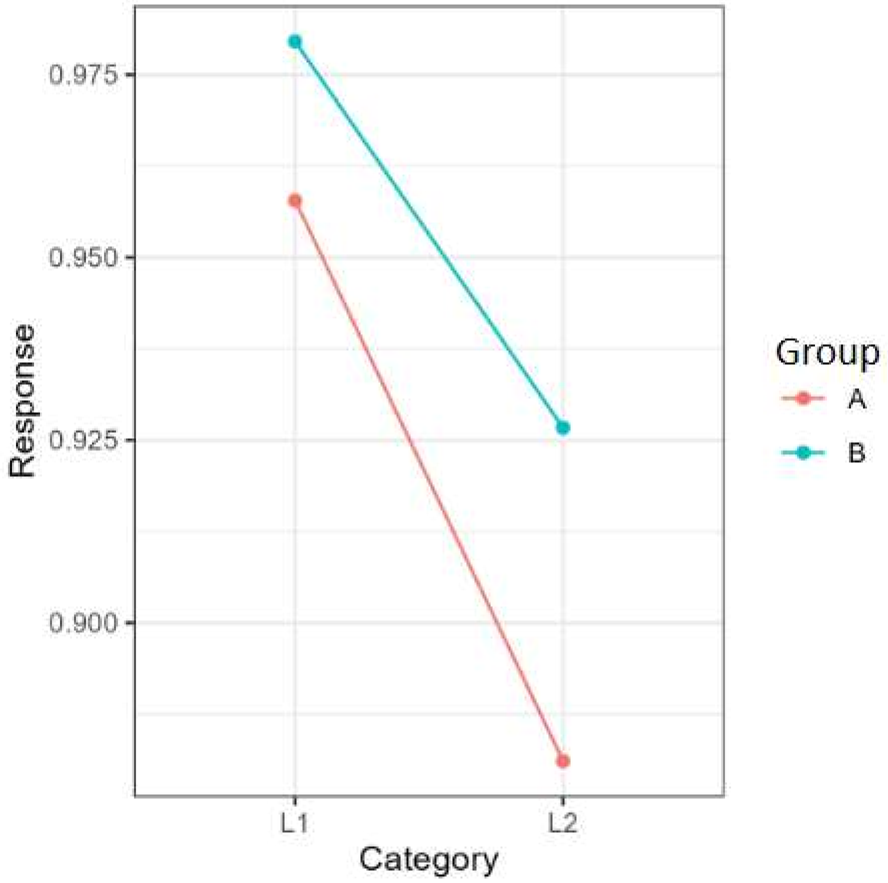

Figure 8 – Proportion of correct responses based on whether the compared vowels exist only in L2 or L1 and the participant's vocabulary size (Groups A and B).

Summarizing values in Table 1, the accuracy rate was 0.9577 for group A and 0.9795 for group B when categories were also in Portuguese (L1). Among pairs containing an L2-exclusive category, the accuracy rate was 0.8810 for group A and 0.9267 for group B. With these results, two analysis strategies were implemented, logistic regression and mixed models.

4.2. Logistic Regression

Logistic regression (generalized linear model) is a statistical technique that assesses the extent to which the occurrence of a binary event can be predicted by other variables. When the predictive variables are categorical, it is equivalent to multiple tests of proportion difference.

To verify if the accuracy difference is larger in L2 than in L1 depending on vocabulary size, the models below considered "response" (correct = 1 versus incorrect = 0) as the dependent binary variable and "group" (A x B) and "language" (L1 x L2) as predictors. The χ^2 test was used to determine whether a model with more variables is better compared to one with fewer. The tests below indicate that the best model includes both predictor variables.

```
Analysis of Deviance Table

Model 1: score_AX ~ 1
Model 2: score_AX ~ Group
Model 3: score_AX ~ Group + Only_L2
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      6335     3459.4
2      6334     3427.8  1   31.588 1.906e-08 ***
3      6333     3327.1  1  100.668 < 2.2e-16 ***
```

Table 3: Analysis of deviance on R.

The best model, described below, confirms that participants in group B achieve more correct answers than group A ($z = 5.504$, $p < .001$). If the compared pair includes an L2-exclusive category, the proportion of correct answers significantly decreases ($z = -8.927$, $p < .001$).

```
summary(axModel.3)

Call:
glm(formula = score_AX ~ Group + Only_L2, family = binomial(),
    data = ax)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7426   0.2865   0.3863   0.5054   0.5054

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.1728     0.1254  25.297  < 2e-16 ***
GroupB        0.5647     0.1026   5.504 3.72e-08 ***
Only_L2L2    -1.1795     0.1321  -8.927  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 3459.4  on 6335  degrees of freedom
Residual deviance: 3327.1  on 6333  degrees of freedom
AIC: 3333.1

Number of Fisher Scoring iterations: 6
```

Table 4: Analysis of accuracy between groups, on R, according to whether.

The model correctly predicts the accuracy proportions obtained empirically (previously presented in Figure 1), as shown in the table below.

		Observed	Predicted
Group A	L1	0.9578	0.9597
	L2	0.8810	0.8809
Group B	L1	0.9795	0.9767
	L2	0.9267	0.9281

Table 5: accuracy proportions obtained in the AX task.

The model's effects are presented as an ANOVA table, showing only the main effects, or graphically in the figure below.

```
> anova(axModel.3, test="Chisq")
Analysis of Deviance Table

Model: binomial, link: logit

Response: score_AX

Terms added sequentially (first to last)
        Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                     6335     3459.4
Group    1   31.588      6334     3427.8 1.906e-08 ***
Only_L2  1  100.668      6333     3327.1 < 2.2e-16 ***
```

Table 6: The model’s effects in a ANOVA analysis.

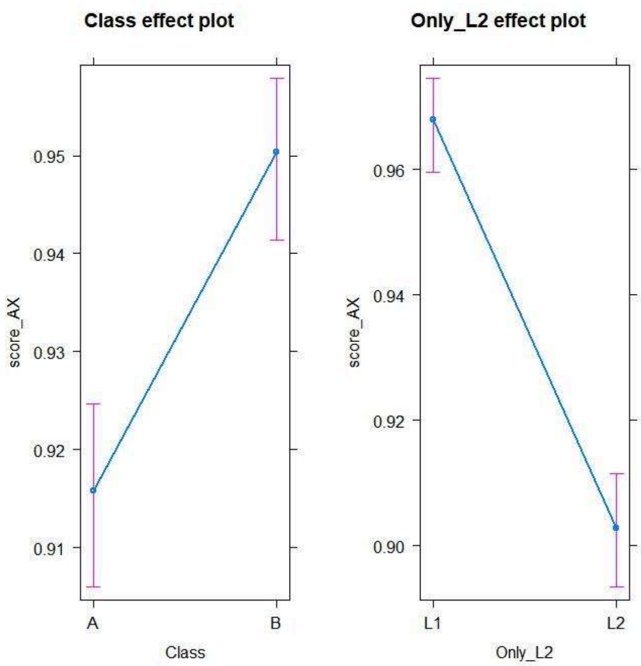

Figure 9: Accuracy rate according to group and whether an L2-exclusive sound was presented.

The results below indicate that the odds for a participant in Group B to answer the discrimination task correctly is 1.75 than for another in Group A.

```
exp(axModel.3$coefficients)
(Intercept)     GroupB   Only_L2L2
  23.873201   1.758973    0.307428
```

Table 7: Odds ratio for accuracy in the task according to group.

4.3. Mixed Models

Mixed models (generalized linear mixed models) are named so due to the fact they allow the modelling of both fixed (the one manipulated during research) and random (not fixed by the researcher) factors. Among these, participants and the items presented in the research are typically considered (Baayen, 2008; et al., 2012, for example).

Field's (2012) incremental models strategy was used, where a simpler model is nested within the next one. This way, the models can be compared using the χ^2 test for the importance of a given factor included in the model to be determined.

```
> anova(ax.glmm1, ax.glmm2, ax.glmm3, ax.glmm4, test="Chisq")
Data: ax
Models:
ax.glmm1: score_AX ~ 1 + (1 | subject)
ax.glmm2: score_AX ~ Group + (1 | subject)
ax.glmm3: score_AX ~ Group + Only_L2 + (1 | subject)
ax.glmm4: score_AX ~ Group + Only_L2 + (1 | subject) + (1 | item)
         npar    AIC    BIC  logLik deviance    Chisq Df Pr(>Chisq)
ax.glmm1    2 3264.9 3278.4 -1630.4   3260.9
ax.glmm2    3 3262.3 3282.6 -1628.2   3256.3   4.5511  1     0.0329 *
ax.glmm3    4 3160.2 3187.3 -1576.1   3152.2 104.0608  1     <2e-16 ***
ax.glmm4    5 2620.5 2654.2 -1305.2   2610.5 541.7811  1     <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 8: Model comparison according to Field's incremental models strategy (2012).

In the above-mentioned models, the list starts with an "empty" model, containing only the intercept (an estimate of the general average probability of correct answers) without random factors. The last model is the most complete one, in which we have a version with random effects of the model already explored through the logistic regression technique. In addition to the fixed factors "Group" and "Only_L2," the model includes the random factors "participant" and "item" (bring an item a pair of stimuli do be compared), The reasoning behind this choice is to suppose that the vocabulary size and the specific item might work differently for each participant when performing the discrimination task. The model summary is presented below

---

```
> summary(ax.glmm4)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [
glmerMod]
 Family: binomial  ( logit )
Formula: score_AX ~ Group + Only_L2 + (1 | subject) + (1 | item)
   Data: ax

    AIC      BIC   logLik deviance df.resid
  2620.5   2654.2  -1305.2   2610.5     6331

Scaled residuals:
    Min      1Q   Median      3Q      Max
-24.1228   0.0855   0.1472   0.2449   1.4787
Random effects:
 Groups  Name        Variance Std.Dev.
 item    (Intercept) 1.745    1.3210
 subject (Intercept) 0.911    0.9545
Number of obs: 6336, groups:  item, 27; subject, 24
Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.5400     0.5789   6.115 9.64e-10 ***
GroupB        0.8995     0.4188   2.148   0.0317 *
Only_L2L2    -0.7348     0.6017  -1.221   0.2220
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 9: Model summary.

The results confirm participants with a more extensive vocabulary are better at the discrimination task when the compared pairs include phonological categories, which also exist in Portuguese ($z = 2.15$, $p < .05$). When the compared phonological categories include L2-exclusive ones, the proportion of correct answers decreases, but the difference is not significant ($z = -0.735$, $p = 0.22$). This means that part of the individual variance was being accounted for as variation due to the "Only_L2" factor, if one of the items in the compared pair exists only in L2. Therefore, the final model should exclude the factor that is not significant.

```
> summary(ax.glmm5)

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [
glmerMod]
 Family: binomial  ( logit )
Formula: score_AX ~ Group + (1 | subject) + (1 | item)
   Data: ax

    AIC      BIC   logLik deviance df.resid
  2619.9   2646.9  -1305.9   2611.9     6332

Scaled residuals:
```

```
    Min       1Q   Median       3Q      Max
-24.5808   0.0855   0.1456   0.2459   1.4782

Random effects:
 Groups  Name        Variance Std.Dev.
 item    (Intercept) 1.8743   1.3691
 subject (Intercept) 0.9115   0.9547
Number of obs: 6336, groups:  item, 27; subject, 24

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.0024     0.3775   7.953 1.82e-15 ***
GroupB        0.8995     0.4189   2.147   0.0318 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 10: The final model should exclude the "Only L2" factor.

The table below demonstrates that the odds for a participant in group B to answer the discrimination task correctly is 2.46 times higher than for another in group A.

```
> fixef(ax.glmm9)
(Intercept)      GroupB
  3.0024481   0.8994559
> exp(fixef(ax.glmm9))
(Intercept)      GroupB
  20.134769    2.458265
```

Table 11: Odds ratio for accuracy in the task according to group.

4.4. Comparing models

The table below demonstrates that both models result in a very similar conclusion, comparable to the observed data.

		Observed	GLM	GLMM
Group A	L1	0.9578	0.9597	0.9587
	L2	0.8810	0.8809	0.8808
Group B	L1	0.9795	0.9767	0.9765
	L2	0.9267	0.9281	0.9285

Table 4: Comparison between the GLM and GLMM models.

The difference between the two models is that L2 turns out to be non-significant in the GLMM, probably because the variance mixed with the fixed factors in the GLM was appropriately estimated by the model as a random effect. The individual variation estimated by the random effects can be seen in the graph below.

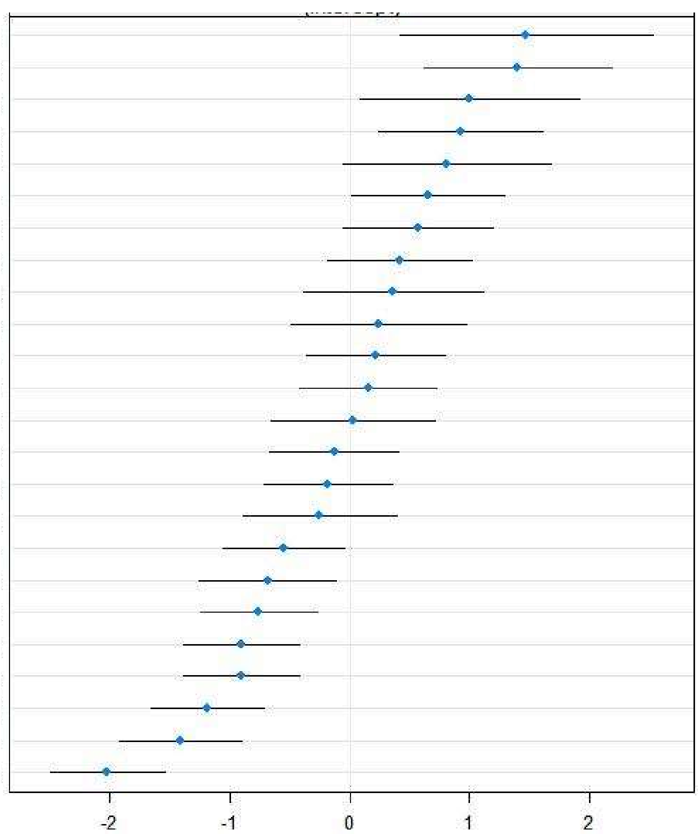

Figure 10: Participant accuracy probability according to random effects

5. Discussion

The results of this study can be explained through the perspectives of the different models described previously. Flege's SLM can effectively justify some of our observations, especially because the pair /ɛ/-/æ/ contrast was notably less accurately identified than the others. The fact that these are perceptually close enables [æ] to be perceived as a possible allophone /ɛ/ for L1 BP speakers.

PAM can also explain many of the study's results. /ɛ/-/æ/ seems to be what the model calls a single-category scenario, in which the two sounds are perceived as the same L1 category. In our case, both [ɛ] and [æ] are perceived as the same, as if they belong to the same category, in a considerable amount of instances. PAM correctly predicts that discrimination is poorer than in other scenarios. [ɛ] and [i] were already expected to be predicted well as they are both L1 sounds. [ɛ] and [ɪ] are what PAM calls a two-category scenario, in which discrimination is expected to be good (and it was indeed). [ɪ] and [i] could be described as a category-goodness scenario since BP speakers could identify both as /i/ (although there is also the possibility that [ɪ] is perceived as [e]). However, they would not be considered equally appropriate since they are stressed differently. PAM predicts mild to very good discrimination, which supports the results.

Unfortunately, due to a technical error in the experiment, the number of pairs for each contrast was not evenly distributed. Therefore, running statistical analysis to test the statistical significance of each pair individually was not possible. The general results, however, are consistent with PAM's predictions: pairs from which both phones are in the L1 were more accurately discriminated than the ones containing L2 sounds. That is predictable according to PAM since the L1-L2 contrasts include single-category and category-goodness scenarios, which hinder discrimination. Results are also coherent with the L2LP (VAN LEUSSEN and ESCUDERO, 2015), since it makes similar predictions, being the *NEW* scenario similar to a single-category scenario.

Apart from satisfactorily predicting L1-L2 contrast discrimination, the L2LP can also explain vocabulary size's identified positive effect on correct contrast discrimination. According to L2LP, perception is message-driven, and this notion is reinforced by our results, which demonstrate that speakers with higher vocabulary are significantly better at discriminating sounds. Interestingly and unexpectedly, the advantage is valid for both L1-L2 and L1-L1 pairs, as we hypothesized this effect would be significant for contrasts containing L2 sounds only. Results showed that group B (Higher Vocabulary) achieved a higher score in

L2-exclusive pairs in the AX task when compared to group A (Lower Vocabulary). However, an analysis through mixed models enabled us to observe that the l2-exclusive factor did not achieve statistical significance, leading us to the fact that this difference is due to random effects. Therefore, the better performance presented by the higher-vocabulary group on both the VLT and the AX categorial task is explained through the difference in vocabulary size only.

An initial objective of this study was to test vocabulary size's positive effect on FL speech perception in initial steps of FL learning, as suggested by the Vocab model (BUNDGAARD-NIELSEN et al., 2011). However, that was not possible since practical issues prevented us from having participants take the experiment twice, before and after a significant interval. That would enable us to analyze results due to vocabulary expansion, which the vocab model goes through. Since it was not possible, we can only analyze our results through the lens of vocabulary size.

Most importantly, our findings show learners inserted in an FL learning context attain and improve the ability to correctly contrast L2-exclusive sounds, and that those with a higher vocabulary are better than those with a lower vocabulary when stimuli are compared to an L1 sound or another L2 sound. However, the field would benefit from further research which is successful at carrying out a task with more individuals with better-selected profiles, that is, being able to statistically compare students from different ages and proficiency levels, ideally two times before and after a significant period, so participants would be able to learn a significant amount of vocabulary and vocabulary expansion could be assessed. Such data could shed light on speech perception development and help lead to either adapting a present model to an FL learning context.

6. References

ALBANO, Eleonora Cavalcante. *O gesto e suas bordas: esboço de fonologia acústico-articulatória do português brasileiro.* Campinas, SP: Mercado de Letras; Associação de Leitura do Brasil; São Paulo, SP: FAPESP, 2001. ISBN 9788585537811.

BERNARDINO, Daniela da Silveira. Identificação das vogais [i-ɪ] do inglês por falantes bilíngues português/inglês e trilíngues (português/inglês/espanhol). Dissertação (Mestrado em Letras) - Instituto de Letras , Universidade Federal do Rio Grande do Sul, Porto Alegre, 2023.

BEST, C. A direct realist view of cross-language speech perception. In: STRANGE, W. (Ed.). *Speech perception and linguistic experience: Issues in cross-language research*. Timonium: York Press, 1995.

BEST, C.; TYLER, M. *Nonnative and second-language speech perception: commonalities and complementarities*. 2007. In: Language experience in second language speech learning: In honor of James Emil Flege. DOI: 10.1075/lllt.17.07bes.

BLOOMFIELD, L. *Language*. New York: Holt, Rinehart & Winston, 1933.

BUNDGAARD-NIELSEN, R. L.; BEST, C. T.; TYLER, M. D. Vocabulary size matters: the assimilation of second-language Australian English vowels to first-language Japanese categories. *Applied Psycholinguistics*, v. 32, p. 51–67, 2011.

BUNDGAARD-NIELSEN, R. L.; BEST, C. T.; KROSS, C.; TYLER, M. D. Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics*, v. 33, 2012. DOI: 10.1017/S0142716411000518.

COLANTONI, L.; STEELE, J.; ESCUDERO, P. *Second language speech: Theory and practice*. Cambridge: Cambridge University Press, 2015.

COOK, V. (Org.). *Effects of the second language on the first*. [S. l.]: Multilingual Matters, 2003.

ELLIS, N. C. Frequency effects in language processing. *Studies in Second Language Acquisition*, v. 24, n. 02, 2002. DOI: 10.1017/s0272263102002024.

FLEGE, J. E. Second language speech learning: theory, findings, and problems. In: STRANGE, W. (Ed.). *Speech perception and linguistic experience: Issues in cross-language research*. Timonium: York Press, 1995. p. 233-277.

FLEGE, J. E. Assessing constraints on second-language segmental production and perception. In: MEYER-KAHLEN, M.; KÄMMERER, K. (Ed.). *Phonetics and phonology in language comprehension and production: differences and similarities*. [S. l.]: [s. n.], 2003.

FLEGE, J.; BOHN, O.-S. The Revised Speech Learning Model (SLM-r). 2021. DOI: 10.1017/9781108886901.002.

GASS, S.; SELINKER, L. *Second language acquisition: an introductory course*. 3. ed. Nova Iorque: Routledge, 2008.

GERRITS, E.; SCHOUTEN, B. Categorical perception depends on the discrimination task. *Perception & Psychophysics*, v. 66, n. 3, p. 363-376, 2004.

GROSJEAN, F. The bilingual as a competent but specific speaker-hearer. *Journal of Multilingual & Multicultural Development*, v. 6, n. 6, p. 467-477, 1985.

KRASHEN, S. *The input hypothesis: issues and implications*. London: Longman, 1985.

KREMMEL, B.; SCHMITT, N. Vocabulary Levels Test. *The TESOL Encyclopedia of English Language Teaching*, p. 1–7, 3 jul. 2017.

LIBERMAN, A. M.; HARRIS, K. S.; HOFFMAN, H. S.; GRIFFITH, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, v. 54, n. 5, p. 358-368, 1957.

MACKEY, W. F. The description of bilingualism. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, v. 7, n. 2, p. 51-85, 1962.

MACNAMARA, J.; KUSHNIR, S. L. Linguistic independence of bilinguals: the input switch. *Journal of Verbal Learning and Verbal Behavior*, v. 10, n. 5, p. 480-487, 1971.

MILLER, G. A.; NICELY, P. E. An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America*, v. 27, n. 2, p. 338-352, 1955.

NATION, P. *Teaching and learning vocabulary*. Boston: Heinle & Heinle, 1990.

NATION, P.; BEGLAR, D. A vocabulary size test. *Language Teacher*, v. 31, p. 9–13, 2007.

OHALA, J. J. Speech perception is hearing sounds, not tongues. *The Journal of the Acoustical Society of America*, v. 99, n. 3, p. 1718–1725, 1996. DOI: 10.1121/1.414696.

PETERSON, G. E.; BARNEY, H. L. Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, v. 24, n. 2, p. 175–184, mar. 1952.

PEROZZO, Reiner; ALVES, Ubiratã. Identificação e discriminação de oclusivas não vozeadas sem soltura de ar audível do inglês: Dados de aprendizes brasileiros. *Fórum Linguístico*, v. 11, p. 22-37, 2014. DOI: 10.5007/1984-8412.2014v11n1p23.

PEROZZO, Reiner Vinicius; ALVES, Ubiratã Kickhöfel. Uma discussão acerca da aplicação do Perceptual Assimilation Model-L2 à percepção fônica de língua estrangeira: questões de pesquisa e desafios teóricos. *Domínios de lingu@gem*. Uberlândia, MG. Vol. 10, n. 2 (abr./jun. 2016), p.733-764, 2016.

PEROZZO, Reiner. Sobre as esferas cognitiva, acústico-articulatória e realista indireta da percepção fônica não nativa: para além do PAM-L2. 2017. Tese (Doutorado em Estudos Linguísticos) - Universidade Federal do Rio Grande do Sul, 2017.

PRAAT: doing phonetics by computer. Disponível em: http://praat.org. Acesso em: 11 mar. 2024.

RAMIRES, H. R. Percepção e produção de vogais coronais do Inglês por falantes nativos do Português Brasileiro: o papel da instrução explícita. Dissertação (Mestrado em Letras) - Universidade Católica de Pelotas, Pelotas, 2016.

SANTOS, Bruna; ALVES, Ubiratã. Desenvolvimento fonético-fonológico bi/multilíngue e atenção – Questões teóricas abordadas no Speech Learning Model (SLM/SLM-r) e desafios futuros para a pesquisa em segunda língua (L2). *Revista X*, v. 16, p. 1203, 2021. DOI: 10.5380/rvx.v16i5.81523.

SCHOUTEN, B.; GERRITS, E.; VAN HESSEN, A. The end of categorical perception as we know it. *Speech Communication*, v. 41, n. 1, p. 71–80, ago. 2003.

SILVA, J. S. Exploring a vocabulary test and a judgment task as diagnoses of early and late bilinguals' L2 proficiency. 1 abr. 2016. Disponível em: <repositorio.ufmg.br>.

SOUZA, H. K. D. Brazilian EFL learners' awareness about L2 phones: is *mall* pronounced as *mal*? *Trabalhos em Linguística Aplicada*, v. 56, n. 1, p. 235-258, 2017.

STOET, G. PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods*, v. 42, n. 4, p. 1096-1104, 2010.

STOET, G. PsyToolkit: a novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, v. 44, n. 1, p. 24-31, 2017.

VAN LEUSSEN, J.-W.; ESCUDERO, P. Learning to perceive and recognize a second language: the L2LP model revised. *Frontiers in Psychology*, v. 6, 2015. DOI: 10.3389/fpsyg.2015.01000.

WEBB, S., SASAO, Y., & BALLANCE, O. (2017). The updated Vocabulary Levels Test. ITL. *International Journal of Applied Linguistics*, 168(1), 33–69.

ZIMMER, M.; BITTENCOURT, H. R. Produção e percepção oral em L2: os processos de transferência do conhecimento grafo fônico-fonológico do português brasileiro (L1) para o inglês (L2) e o desempenho em *listening* (L2). *Cadernos de Estudos Lingüísticos*, v. 50, n. 1, p. 29-44, 2008.