http://dx.doi.org/10.31513/linguistica.2024.v20n3a65367

The production of length contrast by Brazilian L2 English Learners: a Maximum Entropy Model of cue weighting[1]

A produção do contraste de duração por aprendizes brasileiros de inglês como L2: um modelo de Máxima Entropia para ponderação de pistas

Flora Dilza Ngunga[2]

Maria Mendes Cantoni[3]

Wellington Araujo Mendes Jr.[4]

ABSTRACT

This study investigates how Brazilian learners of English as a second language produce the length contrast between long-tense vowels ([i, u]) and short-lax vowels ([ɪ, ʊ]), focusing on cue-weighting strategies. Given that Brazilian Portuguese lacks a phonemic vowel length contrast, a question that arises is how learners process this feature in English. The study evaluates the influence of acoustic cues such as vowel duration, F1, and F2 in the learners' production across different proficiency levels. Using Pillai scores to assess category separation and a Maximum Entropy (MaxEnt) model to estimate cue weights, our analysis reveals how these cues are integrated into learners' phonological systems. The results show that lower proficiency learners heavily rely on vowel duration, while higher proficiency learners incorporate spectral cues (F1, F2) more consistently, especially for front vowel contrasts ([i, ɪ]). For back vowels ([u, ʊ]), however, even advanced learners show limited cue integration, as indicated by significant overlap in their acoustic space. Pillai scores demonstrate greater category separation for advanced learners, particularly in front vowels, but inconsistencies remain in back vowel distinctions. The MaxEnt analysis highlights that duration receives higher weights for back vowel contrasts, while F1 and F2 play more significant roles for front vowel contrasts at higher proficiency levels. These findings suggest that while learners progressively adjust their cue-weighting strategies as proficiency levels increase, L1 transfer effects remain prominent, particularly in the reliance on vowel duration, thus contributing to our understanding of how L2 phonological contrasts are developed.

KEYWORDS: Cue Weighting. Length Contrast. Maximum Entropy Model. L2 Phonology Acquisition. Vowel.

RESUMO

Este estudo investiga como aprendizes brasileiros de inglês como segunda língua produzem o contraste de duração entre vogais longas ([i, u]) e vogais curtas frouxas ([ɪ, ʊ]), com foco em estratégias de ponderação de

[1] We are grateful to Adelino Pinheiro Silva for relevant comments on the statistical methods used in the paper and to two reviewers for helpful suggestions on the data modeling and discussion.

[2] Universidade Federal de Minas Gerais (UFMG), Faculdade de Letras, florangunga@gmail.com, https://orcid.org/0009-0004-1350-4897.

[3] Universidade Federal de Minas Gerais (UFMG), Faculdade de Letras, mmcantoni@gmail.com, https://orcid.org/0000-0001-9515-1802.

[4] Universidade Federal de Uberlândia (UFU), Instituto de Letras e Linguística, wellington.mendes@ufu.br, https://orcid.org/0000-0002-1459-4183.

Recebido em 15/09/2024 - Aceito em 23/01/2025

pistas. Dado que o Português Brasileiro não apresenta contraste fonêmico na duração das vogais, uma questão que se coloca é como os aprendizes adquirem essa propriedade em inglês. O estudo avalia a influência de pistas acústicas como duração da vogal, F1 e F2 na produção dos alunos em diferentes níveis de proficiência. Usando pontuações de Pillai para avaliar a separação de categorias e um modelo de Máxima Entropia (MaxEnt) para estimar os pesos das pistas, a análise revela como essas pistas são integradas nos sistemas fonológicos dos alunos. Os resultados mostram que os alunos com menor proficiência dependem fortemente da duração das vogais, enquanto os alunos com maior proficiência incorporam pistas espectrais (F1, F2) de forma mais consistente, especialmente para contrastes de vogais anteriores ([i, ɪ]). Para vogais posteriores ([u, ʊ]), no entanto, mesmo alunos avançados mostram integração limitada de pistas, conforme indicado pela sobreposição significativa em seu espaço acústico. As pontuações de Pillai demonstram maior separação de categorias para alunos avançados, especialmente nas vogais anteriores, mas permanecem inconsistências nas distinções de vogais posteriores. A análise MaxEnt destaca que a duração recebe pesos mais elevados para contrastes de vogais posteriores, enquanto F1 e F2 desempenham papéis mais significativos para contrastes de vogais anteriores em níveis de proficiência mais elevados. Estas descobertas sugerem que, embora os alunos ajustem progressivamente as suas estratégias de ponderação de pistas à medida que os níveis de proficiência aumentam, os efeitos de transferência de L1 permanecem proeminentes, particularmente na dependência da duração da vogal, contribuindo assim para a nossa compreensão de como os contrastes fonológicos de L2 são desenvolvidos.
PALAVRAS-CHAVE: Ponderação de pistas. Contraste de duração. Modelo de Máxima Entropia. Aquisição de Fonologia de L2. Vogal.

1. Introduction

This study aims to explore how Brazilian learners of English as a second language produce the contrast between long-tense vs. short-lax vowels (e.g. [i] vs. [ɪ], [u] vs. [ʊ]), in which multiple phonetic cues may be used at different extents. Focus will be given on the relative importance of the cues in learners' phonological grammar across proficiency levels.

The influence of first language (L1) patterns on second language (L2) phonological development is a well-documented phenomenon. This impact is seen through various dimensions such as phonotactics (Imaizumi *et al.*, 1999; Gut, 2010), prosody (Li; Post, 2014), speech production (Escudero *et al.*, 2009; Chládková; Escudero, 2012) and speech perception (Rauber, 2009; Rato, 2014). In general, researchers argue that influences arise due to the integrated phonetic space and perceptual mechanisms between the L1 and the L2, where both languages impact each other's phonetic and phonological properties (Nevins; Braun, 2009; Bergmann *et al.*, 2016; Flege; Bohn, 2021).

Besides being influenced by the learners' L1, the acquisition of L2 phonological contrasts is also prone to factors such as proficiency level (Lim; Seo, 2015; Oliveira, 2021), age of acquisition (Darcy; Kruger, 2012; Casillas, 2015), and the specific contexts in which the L2 sounds are encountered (Flege, 1995; Piske *et al.*, 2001). Moreover, the establishment of some L2 contrasts require the perception of multiple phonetic cues, which adds another layer of complexity to the acquisition process (McMurray *et al.*, 2008).

An interesting case of cross-linguistic influence is one in which learners are faced with L2 phonological contrasts that are not present in their L1. This can pose significant challenges, as the learners' perceptual system is often fine-tuned to the phonetic distinctions relevant to their native language, potentially leading to difficulty in discriminating and producing unfamiliar L2 sounds. For Brazilian Portuguese speakers learning English, one such challenge is mastering the vowel length-tenseness contrast (e.g. /i/ vs. /ɪ/, /u/ vs. /ʊ/), which is phonemic in English but allophonic in Portuguese. Indeed, previous studies reported that Brazilian Portuguese L2 learners produce a single category for the front high vowel in English (Gonçalves, Silveira, 2014). Spanish speakers learning English might also struggle with the /i/ and /ɪ/ vowel contrast as it is not present in Spanish (Escudero, 2005). These difficulties are compounded by the learners' tendency to map new L2 sounds onto the closest L1 equivalents, which can lead to difficulties in speech perception and mispronunciations.

Crystal (2011, p. 7) defines phonetic cues as "the acoustic properties of a sound which aid its identification in speech". Examples include fundamental frequency, voice onset time (VOT), amplitude, and harmonic structure. For instance, the perceptual distinction between English plosives /p/ and /b/ is influenced not only by VOT but also by factors such as periodic pulsing at the voice pitch frequency and noise in the frequency range of higher formants (Lisker & Abramson, 1964). Learners whose L1 does not utilize these additional cues might find it challenging to accurately produce and perceive these plosive contrasts, as it is the case of Korean learners of English (Kong; Yoon, 2013). Similarly, Mandarin learners of English often struggle with stress and intonation patterns because Mandarin uses tone to distinguish meaning, whereas these prosodic features interact differently in English (Wang, 2008). Schertz and Clare (2020) argue that this multiplicity of phonetic cues prompts learners to recalibrate their perceptual systems and acquire a new set of phonological rules. This is done by measuring which cues are relevant and determining their relative importance, a process known as cue weighting (Schertz and Clare, *op cit.*).

Research on second language development and cue weighting points to the significance of both crosslinguistic and developmental factors. For example, in differentiating the English vowels [i] and [ɪ], American speakers primarily use spectral cues, such as formant frequencies (Bohn, 1995; Flege *et al.*, 1997). In contrast, Mandarin learners (Wang, 2006) — as well as Japanese learners (Yazawa *et al.*, 2020) — of L2 English heavily rely on temporal cues alone. Interestingly, studies on Spanish learners of L2 English have yielded mixed results, with some findings indicating a reliance on duration alone (Casillas, 2015), while others suggest a combination of both spectral and temporal cues (Flege *et al.*, 1997).

In terms of developmental influence, research has demonstrated that learners adjust cue weights as their experience with the L2 grows. For instance, Kong and Yoon (2013) investigated how Korean learners of English perceive the voicing contrast in English alveolar stops (e.g., /t/ vs. /d/), with a focus on how the learners' proficiency levels influence their cue-weighting strategies. Results found that advanced listeners were better at inhibiting less relevant acoustic dimensions, focusing more effectively on VOT — the primary cue in English for distinguishing /t/ and /d/. Conversely, basic listeners tended to rely more on F0, which is less crucial for distinguishing English stops.

On a similar note, Morrison (2008) investigated how Spanish learners of L2 Canadian English perceive the /i/ and /ɪ/ vowel contrast. Initially, these learners rely heavily on duration cues rather than spectral cues to distinguish between the vowels. This reliance on duration is part of a developmental process where learners progress through stages: initially not distinguishing the vowels (Stage 0), using duration cues (Stage 1), combining duration and spectral cues (Stage 2), and finally adopting a native-like spectral-based perception (Stage 3). Morrison (2008) theorizes that learners initially use a multidimensional category-goodness-difference assimilation strategy, identifying English vowels based on their similarity to Spanish /i/ in terms of duration and spectral properties. Over time, exposure to English leads to an increased use of spectral cues and a reduced reliance on duration cues, helping learners bootstrap towards native-like vowel perception.

In Brazilian Portuguese, the alternation between [i] and [ɪ] and between [u] and [ʊ] occurs as part of an allophonic process, where the long-tense [i] or [u] vowel occur in stressed syllables and the short-lax [ɪ] or [ʊ] vowel arise in post-stressed syllables (e.g. [pah.ˈtʃi] *parti* "(I) left)" vs. [ˈpah.tʃɪ] *parte* "(He/she) breaks", and [ˈbãbʊ] *bambo* "loose" vs. [bãˈbu] *bambu* "bamboo"). Production studies on L1 Brazilian Portuguese showed that the distinction between [i] and [ɪ] is multiply cued by both quality (F1 and F2 space) (Escudero et. al, 2009; Chládková; Escudero, 2012) and duration (Vieira; Lopes, 2019).

Since Brazilian Portuguese lacks the length-tense phonological contrast, it is expected that early learners do not differentiate vowels using either phonetic cue. In a latter stage of the learning process, one or multiple cues may be recruited to express such contrast in L2. Given the allophonic nature of the tense and lax vowel alternation in Brazilian Portuguese, a question that arises is whether L1 allophonic cues be 'elevated' to contrastive status in the L2. In the study by Martinez *et al.* (2023), the authors investigated how Caribbean Spanish (CS) listeners perceived the oral-nasal vowel contrasts in BP. CS speakers have a unique situation where vowel nasalization can result from the elision of a nasal consonant (e.g., /sin/ -> [sĩ] 'without'). This creates a pseudo-contrastive context, where nasality can distinguish between words, even though it originates from an allophonic process. The study found that CS speakers could accurately perceive the BP contrasts /e/–/ẽ/ and /i/–/ĩ/. This suggests that the [nasal] allophonic feature in their L1, which functions in a pseudo-contrastive manner, can be elevated to fully contrastive status in the L2. However, the study did not consider the specific cues that aid in nasal discrimination, highlighting the need for further investigation into the potential L1-L2 transfer of phonetic cues with allophonic status.

Given the role of both crosslinguistic and learning-related factors on vowel contrasts, this study aims to model the interaction of multiple cues in the production of phonological categories. L2 cue-weighting studies have been mainly restricted to the domain of speech perception, rather than production. However, these complementary domains do not always mirror each other and how they are affected during L2 learning stages is still an open research issue. The Maximum Entropy Model (Hayes, Wilson, 2008) is used to assess the weights of each cue. The potential transfer of phonetic cues from L1 Brazilian Portuguese (BP) to L2 English during the production of tense [i, u] and lax [ɪ, ʊ] vowels is also considered. Additionally, it seeks to determine whether this contrast is influenced by different proficiency levels. To achieve this, analysis will be mainly grounded on the premises of the Maximum Entropy Model (Hayes, Wilson, 2008). Research on the development of L2 phonetic contrasts has been framed in a variety of models, such as Schwartz and Sprouse's (1996) Full Transfer, Full Access framework, Best and Tyler's (2007) Perceptual Assimilation Model-L2, Flege's (1995) Speech Learning Model, and Escudero's (2005) Second Language Perception Model. The choice for the Maximum Entropy Model was guided by the feasibility of accounting for and predicting multiple interactive weights associated to well formedness constraints of a grammar, as well as by the possibility of carrying out integrated analysis of production and perception, via weight comparison.

The paper is organized as follows. The next section explores the theoretical framework employed in this study: the Maximum Entropy Model. The third section describes the adopted methodology. The fourth section presents and discusses the results and is followed by the conclusions.

2. Maximum Entropy Model

The Maximum Entropy Model (MaxEnt) is a probabilistic framework originated in the field of Statistical Mechanics, with the primary aim of modeling the distribution of a system's states while making as few assumptions as possible, other than those dictated by the known constraints. (Halvorsen, 2013). A significant advantage of the model lies in its broad applicability across diverse domains, ranging from Ecology to Physics, Economics, and Linguistics.

The MaxEnt model is rooted in the principle of maximum entropy, which identifies the probability distribution that best represents our current state of knowledge as the one that maximizes entropy—a measure of uncertainty or information spread—while still satisfying the constraints given by the available data. This principle ensures that no additional assumptions are made beyond the provided data, rendering the model as unbiased as possible (Halvorsen, 2013). In the context of Linguistics, this translates to constructing probabilistic models of linguistic phenomena where the likelihood of different linguistic outputs is determined by a set of constraints, each weighted according to its importance (Goldwater; Johnson, 2003).

In a MaxEnt framework, constraints can range from simple phonological rules to complex syntactic structures. These constraints do not determine linguistic outcomes deterministically but rather assign probabilities to different possibilities, allowing for both categorical and gradient patterns. As put by Hayes and Wilson (2008, p. 383):

> Every constraint in the grammar has a *weight*, a nonnegative real number. The weights can be thought of as scaling the importance of one constraint relative to others. Constraints with higher weights have a more powerful effect in lowering the probability of forms that violate them (Hayes; Wilson, 2008, p. 383).

One of the significant contributions of the MaxEnt model to Linguistics, as explored by Hayes and Wilson (2008), is its ability to probabilistically interpret phonotactic well-formedness. The model operates by assigning probabilities to a wide range of universally possible phonological surface forms, where higher probabilities correspond to greater phonotactic well-formedness. This probabilistic approach allows for a nuanced understanding of how native speakers perceive the acceptability of different phonological forms.

In their application of the MaxEnt model to English syllable onsets, Hayes and Wilson (2008) effectively demonstrated its predictive power in capturing both categorical and gradient phonotactic patterns. For example, it successfully identified [pl] as a well-formed onset, which follows the sonority sequencing principle, where sonority increases from the beginning to the end of the cluster. Conversely, the model identified [rt] as an ill-formed onset, reflecting native speakers' rejection due to its violation of the sonority sequencing principle, where [t] (a stop) is less sonorous than [r] (a liquid), leading to a sonority slope that decreases rather than increases.

Beyond these clear-cut cases, Hayes and Wilson (2008) demonstrated that the MaxEnt model also predicted experimental judgments on syllable goodness, handling less common onsets like [zw], [sf], and [pw] with remarkable accuracy. While [zw] and [pw] are relatively rare in English, the model recognized their marginal acceptability, acknowledging that they do not violate key phonotactic principles despite their uncommonness. Similarly, the sequence [sf] was judged more acceptable than outright illegal clusters like [tl] or [dl], illustrating the model's sensitivity to the probabilistic nature of phonological knowledge, where frequency and statistical tendencies play a role in shaping native speakers' intuitions.

Another point addressed in Hayes and Wilson (2008) was the analysis of Shona vowel Harmony. Using the MaxEnt model, the authors demonstrated its capability to capture complex phonotactic patterns, particularly nonlocal dependencies. In Shona, vowels such as [e] and [o] can only appear in non-initial syllables if the preceding vowel is harmonic — [e] following another [e] or [o], and [o] only following [o]. Application of the MaxEnt model allowed the identification of constraints that penalize disharmonic sequences — like [e] following [i] — assigning higher probabilities to well-formed harmonic sequences such as *cher*[e]*nga* ('scratch') and *fov*[e]*dza* ('dent') over disharmonic ones. Through iterative learning, the model adjusted the weights of these constraints to reflect the natural phonotactic rules of Shona, accurately predicting the well-formedness of vowel sequences. This analysis highlights the model's flexibility in handling both local and nonlocal phonotactic constraints, making it a powerful tool for understanding the gradient acceptability of phonological forms in languages.

The MaxEnt model's ability to handle both clear-cut and gradient phonotactic patterns, as illustrated by its application to English syllable onsets and Shona vowel harmony, underscores its flexibility in capturing the nuances of phonological representations. Although it does not originate in Optimality Theory (OT), MaxEnt builds on a well-established tradition of constraint-based approaches (e.g., Zuraw, 2000; Goldwater; Johnson, 2003). However, as put by Hayes and Wilson (2008), while OT employs a strict ranking of constraints, the MaxEnt model introduces a more nuanced mechanism by assigning weights to each constraint, allowing for a more refined calculation of well-formedness based on the cumulative effect of constraint violations. In this framework, grammar is viewed as a set of weighted constraints whose summed violations predict well-formedness, allowing both categorical and gradient outcomes (Albright; Hayes, 2011).

Reflecting contemporary trends in phonological theory, Alderete and Finley (2023) emphasize the importance of probabilistic models, including MaxEnt, for handling variable and gradient phenomena. Such models suggest that phonological knowledge involves probability distributions, not simply binary acceptability. This convergence spans both formal and functional approaches, including generative frameworks, exemplar-based theories, and connectionist accounts, all of which stress probability as an essential element. MaxEnt's ability to integrate linguistic theory with computational rigor exemplifies the notion of "probabilistic phonology" shaping current research.

A key point of contention is whether the MaxEnt model should be viewed purely as a statistical tool or as a theoretical model of grammar. We argue that MaxEnt is inherently a theoretical-probabilistic model, as it both encodes hypotheses about the nature of grammar and represents them through weighted constraints. The use of probability to address linguistic variation does not render MaxEnt merely a statistical approach. Rather, the crucial step lies in identifying which constraints matter, a process guided by linguistic theory rather than by data alone. As outlined in this section, Hayes and Wilson (2008) demonstrate this theoretical grounding by proposing linguistically motivated constraints (e.g., segmental co-occurrence restrictions) to model phonotactic knowledge. While the final constraint weights are learned algorithmically from data, the original choice and formulation of these constraints reflect prior theoretical beliefs about phonological relevance. Goldwater and Johnson (2003) likewise apply the MaxEnt model to phonological data involving variation, such as Finnish genitive plurals, demonstrating that the model can successfully learn distributions of variable outputs. By leveraging linguistically motivated constraints (e.g., *LAPSE, *H.H for prosodic structure), the MaxEnt model predicts the probabilistic distribution of endings (-jen vs. -iden) observed in empirical data. This outcome highlights the dual nature of the MaxEnt model: it operates as a statistical framework for learning while retaining its grounding in linguistic theory.

We assume that applying the MaxEnt Model to our data offers the following advantages. First, since the model handles gradient phonological patterns, it seems ideal for capturing the probabilistic nature of L2 learners' reliance on phonetic cues to produce vowel contrasts, which may not be strictly categorical. Second, since the current study also examines the role of proficiency levels, the MaxEnt Model could help in tracking how cue-weighting strategies evolve as learners gain more experience with L2.

3. Methodology

3.1. The dataset

The data used in this study was gathered by Ngunga (2024) in an experiment aimed at describing acoustic characteristics of L2 English learners' vowels and how Brazilian Portuguese L1 impacts the production of vowel contrasts. In the following subsections we briefly outline the experimental protocol and the data summary.

3.1.1. Data gathering and analysis

The experimental group consisted of 17 Brazilian learners of English as a Second Language. The group was further divided based on English proficiency levels, determined by the Kaplan Placement Test (Kaplan International, 2023): 9 participants were classified at a basic level, while 8 participants were at an advanced level. To provide a baseline for comparison, a control group of 8 native speakers of American English living in Brazil was included in the study, with no regard to language variety[5].

Materials consist of 48 monosyllabic words comprising long-tense and short-lax high vowels (e.g., "seek" [i], "loop" [u], "sick" [ɪ], "look" [ʊ]), as depicted in table 1.

Table 1: Target vowels and words.

i	u	ɪ	ʊ
creep	lose	ship	put
meet	choose	bit	good
seek	loop	sick	look
seed	group	stiff	foot
cheese	proof	this	cook
speak	mood	rib	wood
leaf	shoot	dig	bush
leave	prove	give	push
please	loose	his	hood
bead	roof	crib	would
beef	move	lid	took

Source: Elaborated by the authors

As seen in table 1, each of the vowels [u], [i], [ʊ], [ɪ] showed in 12 different words. They are commonly occuring monosyllabic words, with varying onset and coda consonants. These words were embedded in a prosodically controlled phrase to ensure consistent intonation and stress patterns. The lengthening effect of syllable coda voicing (Peterson; Lehiste, 1960) was balanced out by using

[5] Although living in Brazil for some time (on average 8 years) before the data collection, all the English native speakers use English in their daily lives and keep close ties to the English speaking community in Brazil and in the United States. Therefore, the influence of Portuguese allophonic patterns as an L2 on their L1 production would be unlikely.

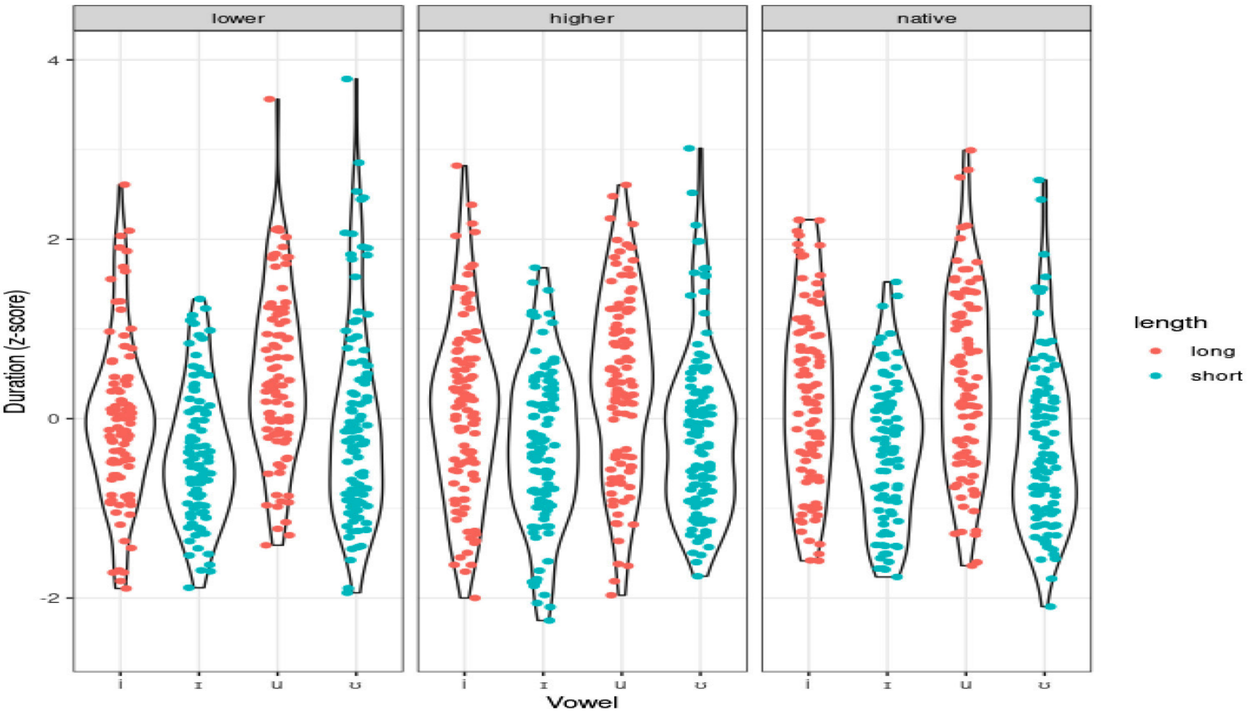
voiced and unvoiced consonants in the coda at near proportions. The trial consisted of the production of the target words inserted into the following carrier phrase: "Say [target word] nicely." Each speaker produced every word once. The analyzed acoustic parameters included duration, F1, and F2.

The acoustic measurements were originally carried out in Praat, using the script *getformants_pitch_amp* (Styler, 2012). The dataset was processed in R using *phonR* (McCloy, 2016) package for vowel normalization and F1-F2 graphs and the *manova* function from *stats* package for statistical analysis. The following section discusses the distribution of these parameters.

3.1.2. Data summary

The dataset comprises 1194 observations, with near 100 data points for each vowel pair-proficiency combination. Figure 1 displays the vowel durations for [i], [ɪ], [u] and [ʊ] across the two proficiency levels and the control group.

Figure 1: Normalized duration of long-tense and short-lax vowels by proficiency level



Source: Elaborated by the authors

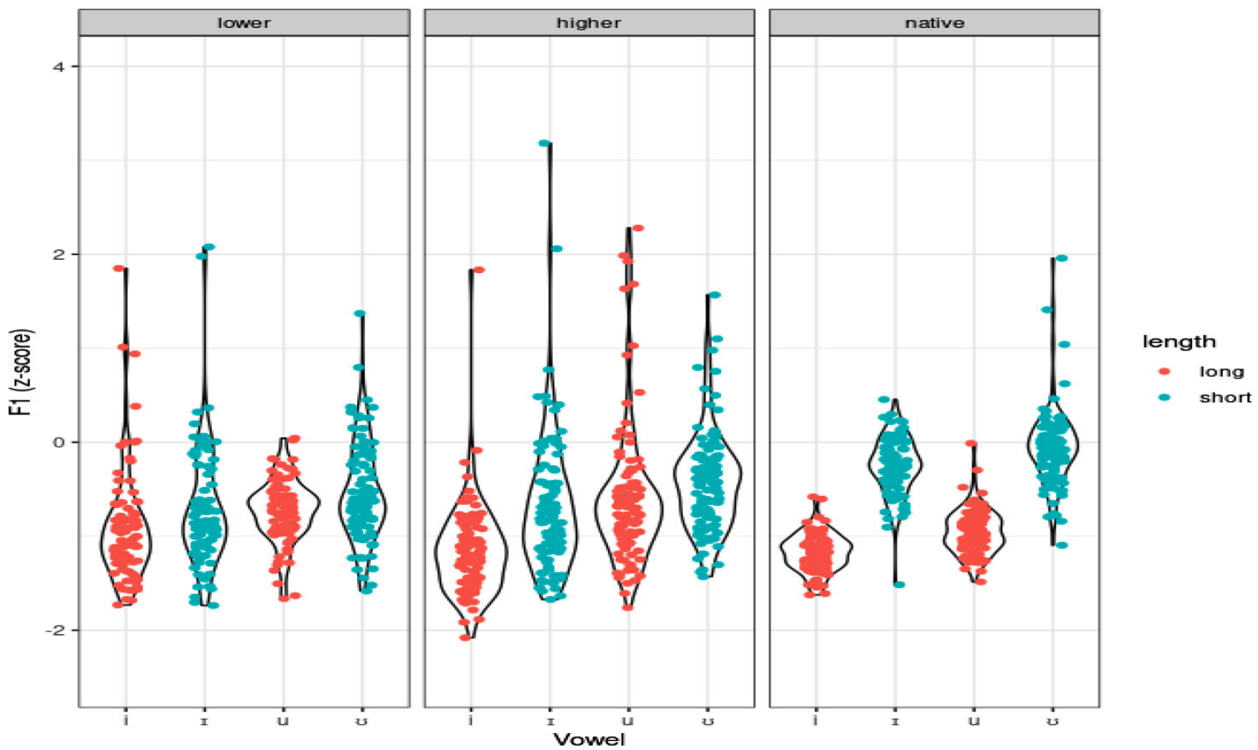
Ngunga's (2024) results indicate that all groups—including both proficiency levels of L2 learners (lower and higher) and American native speakers—exhibited a significant duration contrast for the vowels [i], [ɪ], [u], and [ʊ]. This finding was confirmed by a Kruskal-Wallis test followed by Dunn's post-hoc test with Bonferroni adjustment for multiple comparisons. For the [i] vs. [ɪ] comparison, significant differences were found across all groups: lower proficiency (χ^2 = 9.95, p =

0.00161), higher proficiency ($\chi^2 = 7.56$, $p = 0.00596$), and native speakers ($\chi^2 = 14.11$, $p = 0.00017$). Similarly, the [u] vs. [ʊ] contrast also showed significant differences: low proficiency ($\chi^2 = 11.32$, $p = 0.00077$), high proficiency ($\chi^2 = 12.05$, $p = 0.00052$), and native speakers ($\chi^2 = 20.30$, $p < 0.00001$). This consistent pattern across proficiency levels suggests that L2 learners, regardless of proficiency, are sensitive to the duration contrast feature in English.

The presence of a vowel length contrast in basic level learners' production, further maintained by advanced level learners, suggests that these learners may have developed a duration-based distinction in their English interlanguage, rather than simply transferring it from BP. Because BP does not rely on vowel length contrasts in stressed contexts (and thus does not provide a direct one-to-one cue for length in English), the emergence of this length distinction is better viewed as a new acquisition rather than evidence of full L1 transfer. Consequently, the advanced learners have likely "noticed" the duration cue in English and elevated what might have been an allophonic (or context-dependent) variation in BP to a phonemic-level contrast in their L2 system (Martinez *et al.*, 2023). If this is the case, it implies that L2 learners are actively adapting their phonological representations, moving beyond mere transfer of L1 cues to align more closely with the target language's vowel inventory and phonetic contrasts.

Figure 2 displays F1 values, which correlates with vowel height, of tense and lax vowels across two proficiency levels and the control group.

Figure 2: Normalized F1 of long-tense and short-lax vowels by proficiency level

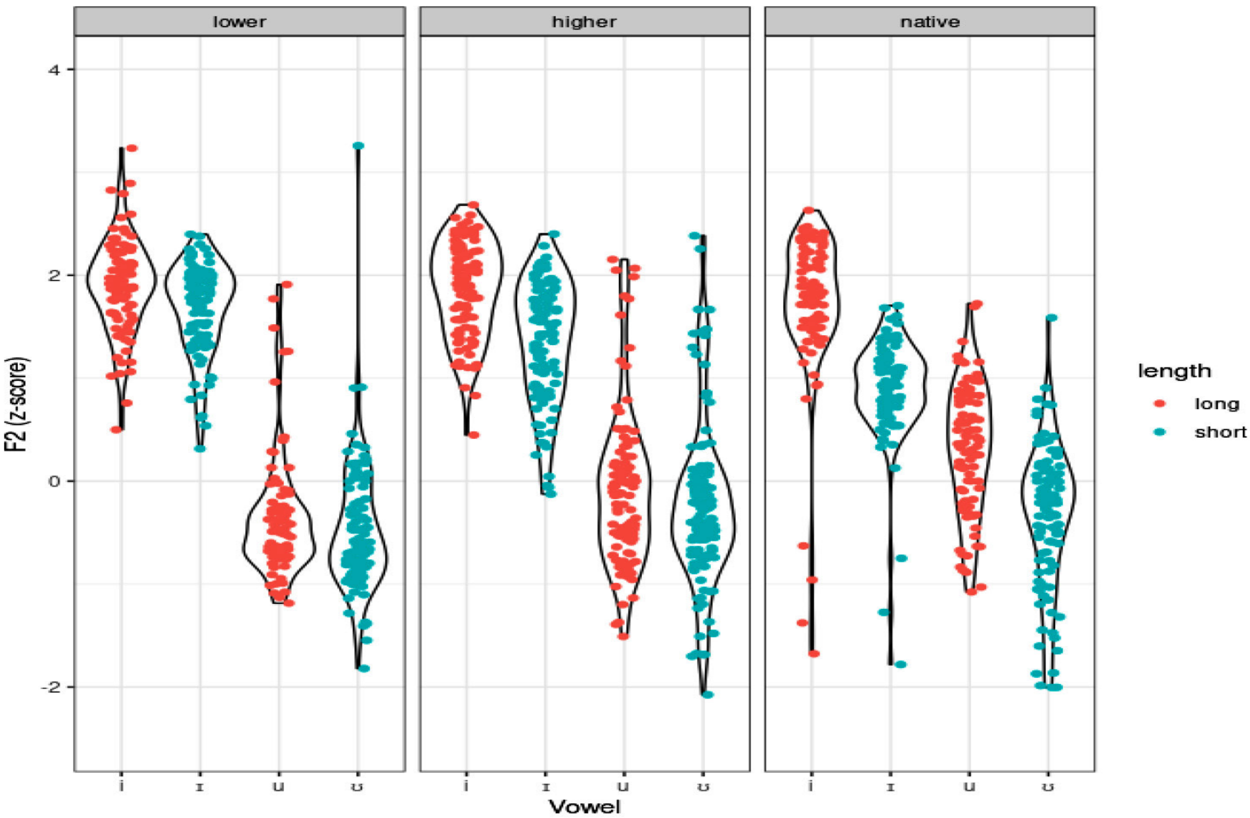


Source: Elaborated by the authors

Results show that no significant differences were observed between tense and lax vowels for both [i] vs [ɪ] ($\chi^2 = 8.47$, $p = 0.0362$) and [u] vs [ʊ] ($\chi^2 = 6.14$, $p = 0.032$) in the lower proficiency group. In contrast, the higher proficiency group displayed significant differences for both vowel pairs—[i] vs [ɪ] ($\chi^2 = 29.23$, $p < 0.00001$) and [u] vs [ʊ] ($\chi^2 = 13.64$, $p = 0.00022$)—indicating improved vowel distinction with increased proficiency. Native speakers also showed significant differences for both [i] vs [ɪ] ($\chi^2 = 127.85$, $p < 0.00001$) and [u] vs [ʊ] ($\chi^2 = 129.06$, $p < 0.00001$), reflecting a clear distinction in vowel contrasts as an acoustic cue in L1 English.

Figure 3 displays F2 values, which correlates with vowel frontness, of long-tense and short-lax vowels across two proficiency levels and the control group.

Figure 3: Normalized F2 of long-tense and short-lax vowels by proficiency level



Source: Elaborated by the authors

For the lower proficiency group, no significant differences were observed between tense and lax vowels for both [i] vs [ɪ] ($\chi2 = 8.31$, $p = 0.393$) and [u] vs [ʊ] ($\chi2 = 2.62$, $p = 0.053$). In the higher proficiency group, a significant difference was found for [i] vs [ɪ] ($\chi2 = 27.05$, $p < 0.00001$), indicating an improved ability to distinguish between these vowels, while no significant difference was found for [u] vs [ʊ] ($\chi2 = 3.29$, $p = 0.0961$). Once again, native speakers exhibited clear distinctions across vowel categories, showing significant differences for both [i] vs [ɪ] ($\chi2 = 88.15$, $p < 0.00001$) and [u] vs [ʊ] ($\chi2 = 42.94$, $p < 0.00001$).

Overall, this suggests that increased proficiency partially improves Brazilian Portuguese learners' ability to distinguish between English long-tense and short-lax vowels, indicating a still emerging yet inconsistent length contrast.

Yet, while learners may show gradual gains in certain vowel distinctions, specific challenges remain – particularly for vowels that do not neatly correspond to BP phonological categories. In the literature, these challenges have been noted in the production and perception of high back vowels, as discussed in the following studies. Rauber (2006) observed that BP speakers tend to have difficulty producing the vowel [ʊ]. For the high back vowels [u] and [ʊ], there are not as many minimal pairs in English when compared to the high front vowels [i] and [ɪ] (Wang, Munro 2004; Lima Jr, 2019). As a result, much of the literature on contrastive vowel production has used words containing vowels with different phonological contexts. For example, Lima Jr (2019), in a longitudinal study on the production of the contrast between [u] and [ʊ], analyzed the production of Brazilian speakers in two semesters. The corpus consisted of words containing the target vowels in a consonantal context similar as in "boot", "poop", and "toot" for [u] and "book", "put" and "took" for [ʊ]. It was possible to verify that some of the students produced distinct vowel pairs, primarily [i] and [ɪ], right from the first two recordings. But, when it comes to the vowels [u] and [ʊ], they were produced as the same in the first recording.

Nobre-Oliveira (2007) assessed the perception and production of the contrast of the long and short vowel pairs /i/ vs. /ɪ/, /æ/ vs. /ɛ/ and /ʊ/ vs. /u/. The author attempted to present target vowels in similar contexts, but found it impossible to do so with the back vowels. The corpus consisted of wTo provide stimuli with similar contexts, the author created "nonsense" words, such as "tuke," to ensure nearly identical consonantal contexts for all target vowels. It is also observed in the corpus used by Rauber (2006) when investigating the production of English vowels [i ɪ ɛ æ u ʊ], since it was impossible to find minimal sets of real words for all contexts.

The production results showed that the participants' production was better for /i/ vs. /ɪ/, followed by /æ/ vs. /ɛ/ and /ʊ/ vs. /u/. Nobre-Oliveira (2007) assumed that learners found the contrast between /ʊ/ and /u/ to be the most challenging. The results further indicated that learners' perception of /i/ vs /ɪ/ was already highly accurate before the training. It is likely that they focused their attention on this contrast during the training sessions, as the pretest results revealed the lowest accuracy in identifying back vowels. Consequently, /ʊ/ and /u/ had the greatest potential for improvement compared to the contrast /i/ vs. /ɪ/. However, the learners experienced difficulty articulating the high back vowels, even after understanding that /ʊ/ and /u/ are distinct and after being able to perceive these vowels accurately. The author suggests that this difficulty may come from the fossilization of speech gestures and the lack of targeted training on this vowel pair throughout the individual's acquisition process. Additionally, while the perception of the /ʊ/ and /u/ contrast showed significant improvement, no corresponding enhancement was observed in the production of /ʊ/ and /u/. We would take this argument further by suggesting that the limited number of minimal pairs involving this vowel pair likely reduces the salience of the back vowel contrast for learners.

3.2. Modelling

This study examines the phonetic and phonological contrasts between long-tense vowels [i, u] and short-lax vowels [ɪ, ʊ], focusing on how learners weigh acoustic cues such as vowel duration, F1, and F2 in speech production. The MaxEnt model allows us to quantify the probability of a particular vowel form based on its adherence to or violation of the relevant phonological conditions and their respective weights. As mentioned in section 2, each MaxEnt condition is assigned a weight, which, in our case, reflects its influence on the overall probability of vowel categorization. The more a vowel violates these conditions and the higher the weight assigned to this condition, the lower its probability of occurrence. According to Hayes and Wilson (2008, p. 383), in the MaxEnt model, the probability of a form given the set of contrasts of a grammar can be calculated as follows:

(1) $$P(x) = e^{-h(x)}$$

where *h(x)* is the score of the form *x*, which is defined as

(2) $$h(x) = \sum_{i=1}^{N} w_i C_i(x)$$

In equation (2), the score *h(x)* of the phonological form *x* is a function of the weight w_i assigned to the i^{th} condition and it respective number of violations $C_i(x)$. Finally, *N* stands for the total number of constraints applied in the MaxEnt model. Together, these components calculate the weighted sum of violations, determining the overall score *h(x)* for the form in question.

In order to estimate weights of phonetic cues, we depart from Equation (1) in the inverse form:

(3) $$h(x) = -ln\big(P(x)\big)$$

The next necessary step is to identify the conditions relevant to the occurrence of the long-tense and short-lax vowels. Building on the observation that typologically frequent vowels—namely [i, e, a, o, u]—are located at the extremes of the trapezoidal vocalic space and are generally unreduced (Ladefoged; Maddieson, 1990), we propose that the relevant conditions for distinguishing long-tense and short-lax vowels can be captured by a set of phonological features that prohibit vowel centralization and temporal reduction. These features, or restrictions, include: the absence of centralized vowels in terms of both height and frontness, and the absence of vowel shortening. Thus, the relevant conditions are formalized as: (1) $\#V_{short}$ ("no short vowel"), (2) $\#V_{cHeight}$ ("no vowel centralized in height"), and (3) $\#V_{cFront}$ ("no vowel centralized in frontness"). Each of these conditions is mapped onto a corresponding phonetic cue: vowel duration is linked to $\#V_{short}$, F1 to $\#V_{cHeight}$, and F2 corresponds to $\#V_{cHeight}$.

In this framework, long-tense vowels such as [i, u] are characterized by non-centralization and no reduction, and would therefore produce no violations of the conditions (C = 0 for all 3 conditions), while short-lax vowels like [ɪ, ʊ], being shorter and more centralized in both height and frontness, would violate each condition once (C = 1 for all 3 conditions).

As a consequence of the conditions selected, long-tense vowels would have a probability of occurrence *P(x)* equal to 1, since *h(x)* sums to 0. In the case of short-lax vowels, we adopted Pillai scores as an index of the probability of occurrence. Pillai scores, which can be derived from MANOVA (Multivariate Analysis of Variance), has been successfully used to measure the degree of separation between two sound categories, such as tense and lax vowels, based on acoustic properties like formant frequencies (F1, F2) and vowel duration (cf. Mairano *et al.*, 2020). Ranging from 0 to 1, a higher Pillai score indicates greater separation between categories, while a lower score suggests more overlap. In our analysis, we interpret the Pillai score as a measure of the probability of occurrence of the short-lax vowels, based on the fact that they are absent in Portuguese L1 in stressed syllables. Therefore, a high Pillai score for the pair [i, ɪ], indicating clear category separation, would correspond to a higher probability of [ɪ], meaning the distinction between categories is more strongly maintained. Conversely, a lower Pillai score would lead to a lower probability of [ɪ], reflecting less separation between the categories.

For each vowel pair [i, ɪ] and [u, ʊ] and for each speaker, the value of $P(x)_{F1F2\mathrm{Dur}}$ was set as the Pillai score of a MANOVA model having as dependent variables F1, F2 and duration, previously normalized by speaker. Then, Equation (3) can be restated as

$$(4) \qquad h_{F1F2Dur}(x) = -ln\big(pillai_{F1F2Dur}(x)\big)$$

Departing from Equation (2), the score $h_{F1F2\mathrm{Dur}}(x)$ of a short-lax vowel form *x* from the combined of F1, F2 and duration can be defined by

$$(5) \qquad h_{F1F2Dur}(x) = w_{F1} + w_{F2} + w_{Dur}$$

Integrating Equations (4) and (5), the following relation is obtained:

$$(6) \qquad -ln\big(pillai_{F1F2Dur}(x)\big) = w_{F1} + w_{F2} + w_{Dur}$$

where w_{F1}, w_{F2}, and w_{Dur} are the weights assigned to violations of conditions associated to F1, F2, and duration, respectively, and the number of violations C_{F1}, C_{F2}, and C_{Dur} are 1.

The problem of estimating the weights w_i of each condition in Equation (6) was tackled by pairwise comparison: the same procedure outlined by Equations (4)-(6) was followed for three partial models containing different combinations of two acoustic parameters, thus yielding pillai scores $pillai_{F1F2}$, $pillai_{F1Dur}$ and $pillai_{F2Dur}$ to be associated with h_{F1F2}, h_{F1Dur} and h_{F2Dur} in the following linear system, which can then be solved for the weights:

(7)

$$h_{F1F2} = w_{F1} + w_{F2}$$

$$h_{F1Dur} = w_{F1} + w_{Dur}$$

$$h_{F2Dur} = w_{F2} + w_{Dur}$$

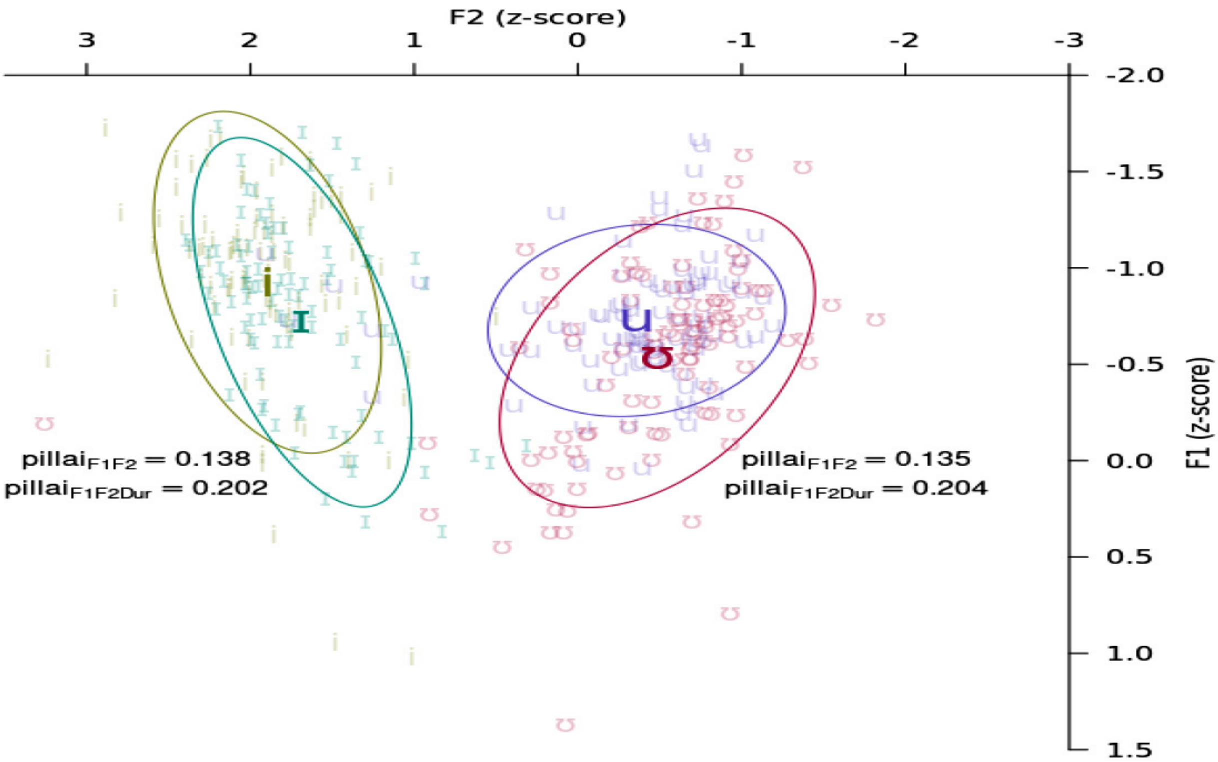
$$\rightarrow w_{Dur} = \frac{h_{F2Dur} - h_{F1F2} + h_{F1Dur}}{2}; w_{F1} = h_{F1Dur} - w_{Dur}; w_{F2} = h_{F1F2} - w_{F1}$$

These weights allow us to break down the interaction of the acoustic properties and identify how learners weigh each cue when producing vowel contrasts.

4. Results

Figure 4 displays an F1-F2 scatterplot of vowel productions by low proficient L2 learners, illustrating the acoustical spaces for tense and lax vowels [i] vs. [ɪ] and [u] vs. [ʊ].

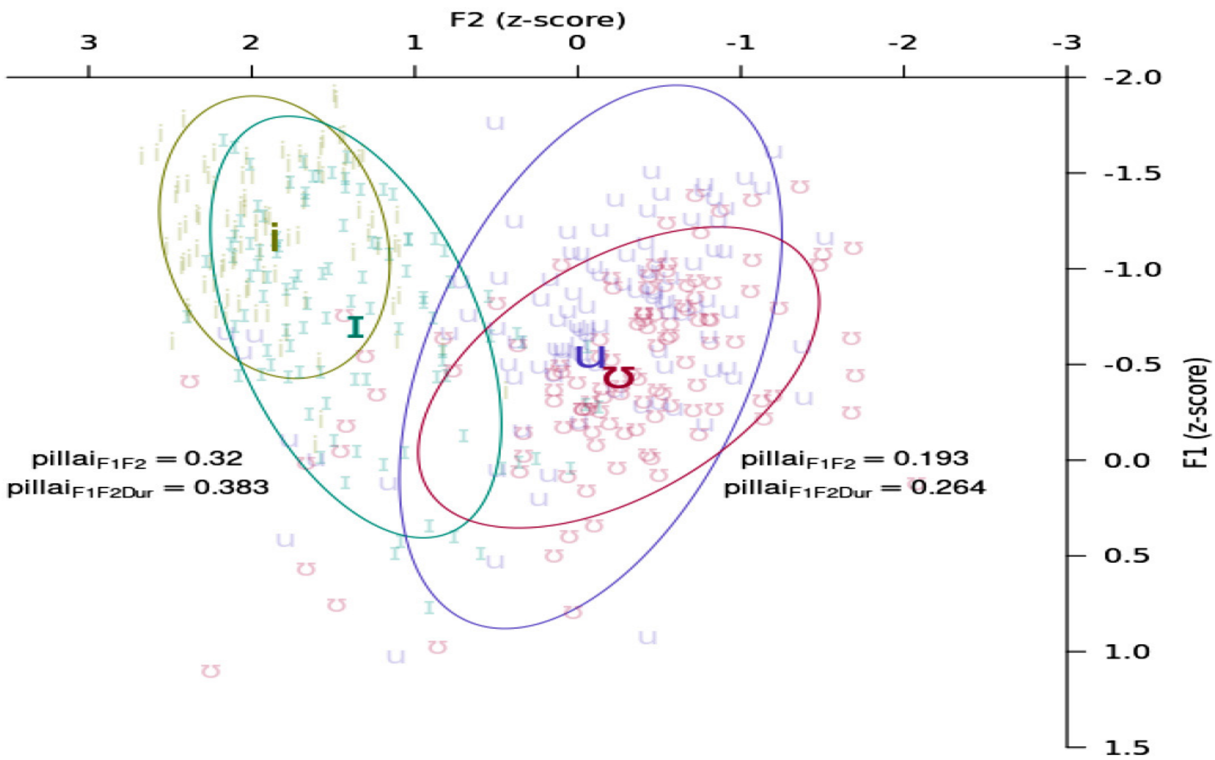
Figure 4: Normalized F1 vs. F2 of long-tense and short-lax vowels for learners with lower proficiency level. For each pair of vowels, mean pillai scores for two combinations of dependent variables are overplotted.



Source: Elaborated by the authors

In figure 4, each ellipse represents plus-or-minus one standard deviation from the mean of vowel tokens for a particular vowel, with green for [i], yellow for [ɪ], red for [u], and blue for [ʊ]. Noticeable overlaps between the ellipses of [i] and [ɪ], as well as [u] and [ʊ], corroborate the data from Figures 1-3 and suggest that learners' productions of these vowels are not distinctly separated, highlighting difficulties in differentiating between the long-tense and short-lax pairs. The variability within each vowel category further indicates inconsistency in vowel production typical among L2 learners. For evaluation purposes, the graphic is annotated with mean Pillai scores for each pair of vowels (front vowels to the left and back vowels to the right), for two combinations of dependent variables (F1 and F2 only and F1, F2 and duration). Pillai scores agree with the area of the F1-F2 graphic that is shared between ellipses. For the learners of lower proficiency, the mean Pillai scores are similar for the two pairs of vowels and their low value – that is, closer to zero than to 1 – indicates a low degree of separability of the categories, either considering F1 and F2 alone or including duration. In the MaxEnt framework adopted in this study, this predicts that the probability of occurrence of short-lax vowels is small for learners of lower proficiency.

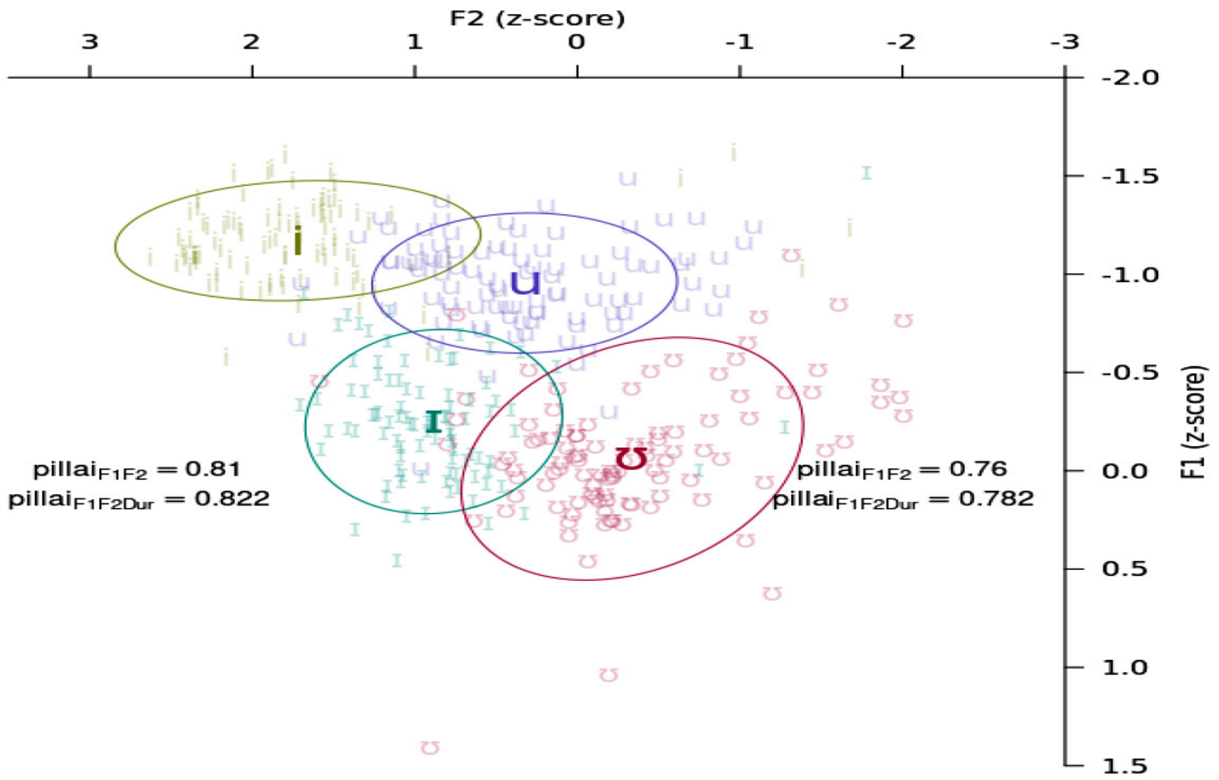
Figure 5: Normalized F1 vs. F2 of long-tense and short-lax vowels for learners with higher proficiency level. For each pair of vowels, mean pillai scores for two combinations of dependent variables are overplotted.



Source: Elaborated by the authors

Figure 5 displays an F1-F2 scatterplot of vowel productions by higher proficiency L2 English learners. The plot shows that [i] has considerably lower F1 and higher F2 values, while [ɪ] exhibits higher F1 and lower F2 values, indicating a significant difference between these vowels and an improved ability to distinguish them compared to lower proficiency learners. In contrast, the back vowels [u] and [ʊ] display substantial overlap, with [u] characterized by low F1 and F2 values, and [ʊ] by slightly higher F1 and F2 values, suggesting no significant difference in the ability to distinguish these vowels. Once again, this indicates that while higher proficiency learners have developed a more consistent tense-lax contrast for front vowels, their distinction between back vowels remains emergent and inconsistent. Overplotted Pillai scores agree with the area of the F1-F2 graphic that is shared between ellipses. For the learners of higher proficiency, the mean Pillai scores for front vowels are higher than those for back vowels, indicating a relative higher separation of from vowels, either considering F1 and F2 alone or including duration, but yet with a low degree of separability of the categories. Therefore, learners of higher proficiency also display a low probability of occurrence of short-lax vowels, but with an increase for the front vowel [ɪ] relative to the lower proficiency group.

Figure 6: Normalized F1 vs. F2 of long-tense and short-lax vowels for native speakers of American English. For each pair of vowels, mean pillai scores for two combinations of dependent variables are overplotted.

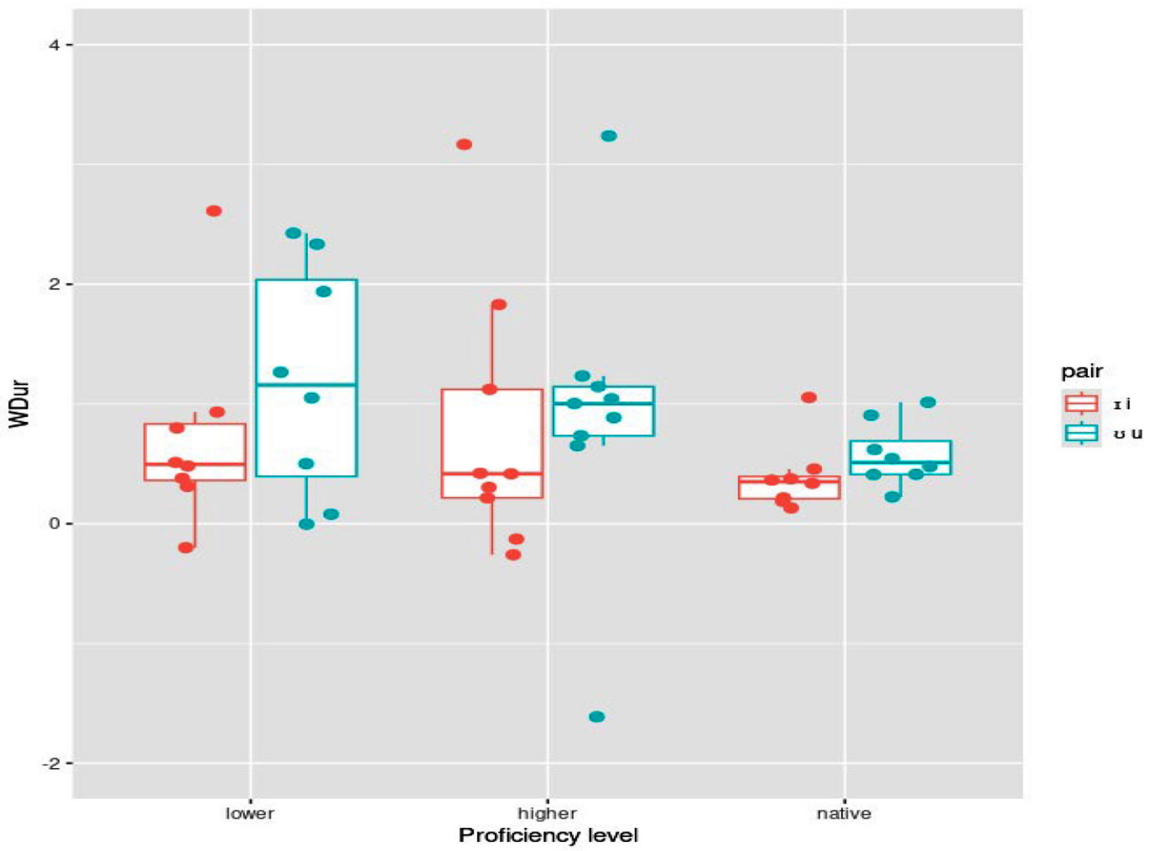


Source: Elaborated by the authors

Figure 6 shows an F1-F2 plot of vowel productions by American English speakers, showing the acoustic spaces for the tense-lax vowel pairs [i] vs. [ɪ] and [u] vs. [ʊ]. The plot demonstrates a clear distinction between these vowel pairs. The ellipses for [i] and [ɪ] are well separated, with [i] having lower F1 and higher F2 values, while [ɪ] has higher F1 and lower F2 values, indicating it is a lower and more central front vowel. Similarly, the ellipses for [u] and [ʊ] are also distinct, with [u] characterized by low F1 and F2 values, and [ʊ] having slightly higher F1 and F2 values, placing it as a lower and more central back vowel. The clear-cut separation between tense and lax vowels in native speakers underscores the distinctive production of these vowel contrasts. Once again the overplotted Pillai scores agree with the area of the F1-F2 graphic that is shared between ellipses. For the native speakers, front vowels have slightly higher mean Pillai scores than back vowels, either considering F1 and F2 alone or including duration, but more importantly the scores are closer to 1, pointing to a high degree of separability of the categories. Hence, as expected, the probability of short-lax vowels occurring in this group is predictably high.

The weights for each of the three conditions estimated using the MaxEnt model are depicted in figures 7-9. Points show the estimates per speaker, for each vowel pair. Starting in figure 7, the weight of the condition $\#V_{short}$ associated with duration is shown.

Figure 7: Estimated weights associated with duration for each vowel pair, grouped by proficiency level.

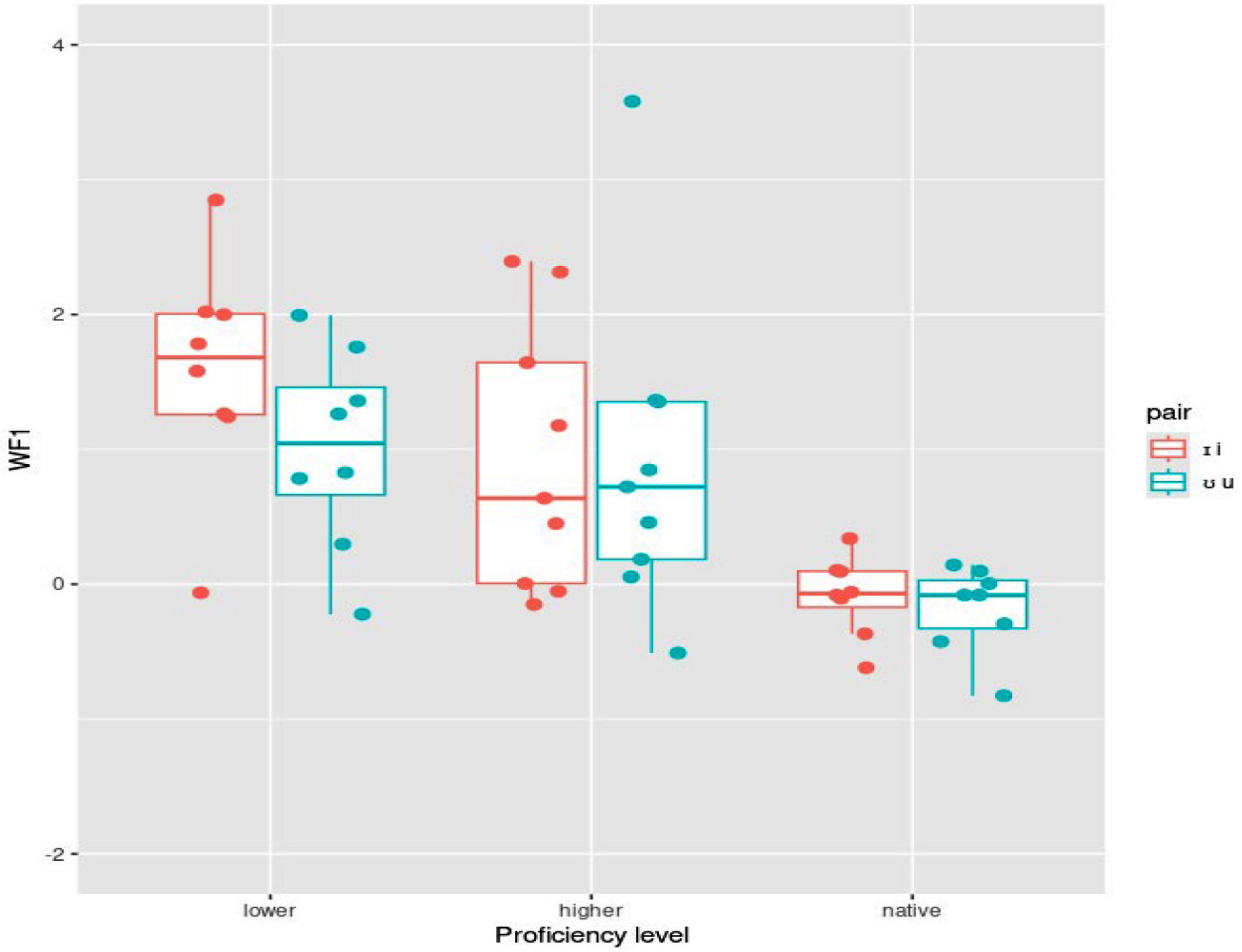


Source: Elaborated by the authors

The weight associated with duration lies mostly between 0 and 2. Learners present higher values than native speakers, with no clear difference between proficiency levels amongst learners. The back vowel pair presents higher weights than the front vowel pair in all proficiency levels, and is more variable for lower proficiency learners.

In figure 8, the weight of the condition $\#V_{cHeight}$ associated with F1 is depicted.

Figure 8: Estimated weight associated with F1 for each vowel pair, grouped by proficiency level.

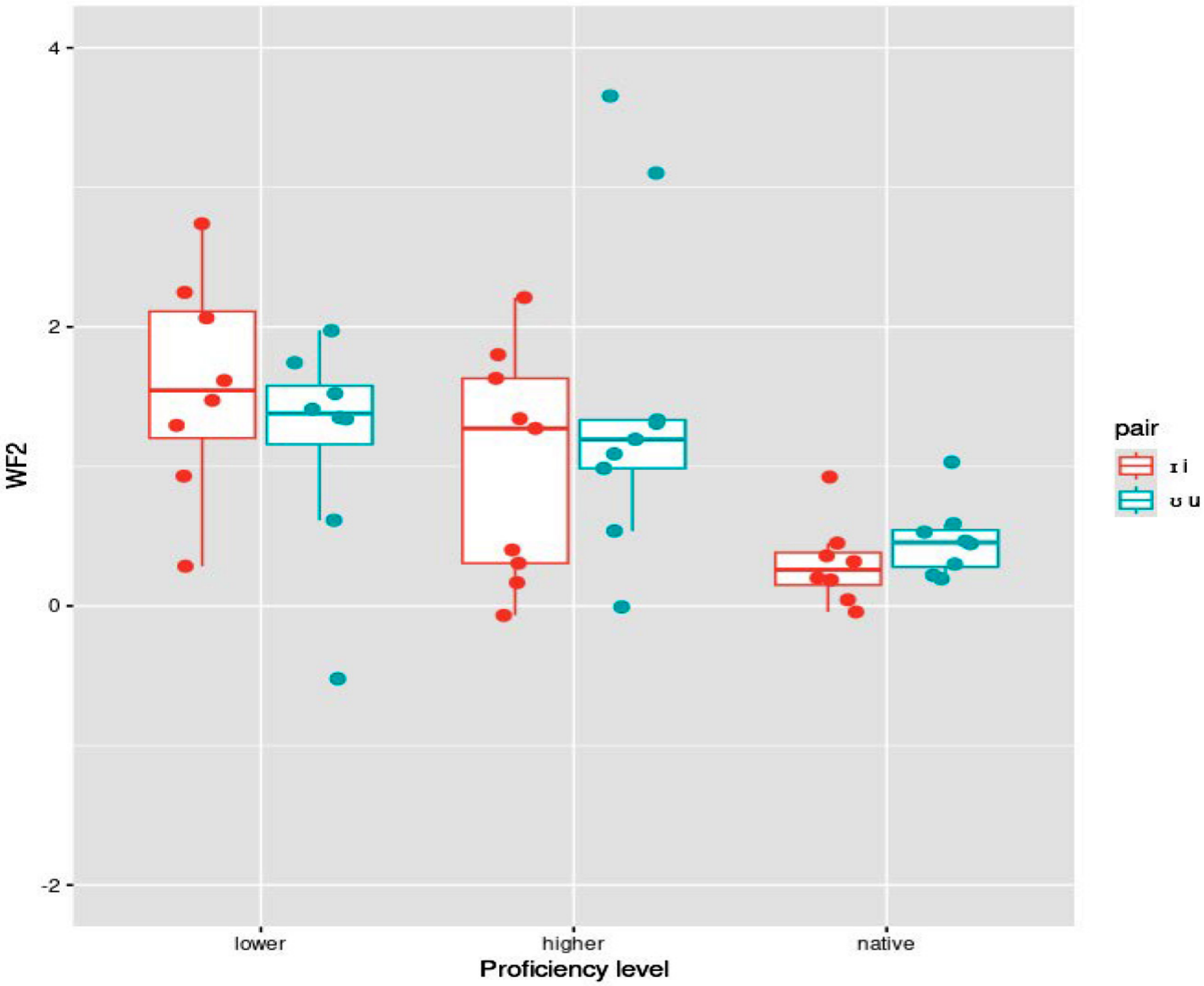


Source: Elaborated by the authors

Figure 8 shows that lower proficiency speakers have greater variability in using F1 to distinguish [i, ɪ] and [u, ʊ] pairs, with higher weights for the [i, ɪ] pair. As proficiency increases, the distribution narrows, especially for the [u, ʊ] pair, indicating more consistent cue usage. Native speakers display the lowest and most uniform weights, suggesting less reliance on F1 for vowel contrast.

Figure 9 shows the weight of the condition $\#V_{cFront}$ associated with F2.

Figure 9: Estimated weight associated with F2 for each vowel pair, grouped by proficiency level.



Source: Elaborated by the authors

Figure 9 shows that lower proficiency speakers exhibit considerable variability in using F2 to distinguish both [i, ɪ] and [u, ʊ] pairs, with higher weights for the [i, ɪ] pair. As proficiency increases, F2 usage becomes more stable, particularly for the [u, ʊ] pair, where weights are more uniform. Native speakers have the lowest and most consistent weights for both pairs, indicating minimal reliance on F2 for vowel contrast. This suggests that as proficiency improves, learners become more consistent and reduce reliance on F2 as a distinguishing cue.

Table 2 displays the mean values and standard deviations for the vowel pairs in each proficiency level. In the MaxEnt framework, high rated weights correspond to conditions that may not be violable in a grammar, since they would severely impact the probability of the form. Lower weights, which have lesser impact on probability, correspond to conditions that may be either violable because they are associated to low frequency forms, or because they form a bundle (or "gang") with other conditions to rule out the same form.

Table 2: Descriptive group values for the weights associated with duration, F1 and F2

Proficiency	Vowel pair	W_{Dur}		W_{F1}		W_{F2}	
		mean	s.d.	mean	s.d.	mean	s.d.
lower	ɪ i	0.729	0.833	1.584	0.840	1.581	0.776
	ʊ u	1.119	0.968	1.008	0.740	1.179	0.792
higher	ɪ i	0.788	1.097	0.935	0.996	1.007	0.819
	ʊ u	0.925	1.228	0.895	1.175	1.467	1.172
native	ɪ i	0.390	0.290	-0.087	0.296	0.305	0.298
	ʊ u	0.576	0.265	-0.183	0.322	0.472	0.267

Source: Elaborated by the authors

Learners with lower proficiency level present higher and similar weights for F1 and F2 in the pair [ɪ i] and very close weights in the pair [ʊ u]. Learners with higher proficiency level also display similar weights in the pair [ɪ i] and higher value for F2 in the pair [ʊ u]. Native speakers show higher and similar weights for F2 and duration in both pairs. In general, higher proficiency learners' weights present higher dispersion and, as expected, native speakers present the lowest dispersion values.

5. Discussion

Ngunga's (2024) study found that, across different proficiency levels, Brazilian learners consistently use vowel duration as a primary cue when distinguishing between tense and lax vowels, particularly in the [i] vs. [ɪ] and [u] vs. [ʊ] contrasts. However, sensitivity to other cues, such as F1 and F2, differed across proficiency levels. While low-proficiency learners showed minimal distinction in height and frontness, higher-proficiency learners displayed significant variability in their ability to differentiate these acoustic dimensions, especially in distinguishing [i] from [ɪ]. Native English speakers, on the other hand, consistently showed significant contrasts across all acoustic dimensions, confirming that F1 and F2, alongside duration, are significant for categorizing vowel length contrasts in English (c.f. section 3.1.2). These findings highlight the complex interaction between L1 transfer and L2 acquisition. Because BP lacks a tense-lax vowel contrast, learners initially produce little to no length distinction in English. Over time, however, they begin to rely on duration as a second developmental stage, treating it as a salient cue for distinguishing vowel contrasts. As proficiency increases further, learners gradually incorporate secondary acoustic cues—such as F1 and F2—though sometimes inconsistently. This progression from no length differences to reliance on duration, and eventually to the integration of spectral cues, underscores both the learners' adaptation to L2 phonological patterns and the continued influence of their L1 phonology. Interestingly, learners with advanced proficiency show higher variability in distinguishing these contrasts, suggesting that they manage multiple competing parameters—such as duration, height, and frontness—while developing a more target-like phonological system. Additionally, even with similar proficiency levels, learners' may be at different stages in the contrast acquisition process (cf. Lima Jr., 2019 for a longitudinal evaluation of vowel inventory of Brazilian L2 English learners), which could also contribute to the attested variability.

While cue weighting has been studied extensively in perception (Schertz & Clare, 2019), its application in production studies, especially in L2 contexts, remains underexplored, making it crucial for understanding how learners balance these cues across modalities. The MaxEnt model employed in this study allowed for a comparison between the multiple cues employed in the emerging L2 contrast. Whereas conventional statistical analyses clearly showed significant differences in vowel duration between tense and lax vowels (cf. section 3.1.2), MaxEnt indicated that duration was not as heavily weighted in the learners' phonological grammar. This result suggests that, while learners use duration as a salient cue in their productions, it does not carry the same weight in their underlying grammatical system. MaxEnt modeling, by accounting for the probabilistic nature of cue use, enabled us to handle multiple dependent variables simultaneously, providing a more nuanced understanding of the relative importance of each cue within the learners' interlanguage phonological system.

6. Conclusion

Cue weighting is an important theme in L2 phonology, since it represents steps of the acquisition of grammatical knowledge by learners. In this study, a case of competing phonetic cues was addressed and modelled using the Maximum Entropy Model (Hayes and Collins, 2008): the production of quantity-tenseness contrast in L2 English by Brazilian learners, whose L1 Portuguese lack this opposition. The competing phonetic cues – duration, F1 and F2 – were associated with phonological conditions avoiding short-lax vowels, which are violable in English. The data modelling was able to separate the weights of each condition, indicating an increase in the reliance on F1 and mostly F2 from initial to intermediate stages of the acquisition process. Furthermore, there was an increased dispersion of the weights in the higher proficiency learners, possibly due to instability during the learning process. Besides being able to assess cue weighting in production, the model could allow for an integrated analysis of production and perception, via weight comparison. The evaluation on production can be supplemented with a perception study on well-formedness judgments, to test the agreement between cue weighting in both dimensions. Lastly, the present work was restricted to a single instance of contrast involving L2 vowels. In future developments, a probabilistic analysis of the whole L2 vowel inventory may offer a broader understanding of the development of phonological categories and contrasts.

References

ALBRIGHT, Adam; HAYES, Bruce. Learning and learnability in phonology. The handbook of phonological theory, pp. 661-690, 2011.

ALDERETE, John; FINLEY, Sara. Probabilistic phonology: A review of theoretical perspectives, applications, and problems. Language and Linguistics, v. 24, n. 4, pp. 565-610, 2023. DOI: https://doi.org/10.1075/lali.00141.ald

BERGMANN, C.; NOTA, A.; SPRENGER, S.; SCHMID, M. L2 immersion causes non-native-like L1 pronunciation in German attriters. *J. Phonetics*, v. 58, pp. 71-86, 2016. DOI: https://doi.org/10.1016/j.wocn.2016.07.001.

BERTI, L. C.; PAGLIUSO, A.; LACAVA, F. Instrumento de avaliação de fala para análise acústica (IAFAC) baseado em critérios linguísticos. *Revista da Sociedade Brasileira de Fonoaudiologia*, v. 14, pp. 305-314, 2009.

BEST, C. T.; TYLER, M. D. Nonnative and second-language speech perception: Commonalities and complementarities. *In*: M. J. Munro & O.-S. Bohn (eds.). *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam: John Benjamins, 2007. pp. 13-34.

BOERSMA, P.; WEENINK, D. Praat: doing phonetics by computer [Computer program]. Version 6.2.06. 1992–2022. Disponível em: https://www.praat.org. Acesso em: 23 maio 2023.

CASILLAS, J. Production and Perception of the /i/-/ɪ/ Vowel Contrast: The Case of L2-Dominant Early Learners of English. *Phonetica*, v. 72, pp. 182-205, 2015. DOI: https://doi.org/10.1159/000431101.

CHLÁDKOVÁ, K.; ESCUDERO, P. Comparing vowel perception and production in Spanish and Portuguese: European versus Latin American dialects. *The Journal of the Acoustical Society of America*, v. 131, n. 2, pp. EL119-EL125, 2012.

CRYSTAL, David. *A dictionary of linguistics and phonetics*. John Wiley & Sons, 2011.

DARCY, I.; KRÜGER, F. Vowel perception and production in Turkish children acquiring L2 German. *J. Phonetics*, v. 40, pp. 568-581, 2012. DOI: https://doi.org/10.1016/j.wocn.2012.05.001.

ESCUDERO, P. *et al*. A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *The Journal of the Acoustical Society of America*, v. 126, n. 3, pp. 1379-1393, 2009.

ESCUDERO, Paola. *Linguistic perception and second language acquisition*. LOT Landelijke Onderzoekschool Taalwetenschap, 2005.

FLEGE, James Emil. Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, v. 92, pp. 233-277, 1995.

FLEGE, James Emil; BOHN, Ocke-Schwen. The revised speech learning model (SLM-r). *Second language speech learning: Theoretical and empirical progress*, v. 10, 2021.

FLEGE, James Emil; BOHN, Ocke-Schwen; JANG, Sunyoung. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, v. 25, n. 4, pp. 437-470, 1997.

GOAD, Heather; WHITE, Lydia. Prosodic transfer and the representation of determiners in Turkish-English interlanguage. *Representational deficits in SLA: Studies in honor of Roger Hawkins*, 2009.

GOLDWATER, Sharon; JOHNSON, Mark. Learning OT constraint rankings using a maximum entropy model. *In*: *Proceedings of the workshop on variation within Optimality Theory*, 2003. pp. 111-120.

GONÇALVES, Alison; SILVEIRA, Rosane. A produção das vogais altas anteriores do inglês por aprendizes brasileiros. *Fórum Linguístico*, v. 11, n. 1, pp. 9-22, 2014.

GUT, U. Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism*, v. 7, pp. 19-38, 2010. DOI: https://doi.org/10.1080/14790710902972248.

HAYES, Bruce; WILSON, Colin. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry*, v. 39, n. 3, pp. 379-440, 2008.

IMAIZUMI, S. *et al*. Effects of L1 phonotactic constraints on L2 speech perception and production. *Journal of the Acoustical Society of America*, v. 105, pp. 1094-1094, 1999. DOI: https://doi.org/10.1121/1.425128.

Kaplan International. Kaplan Placement Test. Disponível em: https://www.kaplaninternational.com. Acesso em: 16 dez. 2023.

KONG, Eun Jong; YOON, In Hee. L2 Proficiency Effect on the Acoustic Cue-Weighting Pattern by Korean L2 Learners of English: Production and Perception of English Stops. 말소리와 음성과학, v. 5, n. 4, pp. 81-90, 2013.

LADEFOGED, Peter; MADDIESON, Ian. Vowels of the world's languages. *Journal of Phonetics*, v. 18, n. 2, pp. 93-122, 1990.

LI, A.; POST, B. L2 Acquisition of Prosodic Properties of Speech Rhythm. *Studies in Second Language Acquisition*, v. 36, pp. 223-255, 2014. DOI: https://doi.org/10.1017/S0272263113000752.

LIM, J.; SEO, M. The Influence of L1 Constraints on L2 Sound Perception. *Foreign Languages Education*, v. 22, pp. 23-42, 2015. DOI: https://doi.org/10.15334/FLE.2015.22.1.23.

LIMA Jr, R. M. A longitudinal study on the acquisition of six English vowels by Brazilian learners. *Proceedings of the 19th International Congress of Phonetic Sciences*, pp. 3180-3184, 2019.

LISKER, L.; ABRAMSON, A. S. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, v. 20, n. 3, pp. 384-422, 1964.

MAIRANO, P. *et al*. Acoustic distances, Pillai scores and LDA classification scores as metrics of L2 comprehensibility and nativelikeness. *In*: *Proceedings of the 2019 International Congress of Phonetic Sciences*, 2019.

MARTINEZ, Ruth Maria; GOAD, Heather; DOW, Michael. L1 phonological effects on L2 (non-) naïve perception: A cross-language investigation of the oral–nasal vowel contrast in Brazilian Portuguese. *Second Language Research*, v. 39, n. 2, pp. 387-423, 2023. DOI: https://doi.org/10.1177/02676583211044953.

McCLOY, Daniel R. "*phonR*: Tools for phoneticians and phonologists." R package version 1.0-7. Disponível em: http://drammock.github.io/phonR/, 2016.

McMURRAY, B. *et al*. Gradient sensitivity to acoustic detail and temporal integration of phonetic cues. *Journal of the Acoustical Society of America*, v. 115, pp. 2631-2631, 2004. DOI: https://doi.org/10.1121/1.4784865.

MOBARAKI, M. *et al.* The status of subjects in early child L2 English. *In*: *Current trends in child second language acquisition: A generative perspective*, pp. 209-235, 2008.

MORRISON, G. L1-Spanish Speakers' Acquisition of the English /i/-/ɪ/ Contrast: Duration-based Perception is Not the Initial Developmental Stage. *Language and Speech*, v. 51, pp. 285-315, 2008. DOI: https://doi.org/10.1177/0023830908099067.

NEVINS, A.; BRAUN, D. The role of underlying representations in L2 Brazilian English. *Loan Phonology*, v. 307, p. 181, 2009.

NGUNGA, Flora. *A produção das vogais breves e longas do inglês por aprendizes do português brasileiro*. Dissertação (Mestrado em Estudos Linguísticos) - Universidade Federal de Minas Gerais, Belo Horizonte, 2024.

NOBRE-OLIVEIRA, D. *The effect of perceptual training on the learning of English vowels by Brazilian Portuguese speakers*. Unpublished Doctoral Dissertation. Florianópolis, SC: Universidade Federal de Santa Catarina, 2007.

OLIVEIRA, N. *Representação fonológica, instrução explícita e produção sonora do espanhol como L2 por aprendizes brasileiros*. 2021. Tese (Doutorado em Estudos Linguísticos) - Universidade Federal de Minas Gerais, Belo Horizonte, 2021.

ÖZÇELIK, Öner. Universal Grammar and second language phonology: Full Transfer/Prevalent Access in the L2 acquisition of Turkish "stress" by English and French speakers. *Language Acquisition*, v. 25, n. 3, pp. 231-267, 2018.

PETERSON, G. E.; LEHISTE, I. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, v. 32, n. 6, pp. 693-703, 1960. DOI: https://doi.org/10.1121/1.1908183.

PISKE, Thorsten; MACKAY, Ian RA; FLEGE, James E. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, v. 29, n. 2, pp. 191-215, 2001.

RATO, A. A. S. *Cross-language perception and production of English vowels by Portuguese learners*: The effects of perceptual training. 2014.

RAUBER, A. English vowel perception by native speakers of European Portuguese and Mandarin. *Journal of the Acoustical Society of America*, v. 125, pp. 2771-2771, 2009. DOI: https://doi.org/10.1121/1.4784724.

RAUBER, A. S. *Perception and production of English vowels by Brazilian EFL speakers,* 2006 (Doctoral dissertation, Universidade Federal de Santa Catarina, Centro de Comunicação e Expressão. Programa de Pós-Graduação em Letras/Inglês e Literatura Correspondente.) 2006.

SCHERTZ, Jessamyn; CLARE, Emily J. Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*, v. 11, n. 2, p. e1521, 2020.

SCHWARTZ, Bonnie D.; SPROUSE, Rex A. L2 cognitive states and the full transfer/full access model. *Second Language Research*, v. 12, n. 1, pp. 40-72, 1996.

STYLER, Will. *getformants_pitch_amp.praat*. Computer program. Available at: https://github.com/stylerw/styler_praat_scripts. 2012.

VIEIRA, Maria José Blaskovski; LOPES, Fernanda Peres. Características acústicas da vogal átona final. *Caderno de Letras*, n. 33, pp. 69-84, 2019.

WANG, X. Mandarin listeners' perception of English vowels: Problems and strategies. *Canadian Acoustics*, v. 34, pp. 15-26, 2006.

WANG, X., & Munro, M. J. Computer-based training for learning English vowel contrasts. *System*, v. 32, n. 4, pp. 539-552, 2004.

YAZAWA, K. *et al.* Language-dependent cue weighting: An investigation of perception modes in L2 learning. *Second Language Research*, v. 36, pp. 557-581, 2020. DOI: https://doi.org/10.1177/0267658319832645.

ZURAW, Kie Ross. *Patterned exceptions in phonology*. University of California, Los Angeles, 2000.