



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE COMUNICAÇÃO E EXPRESSÃO
DEPARTAMENTO DE LÍNGUA E LITERATURA ESTRANGEIRAS
BACHARELADO EM LETRAS INGLÊS

Rubens Costa Güths

**Going Beyond Standard Accents in L2: a Layered Model of accents for speech
production**

Florianópolis, Santa Catarina

2024

Rubens Costa Güths

Going Beyond Standard Accents in L2: a Layered Model of accents for speech production

Trabalho de Conclusão de Curso submetido ao curso de Letras Inglês do Centro de Comunicação e Expressão da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Bacharel em Letras Inglês.

Orientador(a): Profa. Dra. Rosane Silveira

Florianópolis, Santa Catarina

2024

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

Güths, Rubens Costa

Going Beyond Standard Accents in L2 : a Layered Model
of accents for speech production / Rubens Costa Güths ;
orientadora, Rosane Silveira, 2024.

57 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro de
Comunicação e Expressão, Graduação em Letras - Língua
Inglesa, Florianópolis, 2024.

Inclui referências.

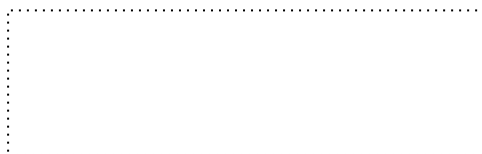
1. Letras - Língua Inglesa. 2. Accent learning. 3.
Pronunciation teaching. 4. Phonetics and Phonology. 5. L2
English. I. Silveira, Rosane. II. Universidade Federal de
Santa Catarina. Graduação em Letras - Língua Inglesa. III.
Título.

Rubens Costa Güths

Going Beyond Standard Accents in L2: a Layered Model of accents for speech production

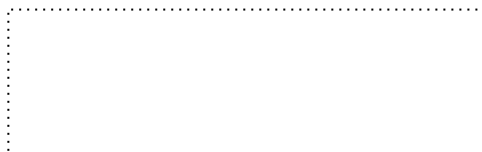
Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do título de Bacharel em Letras Inglês e aprovado em sua forma final pelo Curso de Graduação em Letras Inglês.

Local Florianópolis, 27 de março de 2024.



Coordenação do Curso

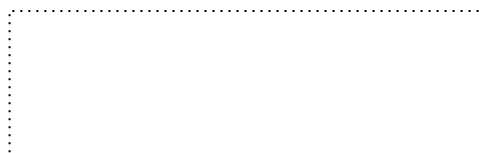
Banca examinadora



Profa. Orientadora Dra. Rosane Silveira



Dr. Alison Roberto Gonçalves
Universidade Federal do Paraná



Doutoranda Aline Almira Morbach
Universidade Federal de Santa Catarina

Florianópolis, 2024.

*A word is dead
When it is said,
Some say —
I say it just
Begins to live
That day.*

Emily Dickinson

ABSTRACT

With the aim of going beyond standard accents of second language (L2) textbooks, this study presents a Layered Model for L2-accent learning. A wide review of literature was conducted combining knowledge from several fields, such as second language learning, Speech-language pathology, anthropology as well as theories in first and second language acquisition and singing pedagogy. The resulting model was then developed and is composed of nine layers. Each layer refers to isolated elements of the accent's aesthetic that can be worked separately during L2-accent learning. The first three layers refer to vocal aspects of accents, which includes vocal tract setting, voice tone and vocal biomechanical articulation. Layers number four, five and six refer to how to connect the vocal abilities from the previous layers into language, which involves practice of speech sounds, consonant clusters and words in general. Layer number seven explores the relations between sounds and letters while layer number eight focuses on adjustments in connected speech, the latter includes the phonological processes. Finally, the ninth layer refers to the skill of breaking down accents into layers in order to enable the learner to develop a panoramic perspective of accents' aesthetic features to learn other accents as well. In order to demonstrate how to apply the Layered Model, English was established as the target L2. Thus, possible segments, clusters and phonological processes present in English were organized into three training-stages, each divided into two minor groups, totaling six groups altogether. Each group displays the English phonetics and phonological elements into smaller feasible components with the aim to guide the L2-accent learning throughout the Layered Model. Different from traditional textbooks, the Layered Model has non-standard accents as the ultimate goal in L2-accent learning.

Keywords: Accent learning; Pronunciation teaching; Phonetics and Phonology; L2 English.

RESUMO

Com o objetivo de ir além dos sotaques padrões dos livros didáticos de segunda língua (L2), este estudo apresenta um Modelo em Camadas para a o estudo de sotaques em L2. Foi realizada uma ampla revisão da literatura, combinando conhecimentos de diversas áreas, como aprendizagem de segunda língua, fonoaudiologia, antropologia, assim como teorias em aquisição de primeira e segunda língua e pedagogia vocal para o canto. O modelo resultante foi, então, desenvolvido e é composto por nove camadas. Cada camada refere-se a elementos isolados da estética dos sotaques que podem ser trabalhados separadamente durante a aprendizagem de sotaques em L2. As três primeiras camadas referem-se a aspectos vocais dos sotaques, que incluem configuração do trato vocal, tom da voz e biomecânica articulatória vocal. As camadas quatro, cinco e seis referem-se a como conectar as habilidades vocais das camadas anteriores à linguagem, o que envolve a prática de sons da fala, agrupamentos de consoantes e palavras em geral. A camada sete explora as relações entre sons e letras, enquanto a camada oito foca em ajustes na fala encadeada, que inclui os processos fonológicos. A última camada, a nona, refere-se à habilidade de decompor sotaques em camadas, para permitir ao estudante desenvolver uma perspectiva panorâmica das características estéticas dos sotaques, bem como para aprender outros sotaques. Para demonstrar como aplicar o Modelo em Camadas, o inglês foi estabelecido como o L2 alvo. Assim, segmentos, encontros consonantais e processos fonológicos possíveis do inglês foram organizados em três estágios de treinamento, cada um dividido em dois grupos menores, totalizando seis grupos ao todo. Cada grupo dispõe os elementos fonéticos e fonológicos do inglês em componentes menores palpáveis com o objetivo de orientar a aprendizagem do sotaque em L2 ao longo do Modelo em Camadas. Diferente dos livros didáticos tradicionais, o Modelo em Camadas tem como objetivo final sotaques não padronizados na aprendizagem de L2.

Palavras-chave: Aprendizagem de sotaque; Ensino de pronuncia; Fonética e Fonologia; Inglês L2.

LIST OF FIGURES

Figure 1 - English syllabic structure possibilities. The nucleus is represented in red and the onset and coda position in black. (based on Yavas, 2011, pp. 139-40).	28
Figure 2 - Sonority Sequencing Principle universal model. The nucleus is represented in red and the onset and coda position in black. (based on Carlisle, 2001, p. 4, and Yin, Weijer and Round, 2023, p. 383).....	29
Figure 3 - The composition of an accent aesthetic identity (developed by the authors).	37
Figure 4 - Organization of the nine layers of the Layered Model (developed by the authors).	46

LIST OF TABLES

Table 1 - Phonological Processes expected according to group and age of suppression.....	32
Table 2 - Common Phonological Processes observed in Brazilian learners of English as L2.....	34
Table 3 - Adaptation of the stages of English L1 acquisition into training stages for English L2 in the Layered Model.	40

LIST OF ABBREVIATIONS

GA	General American
L1	First Language
L2	Second Language
RP	Received Pronunciation
SLP	Speech-language pathology
SLPs	Speech-language pathologists
SSP	Sonority Sequencing Principle
V-SSP	violate the Sonority Sequencing Principle

Contents

1.	INTRODUCTION.....	12
2.	REVIEW OF LITERATURE	14
2.1.	Debates On First And Second Language Acquisition	14
2.2.	Defining Foreign Accents	17
2.3.	Defining Standard Accents.....	18
2.4.	Limitations Of Standard Accents	19
2.5.	Relevance Of Gain Knowledge About Accents	20
3.	THE PRESENT STUDY	21
3.1.	How The Brain Got Language.....	22
3.2.	Vocal Control.....	23
3.3.	Acquiring Articulatory Gestures.....	23
3.4.	Acquiring Segments	25
3.5.	Learning Syllable Patterns And Phonotactics	27
3.6.	Phonological Processes In First Language Acquisition.....	31
3.7.	Phonological Processes In Second Language Acquisition.....	33
4.	GOING BEYOND STANDARD ACCENTS: A LAYERED MODEL OF SPEECH PRODUCTION.....	35
4.1.	Components Of An Accent Aesthetic Identity.....	36
4.2.	Training-Stages On The Layered Model	37
4.3.	The Layered Model Of Pronunciation For L2-Accent Learning	41
4.3.1.	1 st Layer: Vocal Tract Setting	41
4.3.2.	2 nd Layer: Voice Tone Patterns	42
4.3.3.	3 rd Layer: Vocal Gestural Or Biomechanical Articulation.....	43
4.3.4.	4 th Layer: Speech-Sounds Production (Scat Singing And Simple Words) ...	43
4.3.5.	5 th Layer: Combining Speech-Sounds According To The Sonority Sequencing Principle	44
4.3.6.	6 th Layer: Combining Speech-Sounds To Violate The Sonority Sequencing Principle	44
4.3.7.	7 th Layer: Connecting Sounds And Letters	44
4.3.8.	8 th Layer: Adjustments In Connected Speech	45
4.3.9.	9 th Layer: Breaking Down Accents	45
5.	CONCLUDING REMARKS.....	47
6.	REFERENCES.....	49

1. INTRODUCTION

Different ways of pronouncing the same language have historically intrigued researchers (Cruttenden, 2008). Although numerous studies on speech sounds have been conducted throughout history, their predominant focus until the 20th century was centered on the acoustic properties of the sounds. These studies strongly demonstrated that people never pronounced what they thought they were pronouncing and never heard what they thought they were hearing. Consequently, there was a clear need for a linguistic foundation to study speech sounds at that time (Vagones, 1980).

According to Joseph (2022), Ferdinand de Saussure's works in the 20th century had a significant impact on establishing linguistic foundations for studying speech sounds. Saussure redirected researchers' focus from material sound features, such as acoustic properties, to the psychological representations of speech sounds. In his *General Course of Linguistics* (1917) it is written that psychological imprints of sounds become evident when “without moving our lips or tongue, we can talk to ourselves or recite mentally a selection of a verse” (Saussure, 1959, pg. 66). This idea aligns with Vygotsky's theory of private speech, which later transforms into inner speech. Mitchell, Myles, and Marsden (2013) note that Vygotsky's concept of private speech is apparent when an individual appears to be speaking to and for themselves. As speech, without any external articulation, evolves to regulate internal thought, it becomes inner speech—essentially 'pronouncing' the language on a mental level.

Linguistic theories have played a pivotal role in bringing about significant changes in the field of Speech-Language Pathology (SLP) during the 20th century. According to Mota and Wiethan (2014), the incorporation of linguistic theories into SLP therapeutic approaches has led to more effective strategies for reorganizing patients' phonological systems, particularly from the 1980s onward. Prior to this period, SLPs primarily focused on articulatory training, starting with the training of individual sounds and progressing to syllables, words, and sentences. However, this approach proved inadequate as it overlooked the structural aspects of language. As highlighted by Ceron and Keske-Soares (2009) and Mota and Wiethan (2014), the current objective of SLPs in phonological therapy is to reorganize the patient's phonological system, marking a shift from the traditional emphasis on isolated sounds to a more holistic consideration of linguistic structures.

The therapeutic approach of second language accents also has changed during the 20th century. Derwing and Munro (2015) present a summary of studies about the goals in pronunciation teaching for foreign students, including the perspective of SLPs. The authors show that in the early 20th century foreign accents were considered as pathological by SLPs and the therapeutic approach focused on eradicating the foreign accent. Later in the same century, foreign accents lost their status of a “speech pathology” and were defined as a “nonpathological speech produced by second language learners that differs in partially systematic ways from the speech characteristic of native speakers of a given dialect” (Munro, 1998, p. 139, *apud* Zárate-Sández, 2017, p. 228).

Currently, in the realm of second language (L2) pronunciation pedagogy, two primary approaches are prevalent: the Intelligibility Principle and Nativeness (Levis, 2020). The former seeks to achieve an intelligible speech more than a native-like pronunciation. The latter seeks to achieve a native-like pronunciation. Nevertheless, both research and pedagogical studies have highlighted that intelligibility is more important than nativeness in L2 context, resulting in the weakening of the emphasis on sounding native-like. The reason for this could be attributed to empirical evidence indicating that L2 adult learners rarely achieve native-like pronunciation and that intelligibility and accentedness have been demonstrated as somewhat distinct concepts. Research has consistently shown that accented speech can be highly intelligible, but that a highly accented speech may affect the amount of effort required from the listener to understand what the speaker is saying, which is a third speech dimension called comprehensibility (Derwing; Munro, 2015). Despite this, pronunciation instruction still takes a secondary role in language teaching in comparison to grammar and semantics (Cruttenden, 2008; Silveira *et al.*, 2022).

With the aim of presenting a state-of-the-art overview of the neurosciences of accents, Moreno-Torres *et al.* (2016) compiled 13 articles on theoretical and experimental research on the topic. One of the studies included in their research explored the relationship between mastering musical skills and its influence on the imitation of foreign accents. This study revealed higher performance among musicians compared to non-musicians. Furthermore, within the group of musicians, vocalists demonstrated superior performance, suggesting that vocal and singing training can make L2 accents acquisition faster.

The aim of this study is to develop a layered model of speech accents that goes beyond the standard accents presented in textbooks through the practical application of phonetics and phonology in teaching and learning contexts. English will be established as the target L2 for this study. My motivation to conduct this study comes from the fact that I have a first degree in SLP and over the course of at least thirteen years, I have acquired an extensive background in vocal arts, including theatrical and singing practices, as well as instruction and research on popular singing. This background has significantly contributed to my L2 language learning process, especially in the realm of pronunciation. Recently, as a *Letras* undergraduate student, I have frequently observed a gap between pronunciation instruction and vocal arts. Additionally, most English textbooks focus on standard accents, ignoring or giving very little attention to non-standard varieties. Therefore, I decided to integrate knowledge from SLP and vocal arts with instruction in phonetics and phonology to develop a Layered Model for L2-accent learning. Differently from traditional textbooks, this Layered Model has non-standard accents as the ultimate goal in L2-accent learning.

2. REVIEW OF LITERATURE

The aim of this literature review is to provide a comprehensive overview of L2 pronunciation acquisition and the main issues discussed in the literature on this topic. Firstly, debates on the acquisition of first language (L1) and L2 will be presented, along with the definition of the concepts of foreign accent and standard accent. Subsequently, the main limitations regarding the concept of standard accents in L2 teaching will be discussed, as well as the advantages of going beyond the pronunciation models offered by textbooks.

2.1. Debates on first and second language acquisition

Several authors (Figueiredo, 1995; Ipek, 2009; Shormani, 2014) highlight that many similarities are observed in the acquisition of L1 and L2. Figueiredo (1995) investigated theories in L2 acquisition to present a panorama of the state-of-the-art of L2 acquisition process. His study revealed that both children and adults employ the same strategies during the acquisition of L1 and L2. He states that these strategies seem to be universal, and the only qualitative difference observed between children

and adults is in pronunciation. In the same vein, the author emphasizes that the influence of L1 in L2 must be considered, as individuals rely on the structure of L1 to produce L2, resulting in potential errors.

Shormani (2014) explored the nature of L1 and L2 acquisition by examining two controversial theories in language acquisition, named behaviorism and mentalism. He observed that L1 acquisition occurs when learners are too young for the abstract process of internalizing linguistic structures. In contrast, L2 learners, whether children or adults, encounter circumstances notably different from those acquiring their first language. According to him, several researchers emphasize that L2 learners are older, more cognitively mature, possess proficiency in at least one language, and likely have different motivations for acquiring L2 compared to learning their L1. Thus, the primary differences between L1 and L2 learners focus on "age and previous linguistic knowledge". He concludes that there are no differences between L1 and L2 acquisition, asserting that both lead to a linguistic system. This premise is rooted in the observation that specific L2 learners, notably adults, might not reach native-like competence due to fossilization, influenced by linguistic and nonlinguistic factors.

When introducing and defining fossilization for the first time, Selinker (1972) states that:

“Fossilizable linguistic phenomena are linguistic items, rules, and subsystems which speakers of a particular NL [native language] will tend to keep in their IL [interlanguage] relative to a particular TL [target language], no matter what the age of the learner or amount of explanation and instruction he receives in the TL” (p. 215)

In a paper attempting to help Chinese students to better understand fossilization, Wei (2008) points out that the concept was redefined on Selinker and Lamendella's (1978, *apud* Wei, 2008, p. 127). According to him, the authors revised the definition of fossilization to describe it as a permanent cessation of interlanguage learning. This occurs before the learner reaches the norms of the target language in all linguistic structures and discourse domains, despite the learner having the ability, opportunity, and motivation to learn and assimilate into the target society. Wei further explains that fossilization may manifest at the phonological level, mainly due to notable differences between languages. Phonological fossilization involves the repetition of

pronunciation errors in the L2, often influenced by the L1. This phenomenon occurs when errors persist and become stable in an incorrect manner. The author concludes that addressing and researching solutions for fossilization problems remains essential to enhance L2 teaching and learning.

Many teachers and L2-focused instructional materials still believe in repetition as the primary solution for L2 learning. Approaches like these have proven to be outdated as they overlook the generative properties of linguistic systems, focusing solely on operant conditioning of behaviorist approach (Figueiredo, 1995; Ipek, 2009; Shormani, 2014). In behaviorism, the acquisition of L2 involves addressing disparities between L1 and L2 systems, often accomplished through a comparative activity known as "contrastive analysis." However, studies comparing universal strategies in L1 and L2 acquisition demonstrate that most errors stem from the inherent development of L2 learning rather than L1 interference. Teachers frequently note that errors predicted by contrastive analysis do not manifest, and many errors are unpredicted. Similar to L1 acquisition, the behaviorist approach has been shown to be, at the very least, "an incomplete explanation of L2 acquisition" (Figueiredo, 1995).

As highlighted by Nolen-Hoeksema *et al.* (2018, p. 254), all languages have three properties. The first is the symbolic capacity for arbitrary representation, enabling humans to communicate about objects, actions, events, feelings, and ideas. The second property is structuring, which refers to the rules governing how symbols can be combined at all language levels. Finally, the third property is called generativity, which applies to all language levels (including speech sounds) and allows rules governing the language to be combined to create (generate) an infinite number of messages. Therefore, the emphasis on repetition found in instructional materials and in the classroom practices of L2 teachers is ineffective in instructing students at the generative level of L2, including generative capabilities related to pronunciation.

With the dual objectives of summarizing existing knowledge and suggesting future directions for teaching English pronunciation to L2 learners, Baker and Murphy (2011) conducted a large literature review on teachers' cognition in pronunciation instruction. They argue that the limited research on teachers' cognition in pronunciation reflects a broader neglect of pronunciation instruction in both teacher preparation programs and classroom-oriented research. Even when pronunciation instruction is included in graduate education, teachers often lack confidence in teaching various aspects of English pronunciation. The authors also observed that teachers find it

challenging to teach pronunciation aspects without referring to native-speaker standard modeling due to external pressures from governments, educational institutions, and partners who prefer native speaker models. They concluded that little progress has been made in the last decade regarding pronunciation teaching in language-teacher education, emphasizing the need for empirical research exploring the connections between teacher cognition, student perceptions, and student learning.

2.2. Defining foreign accents

The influence of L1 phonology may result in pronunciation difficulties and/or facilities in the realm of L2 pronunciation (Archibald, 2021). According to the American Speech-Language-Hearing Association (ASHA) (n.d.), “accents (regional, foreign, or nonnative) are not a communication disorder” but rather an inherent part of spoken language, and “every person has an accent.”. Similarly, foreign accents can be understood as differences in speech at the segmental and/or suprasegmental levels of pronunciation between native and foreign speakers of a target language (e.g., English) (Inzunza-Madrigal, 2021). According to Munro (1998, p. 139, *apud* Zárate-Sández, 2017, p. 228), a foreign accent is defined as “nonpathological speech produced by L2 learners that differs in partially systematic ways from the speech characteristic of native speakers of a given dialect.” Because of several reasons, including avoiding any relation between foreign accents and pathological speech, the concept of foreign accent has been through several changes along time. A recent definition provided by Derwing and Munro (2015) refers to foreign accent as “a particular pattern of pronunciation that is perceived to distinguish members of different speech communities. (p.5)”

Derwing and Munro (2015) state that several studies show that listeners attribute features to unseen speakers — such as lazy/hardworking, unfriendly/friendly, etc — solely based on the perceived accent. Kozłowski (2015) conducted a literature review with the aim of investigating why some accents are considered more attractive or prestigious than others. Her study revealed that foreign accents signal to listeners that the speaker is not a native and belongs to an external group, leading to assumptions about language fluency. Unlike native regional accents, foreign accents often face stigma and discrimination. Additionally, although impacting various aspects of an individual's life, the perception of foreign-accented speech as less truthful than

native speech is attributed to processing difficulties rather than prejudice against foreigners. In general, people tend to prefer accents that sound similar to their own, and the presence of a dissimilar accent can result in discrimination.

2.3. Defining standard accents

Most of the material available about professional accent intervention tends to focus on “reducing” the accent and literature about learning new accents for professional usage — like actors, voice actors, ventriloquists, etc — are almost nonexistent. In addition, as Derwing and Munro (2015) alert, there can be found many so-called “accent reductionists” entrepreneurs on the internet once there is no regulation of this business nor even is required a professional training in the area to do so.

According to Kozlowski (2015), the standard native accent of a particular country is frequently considered more desirable because of its familiarity, indication of ingroup membership, and prevalence among the educated upper class. Furthermore, she highlights the presence of an “accent hierarchy” in many countries. According to the author, accents provide listeners with inherent information about features such as nationality, ethnicity, social status, and group affiliation. Moreover, native regional accents play a significant role in distinguishing one's cultural background, surpassing the significance of someone's appearance. Additionally, individuals who speak with a standard local accent are often perceived as more competent and socially attractive in comparison to those with regional or foreign accents.

There seems to exist a consensus in the linguistic field that the standardization of accents has been an effort by the countries and grammarians to develop a general model of speech considered as ideal for speaking a target language correctly. However, Wolfram and Schilling-Estes (2006) demonstrate that, in the case of English, a standard accent is not standard everywhere and that many standard models exist within the English language. In the same direction, Cruttenden, (2008), states that, in the case of England, the pronunciation of the south-east London region was the first variety that was obtained more social prestige than the others four centuries ago, due to politics, commerce, and the presence of the Court. And, since that time, the notion that English has a socially preferred type of pronunciation has been present in the country. The exclusion of regional accents in some professions as well as the prestige

given to the standard ones may be related to the negative social perception of regionally-accented speech.

2.4. Limitations of standard accents

While the written form of English has largely been standardized for centuries, diversity in spoken language has always existed. Cruttenden (2008) notes that non-standard accents are the most prevalent in societies worldwide, whereas standard accents are commonly associated with television and radio communication. In the same vein, Wolfram and Schilling-Estes (2006) provide an overview of the limitations of standard accents:

“We have repeatedly noted that it is impossible to speak English without speaking some dialect of English. In those cases where dialect choices have to be made, the guiding principle calls for the selection of a form that will be least likely to call attention to itself for the majority of speakers outside of the area because of its dialect uniqueness.” (p. 314).

Bagno (2011) argues that what is commonly referred to as the standardization of a language, in both its written and spoken forms, actually represents the varieties associated with centers of economic power or those of high society. In accordance with this, Cruttenden (2008) cites Standard Australian English and Standard Scottish English as examples of these standard variations. This reinforces Bagno's (2011) assertion that standardization is rooted in ideology since a pure standard model is nonexistent in any social stratum, and variation is a natural occurrence in a living language.

However, in the realm of accent instruction, the majority of L2 methods focus on Standard Models, with General American (GA) and Received Pronunciation (RP) being the most common (Wolfram and Schilling-Estes, 2006). Cruttenden (2008) argues that standard accents were established as the target models for teachers in L2 learning programs due to implicit assumptions. Frequently, teachers are only familiar with RP and/or GA, assuming that learners will naturally acquire other variations by exposure to English and/or American TV and movies without additional instruction. Bagno (2001) adds that while non-standard accents may be present in some methods, the approach they receive tends to be picturesque, folkloric, or anecdotal in nature.

2.5. Relevance of gain knowledge about accents

In a study aiming to investigate the perspective on English accents among Indonesian learners of English, Harendita (2014) found that many learners still aspire to achieve a native-like accent, despite the emergence of World Englishes. She argues that the findings could be justified due to the wide exposure learners still have to native models in classrooms and proposes that additional research should explore ways to offer learners greater exposure to various English varieties. In a similar vein, Derwing and Munro (2015) present data from several studies showing that learners express greater interest in acquiring a native-like accent compared to their teachers. However, most people consider dedicating effort to achieve more than an intelligible pronunciation level to be a futile endeavor.

Perhaps the initial impression one has when asked about the benefits of gaining knowledge of different accents in L2 is an increase in comprehension and intelligibility. In an effort to assist English instructors to enhance their students' skills to understand various English accents, Johansen (2019) adopted a global approach to introduce several tools for improving learners' ability to understand unfamiliar accents. The author found that global approaches that highlight prosodic and intonational differences appear to be more effective than those focusing solely on phonemic segmental differences. Additionally, learners demonstrated improved understanding of unfamiliar English accents after exposure to various English varieties, despite only listening activities were employed as training in the study. The findings of Johansen further support Harendita's (2014) suggestion that researchers should explore a wider array of language varieties.

Among professionals, artists are likely the ones that mostly demand going beyond standard accents. According to Berthold (2006), in the Turkish Shadow Theater of *Karagöz*, a primary hero of this puppet art, foreign accents and dialects were commonly used. The author describes the experience of Pietro Della Valle, who declared, after watching a *Karagöz* theater piece, that "those who manipulate the puppets also make them speak, or rather, speak through them, keeping themselves hidden and imitating various languages with all kinds of jokes" (p. 28). Berthold also states that dialects became one of the most prominent characteristics of the *Commedia dell'arte* during the Baroque period (p. 353). As stated by Kozlowski (2015) and Levis

and Zhou (2018), one's accent reveals more inherent information about the speaker's sociocultural features than their appearance does. Therefore, it is reasonable to understand why actors demand the implementation of different accents in the character construction processes.

Learning non-standard accents can also enhance the cultural immersion experience in L2 when engaging with literary texts. The presence of regional or foreign accents is inherent in literary works such as *Lyrical Ballads* (1800) by William Wordsworth and Samuel Taylor Coleridge. In this work, the authors propose to explore common and everyday language to compose their poems. However, the most significant use of language varieties, perhaps, is evident in Mark Twain's *The Adventures of Huckleberry Finn*, published in 1884, where seven English varieties are employed throughout the narrative to highlight the sociocultural differences of the characters (O'Shea, 2018). According to Southard and Muller (1993), in the case of Twain's "*Huck Finn*," most teachers commonly overlook the dialects of the novel, typically focusing on aspects such as plot, characters, theme, and structure. Those who do examine the dialects often approach it from a literary perspective of language usage rather than a linguistic perspective. The authors point out that a language-centered approach to "*Huck Finn*" can benefit students in appreciating language varieties.

3. THE PRESENT STUDY

Having reviewed relevant literature in the field, I turn now to this section that focuses on providing the theoretical basis for the development of a Layered Model of speech. This is an exploratory and qualitative study by its nature which aims to develop a layered model of speech accents that goes beyond the standard accents presented in textbooks. English will be established as the target L2 for this study to demonstrate the practical application of phonetics and phonology in teaching and learning contexts.

To develop a Layered Model of speech production, an investigation into both the anthropological emergence of spoken language and the stages of L1 and L2 phonological acquisition will be conducted initially through the bibliographic method. This initial step will help in comprehending the emergency of phonetic and phonological structures of spoken language along the hominization process, as well as to identify

similarities and differences with L2 pronunciation features described in the existing literature.

A second step involves breaking down the phonetic and phonological features of accents into smaller, independent fragments — called here "layers", which will be used to create the layered model of speech accents throughout section 4. These layers can be practiced individually and later combined, enabling the user or instructor to achieve a level of L2 pronunciation performance that goes beyond standard models.

3.1. How the brain got language

In a study aiming to present a research roadmap in Comparative Neuroprimatology, Arbib *et al.* (2018) propose an enigma around the question of how the human brain acquired language:

“When people do speak, their speech is complemented by cospeech gestures of the hands as well as facial expressions. The puzzle is this: Nonhuman primates exhibit very little in the way of vocal control but do exhibit dexterous manual control. Why, then, did vocal control evolve as part of the human brain’s distinctive capabilities, since language could “manage without it”? And how relevant does manual action remain in understanding the brain mechanisms of language?” (Arbib *et al.*, 2018, p. 373).

Although this enigma still lacks an answer, the authors emphasize that, due to the importance of spoken language, it is also necessary to understand the evolution of human vocal control. There appears to be evidence pointing in the direction that spoken language would be somehow an anthropological unfolding of the human voice. Consequently, vocal control is, apparently, the initial stage of spoken language development. This idea can be reinforced by neuroscientific findings (Gazzaniga; Ivry; Mangun, 2014, p. 498) suggesting that after lexical identification, voice modulations, such as inflexions, are activated before the phonological ones. In other words, voice modulations could have been the first form of anthropological vocal manifestations for speech purposes.

3.2. Vocal control

Two studies on primate language learning experiments shed light on the significance of vocal control in language development. The first, conducted in 1931 by Professor Winthrop Kellogg and Luella, involved raising a chimpanzee named Gua alongside their infant son, Donald, for 9 months. Gua did not undergo any language teaching process. Despite Gua exhibiting advanced motor development in comparison to Donald, she did not engage in babbling or learn any words throughout the experiment. In the 1940s, psychologists Catherine and Keith Hayes conducted the second experiment, raising a baby chimpanzee named Viki as if she were their own child and attempting to teach her language. After 6 years of training, Viki showed significant understanding but struggled to produce only four words — *mama*, *papa*, *cup*, and *up*. Viki never progressed beyond these words, and for the /p/ sound, she had to use her fingers to hold her lips. The experiments with Gua and Viki collectively highlighted that chimpanzees lack the specialized articulatory and physiological abilities necessary for spoken language (Gleason and Ratner, 2016, p. 12).

On the other hand, through the utilization of x-ray videos to assess vocal tract dynamics in live macaques during vocalizations, facial displays, and feeding, Fitch *et al.* (2016) found out that macaques possess a speech-ready vocal tract, yet they lack a speech-ready brain, implying that the evolution of human speech was more dependent on neural adaptations than vocal tract anatomical changes. In addition, Aguasaco (2014) argues that, although vocal control plays an important role in language, the development of neocortical control of the laryngeal muscles, in order to communicate in its most primitive form, seems to be somewhat related to the development of poetry, singing and speaking itself.

Indeed, while mastering vocal mechanisms is fundamental, the evolution of spoken language surpasses mere articulatory predispositions and necessitates the development of further neurological abilities.

3.3. Acquiring articulatory gestures

Several authors (Rajabi; Hosseni; Hanifi, 2017; Shormani, 2014; Gleason and Ratner, 2016) consider babbling the first phonological manifestations, and infants across all languages produce similar babbling sounds. Preceding babbling, infants

undergo vocal development stages with distinct vocal behaviors. According to Gleason and Ratner (2016, p. 50), the first stage is reflexive vocalizations (birth to 2 months), where vowel-like sounds may be produced, with crying, fussing, coughing, burping, and sneezing being the predominant vocal behaviors of this stage. Cooining and laughter represents the second stage of vocal development (2 to 4 months). Velar sound seems to be produced in “cooining” and “gooining” during comfort-state vocal behaviors. The third stage is vocal play (4 to 6 months), during which infants test their vocal apparatus, producing loud and soft sounds (yells and whispers), high and low sounds (squeals and growls), lip drills, sustained vowels, and occasionally rudimentary consonant-plus-vowel syllables. Stage four is canonical babbling (6 months and older) when infants begin to produce sequences of consonant-vowel syllables with adult-like timing. In the initial phase of this stage, syllable sequences typically involve repetitive sounds (reduplicated babbles, such as [bababa]), evolving into diverse combinations of consonants and vowels (variegated babbles, like [bagidabu]) as the stage progresses. The final phase in vocal development is the jargon stage (10 months and older), often coinciding with the early phase of meaningful speech. This stage is characterized by sequences of sounds and syllables articulated with a diverse array of stress and intonational patterns. By the conclusion of the first year, most infants generate sounds that mirror the linguistic patterns of the language they have been exposed to.

According to Rajabi, Hosseni and Hanifi (2017), although some features may vary in time and organization, language acquisition occurs in several stages from birth (or even before that) until, at least, adolescence. Also, these stages are considered to be universal to every child, independent of the language that is being acquired. A classification presented in the literature organizes the acquisition stages as following: prelinguistic stage (from birth to about 6 months); babbling, first phonological manifestation (around 6-8 months); first words (around 10-12 months); two-words stage (around 20-24 months); content-word sentences or telegraphic speech (around 36-40 months); and complete acquisition (for further reading about this classification see Hamann, 2002, *apud* Rajabi; Hosseni; Hanifi, 2017, p. 429). The author states that babbling can be considered the forerunner of language development. In the same direction, Gleason and Ratner (2016, p. 52) state that the connection of babbling and the emergence of first words becomes particularly evident in the vocabularies of

numerous children learning English. The sound of words such as *daddy*, *mommy*, *baby*, and *hi* closely resemble the (nonmeaningful) babbling sounds like [mama] and [dada].

Phonological development can be examined through the lenses of vocal articulatory gestures. Brownman and Goldstein (1992) propose that articulatory phonology aims to explain the vocal gestures involved in the process of speech production. Vocal gestures differ from segments or features and are regarded as primitive phonological units. To illustrate, the [+nasal] feature of a phoneme corresponds to the act of raising the soft palate. Consequently, these vocal gestures can be understood as biomechanical movements of the vocal tract that, while not directly generating sound, provide the foundation for the production of speech sounds. Furthermore, Brownman and Goldstein (1992) and Laver (1980) state that the vocal tract, characterized as a "long-term muscular adjustment," contributes to the auditory "coloring" or voice quality of speech, which in turn influences the production of speech sounds. Both Brownman and Goldstein's (1992) and Laver's (1980) contributions highlight the relationship between articulation and the vocal tract, suggesting that vocal gestures (and subsequently, segments) of a language are produced following an initial adjustment of the vocal tract sizes, through which the speech sounds flow during utterance. Variations in vocal tract sizes are frequently observed among languages and significantly impact the articulation of speech.

3.4. Acquiring segments

Gleason and Ratner (2016, p. 51) emphasize that longitudinal studies on prelinguistic vocalizations suggest that children's early words often incorporate the same sounds and sound sequences as those preferred during babbling. This is because children can recognize that these words align with the vocal motor schemes that they have successfully brought under voluntary control. Additionally, the authors state that "the sounds and syllables of babble form the building blocks for word productions in young children" (p. 52).

With the aim to identify characteristics of studies on consonant acquisition and to describe general principles of consonant acquisition, McLeod and Crowe (2018) conducted a systematic cross-linguistic review of literature comparing children's consonant acquisition in 27 languages around the world. According to them, three methodological approaches have been adopted to describe general principles of

consonant acquisition: (1) consonants within one language; (2) features across languages; and (3) consonants across languages. Considering place of articulation, bilabials, labiodentals, palatals, velars, uvulars, pharyngeals, epiglottals and glottals ones are acquired earlier than dentals, alveolars, post-alveolars and retroflexes. In other words, anterior lingual placement consonants are acquired later. Considering manner of articulation, plosives, nasals and non-pulmonic (e.g., clicks) are acquired earlier than trills, flaps, fricatives and affricates. Also, nasal consonants are the earliest acquired ones, and plosives are acquired earlier than fricatives. They concluded that, although individual variations should be considered, at the age of five children in the 27 languages reviewed in their study have already acquired most consonants of their first language.

In the case of English, Girona and Fabus (2011) state that vowels are acquired earlier than consonants. By the age of two, the vowel /ʌ/ is mastered and until the age of three the vowels /i/, /u/, /o/, /ɛ/, /ɑ/ and the diphthongs /aɪ/ /aʊ/ and /ɔɪ/ are mastered. By the age of three many other vowels can be observed in the child's speech; however, complete mastery occurs at different ages. The vowel /ɪ/ is mastered by the age of four, /æ/ and /e/ by the age of five, and /ə/, /ɜ/ and /ə/ by the age of six. The later mastery can be related to a higher articulatory refinement required for the complete acquisition of vowels such as /ə/, /ɜ/ and /ə/. This refinement might be related even to the amount of gestural articulatory training demanded as well as to nervous system maturation.

After reviewing 15 studies describing English consonant acquisition order for children, McLeod and Crowe (2018) propose the following classification: early consonants (by the age 2;0-3;11) are /p/, /b/, /m/, /t/, /d/, /n/, /k/, /g/, /ŋ/, /w/, /j/, /f/ and /h/; middle consonants (by the age 4;0-4;11) are /v/, /s/, /z/, /ʃ/, /tʃ/, /dʒ/, and /l/; and late consonants (by the age of 5;0-6;11) are /ʒ/, /ɹ/, /ð/ and /θ/. It is important to highlight that /ð/ and /θ/ are the last two consonants to be acquired, being /θ/ the latest one. The authors also found that some English consonants apparently follow a universal order for consonantal acquisition. They explain that nasals (/m/, /n/ and /ŋ/), plosives (/p/, /b/, /t/, /d/, /k/ and /g/) and laterals (/w/, /l/ and /j/) are acquired earlier than most fricatives (/f/, /v/, /s/, /z/, /ʃ/, /ʒ/, /h/, and the latest two sounds /ð/, /θ/) and affricates (/tʃ/ and /dʒ/). Moreover, sounds produced by lips (bilabials and labiodentals → /m/, /p/, /b/, /f/, /v/ and /w/), pharynx (pharyngeal, epiglottal and glottal → /h/) and the back of the tongue (palatal, velar and uvular → /ŋ/, /k/ and /g/) were acquired earlier than consonants

produced with anterior tongue placement ((inter)dentals, alveolars, postalveolars and retroflexes → /n/, /t/, /d/, /l/, /s/, /z/, /ʃ/, /ʒ/, /tʃ/, /dʒ/, /ð/ and /θ/), but an overall interaction between place and manner was observed.

3.5. Learning syllable patterns and phonotactics

While the age of acquisition of speech sounds is related to the emergence of first words, the words of a language are not formed by random arrangements of speech sounds. Therefore, there are phonotactic constraints that govern the possible sequences of sounds to form words in every language system (Gleason and Ratner, 2016). As a general rule for all languages, vowels are the main components of a syllable. Because of this, vowels are considered the **nucleus** of a syllable. Consonants, on the other hand, can be allocated to the margins of the nucleus. When consonants occur before the nucleus of a syllable, they are in the syllabic **onset** position, and when they occur after the nucleus, they are in the syllabic **coda** position. The number of consonants that can be allocated in the onset or coda position varies according to the language (Carlisle, 2001). Gleason and Ratner (2016) state that the majority of phonotactic constraints are related to consonant clusters and their possible positions in words. According to them, for both L1 and L2 “mastering a new cluster or a new word position for a familiar sound may require as much work as mastering a new sound” (p. 49).

According to Yavas (2011), in the case of English, the nucleus — represented as V (i.e. vowel) — can be composed by single vowels or diphthongs. The onset position allows a maximum sequence of three consonants — represented as C (i.e. consonant) — and a minimum of zero. The coda position allows a maximum sequence of four consonants and a minimum of zero. The possible syllabic structures of English can be observed below (Figure 1):

V	—	I
CV	—	he
CCV	—	play
CCCV	—	spray
VC	—	at
VCC	—	ask
VCCC	—	asks
CVC	—	cat
CCVC	—	break
CCCV	—	strike
CVCC	—	pact
CVCCC	—	text
CVCCCC	—	worlds
CCVCC	—	breathed
CCCVCC	—	strength
CCVCCC	—	trumped
CCCVCCC	—	strengths

Figure 1 - English syllabic structure possibilities. The nucleus is represented in red and the onset and coda position in black. (based on Yavas, 2011, pp. 139-40).

Although the same syllabic structures can occur in two different languages, such as English and Portuguese, the consonant sequences in the onset and coda positions may differ. For instance, the arrangement CCVC exists in both languages; however, the cluster /st/, as found in the word *stop* (Eng), is not permitted at the onset position in Portuguese. Thus, English and Portuguese are governed by distinct phonotactic constraints, and the frequency of certain clusters, as well as their age of acquisition, can vary.

Clemens (1990, *apud* Carlisle, 2001, p. 4) proposed a principle known as the Sonority Sequencing Principle (SSP) to identify universal features in the arrangement of consonant clusters in languages. According to Carlisle (2001), this principle outlines a strong universal tendency in organizing sounds within the syllabic structure. Across all languages, the preferred syllable type is one where the nucleus is composed of vowels (or diphthongs) due to their flowing sonorant sounds. On the other hand, consonants play a role in either increasing or decreasing the sonority quality of the nucleus. As a result, consonants in the onset gradually increase in sonority quality from the farthest margin until they reach the nucleus. Consonants in the coda exhibit the opposite pattern, with a gradual decrease in sonority quality from the closest margin to

the presence or absence of stimuli for a target cluster — for example, /st/ is not a cluster in Brazilian Portuguese, but it is a cluster in English. The presence or absence, as well as the frequency of stimuli of a given cluster among languages, may influence the period of mastery of that consonant cluster as well.

As exposed by Templin (1957 *apud* Gironda and Fabus, 2011, p. 147), considering the acquisition of L1 English, at the age of four the clusters /bl/, /pl/, /gl/, /kl/, /br/, /pr/, /kr/, /dr/, /tr/, /sm/, /sp/, /st/, /sn/, /kw/ and /tw/ are acquired on the syllabic onset position and the clusters /mp/, /pt/, /ks/, /ft/, /lp/, /lt/, /mpt/, and /mps/ are acquired on coda position. At the age of five the clusters /gr/, /fl/, /fr/ and /str/ are acquired on the onset position and /lb/, /lf/, /rd/, /rf/ and /rn/ on the coda position. At the age of six, /skw/ is acquired on the onset position and /nd/, /nt/, /lk/, /lf/, /rb/, /rg/, /rst/, /nθ/ and /rθ/ on the coda position. At the age of seven, when children are expected to complete the acquisition of the last speech sounds of English, three-consonant clusters become more frequent in the acquisition. On the onset position the clusters /sw/, /sl/, /spl/, /spr/, /skr/, /jr/ and /θr/ and on the coda position /lz/, /sk/, /st/ and /lθ/ are acquired. The latest clusters are /sp/ and /kt/ on the coda position, acquired at the age of eight.

In a review of literature encompassing studies from 1930-2000, McLeod, Van Doorn and Reed (2001) investigated data about children's normal acquisition of consonant clusters from English encompassing supplementary examples of other languages. Their study revealed that, by the age of two, children demonstrate the ability to produce consonant clusters. However, the clusters produced at this age may not correspond with those observed in the surrounding language. The authors highlight that throughout the acquisition process, it is common for consonant clusters to undergo reduction and simplification. Notably, two-element consonant clusters tend to be produced and mastered earlier than those with three-element, and consonant clusters containing stops are generally acquired before those containing fricatives. Additionally, word-final consonant clusters typically appear in inventories earlier than word-initial clusters, possibly due to the emergence of grammatical morphemes, such as plural and past tense. These morphemes result in the creation of morphophonological consonant clusters. Despite following a typical developmental sequence, it is expected that reversals and revisions will occur due to individual variations.

3.6. Phonological processes in first language acquisition

More than pronouncing individual sounds, syllables, words and sentences, speech also involves adjustments in pronunciation within and between words caused by the coarticulation of speech sounds within the utterance. These adjustments are frequently called **Adjustments in Connected Speech** by pronunciation textbooks authors (Celce-Murcia; Brinton; Goodwin, 1996; Collins; Mees; Carley, 2019), but are also called **Phonological Processes** (Gironda and Fabus, 2011) in the field of SLP; however, the latter also includes disorder patterns observed in the phonological system. For the purposes of this study, the term phonological processes will be adopted for both normal and pathological contexts in order to provide a broader overview of adjustments done in connected speech.

Different from isolated sounds, when speech sounds are combined to form words and sentences in connected speech, they are adjusted to promote the regularity of rhythm (Celce-Murcia; Brinton; Goodwin, 1996; Collins; Mees; Carley, 2019). According to Gironda and Fabus (2011), phonological processes are classified in three groups: syllable structure processes, substitution processes and assimilation processes. Syllable structure processes occur when a reduction, a deletion or an expansion of one or more sounds affect the syllable structure. Substitution processes occur when one sound replaces the other, often resulting in changes in place or manner of articulation. Assimilation processes occur when two sounds become more similar in some features, and they can occur in a regressive (e.g. *doggy* = “*goggy*”) or progressive (e.g. *doggy* = “*doddy*”) manner (Gironda and Fabus, 2011). A wider overview of phonological processes can be observed in Table 1.

Table 1 - Phonological Processes expected according to group and age of suppression.

SUBSTITUTION:			
Alveolarization	substitution of an alveolar sound for a linguadental or labial sound	pan = tan	6;0
Deaffrication	an affricate manner changed to a fricative	shoe = chew /ʃu/ = /tʃu/	4;0
Depalatalization	substitution of an alveolar fricative or affricate for a palatal fricative or affricate	fish = fis /fɪʃ/ = /fis/	4;0
Derhotacization	omission of the r-coloring for the consonant /r/, and for the central vowels with r-coloring	zipper = zipp	4;0
Gliding	substitution of a glide for a liquid	run = wun	5;0-7;0
Labialization	substitution of a labial sound for an alveolar sound	dog = bog	6;0
Liquid simplification	substitution of another sound for a liquid	lake = take	5;0
Stopping	substitution of a stop consonant for a fricative or affricate	catch = cat /catʃ/ = /cat/	3;0-5;0
Stridency deletion	omission or substitution of another sound for a fricative	soap = oap	6;0
Velar fronting	substitution of sounds in the front of the mouth, usually alveolar sounds, for velar or palatal sounds	candy = tandy	3;6
Vocalization	substitution of a vowel for a final-position liquid sound	people = peopo	7;0
ASSIMILATION:			
Alveolar assimilation	alveolar sound influences a nearby sound	/lelo/ = yellow	3;6
Final devoicing	alteration in voicing affected by a nearby sound	bake = bag	3;0
Labial assimilation	labial sound influences a nearby sound	/pebo/ = table	3;6
Nasal assimilation	nasal sound influences a nearby sound	/nun/ = /spun/	3;6
Prevocalic voicing	voicing of an initial voiceless consonant in a word	kup = gup	3;0
Velar assimilation	velar sound influences a nearby sound	/gogi/ = doggie	3;6

(source: Girona and Fabus, 2011, pp. 150-1)

According to Celce-Murcia, Brinton and Goodwin (1996), in the case of English, epenthesis will be used to split consonant clusters in regular plural and past tenses. For example, the word *place* in its plural form, *places*, has a schwa sound added after the /s/ to break the cluster of sibilants sounds. The same happens with the word *plant* in its past form, *planted*, to split the alveolar cluster of the stop sounds /t/ and /d/. By the same token, assimilation may happen in three forms: regressive, progressive and coalescent. In regressive assimilation the continuing sound is affected by its preceding one. For example, the /v/ in *have to* (/hæv/ + /tuw/) is pronounced as /f/ because it is followed by an unvoiced /t/ → /hæftə/. In progressive assimilation the preceding sound affects the continuing one. For example, the <s> in the word *backs* is pronounced as

/s/ while in the word *bags* it is pronounced as /z/ because of the voice assimilation of the preceding /g/. In coalescent assimilation two sounds are combined to create a third sound that shares the features of both combined sounds. For example, the sounds /s/ and /j/ in the sentence *this year* (/ðɪs/ + /jɪr/) are combined to form the sound /ʃ/ → /ðɪʃɪr/. On the other hand, the deletion process — also known as omission, ellipsis or elision — occurs when a vowel or a consonant sound disappears in certain contexts. For example, when /t/ or /d/ are followed by a consonant in word boundaries, such as in the sentence *east side*.

This study does not aim to describe all existing phonological processes, but rather to present them as an important aspect of spoken languages that must be considered when striving to go beyond standard accents. Therefore, addressing phonological processes for pedagogical purposes may offer valuable insights for both teaching and self-study in the realm of L2 pronunciation improvement and learning about different L2 accents.

3.7. Phonological processes in second language acquisition

It is not the intention of this study to approach foreign accents as defective speech as they were interpreted by SLPs in the first half of the 20th century (Mota and Wiethan, 2014). Therefore, foreign accents here are understood as speech differences instead of speech pathologies. However, foreign accents may present some phonological processes that are similar to those observed in speech pathologies as Articulation Disorders and Phonological Disorders.

According to Gironda and Fabus (2011), difficulties or impairments in producing vocal gestures are observed in Articulation Disorders. Similarly, challenges in understanding the phonological rules of a language are present in Phonological Disorders. These difficulties can also be observed in the speech of L2 learners. It is essential to note, however, that in the field of SLP, phonological disorders result from body and health issues that impact language acquisition. In contrast, in the field of L2 learning, the origins are related to previous language experiences and comparisons with L1 rules. In other words, they differ in etiology.

Ceron and Keske-Soares (2009) emphasize that the primary goal of SLPs in phonological therapy is to reorganize the patient's phonological system. To achieve this, SLPs aim to establish phonological patterns that facilitate generalization.

Generalization is characterized by the extension of using target sounds in contexts or environments not initially part of the training, which can be examined from two perspectives: the functional perspective, which explores how individuals reorganize their sound systems, and the structural perspective, which involves identifying the specific situations where this phenomenon occurs.

As previously mentioned, although they vary in etiology, L2 acquisition problems share some similarities with speech disorders. Table 2 presents common phonological processes Brazilian students of English as L2 produce in speech:

Table 2 - Common Phonological Processes observed in Brazilian learners of English as L2.

Processes	Examples	Processes	Examples
1) syllable simplification	[ɪst] for [st] ex.: <i>start</i> → [ˈɪstɑrt] [pi] for [p] ex.: <i>tape</i> → [ˈtʰeɪpi]	5) delateralization and rounding of lateral liquids in final position	[w] or [ʊ] for [ɫ] or [l] ex.: <i>feel</i> → [fiʊ]
2) consonant change (substitutions)	[h], [x] for [r] ex.: <i>ripe</i> → [haɪp] [t], [s], or [f] for [θ] ex.: <i>think</i> → [tɪŋk] [d], [z] or [v] for [ð] - ex.: <i>this</i> → [dɪs]	6) vocalization of final nasals	[ỹ] or [w̃] for [m] or [n] ex.: <i>team</i> → [tiỹ]; <i>moon</i> → [mūw̃]
		7) velar consonantal paragoge	[ŋg] for [ŋ] ex.: <i>sing</i> → [sɪŋg]
3) deaspiration of voiceless plosives in initial or stressed position	[t] for [tʰ] ex.: <i>tea</i> → [ti]; <i>attend</i> → [əˈtend]	8) vowel assimilation	[ʌ] for [ʊ] ex.: <i>put</i> → [pʰʌt] [ɛ] for [æ] ex.: <i>bad</i> → [bɛd]
4) terminal devoicing in word-final obstruents	[s] for [z] ex.: <i>does</i> → [dʌs] [k] for [g] ex.: <i>dog</i> → [dɔk]		9) interconsonantal epenthesis (-ed morpheme)

(source: adapted from Zimmer; Silveira; Alves, 2009, pp. 20-1)

By comparing the similarities of phonological processes, L2 pronunciation instructors can gain a broader understanding of the difficulties their students are facing. Subsequently, they can develop more precise instructions for each specific case. For instance, in pathological contexts, phonological processes indicate incomplete or problematic acquisition of phonetics, phonology, and phonotactic rules. However, in the L2 context, the same phonological processes might indicate interferences of L1 in the L2 phonological system. Therefore, the phenomenon itself reveals what is happening in the phonological structure of the target language.

Zimmer, Silveira, and Alves (2009) emphasize that the influence of L1 grapho-phonetic-phonological elements can result in phonological processes that are not typical in the target L2. In the context of Brazilian students learning English as L2, the authors observe that the relationship between graphemes and their phonetic representation is clearer in Portuguese than in English, owing to historical aspects of English orthography. The authors state that Brazilian learners often apply the same phonemes they would use when speaking or reading in their L1, influenced by their well-established knowledge of the alphabetic system in their native language. An illustrative example of this phenomenon is the final epenthesis in words such as *take*, which might be pronounced as [tejkɪ] by Brazilian learners of English L2. The authors conclude that deviant phonetic production in the L2 can originate from two distinct sources: phonetic-phonological transfer between the L1 and L2, and grapho-phonetic-phonological transfer between the two languages.

4. GOING BEYOND STANDARD ACCENTS: A LAYERED MODEL OF SPEECH PRODUCTION

After presenting the main features of the anthropological emergence of spoken language, the stages of L1 phonological acquisition, as well as the similarities and differences of both processes with L2 pronunciation features described in the literature, this session will focus on the development of a Layered Model of speech accents based on the findings of the investigation conducted in the previous session.

The primary purpose of the Layered Model is to guide oral production training in L2. Since L2 learners typically have at least one previous language, which is their mother tongue, accents were categorized into two main groups for the purposes of this study. The first group is called the **target accent(s)**, representing the pronunciation goal(s) the student aims to achieve in L2. This group encompasses both native (standard and regional) and nonnative (foreign) varieties of the target L2. The second group is called the **source accent(s)**, representing the native accent(s) of the student's L1. Thus, the Layered Model aims to guide the learner from their source accent(s) to the target accent(s) of L2.

While linguistic varieties encompass lexico-syntactic and phonetic differences (O'Shea, 2018), accents are associated with the diverse ways of pronouncing a language (Levis and Zhou, 2018). Consequently, moving beyond the standard accents

found in textbooks for L2 requires a focus on the various ways a language can sound. In essence, it involves uncovering the features that contribute to the aesthetic properties of accents. These properties will now be referred to as the **aesthetic identity** of an accent. The aesthetic identity of an accent represents the sonorous image of a speaker. It is due to this sonorous image that a person becomes associated with one cultural group or another (Kozlowski, 2015). In the realm of performing arts, such as theater, artists might use different accents to bring their characters closer to a cultural group, for example. The elements that compose the aesthetic identity of accents are discussed below.

4.1. Components of an accent aesthetic identity

As discussed in previous sessions of this study, language acquisition occurs in numerous stages. The aesthetic identity of an accent is acquired simultaneously during L1 acquisition. This acquisition is far from being a random process. For instance, a child learning English as its L1 in the middle of Texas will not speak English with a Cockney accent, even though both are native accents of English. The acquired aesthetic identity will be the closest variety available in the environment or the one taught to the child; in this case, probably some variety of a Texan accent. Therefore, identifying the features that compose the aesthetic identity of accents is an intriguing task. Figure 3 illustrates our attempt to identify these features.

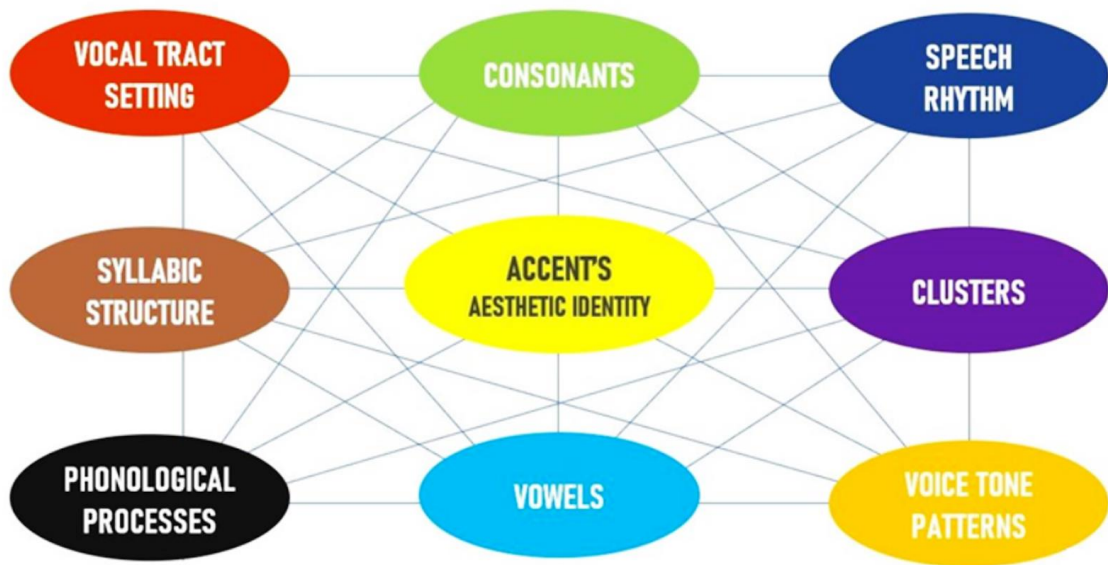


Figure 3 - The composition of an accent aesthetic identity (developed by the authors).

According to Figure 3, the aesthetic identity of an accent emerges by the interaction of multiple linguistic elements. It is important to note that the aesthetic identity is not necessarily connected to the intelligibility potential of an accent. Both foreign and native accents might have a strong aesthetic identity and be unintelligible at the same time. By the same token, some non-native accents may be more intelligible than some regional native accents, which means that intelligibility is neither mandatory nor exclusive of native accents. Therefore, aesthetic training and intelligibility training are two different realms of pronunciation teaching.

Although they share numerous similarities and one might complement the other, aesthetic training is intended to explore the various sonorous features of specific accents during speech production. In contrast, intelligibility training is more likely to focus on assisting L2 learners in achieving a clear pronunciation, enabling successful communication in the L2 (Silveira, 2019; Levis, 2020). The next section demonstrates how the stages of acquisition of a target language can be adapted into training stages for application in the Layered Model.

4.2. Training-stages on the Layered Model

Establishing English as the target L2 for the purposes of this study, the stages of language acquisition described in the literature were reorganized to be applied to the Layered Model. Four steps were followed in this reorganization. First, speech sounds were categorized according to the age of acquisition in L1, resulting in three main training-stages: early, intermediate, and later training-stages. Second, in order to provide a panorama of the evolution of vocal articulatory gestures, the speech sounds of each training-stage were divided into two subgroups. These subgroups were organized based on the order of emergence for vocal articulatory gestures described in the literature. For instance, according to McLeod and Crowe (2018), bilabials and velars are acquired earlier than dentals (placement), and plosives and nasals are acquired earlier than fricatives (manner). Therefore, although /m/ and /n/ are both nasals, /n/ is acquired later than /m/ because it requires anterior lingual placement. Additionally, bilabials such as /p/, /b/, /w/, and velars such as /k/ and /g/ are also acquired earlier than /n/, as these sounds are acquired before anterior lingual placement consonants. The subgroups thus delineate early and later vocal articulatory gestures within each training-stage.

The second step was also applied to the acquisition of vowels described in Girona and Fabus' (2011); however, a different classification was needed for these sounds. Despite English having more consonants than vowels, the acquisition of the latter is commonly extended through several stages. For example, the vowel /ə/ can be observed in the earlier stages of acquisition, but mastery occurs in later stages. Due to this specific aspect, the training-stages of this reorganization encompass two categories of sounds: target segments to be mastered during the training-stage and cross-group segments that demand longer practicing at or through several groups or training-stages. These two categories were included in the subgroups of the training-stages. Consequently, one might expect longer training processes for some vowels than for consonants in the Layered Model, consistent with the apparent pattern in L1 normal acquisition as well.

After organizing consonants and vowels in the previous two steps, the third step involved organizing consonant clusters in the training-stages. Initially, clusters were arranged based on the age of acquisition as described in the literature. A separate session for clusters was created in each training-stage, distinct from the sessions for consonants and vowels. Within the cluster session, it was specified whether the cluster was in the onset or coda syllabic position. Following this initial organization, a second

division was made, considering the SSP. Within each training-stage, clusters were categorized into two subgroups. The first subgroup referred to clusters arranged according to the SSP (A-SSP), while the second subgroup referred to clusters that violated the SSP (V-SSP). This division was implemented because, being a universal tendency, A-SSP clusters are generally considered easier to acquire than V-SSP clusters, which likely require a more extended vocal articulatory gestural training. For example, Brazilian learners of English as an L2 often simplify syllables in V-SSP clusters, such as /st/ in the onset position of the word *start*, which tends to be pronounced as [ˈɪstart] as described by Zimmer, Silveira, and Alves (2009, pp. 20-1). Consequently, both types of clusters are placed in different layers in the Layered Model, with A-SSP on the 5th layer and V-SSP on the 6th layer.

The fourth and final step focused on incorporating phonological processes within the training-stages. The approach followed mirrored that used for consonants and vowels. Initially, phonological processes were organized based on the age of acquisition as described in the literature. A session for phonological processes was introduced in each subgroup of the training-stages then. Secondly, given that the acquisition of phonological processes often extends across multiple stages, similar to vowel acquisition, these processes were also categorized into two groups: target phonological processes to be mastered during the training-stage and target phonological processes to start or continue practicing at or through the training-stages. To simplify the reorganization, the second category for phonological processes was incorporated within the existing category created for vowels. Thus, the category for target sounds to start or continue practicing through the training-stages also encompasses target phonological processes.

The resulting table consists of three training-stages, each divided into two minor groups, totaling six groups altogether. These groups contain specific subdivisions to be trained, categorized as "segments," "A-SSP clusters," "V-SSP clusters," and "phonological processes." It is worth noting that these groups are organized progressively. Therefore, in each training-stage, the second group necessitates training from the first group before proceeding. This sequential progression is consistent across the training-stages. Additionally, certain groups have a supplementary category called "cross-group elements." These elements consist of particular segments, clusters, and/or phonological processes that demand initial training within a specific group, but their mastery extends to further groups or training-

stages. For instance, the vowel /ə/ is introduced in the second group of the Early Training-Stage, but its mastery will occur only in the second group of the Late Training-Stage according to the table. Therefore, cross-group elements represent segments, clusters, and/or phonological processes that necessitate overlapping extended learning across subsequent groups and/or training stages due to the extended longitudinal practice they require from the learner.

After completing the four steps to reorganize the stages of language acquisition for application in the Layered Model, the final result can be observed in Table 3:

Table 3 - Adaptation of the stages of English L1 acquisition into training stages for English L2 in the Layered Model.

Early Training-stage		
Training Order	Target	Segments / Clusters / Phonological Processes
1st Group	Segments	Vowels → /ɪ/ - /i/ - /u/ - /o/ Consonants → /m/ - /ŋ/ - /p/ - /b/ - /h/ - /k/ - /g/ - /w/ - /j/
	Clusters	—
	Phonological Processes	Prevocalic voicing – Final devoicing – Stopping
2nd Group	Segments	Vowels → /e/ - /a/ - /aɪ/ - /aʊ/ - /ɔɪ/ Consonants → /n/ - /t/ - /d/ - /f/
	A-SSP Clusters	/kw/ - /tw/
	Phonological Processes	Alveolar assimilation – Labial assimilation – Nasal assimilation – Velar assimilation – Velar fronting
	Cross-group Elements	Vowels → /ɪ/ - /æ/ - /e/ - /ə/ - /ɜ/ - /ə/ Stopping
Intermediate Training-stage		
Training Order	Target	Segments / Clusters / Phonological Processes
1st Group	Segments	Vowels → /ɪ/ Consonants → /l/ - /v/ - /s/ - /z/ - /tʃ/
	A-SSP Clusters	Onset → /bl/ - /pl/ - /gl/ - /kl/ - /kw/ - /tw/ Coda → /mp/
	V-SSP Clusters	Coda → /mps/ - /mpt/ - /pt/ - /ks/
	Phonological Processes	Deaffrication – Depalatalization – Derhotacization
	Cross-group Elements	Vowels → /æ/ - /e/ - /ə/ - /ɜ/ - /ə/ Stopping
2nd Group	Segments	Vowels → /æ/ - /e/ Consonants → /ʃ/ - /dʒ/
	A-SSP Clusters	Onset → /br/ - /pr/ - /kr/ - /dr/ - /tr/ Coda → /ft/ - /lp/ - /lt/
	V-SSP Clusters	Onset → /sm/ - /sp/ - /sk/ - /sn/ - /st/
	Phonological Processes	—
	Cross-group Elements	Vowels → /ə/ - /ɜ/ - /ə/ Stopping – Deaffrication – Depalatalization – Derhotacization

Late Training-stage		
Training Order	Target	Segments / Clusters / Phonological Processes
1st Group	Segments	Consonants → /ʒ/ - /ʁ/
	A-SSP Clusters	Onset → /gr/ - /fl/ - /fr/ Coda → /lb/ - /lf/ - /rn/ - /rd/ - /rf/
	V-SSP Clusters	Onset → /str/
	Phonological Processes	Liquid simplification – Stopping – Gliding
	Cross-group Elements	Vowels → /ə/ - /ɜ/ - /ɔ/
2nd Group	Segments	Vowels → /ə/ - /ɜ/ - /ɔ/ Consonants → /ð/ - /θ/
	A-SSP Clusters	Coda → /nd/ - /nt/ - /lk/ - /rb/ - /rg/ - /rst/ - /nth/ - /rth/
	V-SSP Clusters	Onset → /skw/
	Phonological Processes	Alveolarization – Labialization – Stridency deletion – Gliding – Vocalization
	Cross-group Elements	Gliding

(source: developed by the authors)

The Layered Model is designed to guide the pronunciation instruction process through feasible components, highlighting the elements that should be emphasized with each student in order to go beyond standard accents found in textbooks. Table 3 is intended to offer a practical approach to improvement and assessment of the learner's progression within the Layered Model. Due to the close relationship between aesthetic training and intelligibility training, this Layered Model is not exclusively oriented to aesthetic training and may be applied for increasing intelligibility too. The following section describes the skills each layer represents and how to use them in accent learning.

4.3. The Layered Model of pronunciation for L2-accent learning

The Layered Model of pronunciation consists of nine different layers that interact with each other as they are progressively learned. The details of each layer are explained below.

4.3.1. 1st layer: vocal tract setting

This layer refers to the gestural adjustment of the vocal tract mentioned in Brownman and Goldstein's (1992) and Laver's (1980). As it is gestural, this layer produces no acoustic signal. An example of how this layer works can be observed in the following scale:

yawning ↓ **flattening**

This scale represents the contrast between two vocal tract settings: yawn and flat. As the cursor moves to the left, the vocal tract setting becomes more yawning-like, and as it moves to the right, the setting becomes more flattening-like. When comparing Brazilian Portuguese and American English, the vocal tract setting in the former tends towards the yawning side, while in the latter, it tends towards the flattening side. This might be one of the reasons why English sounds more nasal to Brazilians than Portuguese, even though Portuguese has more nasal sounds, especially vowels, than English.

Brazilian Portuguese

yawning ↓ **flattening**

American English

yawning ↓ **flattening**

Another intriguing aspect of this layer is its impact on vowel production. Comparing both languages once again, when the vocal tract setting is reduced to fit the pattern of English language, the total intraoral space available becomes smaller. As the roof of the mouth — the upper teeth, hard palate and soft palate — are fixed, they cannot move downwards to reduce the intraoral space. Therefore, the “Lingual Setting”, as described by Laver (1980, p. 43), takes place to reduce the intraoral space, causing the vowel triangle (the vowel chart) to squeeze upwards. This process results in different vowel possibilities for English, such as /ɪ/ and /ʊ/, when compared to Brazilian Portuguese, for example. Additionally, it justifies the need of a bigger jaw dropping to pronounce lower-chart vowels of English, such as /æ/. Consequently, vocal tract setting significantly impacts vowel production.

4.3.2. 2nd layer: voice tone patterns

This layer refers to the tone of the voice itself. While this layer may initially seem relevant only for tonal languages, vocal inflections of languages such as stress and intonation patterns are also part of this layer. An interaction between the 1st and 2nd layers is implicated, as the vocal tract and the larynx interact in voice production. Thus,

emotions such as disgust and surprise demand both layers. Furthermore, other vocal fold adjustments, such as breathiness and creaky voice, are included in this layer. Consequently, this layer produces acoustic signals.

4.3.3. 3rd layer: vocal gestural or biomechanical articulation

This layer refers to the primitive vocal units of articulatory phonology as described in Brownman and Goldstein's (1992). In the same direction of the 1st layer, this layer produces no acoustic signal. The biomechanical control of speech articulators is the main content of this layer. The main focus relies on articulatory coordination.

4.3.4. 4th layer: speech-sounds production (scat singing and simple words)

While it may appear very close to the 3rd layer, the 4th layer involves coordination among the 1st, 2nd, and 3rd layers to combine phonation and articulation, resulting in speech sounds production. Although the place of articulation may be encompassed within the 3rd layer, it is in the 4th layer that both manner of articulation and voicing take place in the speech sounds. In this layer, both vowels and consonants are produced, enabling the formation of simple syllabic combinations without consonant clusters. Speech sounds can be trained in two contexts within this layer: through scat-singing-like combinations and by using words without consonant clusters.

Scat-singing-like combinations in this layer are influenced by a singing style known as scat singing, defined by Stoloff (1999, pg. 6) as 'the vocalization of sounds and syllables that are musical but have no literal translation.' The intention behind scat-singing-like combinations is to simulate the prelinguistic and babbling stages of language acquisition. Therefore, speech sounds are combined to form syllables, but no meaning is produced. Examples such as [pataka] and [vazađaʒa] illustrate scat-singing-like combinations in this layer. It's important to note that, unlike scat singing, which may use clusters, no clusters are admitted in this layer. Scat-singing-like combinations primarily aim to practice speech sounds that might not be present in the learner's L1, making them potentially more challenging to acquire.

Similar to scat-singing-like combinations, words without consonant clusters can also be utilized to train the skills of this layer. However, unlike scat-singing-like

combinations, words without consonant clusters are expected to have meaning. The intention behind words without consonant clusters is to simulate the first words stage of language acquisition. Examples of words suitable for practice in this layer include *my*, *that*, *divine*, *television*, and *communicate*. Scat-singing-like combinations might progress to words without a consonant cluster when a target speech sound demands additional training.

4.3.5. 5th layer: combining speech-sounds according to the sonority sequencing principle

Continuing in the same direction as the previous layer, the 5th layer involves training with scat-singing-like combinations and using words. However, the primary emphasis in this layer is on combining consonants according to the SSP, resulting in the creation of consonant clusters. Scat-singing-like combinations like [praθrakra] and words like *ask*, *blue*, *gray*, *through*, *athlete* and *interpretation* illustrate words for practice in this layer. Thus, consonant clusters that follow the SSP at both the onset and coda positions are the focal point of this layer.

4.3.6. 6th layer: combining speech-sounds to violate the sonority sequencing principle

The aim of this layer is to train consonant clusters that violate the SSP. Scat-singing-like combinations such [spra], [krafθs] and words like *stop*, *fifth* and *clothes* are examples for practicing the skills of this layer. Consonant clusters should be explored in both onset and coda position as well.

4.3.7. 7th layer: connecting sounds and letters

Practically every word in a language can be formed and pronounced using the skills acquired in the previous six layers. However, as highlighted by Zimmer, Silveira, and Alves (2009), grapho-phonetic-phonological transfers may pose challenges in the pronunciation of L2 learners. While writing is not the primary focus of this study, the skill of decoding the pronunciation possibilities behind written words is valuable for going beyond the standard accents of textbooks. An example of this ability would be

recognizing various possible pronunciations of a specific written form, such as the letter <t> in English. Additionally, this ability enables learners to develop a panoramic overview of the target accent, enhancing the potential for generalization and the ability to predict or deduce the expected pronunciation of a word in a given accent.

Another interesting skill enhanced by generalization is the ability to practice different accents during reading tasks. The fragment below illustrates this process in the translation practice of Rosa Freire D'Aguiar:

When I finish a chapter, I read it all over again, Brazilianizing it here and there, comparing it with the original, reading aloud—not like [Gustave] Flaubert, who seems to have read aloud literally, but reading aloud in my mind, if I make myself understood (D'aguiar, 2004, p. 75, our translation).

This passage makes it clear that the translator is mentally pronouncing the words with the aim of incorporating features of non-standard Brazilian Portuguese and checking the reading fluency of the translated text. Consequently, the 7th layer focuses on developing this ability in L2.

4.3.8. 8th layer: adjustments in connected speech

The goal of the 8th layer is to train phonological processes in L2. Phonological processes should be explored both within and between words to understand how speech sounds interact within the utterance. Training strategies should encompass both words and sentences, as phonological processes also occur at word boundaries. The skills developed in the 7th layer may assist learners in comprehending phonological processes during spoken utterances. Written texts can be used during the initial practices of the 8th layer, progressing to listening identification.

4.3.9. 9th layer: breaking down accents

The 9th layer, the final stage of the Layered Model, establishes a loop in the learning processes. The skill involved in this layer refers to the ability to decompose accents into layers to map where each accent can be modified to increase or decrease accentedness. Decomposing accents into layers requires awareness of accent layers,

and as such, learners must be proficient in the previous eight layers to succeed in the 9th layer. Those learners who enjoy or require studying multiple accents, such as actors, should master this layer to continue learning and producing new accents.

Figure 4 illustrates the distribution of the layers in the Layered Model.

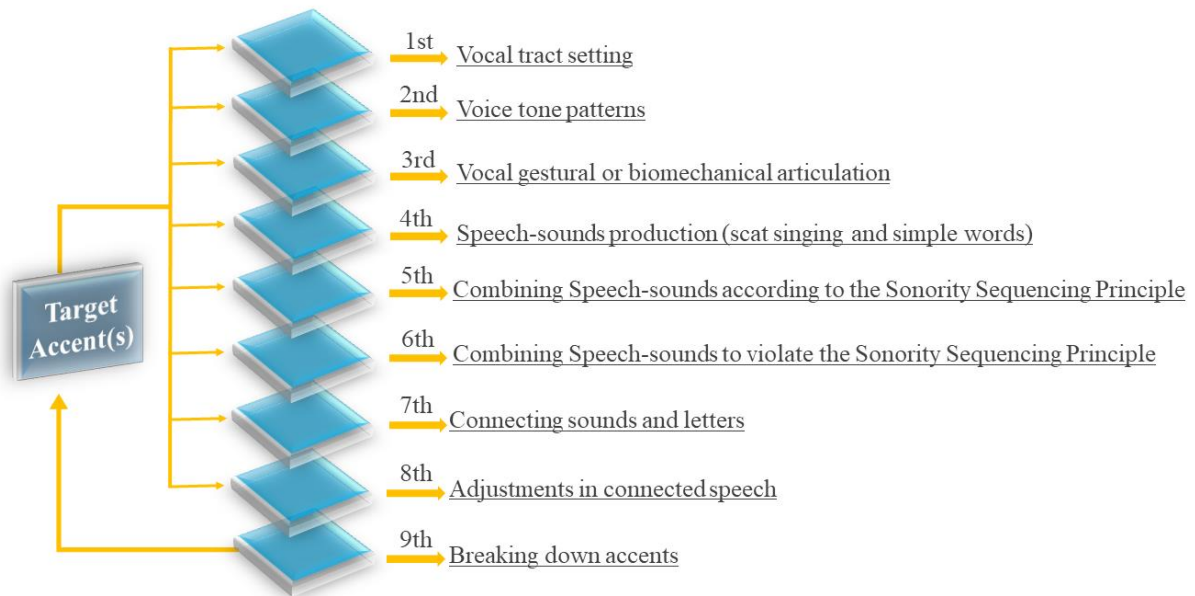


Figure 4 - Organization of the nine layers of the Layered Model (developed by the authors).

The layers are arranged as a progression of skills to be developed; thus, the 1st layer establishes foundations for the 2nd, the 2nd for the 3rd, and so on. Consequently, the application of the Layered Model is a cumulative process. However, this cumulative process does not occur in a single direction and the learner has to go back and forth through the layers in the learning process. This means that while studying, for instance, the 5th layer, one can easily revisit any previous layer to reinforce content or increase the difficulty level.

This property of the Layered Model becomes particularly important when studying consonant clusters. For instance, in accordance with the Layered Model, the cluster /fl/ is placed in the 5th layer, while the cluster /st/ is placed in the 6th layer. However, considering the training-stages for English described in Table 3, the consonant cluster /st/, which violates the SSP, is acquired during the intermediate training-stage, whereas the cluster /fl/, which follows the SSP, is acquired during the late training-stage. Thus, in this case, a cluster from the 6th is acquired before a cluster from the 5th layer. The reason for such changes in layer order in comparison to the

training-stages lies in the Layered Model's approach. The Layered Model initially focuses on universal principles for language acquisition and then progresses to target language specific features. Considering that the SSP represents a strong universal tendency according to Carlisle's (2001), it seems reasonable for the Layered Model to address clusters that follow the SSP in a previous layer than clusters that violate the SSP.

The flexibility to move back and forth among layers enables learners to emphasize skills that require more attention based on their individual needs. This means that while studying, for example, the 5th layer, one can easily revisit any previous layer to reinforce content or increase the difficulty level. It is not necessary to fully master one layer before delving into another. Therefore, the cumulative process involves not only acquiring higher layers but also enhancing one's competence in each layer. For instance, a learner might have high performance on the 6th layer and low performance on the 2nd layer, highlighting a weak aspect of their accent pronunciation. As each layer can be independently emphasized, the learner can thus focus on the 2nd layer to improve their accent skills.

5. CONCLUDING REMARKS

By examining the phonetic and phonological acquisition of L1, as well as the anthropological features of spoken language emergence, we identified both similarities and differences with L2 pronunciation. These similarities and differences were then utilized to create a model of pronunciation for L2-accent learning designed to guide learners in going beyond standard accents of textbooks. Firstly, elements related to the aesthetic features of accents were highlighted to differentiate aesthetic training from intelligibility training in pronunciation. Secondly, by establishing English as the target L2 for this study, stages of English acquisition as L1 were adapted into training-stages for use in the Layered Model of pronunciation for L2-accent learning.

The Layered Model developed consists of nine layers designed to guide accent training in L2, but its application for accent assessment purposes also seems possible. While the layers may be presented as isolated items for pedagogical reasons, it is fundamental to state that the layers work simultaneously in speech production. As a result, a learner might have different performance levels on the layers in a non-progressive way, possibly due to L1 transfers. This particularity demonstrates that it is

not necessary to fully acquire one layer before delving into another, and learners can move back and forth among layers. Thus, the Layered Model can be personalized according to the learner's accent needs.

By applying the training-stages presented in Table 3, learners can progress through the Layered Model using phonetics and phonology. If, unlike in this study, English is not the learner's target L2, only Table 3 needs to be adapted to the target L2, while the Layered Model remains the same. Table 3 represents the guidelines in phonetics and phonology that should be approached when progressing through the Layered Model. For instance, in the 1st group of the early training-stage of Table 3, no consonant clusters are expected. Therefore, learners may ignore the 5th and 6th layers and work with all the remaining ones. Pedagogical and self-study strategies should be designed with the Layered Model to avoid confusion for the learner throughout the process.

The Layered Model enable both instructors and students to work with accents even in classroom contexts, regardless of the applied method. Although it might initially seem that the learner must first acquire a standard accent to reach the 9th layer and then begin going beyond standard accents, it is important to note that the first target accent to be learned can be any accent. This is fundamental when considering that not all learners aim to learn more than one accent and may have few demands to master the 9th layer. They are likely to prefer mastering or automating the first eight layers of the model to achieve their desired target accent.

The limitations of this study include the lack of longitudinal data from the model, since field application in different teaching-learning contexts was not carried out. The absence of longitudinal experiments is justified due to the limited space for conducting this study, which was subordinated to the curricular provisions of the referred undergraduate course — i.e., Bachelor in English Language. In this sense, it became unfeasible to train instructors to apply the model in the classroom and subsequently evaluate the results obtained after applying the model. However, in future longitudinal studies, it is expected that, despite the model providing guidelines for working with accents, the difficulty in choosing activities targeted for each layer may be observed due to the close relationship with the lack of teachers' cognition about pronunciation instruction described in the literature.

In conclusion, the Layered Model offers guidelines for L2-accent learning, empowering learners to go beyond the standard accents found in textbooks. The

model explores various aspects of accent learning, extending beyond mere articulatory training or theoretical instruction in phonetics and phonology. Its focus is on practical training rather than theoretical instruction. Moreover, the model is versatile and can be incorporated into reading and listening tasks, broadening its applicability in association with diverse language teaching manuals. As the Layered Model is introduced in this study, further research is still needed and encouraged by the authors. It is crucial for researchers to understand that the Layered Model does not aim to explain L2 acquisition but rather offers a practical approach to L2 pronunciation instruction, primarily focusing on providing guidelines to go beyond standard accents found in textbooks.

6. REFERENCES

AGUASACO, Carlos. On the Origin and Future of Poetry: Notes Towards an Investigation. 2014. CUNY City College. **Publicações e Pesquisa do City College of New York**. Disponível em: [https://academicworks.cuny.edu/cc_pubs/809].

ARBIB, Michael A.; ABOITIZ, Francisco; BURKART, Judith M.; CORBALLIS, Michael; COUDÉ, Gino; HECHT, Erin; LIEBAL, Katja; MYOWA-YAMAKOSHI, Masako; PUSTEJOVSKY, James; PUTT, Shelby. The comparative neuroprimatology 2018 (CNP-2018) road map for research on How the Brain Got Language. **Interaction Studies**, [S.L.], v. 19, n. 1-2, p. 370-387, 17 Sep. 2018. John Benjamins Publishing Company. <http://dx.doi.org/10.1075/is.18013.arb>.

ARCHIBALD, John. Ease and Difficulty in L2 Phonology: a mini-review. **Frontiers In Communication**, [S.L.], v. 6, n. 1, p. 1-7, 3 mar. 2021. Frontiers Media SA. <http://dx.doi.org/10.3389/fcomm.2021.626529>.

ASHA, American Speech-Language-Hearing Association. (n.d.). Accent modification [Practice portal]. Retrieved oct 22nd, 2023, from www.asha.org/Practice-Portal/Professional-Issues/Accent-Modification/

BAGNO, Marcos. O que é uma língua?: imaginário, ciência & hipóstase. In: LAGARES, Xoán; BAGNO, Marcos (org.). **Políticas da norma e conflitos linguísticos**. São Paulo: Parábola Editorial, 2011. Cap. 13. p. 355-387. (Lingua[gem]). 1ª reimpressão 2019. Tradução de Marcos Bagno. ISBN: 978-85-7934-039-0.

BAKER, Amanda; MURPHY, John. Knowledge Base of Pronunciation Teaching: staking out the territory. **TESL Canada Journal**. Canadá, p. 29-50. 2011.

BEHLAU, Mara (org.). **Voz: o livro do especialista**. Rio de Janeiro: Revinter, 2013. 1 v. Terceira Reimpressão.

BERTHOLD, Margot. **História Mundial do Teatro**. 3. ed. São Paulo: Perspectiva, 2006. Título Original: Weltgeschichte des Theaters. Tradução de Maria Paula V. Zurawski, J. Guinsburg, Sérgio Coelho e Clóvis Garcia. ISBN 85-273-0228-4.

BROWMAN, Catherine P.; GOLDSTEIN, Louis. Articulatory Phonology: an overview. **Phonetica**, [S.L.], v. 49, n. 3-4, p. 155-180, 1 May 1992. Walter de Gruyter GmbH. <http://dx.doi.org/10.1159/000261913>.

CARLISLE, Robert S.. Syllable Structure Universals and Second Language Acquisition. **International Journal of English Studies**, Universidad de Murcia, v. 1, n. 1, p. 1-19. 2001.

CELCE-MURCIA, Marianne; BRINTON, Donna M.; GOODWIN, Janet M.. **Teaching Pronunciation**: a reference for teachers of english to speakers of other languages. University Of Cambridge: Cambridge University Press, 1996.

CERON, Marizete Ilha; KESKE-SOARES, Márcia. Terapia fonológica: a generalização para outra posição na palavra. **Revista Cefac**, [S.L.], v. 11, n. 2, p. 199-206, jun. 2009. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s1516-18462009000200004>.

COLLINS, Beverley; MEES, Inger M.; CARLEY, Paul. **Practical English phonetics and phonology**: a resource book for students. 4. ed. New York: Routledge Taylor & Francis Group, 2019. (Routledge English language introductions).

CRUTTENDEN, Alan. **Gimson's Pronunciation of English**, revised. 7th ed. London: Hodder Education an Hachette UK Company, 2008. ISBN 978-0-340-958773.

D'AGUIAR, Rosa Freire. **Memória de Tradutora**. Florianópolis: Escritório do Livro: Nut/Ufsc, 2004. 112 p. Entrevista a Marlova Aseff e Dorothée de Bruchard. Edição de texto de Marlova Aseff. Prefácio de Clélia Piza.

DERWING, T. M.; MUNRO, M. J. **Pronunciation Fundamentals**: Evidence-based Perspectives for L2 Teaching and Research. Language Learning & Language Teaching Series. John Benjamins Publishing Company, 2015. DOI: 10.1075/llt.42.

FAROOQ, Aneesa. The Anatomy of Language in Human Brain: a review. **Universe International Journal of Interdisciplinary Research**, India, v. 3, n. 9, p. 79-85, fev. 2023. An International Peer Reviewed Refereed Journal.

FIGUEIREDO, Francisco José Quaresma de. Aquisição E Aprendizagem De Segunda Língua. **Signótica**, [S.L.], v. 7, n. 1, p. 39-57, 11 set. 2009. Universidade Federal de Goiás. <http://dx.doi.org/10.5216/sig.v7i1.7380>.

FITCH, W. Tecumseh; BOER, Bart de; MATHUR, Neil; GHAZANFAR, Asif A.. Monkey vocal tracts are speech-ready. **Science Advances**, [S.L.], v. 2, n. 12, p. 1-7, 2 dez. 2016. American Association for the Advancement of Science (AAAS). <http://dx.doi.org/10.1126/sciadv.1600723>.

GAZZANIGA, Michael S.; IVRY, Richard B.; MANGUN, George R.. Language. In: GAZZANIGA, Michael S.; IVRY, Richard B.; MANGUN, George R.. **Cognitive Neuroscience: the biology of the mind**. 4. ed. New York And London: W. W. Norton & Company, 2014. Cap. 11. p. 468-505.

GIRONDA, Felicia; FABUS, Renee. Assessment of Articulation and Phonological Disorders. In: STEIN-RUBIN, Cyndi; FABUS, Renee. **A guide to clinical assessment and professional report writing in speech-language pathology**. New York: Delmar Publications, 2011. Cap. 7. p. 139-176.

GLEASON, Jean Berko; RATNER, Nan Berstein. **The Development of Language**. 9. ed. Boston: Pearson Education, 2016.

HARENDITA, Monica Ella. "Native-like Accent: A Myth or a Goal? A Study on Learner Preferences for English Accents." In: **Proceedings of the 61st Teflin International Conference 2014**, October 7-9, 2014, English Education Department, Teacher Training and Education Faculty, Sebelas Maret University Solo.

INZUNZA-MADRIGAL, Rosella. **Attitudes Towards Nonnative Accents in the Field of Speech-Language Pathology**. 2021. Thesis - Master of Science in Speech-Language Pathology - California State University San Marcos, 2021.

IPEK, Hulya. Comparing and Contrasting First and Second Language Acquisition: implications for language teachers. **English Language Teaching**, [S.L.], v. 2, n. 2, p. 157-163, 19 maio 2009. Canadian Center of Science and Education. <http://dx.doi.org/10.5539/elt.v2n2p155>.

JOHANSEN, Aimée. Accent on Accents: helping learners better understand English spoken by speakers having a variety of accents. **WoPaLP**, Hal Open Science, v. 13, p. 1-13, 2019. <https://hal.archives-ouvertes.fr/hal-02733357>.

JOSEPH, John E.. Saussure and structural phonology. In: DRESHER, B. Elan; HULST, Harry van Der. **The Oxford History of Phonology**. Oxford, United Kingdom: Oxford University Press, 2022. Cap. 10. p. 203-220. Part II: The Founders of Phonology.

KOZLOWSKI, Alexandra. The Influence of Accents on Social Perception. **Inkblot**, [S.I.], v. 4, p. 16-16, set. 2015.

LAVIER, John. **The Phonetic Description of Voice Quality**. Cambridge: Cambridge University Press, 1980.

LEVIS, John. M. Revisiting the intelligibility and nativeness principles. **Journal of Second Language Pronunciation**, v. 6, n. 3, p. 310-328, 2020.

LEVIS, John M.; ZHOU, Ziwei. Accent. **The Tesol Encyclopedia of English Language Teaching**, [S.L.], p. 1-5, 18 jan. 2018. Wiley. <http://dx.doi.org/10.1002/9781118784235.eelt0002>.

MCLEOD, Sharynne; CROWE, Kathryn. Children's Consonant Acquisition in 27 Languages: a cross-linguistic review. **American Journal of Speech-Language Pathology**, [S.L.], v. 27, n. 4, p. 1546-1571, 21 nov. 2018. American Speech Language Hearing Association. http://dx.doi.org/10.1044/2018_ajslp-17-0100.

MCLEOD, Sharynne; VAN DOORN, Jan; REED, Vicki A.. Normal Acquisition of Consonant Clusters. **American Journal of Speech-Language Pathology**, [S.L.], v. 10, n. 2, p. 99-110, maio 2001. American Speech Language Hearing Association. [http://dx.doi.org/10.1044/1058-0360\(2001/011\)](http://dx.doi.org/10.1044/1058-0360(2001/011)).

MITCHELL, Rosamond; MYLES, Florence; MARSDEN, Emma. Sociocultural perspectives on second language learning. In: MITCHELL, Rosamond; MYLES, Florence; MARSDEN, Emma. **Second Language Learning Theories**. 3. ed. London And New York: Routledge Taylor & Francis Group, 2013. Cap. 8. p. 220-249.

MORENO-TORRES, Ignacio; MARIËN, Peter; DÁVILA, Guadalupe; BERTHIER, Marcelo L.. Editorial: language beyond words. **Frontiers In Human Neuroscience**, [S.L.], v. 10, p. 1-3, 20 dec. 2016. Frontiers Media SA. <http://dx.doi.org/10.3389/fnhum.2016.00639>.

MOTA, Helena Bolli; WIETHAN, Fernanda Marafiga. Fonologia: intervenção. In: MARCHESAN, Irene Queiroz; SILVA, Hilton Justino da; TOMÉ, Marileda Cattelan. **Tratado das especialidades em fonoaudiologia**. São Paulo: Guanabara Koogan, 2014. Cap. 72. p. 600-608. ISBN: 978-85-277-2641-2.

NOLEN-HOEKSEMA, Susan *et al.* Linguagem e Pensamento. In: NOLEN-HOEKSEMA, Susan *et al.* **Atkinson & Hilgard Introdução à psicologia**. 16. ed. São Paulo: Cengage Learning, 2018. Cap. 5. p. 253-285. Tradução de Priscila Rodrigues da Silva e Lopes.

O'SHEA, José Roberto. Traduzindo *Huckleberry Finn*: aventuras da variedade linguística. **Scripta Uniandrade**, Curitiba, Paraná, Brasil, v. 16, n. 3, p. 31-42, 11 nov. 2018.

RAJABI, Peyman; HOSSEINI, Seyed Ali; HANIFI, Rouzbeh. Linguistic Stages in First Language Acquisition (IJELR): a critical analysis. **International Journal of English Language, Literature And Translation Studies**, India, v. 4, n. 3, p. 428-435, jul./set. 2017.

SAUSSURE, Ferdinand de. **Course in General Linguistics**. New York: Philosophical Library, 1959. Edited by Charles Bally and Albert Sechehaye. In collaboration with Albert Reidlinger. Translated from the French by Wade Baskin.

SELINKER, Larry. Interlanguage. **Iral - International Review of Applied Linguistics in Language Teaching**, [S.L.], v. 10, n. 1-4, p. 209-231, 1972. Walter de Gruyter GmbH. <http://dx.doi.org/10.1515/iral.1972.10.1-4.209>.

SHORMANI, Mohammed Q.. The Nature of Language Acquisition: where L1 and L2 acquisition meet?. **Journal of Literature, Languages and Linguistics**, v. 4, p. 24-34, 2014.

SILVEIRA, Rosane. **Contribuições da fonética para o estudo da aprendizagem do componente sonoro em L2**. Fórum Linguístico (UFSC. Impresso), v. 16, p. 3861-3872, 2019.

SILVEIRA, R.; ZANCHET, C. E.; PEREIRA, M. H. Affordances of digital technology for English pronunciation teaching. **Veredas - Revista de Estudos Linguísticos**. Vol.26(2), p. 1-25, 2022.

SOUTHARD, Bruce; MULLER, Al. Blame it on Twain: reading american dialects in 'the adventures of huckleberry finn'. **Journal of Reading**, [S. l.], v. 36, n. 8, p. 630-364, 1993. Disponível em: <http://www.jstor.org/stable/40033380>. Acesso em: 3 fev. 2024.

STOLOFF, Bob. **Scat!**: vocal improvisation techniques. Brooklyn, New York: Gerard & Sarzin Publishing Co., 1999. Forward and Introduction by Evan Sarzin. Compact Disc Included. ISBN 0-9628467-5-9.

VAGONES, E. W. A fonética e seus precursores. **Alia**, São Paulo, n. 24, p. 179-85, 1980.

WEI, Xueping. Implication of IL Fossilization in Second Language Acquisition. **English Language Teaching**, [S.L.], v. 1, n. 1, p. 127-131, 1 jun. 2008. Canadian Center of Science and Education. <http://dx.doi.org/10.5539/elt.v1n1p127>.

WIESE, R.. Phonology: overview. **Encyclopedia of Language & Linguistics**, [S.L.], p. 562-564, 2006. Elsevier. <http://dx.doi.org/10.1016/b0-08-044854-2/00035-3>.

WOLFRAM, Walt; SCHILLING-ESTES, Natalie. On the applications of Dialect study. In: WOLFRAM, Walt; SCHILLING-ESTES, Natalie. **American English**::: dialects and variation, 2nd edition. 2. ed. Malden, Ma, Usa: Blackwell Publishing, 2006. Cap. 10. p. 312-328. (Language in Society, Vol. 25). ISBN 978-1-4051-1265-9.

WORDSWORTH, William; COLERIDGE, Samuel Taylor. **Lyrical Ballads**. London: T. N. Longman and O. Rees, Paternoster-Row, By Biggs And Co. Bristol., 1800.

YAVAS, Mehmet. **Applied English Phonology**. 2. ed. United Kingdom: Wiley-Blackwell, 2011. ISBN: 978-1-444-33322-0.

YIN, Ruihua; WEIJER, Jeroen van de; ROUND, Erich R.. Frequent violation of the sonority sequencing principle in hundreds of languages: how often and by which sequences?. **Linguistic Typology**, [S.L.], v. 27, n. 2, p. 381-403, 12 jan. 2023. Walter de Gruyter GmbH. <http://dx.doi.org/10.1515/lingty-2022-0038>.

ZARATE, Jean Mary. The neural control of singing. **Frontiers in Human Neuroscience**, [S.L.], v. 7, jun. 2013. Frontiers Media SA. <http://dx.doi.org/10.3389/fnhum.2013.00237>.

ZIMMER, Márcia; SILVEIRA, Rosane; ALVES, Ubiratã Kickhöfel. **Pronunciation Instruction for Brazilians**: bringing theory and practice together. United Kingdom: Cambridge Scholars Publishing, 2009. 220 p.