University of Campinas
Institute of Language Studies
Graduate Program in Linguistics

Master's Qualifying Report

The prosody of speech in a dialect contact situation

A sociophonetic study of the speech of Alagoan migrants in São Paulo

Gustavo de Campos Pinheiro da Silveira

Research report presented to the Graduate Program in Linguistics of the University of Campinas in partial fulfillment of the requirements for the Master's degree, in the area of Sociolinguistics.

Advisor: Prof. Dr. Livia Oushiro

Campinas
2021

Contents

List of Tables

List of Figures

Foreword

This report refers to my academic activities from March 2020 to July 2021, in which I have been developing my master's research project "The prosody of speech in a dialect contact situation: A sociophonetic study in the speech of Alagoan migrants in São Paulo" at the Institute of Language Studies of University of Campinas. I have been conducting these activities under the supervision of Dr. Livia Oushiro with a fellowship from CAPES (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*; Process: 88887.495348/2020-00).

This master's research is a sociophonetic study on the prosody of speech in a dialect contact situation. I examine variation in the rhythm and intonation in the speech of Alagoan migrants living in the Campinas Metropolitan Area, a region in the state of São Paulo, in Southeast Brazil. The study adopts the perspective of Variationist Sociolinguistics (Labov, 1972), and it is also theoretically and methodologically informed by Acoustic Phonetics (Barbosa and Madureira, 2015; Johnson, 2012; Ladefoged, 1996). Specifically, the question to be addressed here is: to what degree can the prosody of mobile speakers suffer modifications due to the exposure to the prosody of other regional varieties? I analyze the distribution of 23 acoustic parameters related to rhythm and intonation in a speech sample of 21 adult Alagoan migrants, compared to a control sample of the speech of 60 lifelong native residents of the city of São Paulo. To determine if the migrants' prosody has changed as a result of the contact with the Paulista variety, I examine correlations of the prosodic variables with gender, age of arrival, and length of residence in the state of São Paulo. Partial results show that Alagoan migrants have a more stress-timed rhythm than native

Paulistanos. Male Alagoan migrants differ from male native Paulistanos in 9 of the 13 intonation measures, showing a more high-pitched and variable fundamental frequency, but only 2 of these measures distinguish between female Alagoan migrants and female natives from São Paulo. Analyses on the effects of age of arrival and length of residence suggest that prosody is less susceptible to contact-induced modifications than segmental features of speech.

In what follows, I report the developments, the partial results, and the future steps of this research. In Chapter 1, I situate this research within the broader context of the sociolinguistic studies on dialect contact involving migration. Then, in Chapter 2, I describe the corpus and the methodological procedures that allowed me to collect, process, and analyze prosodic measurements in the speech of Alagoan migrants. In Chapter 3, I present the partial results achieved so far. Finally, I end this report with Chapter 4, in which I describe the tasks to be carried out in the following months.

Chapter 1

Background

Meta-comentário:
Para a versão final da dissertação, criarei um capítulo de Introdução (que substituirá o Foreword), em que farei uma apresentação mais extensa do projeto de pesquisa.

Migration is an ever-present phenomenon in all societies, and its consequences extend to many dimensions of community life. Large-scale displacements of populations affect the languages and the social relations that people establish through linguistic interaction. One of the most common outcomes of migration is the long-term contact between speakers that share the same native language but not the same variety. When people move to regions where a different regional dialect is spoken, their speech and linguistic repertoire can be modified by the daily interactions they start to have with native speakers from these regions. Dialect contact situations are intriguing for sociolinguistic inquiry as the study of the speech of migrants in this scenario is a fertile ground for the understanding of broader questions about the functioning of the human language, such as the ones that relate to the stability of the speaker's grammar (Oushiro, 2018) and to the effects of the population dynamics on the historical evolution of regional varieties (Britain, 2013).

This chapter reviews previous studies on dialect contact and prosody. In Section 1.1, I discuss the pioneering works that have established the main conceptual framework widely adopted by most of the subsequent

studies concerned with sociolinguistic phenomena emerging from dialect contact settings. In Section 1.2, I review recent research about the speech of internal migrants in the state of São Paulo. Section 1.3 discusses the few sociolinguistic investigations conducted so far about prosodic variation in speech communities. Finally, in Section 1.4, I briefly describe some of the features that characterize the prosody of Brazilian Portuguese, according to some of the most influential studies on this topic.

1.1 Migrants in sociolinguistic studies

Since the mid-1970s, sociolinguists have investigated the mechanisms of the process of accommodation to regional varieties in the speech of migrants (Payne, 1976; Trudgill, 1986). In this context, *accommodation* can be understood as the process through which speech features of the migrants converge (i.e., they become more similar) to those of the dialect of the host community (i.e., the community to which they have moved). The central question that the studies in this line of research try to answer is the following: which social and linguistic factors condition the contact-induced modifications in the speech of migrants that move to a different dialectal region? Many studies have already contributed to answering this question, with empirical analyses of specific cases of dialect contact (Payne, 1976; Bortoni-Ricardo, 1985; Chambers, 1992; Kerswill, 1993; Dodsworth, 2017), as well as with important theoretical advances (Chambers, 1992; Trudgill, 1986). However, most contributions are restricted to lexical and segmental variables, while other dimensions of dialect accommodation remain unexplored. In particular, we need more studies on the effects of dialect contact on the prosody, the morphology, and the syntax of migrants' speech (Torgersen and Szakay, 2012). Furthermore, most studies focus on English varieties, and there are too few works concerning the contact between dialects of Brazilian Portuguese (Oushiro, 2016).

Traditionally, sociolinguistic studies have privileged the analysis of the speech of the "prototypical" members of the communities: the lifelong

residents born and raised in the community, especially those whose parents are also natives (Britain, 2018; Oushiro, 2016). Kerswill (1993) argues that the nativeness of the speakers serves in variationist tradition as one of the criteria for defining a speech community. One of the motivations to exclude migrants from the sociolinguistic analysis relates to Lenneberg's (1967) critical period hypothesis. According to this hypothesis, around the first years of puberty, the speaker's linguistic system stabilizes, and, from then on, it will not change in face of new linguistic stimuli. In short, the process of natural acquisition stops.

The argument usually used by sociolinguists to justify the nativeness criterion goes as follows: if the critical period hypothesis is assumed to be true, then the study of the speech community must focus on native members, because the linguistic patterns of adult migrants will not reflect the ones of the community of interest, but of the community where the migrants were raised – that is, patterns they acquired during childhood before migrating. Based on this argument, many sociolinguistic researchers have decided to exclude from their samples the migrants who arrived there after a certain age. For instance, Labov (2006, p. 111) decided to remove from his study of English in New York City any migrant who arrived after the age of eight.

However, many studies have shown the importance of considering migrants in the dynamics of the speech community. Among them, the works of Bortoni-Ricardo (1985), Trudgill (1986), and Chambers (1992) stand out as pioneers in the analysis of migrants' speech from a variationist perspective, as well as for showing that many sociolinguistic phenomena (like linguistic diffusion, reallocation, and leveling) can emerge from the contact between migrants and a community's native population.

Meta-comentário:
Pretendo incluir nesta parte também os trabalhos de Payne (1976; 1980), sobre aquisição dialetal na Filadélfia (EUA), e de Kerswill (1985), sobre migrantes rurais em Bergen (Noruega).

From the perspective of social networks (Milroy, 1980), Bortoni-Ricardo (1985) systematically described the process of dialectal diffusion in the speech of 33 migrants from the rural area of Minas Gerais, who had

moved to Brazlândia, an urban city in the outskirts of Brasília, Brazil's federal capital in the central-western region of the country. Bortoni-Ricardo (1985) analyzed linguistic features specific to the *caipira* dialect, a variety that is traditionally spoken in rural areas of the state of São Paulo and adjacent regions of the neighboring states of Minas Gerais, Mato Grosso do Sul and Goiás: (i) the delateralization (replacement of a lateral consonant by a central consonant) of the palatal lateral [ʎ], replaced by the palatal approximant [j] (e.g., *velho* 'old' pronounced [ˈvɛjʊ] instead of the prestige form [ˈvɛʎʊ]); (ii) the reduction of rising glides (e.g., *alívio* 'relief' pronounced [aˈlivʊ] instead of the prestige form [aˈlivjʊ]); and (iii) number agreement between verb and subject (e.g., *Nós queria ir* 'we wanted to go' instead of the prestige form "Nós queríamos ir"). The author analyzed these linguistic variables in correlation with the speakers' social networks and demographic categories (gender, age, exposure to media, and occupation).

Bortoni-Ricardo (1985) showed that the migrants with more integrated social networks used the prestige forms of the host community more frequently, and male speakers were more advanced than females in this process of accommodation to the urban speech. According to her, gender differences in occupation and the social networks these occupations entail can explain the gender differences in speech patterns. While male speakers performed activities involving interpersonal relations in the public sphere, exposing them to supra-local prestige forms, female speakers tended to keep social networks more restricted to the domestic and neighborhood domains, which favored the preservation of the rural stigmatized variants.

Trudgill (1986) developed a broader theoretical framework to analyze the linguistic changes resulting from dialect contact. One of his contributions was to explain dialect contact phenomena from the perspective of Accommodation Theory, originally developed by Giles (1973). Trudgill (1986) describes the modifications in the speech of migrants as the result of the long-term accommodation process to the variety of the host community. He also argues that many linguistic and social variables can influence the rate and course of this process, acting as inhibitors or catalysts of the

accommodation. For instance, speakers tend to accommodate faster to a linguistic variant when not using this variant impairs mutual intelligibility. Trudgill (1986) illustrates this point with some examples of linguistic variables from American and British English. A native speaker of British English living in the United States can make an effort to accommodate faster to the American pronunciation of /t/ as [ɾ] in intervocalic position, since the British pronunciation of this phoneme, the aspirated plosive [t^h], can confuse American speakers, making them wrongly understand "pizza" instead of "Peter" (Trudgill, 1986, p. 16). On the other hand, accommodation tends to be slower with phonological variables under phonotactic restrictions. For example, British speakers have difficulty acquiring the American pronunciation of non-prevocalic /r/ (as in "part"), in spite of its salience. According to Trudgill (1986, p. 16), the British phonotactic restrictions on this variable, which allow for the realization of /r/ only in prevocalic position, make it difficult for British speakers to accommodate to the American form.

Chambers (1992) is an influential study on dialect contact, largely responsible for spreading the use of the term *dialect acquisition* in place of Trudgill's *dialect accommodation*. He justifies his terminological proposal arguing that the consequences of Trudgill's long-term accommodation to a regional variety are usually not ephemeral adjustments in the migrant's speech, but permanent acquisitions of new variants. The importance of this conceptual issue is evident since it concerns how dialect contact impacts the speaker's linguistic knowledge. However, in the absence of procedures to reliably determine whether a modification in the speaker's repertoire is permanent or not, the two terms, most of the time, refer to the same thing: new patterns in the speech of a person that emerged after she moved to a new dialectal region. In this research, I use both interchangeably.

Chambers (1992) analyzed the speech of six Canadian children who moved to the south of England at the beginning of the 1980s. He examined the acquisition of lexical (e.g., the acquisition of the British form "coach" in place of the Canadian variant "bus") and pronunciation variables (e.g., the acquisition of the aspirated plosive [t^h] in the place of the tap [ɾ], in

words like "putting"). Based on the patterns identified in these children's speech, he proposes eight generalizations concerning the acquisition of new dialectal features in contact situations. For instance, he argues that lexical substitutions occur quickly in the early stages and slow down later on. On the other hand, the acquisition of phonological variants tends to be slower when involving more complex rules – a statement that reinforces Trudgill's (1986) position about phonotactic restrictions.

The works discussed above show that Variationist Sociolinguistics' (Labov, 2006) theoretical model and methodological procedures can be employed to investigate the complex phenomenon of the acquisition of dialectal features by migrants. With quantitative analysis techniques, they were able to model the complex process of dialect accommodation in order to compare the degree and rate of the modifications in the migrants' speech. Combined with ethnographic analysis of the social context of the speakers, the quantification allowed for the identification of linguistic and social patterns of dialect acquisition. It became clearer from these studies that the modifications in the speech of mobile speakers vary according to the influence of multiple variables, such as gender and social network (Bortoni-Ricardo, 1985), mutual intelligibility (Trudgill, 1986), and phonological complexity (Chambers, 1992; Trudgill, 1986). These studies also gathered evidence for not taking the critical period hypothesis (Lenneberg, 1967) in absolute terms since some modifications in the speech of adult speakers appear to be possible in face of dialect contact.

Oushiro (2016) argues that one of the challenges of dialect accommodation research is to deal with multiple linguistic and social variables. She emphasizes that "the fact that an individual accommodates to a linguistic feature of the host community does not necessarily imply that a global process of accommodation has taken place" (Oushiro, 2016, p. 10, my translation). As already pointed out above, excepting for a few studies (as the work of Bortoni-Ricardo (1985), which analyzes the morphosyntactic variable number agreement in Brazilian Portuguese), most investigate accommodation to segmental features, mainly in English varieties — as it is the case of Chambers (1992) and Trudgill (1986). For a broader understanding

of the linguistic and social conditioning of dialect accommodation, it is crucial to expand the studies to other linguistic levels of analysis, such as the prosodic and the morphosyntactic ones, and also to other languages, such as Brazilian Portuguese.

1.2 The speech of migrants in São Paulo

Meta-comentário:
Pretendo ampliar esta seção para incluir estudos sobre acomodação dialetal no português falado em outras regiões do Brasil além de São Paulo.

The present research is concerned with the expansion of sociolinguistic studies on the speech of internal migrants in the state of São Paulo. The *Laboratório Variação, Identidade, Estilo e Mudança* (VARIEM), under the coordination of Oushiro (2018), has recently conducted one of the first systematic studies about the speech of internal migrants in Brazil. The project "Processes of dialect accommodation in the speech of Northeastern residents in the state of São Paulo" (henceforth "Accommodation Project") analyzed two speech samples of migrants from the states of Alagoas and Paraíba, both located in Northeast Brazil, who moved to Campinas in the Southeastern state of São Paulo. The project aimed to determine to what degree their speech changed due to the contact with the Paulista variety (the variety spoken in the state of São Paulo). The analysis focused on the effects of age of arrival and length of residence in the host community (two social variables that stratified the samples) in relation to five linguistic variables: (i) the realization of coda /r/ (e.g., *porta* 'door' pronounced with fricative or tap/approximant rhotic); (ii) the realization of /t/ and /d/ before [i] (e.g., *tia* 'aunt' and *dia* 'day', as plosives or affricates); (iii) the height of the pretonic mid vowels /e/ and /o/ in Hertz (e.g., *relógio* 'clock' and *roseira* 'rose bush'); (iv) number agreement in noun phrases (e.g., *os meninos* and *os menino* 'the boys'); and (v) sentence negation (e.g., *não vi* or *não vi não/vi não* 'I did not see'). The results indicate that the age of the speakers when they moved to São Paulo performed an important role in the

accommodation to the phonetic variants, but not to the morphosyntactic ones. On the other hand, the length of residence in the host community correlated only to coda /r/ (Oushiro, 2020).

The Accommodation Project helped to consolidate a research agenda on the speech of migrants in the state of São Paulo. Other researchers have recently contributed to the topic. In her master's thesis, Santana (2018) analyzed the speech of migrants from Sergipe, a state also located in the Northeastern region of Brazil, with a focus on the height of the pretonic mid vowels /e/ and /o/. She observed accommodation to the Paulista pronunciation of the anterior mid vowel /e/, but not of the posterior /o/, which suggests that the process of phonetic accommodation does not necessarily follow the principle of parallelism, a result also reported by Oushiro (2019c). Both Souza (2019) and Oliveira (2019) analyze the speech of migrants from the Northeastern state of Bahia. Souza (2019) studies dialect accommodation of Baianos in the Metropolitan Area of São Paulo, focusing on four sociolinguistic variables: (i) the realization of coda /r/; (ii) the height of pretonic mid vowels /e/ and /o/; (iii) sentence negation; and (iv) the use of definite article before proper nouns (e.g., *O João falou com a Flávia* or *João falou com Flávia* 'João spoke to Flávia'). In turn, Oliveira (2019) examines the accommodation to Paulista's variants of coda /r/ (alveolar tap and approximant) in the speech of Baianos living in Bauru, a city in the western outskirts of São Paulo city.

Meta-comentário:
Pretendo discutir aqui os resultados de Souza (2019) e Oliveira (2019).

All these studies have significantly contributed to the research agenda started by the Accommodation Project, but there remain many pending questions. Among the questions delineated by Oushiro (2019a), one that has not yet received attention concerns migrants' prosody and the effects that dialect contact has upon this level of phonetic analysis. Actually, not only in the context of the studies on migrants' speech but also in sociolinguistic research more generally, there are still too few works dedicated to prosodic variation and prosodic accommodation when compared to the ones focusing on consonantal and vocalic segments (Thomas, 2013).

For instance, we do not have enough studies on prosodic accommodation to know if Chambers' (1992) generalization is restricted to segmental variables or can be expanded to suprasegmental features of speech, like rhythm and intonation. Other questions also await an answer: is there a difference in the accommodation rate to prosodic patterns compared to vocalic and consonantal features? Are the linguistic and social variables that condition the course and rate of prosodic accommodation the same ones that affect the phonetic segments? The next section reviews some variationist studies concerning dialect accommodation in the rhythm and intonation of English and Spanish.

1.3 The prosody of speech in dialect contact situations

Before reviewing some sociophonetic studies on prosody, first I will try to clarify what I mean by prosody. In this research, I adopt a phonetic perspective of prosody. By "phonetic", I mean an approach to prosody as a phenomenon that can be empirically observed and examined by instrumental analysis of articulatory, acoustic, and auditory data.

From an acoustic point of view, I understand prosody as the simultaneous variation in time of fundamental frequency (f_0), duration, and intensity, resulting from the coordinated motion of the speech organs. When the sound wave stimulates the nerve endings of the auditory system, these acoustic parameters provoke sensations in the hearer that can be classified in three psychoacoustic categories: pitch, which refers to the sensations of a sound being perceived as low or high; length, which is the perception of a sound as being short or long; and volume, by which the hearers distinguish strong and weak sounds. The various combinations of these three parameters through the speech flow give rise to the perception of rhythm and intonation, the two dimensions of prosody, and also of other phenomena that are suprasegmental, but not prosodic, such as lexical stress and tone.

Intonation is the melody of speech; that is, the melodic contour that emerges from the variable sequence of high and low-pitched sounds through the speech chain. While f_0 is the primary acoustic correlate of intonation, the duration is the primary correlate of rhythm (Barbosa, 2019; Fletcher, 2010). Rhythm can be understood as the temporal structuring of speech by the sequencing of syllables of variable duration. According to Barbosa (2006), speech rhythm emerges from the conjunction of two opposing tendencies: regularity and alternation in syllable duration. In this sense, rhythm can be thought of as emerging from the overlaying of two different patterns of syllable duration. In one pattern, syllables tend to have approximately the same duration; in the other, they tend to have different duration, with longer syllables alternating with shorter ones. The result of the superposition of these patterns is a sequence in which syllables with more regular duration are recurrently "disturbed" by longer syllables (the stressed ones).

Meta-comentário:
Pretendo ampliar essa apresentação sobre prosódia da fala, explicitando melhor quais são minhas referências teóricas.

Most sociophonetic studies on prosody analyze either intonation or rhythm, but not both. The works on intonation can be distinguished into two groups. One comprises studies about stylistic uses of pitch, such as Podesva's (2006) study on the f_0 contour in the speech of three gay professionals in different social situations. Concerning Brazilian Portuguese, this line of research has been pursued by Mendes and Sene (2020), who investigated how pitch variation affects the perception of male speakers as sounding gay. The other group involves studies about change in the intonation patterns of speech communities, as Britain's (1992) study on linguistic change in the intonation contour of declarative clauses in New Zealand English.

As for the rhythm, most sociolinguistic studies conducted so far focus on the analysis of rhythm metrics; i.e., mathematical formulas used to compare the rhythm of languages and varieties. These rhythm metrics are closely related to a rhythmic typology popularized in the 1970s, by the

influence of Pike (1945) and Abercrombie's (1967) works. This typology emphasizes the regularity rather than the alternation in rhythm, separating the languages of the world into two rhythmic categories: syllable-timing and stress-timing. A perfect syllable-timed language would be one in which the duration of all syllables is isochronous; that is, all syllables would have approximately the same duration. On the other hand, in a perfect stress-timed language, not the syllables, but the intervals between stressed syllables would show isochrony. Some studies refer to these inter-stress intervals by "feet", but the current phonetic literature has preferred the term "stress group" (Barbosa, 2019; Fuchs, 2016).

It did not take long for phoneticians to empirically show that isochrony in speech production does not exist and that no language in the world has a perfect syllable-timed or stress-timed rhythm (Barbosa, 2000). The failure of the quest for isochrony led many researchers to abandon the categorical perspective of rhythm typology. Instead, since the 1990s, they have shifted to a continuous conception of rhythm, in which syllable- and stress-timing are opposite poles of a continuum, along which different languages stand. Some of them can be closer to the stress-timing pole, others can be near the syllable-timing pole, and some can even be located in the middle, showing a hybrid rhythm (Fuchs, 2016).

The rise of the continuous conception of rhythm typology fostered the emergence of rhythm metrics. These metrics refer to procedures to quantify the degree of syllable- and stress-timing of languages. Currently, there are more than two dozen of these metrics. However, all of them are alternative ways of calculating the same thing: the degree of variation in the duration of vowels, consonants, and syllables in utterances (Fuchs, 2016). Roughly speaking, the greater the variation the closer the language is to the stress-timed pole, and the smaller the variation the closer it is to the syllable-timed pole. I explain the technicalities of these metrics in greater detail in Section 2.4.

As discussed above, the majority of the sociolinguistic studies on rhythm consists of analyses of rhythm metrics. One of the first is the study of

Szakay (2006), in which rhythmic variation is analyzed in English spoken by 36 New Zealanders, of which 24 were Maoris and the others had European ascendancy, called *Pakeha* in Maoris' culture. The author calculated the Pairwise Variability Index (PVI), a metric developed by Low, Grabe, and Nolan (2000) and explained in section 2.4. Szakay (2006) showed that Maori English has a significantly more syllable-timed rhythm than *Pakeha* English, which tends to be more stress-timed. In another study, Szakay (2008) reports that younger speakers of both ethnicities show a more syllable-timed rhythm than older ones, a result that suggests that the contact between the Maori language and the Maori variety of English with *Pakeha* English is causing a rhythmic change in New Zealand.

Nokes and Hay (2012) also analyzed the change in rhythm in New Zealand English. The authors conducted extensive diachronic research involving more than 500 New Zealanders born between 1851 and 1988. The research also relies on calculations of the PVI, and the results reinforce the thesis about the rhythmic change proposed in Szakay's (2006; 2008) studies. According to Nokes and Hay (2012), variation in the vocalic duration gets smaller in the course of time, and New Zealanders no longer distinguish stressed and unstressed vowels based on duration. Furthermore, syllable duration no longer plays a significant role in prominence marking. The authors point out that contact with the Maori population is one of the factors that can explain these changes in New Zealand English.

Thomas and Carter (2006) also analyze rhythm from the sociolinguistic perspective. They investigated rhythmic differences between two English varieties in the United States: one spoken by African Americans (a variety known as African American Vernacular English) and the other by white Americans (also called White American English). The authors calculated the PVI in a sample of 40 speakers balanced for ethnic ancestry and stratified by three periods of birth: before the American Civil War; between 1869 and 1960; and between 1961 and 1985. They did not observe significant rhythmic differences among younger speakers. On the other hand, such a difference was present among older speakers born before the civil war. According to the authors, the results suggest that African

American English had a more syllable-timed rhythm, but it became more stress-timed in the course of the 20th century, probably by contact with White American English. According to Torgersen and Szakay (2012, p. 925), the study "clearly shows the effect of dialect contact: where there is long-term contact, rhythmic patterns become more similar".

More recently, Troncoso-Ruiz and Elordieta (2018) investigated the process of prosodic accommodation in the speech of Andalusian speakers (Southeast Spain) that had moved to Asturias (Northeast Spain). More specifically, they analyzed a corpus with the spontaneous speech of a group of Andalusian migrants living in Asturias and two control groups with non-migrant Andalusian and Asturian speakers, to investigate potential changes in the intonation patterns of migrants due to the contact with Asturian speech. The results show that the most salient intonation feature of Asturias to Andalusian hearers — the final falling contour in absolute interrogatives, which contrast to the raising contour of the Andalusian variety — was the most acquired feature by the migrants. This result is thought-provoking since it suggests that salience, seen by Trudgill (1986) as the main conditioning factor for the accommodation to lexical and segmental variables, can also influence the accommodation to suprasegmental features.

Meta-comentário:
Eu pretendo ampliar essa parte da revisão bibliográfica, incluindo mais estudos sociofonéticos sobre variação entoacional.

1.4 Prosodic variation in Brazilian Portuguese

Within Pike's (1945) and Abercrombie's (1967) framework of the rhythmic typology, most studies on Portuguese describe the Brazilian variety as having a hybrid rhythm (Moraes and Leite, 2002; Barbosa, 2000; Frota and Vigário, 2000; Cagliari and Abaurre, 1986; Abaurre-Gnerre, 1981), but a few studies argue in favor of classifying this variety as having a strict stress-timing, such as the works of Major (July 1, 1981; 1985) and, more

recently, Migliorini and Massini-Cagliari (2010) and Cagliari (2013).

Excepting for Barbosa (2000) and Frota and Vigário (2000), who conduct phonetic analyses, the studies that argue for a hybrid rhythm in Brazilian Portuguese rely on phonological considerations. Abaurre-Gnerre's (1981) is probably the first and the most representative of this phonological approach. According to her, the hybrid rhythm of Brazilian Portuguese emerges from the acting of rhythmically opposing phonological processes. The author focuses on two processes of vowel harmony: pretonic mid vowels raising and lowering. An example of pretonic vowel raising in Brazilian Portuguese is the pronunciation of the word /feˈɾidɐ/ 'wound' with the pretonic /e/ as [i] instead of [e]. Abaurre-Gnerre (1981) argues that the raising favors vowel reduction, and may cause the word /feˈɾidɐ/, for instance, be pronounced as /ˈfɾidɐ/, with the complete suppression of the pretonic vowel. An example of pretonic vowel lowering is the pronunciation of the word /foˈfɔkɐ/ 'gossip' with the pretonic /o/ as [ɔ] instead of [o]. The author argues that the lowering has the opposite effect of vowel reduction since the bigger opening of the vocal tract in lower vowels increases the sonority.

Vowel reduction is often called on to explain rhythmic differences between languages and varieties. The idea is that the higher incidence of vowel reduction favors the occurrence of consonant clusters and more complex syllabic types, with complex onsets and codas, which, in turn, increases variation in syllable duration, a stress-timing feature. On the other hand, a higher incidence of processes that preserve vowels and break up consonant clusters (such as vowel epenthesis in Brazilian Portuguese) decreases variation in syllable duration, a syllable-timing feature.

According to Abaurre-Gnerre (1981), the factors that mostly affect the frequency of these rhythmically opposing phonological processes in Brazilian Portuguese are style (formal and informal) and the speakers' geographical origin. She suggests that a higher incidence of pretonic mid vowel lowering causes the speech rhythm in the state of Bahia (Northeast Brazil) to be more syllable-timed. On the other hand, the speech in the state of

Espírito Santo (Southeast Brazil) is more stressed-timed, according to the author, since the frequency of pretonic mid vowel raising is higher. If we follow this reasoning, then we can expect that the speech rhythm in Alagoas is more syllable-timed than that in the state of São Paulo, because Northeastern varieties of Brazilian Portuguese show significantly higher rates of pretonic mid vowel lowering. Data from the Project *Norma Urbana Linguística Culta* (NURC Project) shows 60% of lowered pretonic vowels in Salvador (capital of Bahia) and 47% in Recife (capital of Pernambuco, another Northeastern state of Brazil), while in Rio de Janeiro and São Paulo the frequencies are, respectively, 5% and 0% (Barbosa, Papa, et al., 2019). Data from the Accommodation Project show that the pretonic mid vowels /e/ and /o/ are significantly lower in the speech of Alagoan migrants than in the speech of lifelong native Paulistas (Oushiro, 2019b, p. 687).

Metalinguistic evaluations by Northeastern migrants living in São Paulo also point to rhythmic differences between Alagoans and Paulistas. Some examples are found in sociolinguistic interviews conducted by the Accommodation Project (Oushiro, 2018):[1]

(1) D1: *mas eles reconheciam que você era nordestino ou alagoano mesmo?*
S1: *é... eles falavam e só quando a gente conversa eles percebe mesmo que a gente é alagoano... porque muda o jeito de falar né... é diferente do dos modo (dos) cada um cada um tem um ritmo eu tenho o meu quem é baiano já tem o jeito diferente né [...] só que tem muitos alagoanos aqui [em São Paulo] que já já puxa pro lado paulista já...* (MB-, RicardoR)[2]

(2) S1: *é porque a gente paraibano a gente se conhece porque aqui em Paulínia mesmo é quase (acho que) tudo lá da minha cidade então a gente se conhece todos e eu conheço né é ah um paraibano você conhece de longe quando você se você for um paraibano é abriu a boca filha o paraibano fala muito alto e muito rápido é sem falar no o sotaque o jeito de falar né?* (FB+, MarisaR)[3]

[1] All excerpts from this corpus identify the interviewer by "D1" and the speaker by "S1". Speakers are identified by a pseudonym and their sociolinguistic profile: F = female; M = male; A = speaker migrated before 20 years old; B = speaker migrated when she was 20 years old or older; - = less than 10 years in the host community; + = 10 years or more in the host community.

[2] Translation of example (1): 'S1: but did they recognize that you were from Northeast Brazil or Alagoas? D1: they spoke and only when we talk they notice that we are from Alagoas... because the way of speaking changes right? it is different from the the way (the) each one has a rhythm... I have mine... who is Baiano has a different way of speaking right? [...] but many Alagoans here [in São Paulo] are already speaking like Paulistas...'

[3] Translation of example (2): 'it is because we from Paraíba we know each other because here in

(3) D1: *tem algum outro sotaque do Brasil aqui que você conhece?*
S1: *outro sotaque do Brasil? acho que um baiano poderia puxar mais falar mais... algumas palavras aqui deixa eu ver... ele fala mais devagar o baiano... mais é só na velocidade da pronúncia falaria mais devagar... já de Alagoas já é um sotaque mais acelerado* (FB-, YasminS)[4]

In these excerpts, Alagoan and Paraiban migrants that live in São Paulo comment on the differences between Northeastern and Paulista varieties referring to rhythm (RicardoR), speech rate (MarisaR and YasminR), and speech intensity (MarisaR). RicardoR mentions the process of prosodic accommodation to Paulista speech when he says: "só que tem muitos alagoanos aqui [em São Paulo] que já já puxa pro lado paulista já..." ("but many Alagoans here [in São Paulo] are already speaking like Paulistas").

Meta-comentário:
Pretendo melhorar o fechamento deste capítulo.

Paulínia almost (I think that) most are all from there my city then we know each other everyone and I know right? a Paraiban you know by far when you if you are from Paraíba it is just to start to talk... Paraibans talk too loud and too fast... not to mention the sotaque the way of speaking right?'

[4]Translation of example (3): 'D1: is there some other sotaque in Brazil that you know? S1: another sotaque in Brazil? I think that Baianos could speak more like to speak more... some words here let me see... they speak slower, the Baianos... but it is only in the pronunciation speed they would speak slower... as for Alagoas the sotaque is more accelerated'

Chapter 2

Data and methods

In this chapter, I outline the methodological procedures used in this research. I begin by discussing some challenges concerning data collection in dialect contact research. Next, in Section 2.2, I present the two speech samples that served as the raw data for this study. Then, in Section 2.3, I explain the scheme of phonetic annotation of speech recordings that allowed me to extract many acoustic measures related to prosody. From phonetically annotated speech recordings, I computed relevant acoustic measures to analyze rhythm and intonation, such as measures of variability in syllable duration and fundamental frequency. In Section 2.4, I explain what each of these measures informs about prosody and how they are mathematically computed. I also take this section to describe in detail the computational techniques used to automate speech annotation and the extraction of these acoustic measures.

2.1 Methodological issues

The study of how migrants' speech changes as a result of their daily interactions with people who speak a regionally different variety presents a twofold challenge when it comes to data collection. First, we have to determine which kind of data is the best to isolate only the patterns in the migrants' speech that emerge from the contact with a new dialect. In other words, the challenge is to find out how we can disentangle the patterns in

the migrants' speech due to long-term accommodation from all other types of speech patterns. After determining the relevant data for dialect contact research, we still have to face the second part of the challenge: how to collect this data? In this section, I will discuss the two approaches mostly used by sociolinguists to deal with this issue concerning data collection: the longitudinal and the cross-sectional approaches (Nycz, 2015).

Longitudinal data consists of repeated observations of the same population over months or years. When the observations come from the same subjects, the longitudinal data is named panel data. Otherwise, if the observations made at each time were elicited from the same population but not from the same subjects, the data is termed trend data (Cukor-Avila and Bailey, 2013; Blondeau, 2013).

Panel data seems to be the ideal type of data for dialect contact research (Nycz, 2015; Britain, 2002). Through a panel study, we can collect many samples of migrants' speech at different points in time, starting while they are still in their home communities. The comparison of samples allows for the researcher to identify the emergence of patterns after the dialect contact started. In other words, panel data enables tracking in real-time the linguistic behavior of a speaker or group of speakers as the exposure to the new dialect increases. Unfortunately, as ideal as it is to analyze dialect accommodation, panel design brings some obstacles that make it an impractical method of data collection for most researchers, and these obstacles are even more significant in dialect contact research.

The first challenge arises in the recruitment of participants. As said before, ideally, the first speech recordings should be made while the participants are still in their home communities. For a more controlled study, it is also advisable to select participants from the same home community that will migrate to the same destination. Most sociolinguistic studies also aim to balance social variables to investigate how they condition the accommodation to the new dialect, making the recruitment even more complex.

It is quite hard to find people that naturally match these recruitment

criteria. How will the researcher find persons that are still not migrants but are planning to migrate soon? Furthermore, the experimental method is not ethically suitable for this kind of research. It must be evident that it is not morally acceptable to induce a group of people to migrate for the sake of experimental control. As noticed by Nycz (2015), these practical difficulties of participant recruitment led most dialect contact studies with a longitudinal approach to be restricted to a small sample with few participants (Tagliamonte and Molfenter, 2007; Chambers, 1992; Trudgill, 1986).

Other challenges of panel design involve the duration of research. To fully explore the consequences of migration in the speech of a group, we should get in contact and repeat the interview with the same participants several times over months and years. Cukor-Avila and Bailey (2013) point out how the mobility of populations makes finding the same speakers again a difficult task. As time goes by, some participants can move to other regions, some can give up participating until the end of the study, and others can pass away. In addition, most researchers do not have such a long time to collect their data, nor the necessary funding to cover the expensive costs involved in a years-long study.

In face of the challenges of longitudinal data collection, many studies on dialect contact have turned to cross-sectional alternative approaches (Oushiro, 2020; A. J. Walker, 2014; Payne, 1976). Cross-sectional data are observations collected at a single point in time. Longitudinal studies compare the behavior of the same population over time. In contrast, cross-sectional studies compare the behavior of two or more different populations at the same point in time. In dialect contact research, cross-sectional studies try to isolate contact-induced patterns in migrants' speech by comparing their speech with that of other groups.

The most common cross-sectional design adopted by sociolinguists involves collecting three speech samples. One of them is the target sample, and the others are control samples. The target sample consists of recordings of the speech of a group of migrants that share the same home and

host communities. In other words, all of them are native speakers of the same regional variety and are in the same dialect contact situation. In turn, the control samples are composed of the speech of non-migrant groups: lifelong residents of the host community and lifelong residents of the migrants' home community. The purpose of control samples is to serve as references to distinguish dialect contact patterns in the target sample.

The strategy of the cross-sectional design is to estimate the degree to which the migrants linguistically accommodate to the host community, assessing the differences and similarities between the migrants' speech (the target sample) and the non-mobile reference groups (the control samples). Accommodation can be estimated from two points of view. One is to examine the frequency of the host community's specific linguistic features in migrants' speech. The recurrent use of these features by migrants is evidence of dialect acquisition. The degree of accommodation can also be estimated not from the perspective of acquisition but of "loss" of linguistic features by examining if linguistic forms specific to migrants' native community are less frequent in their speech than in lifetime residents' speech. The decrease in the frequency of these features is evidence that the migrants favor alternative forms, probably the forms with which they have more contact in the host community.

Length of residence in the host community is a significant variable to analyze dialect accommodation (Siegel, 2010). We can reasonably think that the longer the speaker has lived in the host community, the more likely she will have acquired the linguistic features of this community. Panel studies measure the effects of length of residence on dialect acquisition by tracking the same speakers over the years. Instead, the cross-sectional approach to investigate this variable is to balance the samples recruiting speakers with different lengths of residence. The logic is similar to that of the apparent-time approach to linguistic change (Cukor-Avila and Bailey, 2013). We hypothesize that the patterns synchronously observed among speakers of different lengths of residence mirror the modifications in the speech of the community of migrants over the years (i.e., real-time

modifications).

Another relevant variable is the age of arrival; that is, the speaker's age when she arrives in the host community. Based on considerations related to the critical period hypothesis discussed in section 1.1, we can think that the younger the speaker is when moving to the new community, the bigger her success in acquiring the dialectal features of this community. Both approaches measure the effects of age of arrival recruiting speakers that arrived in the host community at different ages and searching for a correlation between age of arrival and success in dialect acquisition. However, in the panel approach, the recruitment by the age of arrival also conditions another variable, participants' age, making the sampling harder. Since the researcher should select the participants before they move, the age of arrival should be equal to or slightly higher than the participant's age in the first collection.

The cross-sectional approach has undeniable practical advantages over the longitudinal methods of data collection. Time issues are probably the main reasons that lead researchers to opt for this approach. It enables collecting speech recordings in a short time while allowing for the control of relevant variables for dialect accommodation study. Besides these time issues, the primary reason that led me to choose this methodological approach was the availability of the Accommodation Project's cross-sectional sample of the speech of Alagoan migrants, collected specifically for the study of dialect acquisition (see Section 2.2).

The practical advantages do not come without some limitations. Britain (2018) argues that the cross-sectional design is a *post-hoc* study since the phenomenon is analyzed after it already occurred. In the absence of data about the migrants' speech before they moved to the host community, we need to hypothesize that their speech before migration matched the speech of current lifelong residents from the home community. However, Nycz (2013) shows that there can be reasons for this hypothesis not to be true. For instance, being less attached to the local region, mobile speakers could have a less regionally marked speech than their non-mobile peers. Furthermore,

we also assume that the speech patterns among native speakers from the host community reflect the stimuli received by the migrants. Nevertheless, Nycz (2013) argues that it can be the case that the migrants we study have a social network in the new community consisting predominantly of people from other regions.

It is ultimately impossible to solve these issues since we cannot in principle access the participants' speech before migration nor the linguistic input they receive in the host community. But we can still lessen these obstacles by gathering information about speakers' social networks and their identity relations with their host and home communities. This information can serve as a background to more accurately interpret linguistic patterns in migrants' speech. As described in the next section, this research relies on the prosodic analysis of ALCP, a sample of Alagoan migrants' speech in Campinas. In addition to its balance for the age of arrival and length of residence, this sample also has extensive social data about the participants, including information about their social networks and identity relations to the states of São Paulo and Alagoas.

2.2 Samples

In this research, I take the cross-sectional approach for the analysis of prosodic accommodation in the speech of Alagoan migrants in São Paulo. I have analyzed two speech samples, both collected by other researchers in previous projects. One of them is the ALCP sample, collected between 2016 and 2018 by the *Laboratório Variação, Identidade, Estilo e Mudança* (VARIEM) of University of Campinas (UNICAMP), as part of the Accommodation Project (Oushiro, 2018). The other is a large corpus named SP2010 Project (Mendes and Oushiro, 2012), collected between 2012 and 2013 by *Grupo de Estudos e Pesquisa em Sociolinguística* (GESOL) of University of São Paulo (USP). In the following lines, I will describe the main characteristics of these samples.

Meta-comentário:
Originalmente, o projeto iria analisar também o corpus do Projeto PORTAL (coordenado pelo Prof. Alan Jardel da UFAL), com gravações de alagoanos não migrantes, como segunda amostra de controle. No entanto, o CEP não autorizou a utilização desse corpus, pois não foi solicitado aos participantes o consentimento para que os dados pudessem ser usados por outros pesquisadores em pesquisas futuras.

2.2.1 ALCP sample

The ALCP sample is composed of audio recordings of the speech of 23 Brazilian speakers from the Northeastern state of Alagoas who moved to the Metropolitan Area of Campinas, in the Southeastern state of São Paulo. This sample is especially suitable for this research because the Accommodation Project (Oushiro, 2018) collected it specifically for the study of dialect accommodation phenomena in the speech of internal migrants in Brazil.

According to Oushiro (2018), recruitment for the ALCP sample was based on the "friend of a friend" sampling technique (Milroy and Gordon, 2003), a strategy frequently used by sociolinguists. The objective of this technique is to facilitate entering a community. In this method, the researcher uses his own social network to recruit a first potential participant, by asking her friends if they know someone who fits the study's inclusion criteria. By snowballing, the researcher then asks the first participant to recommend another potential participant, and so on. In this way, the researcher approaches new candidates, not as a complete outsider to the community, but as a "friend of a friend", a strategy with the advantage of reducing the rate at which candidates decline to participate (Milroy and Gordon, 2003).

The ALCP sample was collected according to three control variables. To reduce the effects of potential confounding variables (Gries, 2013), Oushiro (2018) controlled participants' profiles in relation to age, education, and geographical area. The author ensured that all participants had lower levels of education (achieving a high school diploma or less), had come from rural

areas, and were between the ages of 19 and 65 (i.e., retirees and persons in school-age were excluded). Table 2.1 lists the 23 participants of the ALCP sample by their pseudonyms, along with their sex, age, and city of origin in the state of Alagoas. The sample has a majority of male speakers (62%) at the age of 33 years old on average. None of the participants has higher education, and many are functionally illiterate, with less than eight years of compulsory schooling in Brazil (37%).

Table 2.1: Social profile of the ALCP sample's participants

Participant	Sex	Age	City of origin	Job
ArthurG	Male	36	Arapiraca	Restaurant assistant
JorgeS	Male	33	Arapiraca	Security guard
JosueA	Male	23	Arapiraca	Cooker
LucimaraF	Female	38	Arapiraca	Housewife
MarleneN	Female	23	Arapiraca	Manicure
RicardoR	Male	27	Arapiraca	Bartender
WellingtonF	Male	21	Arapiraca	Weiter
YasminS	Female	35	Arapiraca	Cashier
AlbertoS	Male	39	Campo Grande	Self-employment
CleoniceC	Female	27	Campo Grande	Painter's assistant
MoniqueR	Female	32	Campo Grande	Factory worker
IsabelaS	Female	60	Coité do Noia	Kitchen assistant
JefersonL	Male	34	Coité do Noia	Bricklayer
JoseO	Male	28	Delmiro Gouveia	Security guard
DanielS	Male	19	Girau do Ponciano	Waiter
FabianoJ	Male	22	Girau do Ponciano	Technician
WalterN	Male	24	Girau do Ponciano	Waiter
EraniceS	Female	25	Inhapi	Attendant
EraldoF	Male	38	Jaramataia	Car mechanic
VeronicaS	Female	42	Marechal Deodoro	Assistant
JoaquimS	Male	42	Not available	Hawker
JosiasP	Male	63	Piaçabuçu	Retired
AdrieleS	Female	38	São José da Lage	Unemployed

Almost all participants have come from the countryside of Alagoas, except for JosiasP from Piaçabuçu and VeronicaS from Marechal Deodoro, two coastal cities. Compared to the Metropolitan Area of Campinas, which has more than 3 million inhabitants (of which 1.2 million are residents of the

city of Campinas), the native cities of the participants are much smaller. The biggest is Arapiraca, with approximately 230,000 inhabitants. All the others have no more than 50,000 residents, with Jaramataia, the native city of EraldoF, being the smallest, with only 5,000.

Besides the city of Campinas, with the biggest population, the Metropolitan Area of Campinas also comprises smaller cities. Of the 23 participants, 14 are residents of the city of Campinas (61%), and the others live in one of the four nearby towns: six speakers in Holambra; two in Sumaré; one in Indaiatuba, and one in Vinhedo.

Apart from LucimaraF, CleoniceC, and AdrieleS, who migrated to live near their relatives, all the other participants are labor migrants (IOM, 2019); that is, they moved to the state of São Paulo seeking employment. At the time the recordings were made, most of them had a low-skilled job in Campinas, working as waiter, manicure, kitchen assistant, hawker, etc., activities that in Brazil make up a large part of the informal sector.

As said before, the ALCP sample was collected with the purpose of studying dialect accommodation. In particular, the collection was made so that the effects of the age of arrival and length of residence could be disentangled (Oushiro, 2020). Studies on dialect contact consider these two variables as important factors to explain the degree of success in the acquisition of new dialect features by migrants (Siegel, 2010; Trudgill, 1986; Chambers, 1992; Bortoni-Ricardo, 1985; Payne, 1976). However, the effects of these two variables can be easily confounded since, as explained by Oushiro (2020), “it is often the case that speakers who arrived earlier are also the ones who have lived the longest in the host community”.

To distinguish the effects of age of arrival and length of residence, the speakers were balanced according to a binary definition of these two variables. Age of arrival divided them into those who arrived in São Paulo before they were 20 years old and those who were 20 or older. In turn, length of residence arranged the participants in those who had been living in São Paulo for less than 10 years and those who had been living there for 10 years or more.

The balancing of these binary variables ensures the orthogonality between them. From the statistical point of view, we can say that two categorical variables are orthogonal if each factor in a variable can co-occur with each factor of the other (Guy and Zilles, 2007). Table 2.2 shows the crossed distribution of the ALCP's speakers by age of arrival, length of residence, and sex. What is important here is to notice that there is no empty cell in the table, with at least two participants filling each combination between the factors of these three variables.

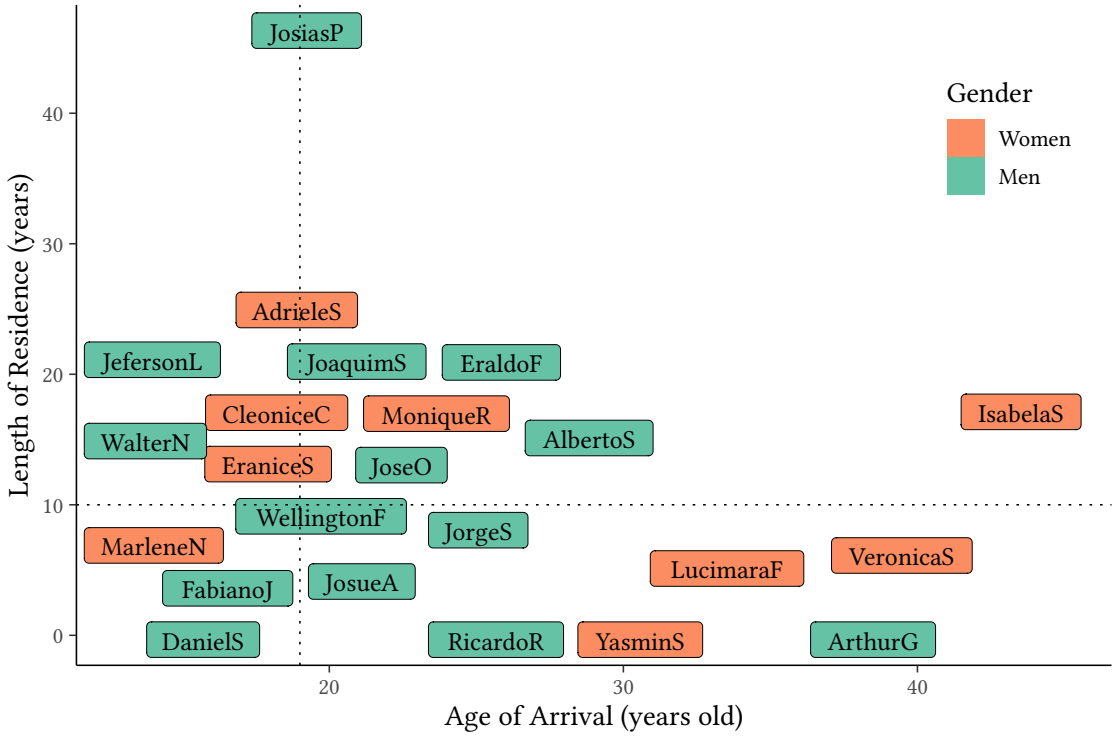
Table 2.2: Distribution of the ALCP sample's participants by sex, age of arrival and length of residence

Length of Residence		Age of Arrival	
		Before 19 years	20 years or more
Less than 9 years	Female	EraniceC MarleneN	LucimaraF VerônicaS YasminS
	Male	DanielS FabianoJ WellingtonF	ArthurG JoséO JosuéA RicardoR
10 years or more	Female	AdrieleS CleoniceC	IsabelaS MoniqueR
	Male	JoaquimS JosiasP WalterN JeffersonL	EraldoF JorgeS AlbertoS

Although the participants were balanced according to the binary definition of age of arrival and length of residence, these variables are also reasonably balanced when they are treated as continuous. Figure 2.1 illustrates the distribution of the speakers by these two variables in continuous scales in years. The difference in sex is also shown. The distribution is not as balanced as in Table 2.2, but, as noted by Oushiro (2018), the imbalance does not appear to be large enough to interfere with their statistical analyses

as continuous.

Figure 2.1: Distribution of the ALCP sample's participants by sex, age of arrival, and length of residence



The recording schedule used for the collection of the ALCP sample comprised a sociolinguistic interview, the reading of a word list, and two questionnaires. Except for a married couple that was recorded together (MoniqueR and AlbertoS), all participants were individually recorded by native-speaking Linguistics students from the state of São Paulo. The recordings were made in different places according to the preference and the availability of the participants. The digital handy recorder Tascam DR-100MKIII and two Shure SM93 lavalier microphones were used, one for the interviewee and the other for the interviewer. The audio files were written in the uncompressed WAVE format and stored in the VARIEM repository at University of Campinas. In total, the sample is composed of 23 hours and 35 minutes of speech recordings.

Before the recording starts, the participants were asked to give their informed consent to participate in the research and for their speech to be recorded. It is also noteworthy that their identities were preserved by changing their names to pseudonyms in the transcripts and deleting from

both the audio files and transcripts any mention of their names or any other information that could potentially be used to recover their identities. These ethical procedures were also adopted in the collection of the SP2010 sample (see section 2.2.2). In this text, I will refer to specific speakers from these samples only by their pseudonyms.

In the first part of the schedule, the participants of the ALCP were recorded in sociolinguistic interviews. Each interview lasted about one hour and was structured in thematic modules. Each module comprised an ordered set of questions about some aspect of the participants' life. According to Oushiro (2020, p. 80), the interviewer asked them "about their current neighborhood in the community of destination, childhood in their home state, family, work and leisure activities, as well as the Metropolitan Area of Campinas".

Besides the primary goal of eliciting more spontaneous speech, by encouraging the participant to narrate some of his personal experiences, the interview questions were also devised to gather relevant information for the study of dialect accommodation. Oushiro (2018) explains that the questions had the special purpose of discovering who the persons with whom the participants interact the most in their daily life are; what their personal evaluations of the Alagoas state, the São Paulo state, and Northeast Brazil are; which sociolinguistic variables are more salient to them and which is below their level of consciousness; what their migration trajectories are (to which locations they have moved, how long they lived in these places, etc.); and what their overt evaluations of regionally different ways of speaking (*sotaques*) in Brazil are.

By the end of the interview, some participants were asked to read a word list. Since some of them were illiterate, this task could not be applied to all of them. In this study, the reading recordings were not analyzed. The participants were also asked to respond to two questionnaires. The first one is the Questionnaire of Social Network, Habits, and Identity, based on Hoffman and Walker's (2010) Ethnic Orientation Questionnaire. It comprised eleven multiple-choice and five scale questions aimed at quan-

tifying how much the participant identifies herself with her community of origin (in the case of the ALCP's speakers, Alagoas) and also with her community of destination (São Paulo), and how strong are her ties with these communities. The second questionnaire was socioeconomic, and optional to be filled out. It was applied by the Accommodation Project to better evaluate possible differences in the accommodation to Paulista speech due to differences concerning the social mobility of the migrants (Oushiro, 2018).

2.2.2 SP2010 SAMPLE

Ideally, the control sample of the speech of the host community should be a sociolinguistic sample with native lifetime inhabitants of the Metropolitan Area of Campinas since it is the region where ALCP's migrants live. However, until now, no one has collected such a sample. For this reason, I have decided to use the SP2010 corpus as the control sample of this research. The SP2010 Project is a robust spoken corpus of the Paulistano Portuguese; that is, the Portuguese variety spoken by native lifelong residents of the city of São Paulo (Mendes and Oushiro, 2012). Closer to the coast, it is approximately 90 kilometers away from Campinas.

The participants of the SP2010 corpus are lifetime residents of the city of São Paulo. The project selected only speakers born in the town or who moved to it before ten years old. The recruitment aimed to balance three social variables: (i) sex (males, females); (ii) age group (20–34 y.o., 35–59 y.o., 60 y.o. or more); and (iii) level of education (up to high school, college). The possible combinations between the factors of these three variables give rise to 12 social profiles (two sex × three age groups × two levels of education). Five speakers were recruited for each social profile, adding up to the 60 speakers of the corpus.

The corpus covers one-hour-long recordings of sociolinguistic interviews. The structure of the interview is similar to that of the ALCP sample, described in section 2.2.1. According to Mendes and Oushiro (2012), the

interview has two parts. The first covers more personal questions about the speakers' daily life (neighborhood, childhood, family, work, and leisure activities) to elicit relatively unmonitored speech, with more narrative forms. The second part deals with more "objective" questions about the city of São Paulo, such as problems of urban life and possible solutions, to elicit a monitored style of speech, in which the speaker adopts a more argumentative discourse. At the end of the interview, the speakers were recorded reading a word list, a newspaper excerpt, and an oral text.

Meta-comentário:
Aqui pretendo discutir possíveis diferenças entre a fala campineira e a paulistana.

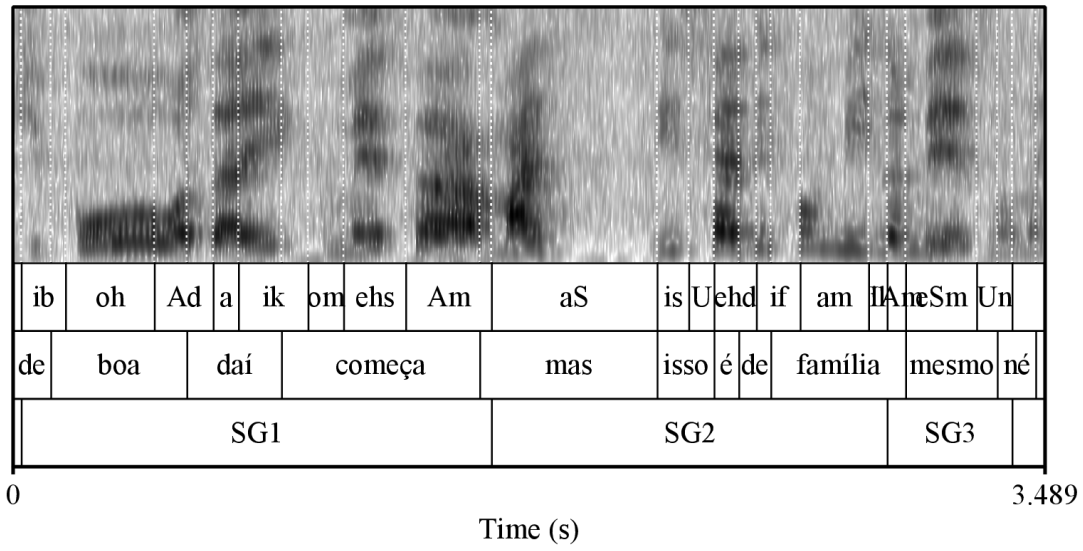
2.3 Speech annotation

In this research, I study prosodic accommodation using acoustic techniques of speech analysis, and annotation is an essential step for this type of analysis. Most relevant variables for phonetic and linguistic study depend on additional information to the audio signal, that is, information involving abstract knowledge about language and speech. Since the speech recordings encode only data concerning oscillations in pressure, the researcher herself needs to provide additional data based on linguistic theory. In this sense, speech annotation is a specialized type of metadata that involves associating descriptive and analytical notations with speech recordings, according to conventional rules grounded by theoretical assumptions (Ide and Pustejovsky, 2016).

In phonetic research, an essential feature of annotation is time alignment. The annotation must be time-aligned to the audio signal so that the researcher can verify to which point in the recording each piece of annotation refers. Time alignment is necessary to automate searches in spoken corpora and extractions of acoustic measures from audio files. With a time-aligned annotated corpus, we can use available computational techniques for finding specific linguistic units and extracting measurements.

Time-aligned linguistic annotation of speech recordings involves computer programs like Praat (Boersma and Weenink, 2021). Paul Boersma and David Weenink (University of Amsterdam) have developed Praat, the most used software for acoustic analysis of speech. Praat has tools for time-aligned annotations and allows for the automatic extraction of acoustic measures. Furthermore, it is free and open-source.

Figure 2.2: Spectrogram of utterance segmented in V-to-V units (first tier), words (second tier) and stress groups (third tier) in Praat



Figures 2.2 is an example of speech annotation in Praat. Annotation consists of textual content arranged in multiple *tiers*. The researcher can use tiers to annotate different types of information. In this figure, I have used the first tier for segmentation into syllable-sized units (see Section 2.3.1), the second for words, and the third to demarcate stress groups (see Section 2.3.2). Although Praat supports the use of IPA symbols, in this research, I opt for a phonetic transcription system in ASCII characters to ensure compatibility between computer programs. The tiers align to the timeline of the audio signal, with word boundaries exactly matching the beginning and the end of the word's occurrence in the audio.

In this research, to extract prosodic data from the ALCP and SP2010 samples, I annotated the speech recordings using computational tools and following a conventional annotation scheme grounded on phonetic

assumptions. Then, I extracted a group of relevant acoustic measures from the annotated recordings. These procedures were not executed manually. Instead, I automated them by developing scripts in Python and Praat's programming languages. I also used scripts developed by other researchers (Kruse and Barbosa, 2020; Silva Jr and Barbosa, 2019; Barbosa, 2020; Gorman, Howell, and M. Wagner, 2011; Young et al., 2015). Without automation, this research would not have been feasible within the time constraints of a master's research[1]. In what follows, I describe the procedures used to annotate ALCP and SP2010 samples.

Speech annotation for sociophonetic analysis is divided into two tasks: segmentation and labeling of annotation units. Annotation units are vowels, consonants, syllables, words, phrases, utterances, and any other unit posited by linguistic and phonetic theories. Segmentation consists of "slicing" the speech signal's continuum into these units by marking their boundaries in the tiers. Then, a specific label is given to each "slice." In this master's research, I adopted an annotation scheme with four units: syllable, word, stress group, and chunk, each annotated on a different tier. The word tier makes it easier to locate specific excerpts in the audio. The other three units were required to extract prosodic measures from the ALCP and SP2010's recordings.

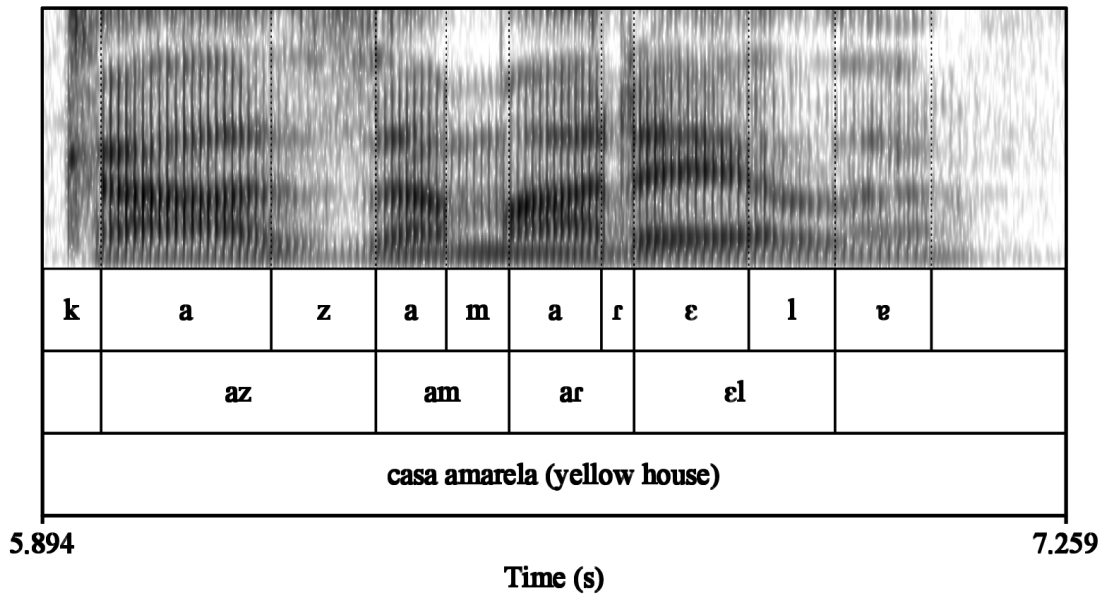
2.3.1 Syllables

This research assumes the syllable as the minimal unit of speech production. This premise draws upon the fact that speakers are not capable of producing constituents smaller than the syllable, and hearers usually cannot correctly identify vowels and consonants unless they hear them

[1]The ALCP and SP2010 samples add up to 83 sociolinguistic interviews altogether, each one with an average duration of one hour. They amount to 4,980 hours of speech recording. The extraction of prosodic variables depends on the demarcation of syllable-sized units. Assuming that people produce on average five syllables per second, we can grossly estimate there would be 89 million syllables to be manually annotated (4,980 hours × 60 minutes × 60 seconds × 5 syllables). Of course, the participant's speech does not fill the entire recording — i.e., there are also silences, the interviewer's speech, and maybe other events that interrupt the conversation. But even if we divide this number in half, it is still too many syllables to be manually annotated.

within their syllabic context (Barbosa, 2006). However, I do not analyze prosody based on the traditional phonological conception of the syllable. In phonology, the syllable is usually a hierarchical structure with a vowel (or a glide) as its nucleus optionally preceded (onset) and followed (coda) by consonants (Zec, 2007; Mendonça, 2003). In place of the phonological syllable, I analyze the V-to-V unit, a syllable-sized unit defined as the acoustic interval between two consecutive vocoid onsets (Barbosa, 2009). In other words, the V-to-V is the interval from a vocoid onset until immediately before the next vocoid's onset. Therefore, despite its name, this unit comprises a single vocoid. Instead of a phonological syllable, the V-to-V never starts with consonants (i.e., consonants always follow the vowel onset).

Figure 2.3: Spectrogram of the noun phrase *casa amarela* 'yellow house' exemplifying speech segmentation in vowels and consoants (first tier) and V-to-V units (second tier) using Praat.



Being a phonetic unit, the V-to-V has to be demarcated based not on the abstract phonological representation of lexical items but on phonetic data of each particular production. Figure 2.3 shows an example of segmentation into vowels and consonants in the first tier and V-to-V units in the second one. We can verify that the syllable /zɐ/, ending the word *casa* 'house', merged with /a/, starting the word *amarela* 'yellow', into one

single syllable. Besides, the vowel /ɐ/ ending the word *amarela* cannot form a new V-to-V unit, since there is not a next vocoid to be used as the ending boundary. Thus, this phrase has only four V-to-V units: [az], [am], [aɾ] and [ɛl].

The V-to-V unit is supported by experimental studies showing that vowel onset is the best candidate to serve as syllable boundary (Barbosa, 2007; Barbosa, Arantes, et al., 2005; Pompino-Marschall, 1989). Since the 1940s, phoneticians have used vowel onsets to demarcate the limits of syllable-sized units and stress groups (Lehiste, 1970). The transition from the low-energy spectral region of consonants to the high-energy of vowels is the most acoustically prominent area for syllable marking (Barbosa, 2006). Figure 2.4 shows the spectrogram of a recording of my speech in which I produce the word *patologia* 'pathology', with segments labeled as consonant (C) or vowel (V). The yellow contour indicates the variation of spectral energy in decibel (dB). We see that the regions where the rate of change in spectral energy is the highest (i.e., where the rising contour becomes steeper) are precisely the transition from a consonant to a vowel.

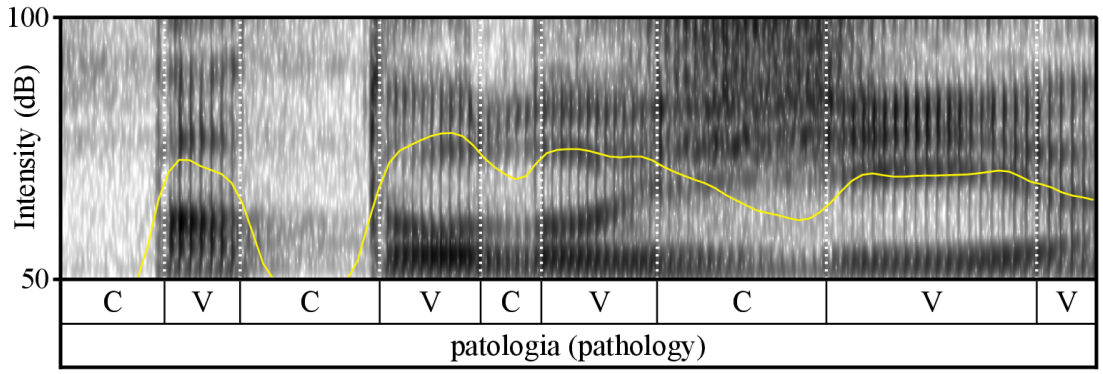
Experimental studies on the perceptual center, such as Pompino-Marschall (1989) and Barbosa, Arantes, et al. (2005), support the V-to-V unit. They show that when asked to produce a sequence of syllables (e.g., "pa, pa, pa, ...") in synchrony with a metronome, speakers try to synchronize the metronome's beats with the vowel onsets. These studies suggest that the high energy contrast of the C-V transition draws the speaker's attention, being used as a reference point of synchronization.

Since this research analyzes the V-to-V unit in place of the phonological syllable, from now on, whenever I refer to syllable, I mean the V-to-V. References to the phonological syllable will be explicitly qualified as so.

2.3.2 Stress groups

The syllable is the minimal but not the only prosodic unit considered in this research. Syllable groupings constitute larger units that also integrate

Figure 2.4: Spectrogram of the production of the word *patologia* 'pathology', segmented in vowels (V) and consonants (C), exemplifying the increase of acoustic energy in the transition from consonant to vowel. The yellow contour indicates the variation in intensity (decibel). Figure produced using Praat.



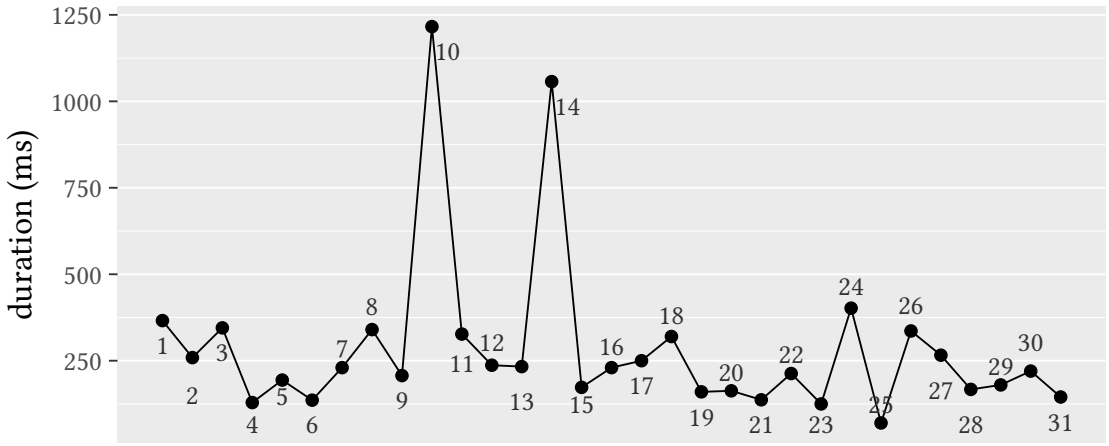
speech prosody, such as the stress group. A stress group comprises one prominent syllable, called phrasal accent, and one or more adjacent non-prominent syllables (Fletcher, 2010; Barbosa, 2006; Crystal and House, 1990). It is noteworthy that phrasal accents here do not refer to lexical stresses, but to *prosodically* stressed syllables. In other words, a phrasal accent is a syllable whose prominence is not restricted to distinguishing words but to emphasizing a syllable in relation to the neighboring ones beyond word boundaries (Barbosa, 2012). In Portuguese, phrasal accents are perceived as closing the stress group, but in some languages, like English, they start it. A stress group ended by a phrasal accent consists of a sequence of non-prominent syllables followed by a single prominent one.

Ideally, the segmentation of stress groups should be guided by perceptual data. For instance, to identify phrasal accents, we can ask native speakers to listen to a speech recording and mark on the orthographic transcription words that sounded more salient to them. Then, the lexically stressed syllables of most marked words are chosen as the best candidates for phrasal accents. As ideal as this approach is, it has the practical disadvantage of requiring the conduction of a perception experiment. Fortunately, Barbosa (2006) developed a normalization procedure of syllable duration that is ca-

pable of identifying phrasal accents with an accuracy rate higher than 80% (Barbosa, 2007, p. 732). That is, more than 80% of stress groups perceived by native hearers are predicted by the normalized syllable duration.

It is easier to understand this normalization procedure through a kind of graphical representation of syllable duration called *duration contour* (Barbosa, 2006). In this type of plot, the horizontal axis does not represent time, but only the order of syllables in the utterance, and the vertical axis indicates the duration of each syllable. When the contour indicates the absolute syllable duration (usually in milliseconds), it is called raw duration contour. Figure 2.5 shows the raw contour of an utterance from one of the ALCP's recordings. However, absolute syllable duration is not a reliable representation of prosodic salience. If we conduct a perception experiment like the one just described above, the phrasal accents perceived by the participants possibly would not match all duration peaks in Figure 2.5.

Figure 2.5: V-to-V raw duration contour of an utterance from Accommodation Project.



The reason for this mismatch is that rhythmic salience is not about if a syllable is long or short in absolute terms, but if it is longer or shorter than expected by the hearer (Barbosa, 2019). Some sounds are intrinsically shorter or longer than others. For example, in Brazilian Portuguese, the voiceless alveolar fricative /s/ is intrinsically longer than many other

consonants. Thus, a Brazilian hearer does not perceive a long /s/ as salient just because it is longer than the voiced alveolar plosive /b/, since /s/ is expected to be longer than /b/[2]. The same goes for syllable duration differences between stressed and unstressed vowels in Portuguese. The lengthening of unstressed vowels is more salient than of stressed ones, since speakers already expect that a stressed vowel shows more variability in duration. For instance, the intrinsic standard deviation of stressed /a/ vowel is 45 msec, while the unstressed one is only 15 msec (Barbosa, Violaro, et al., 1999). In short, the intrinsic duration of segments must be "neutralized" by a normalization procedure.

Barbosa (2007) normalizes syllable duration applying two common statistical techniques: standard scores (z-scores) and weighted moving average of five points. Normalization by z-scores is used to convert absolute values into relative ones. A z-score unit indicates how far each value is from the mean in units of standard deviation. For example, three z-scores means that the value is three standard deviations far from the mean.

The mean and the standard deviation used to calculate z-scores do not need to be from the same sample that will be normalized. Since the salience of syllable duration relates to how far the duration is from an expected value (i.e., the intrinsic duration), we can represent this distance by transforming absolute duration in z-scores, taking the intrinsic mean and standard deviation of each segment as reference. Doing this, the normalized values will indicate not the syllable duration in milliseconds, but how far the duration is from the expected in units of standard deviation. In other words, a long syllable that is expected to be long will have its duration value reduced and a short syllable that is longer than expected will have its value increased.

Equation 2.1 shows the formula to normalize syllable duration by z-scores. The normalization involves applying this equation for each syllable of the utterance. The index i refers to each syllable and j to each segment within the syllable. Thus, d_i is the absolute duration of a syllable i. The variables

[2]In Brazilian Portuguese, the mean intrinsic duration of the voiced alveolar plosive /b/ is 59 msec, against 96 msec of voiceless alveolar fricative /s/ (Barbosa, Violaro, et al., 1999).

μ_j and σ_j refer respectively to the mean and the standard deviation of the duration of each segment j within the syllable i. First, the sum of the means of the intrinsic duration of all segments inside the syllable j is subtracted from the absolute duration of this syllable. The outcome of this step is the absolute difference in milliseconds between observed and expected duration. This difference, then, is divided by the sum of the standard deviations of the intrinsic duration of the same segments. The division transforms the absolute difference in milliseconds into the relative difference in units of standard deviation, that is, in z_i (i.e., z-score of the syllable i).

$$z_i = \frac{d_i - \sum_{j=1}^{n} \mu_j}{\sum_{j=1}^{n} \sigma_j} \tag{2.1}$$

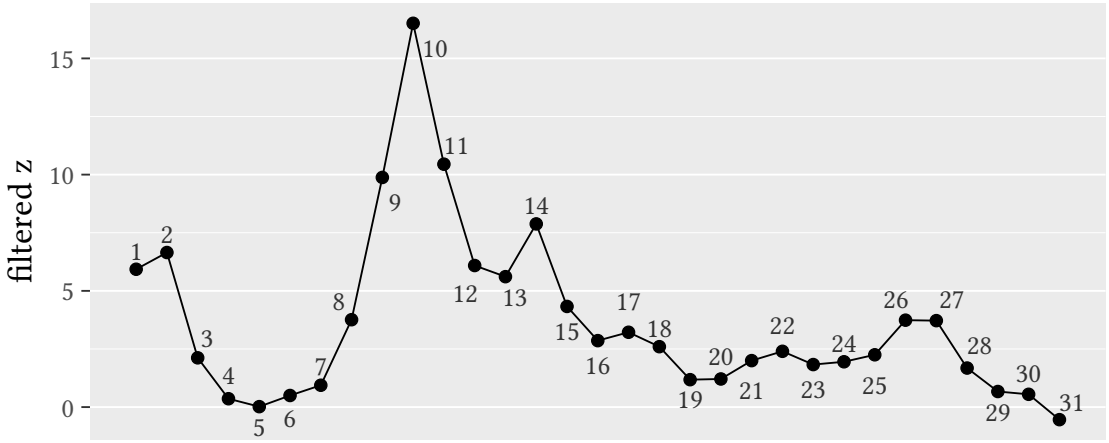
When normalized by z-scores, the duration contour is closer to the perception of rhythmic salience. However, minor fluctuations in syllable duration can still produce small peaks that do not correspond to prominent syllables. For this reason, the contour must be smoothed using a weighted moving average of five points. For each value in the contour, it is necessary to calculate the mean of the current value with the two values on the left and the two others on the right. Then, the target value (the one that is in the middle) is replaced by the outcome. Weighted moving average uses the weighted arithmetic mean in the place of the simple mean. Before computing the mean, the middle value is multiplied by 5, the two adjacent values by 3, and the two others by 1. Then, we sum the values and divide by the sum of the weights (i.e., 5 + 3 + 3 + 1 + 1 = 13). This same procedure is applied to all values of the contour, and that is why it is called *moving* average. The equation 2.2 represents the application of five points moving average to z-scores.

$$\text{smoothed } z_i = \frac{1 \cdot z_{i-2} + 3 \cdot z_{i-1} + 5 \cdot z_i + 3 \cdot z_{i+1} + 1 \cdot z_{i+2}}{13} \tag{2.2}$$

Figure 2.6 shows the normalized and smoothed version of the duration

contour of Figure 2.5. We see that many peaks disappeared. Syllables 3, 5, 8, 18, and 24 appear as peaks in the raw duration contour but not in the normalized one. On the other hand, syllables 2, 17, and 27 do not stand out in the raw duration contour, but, in the normalized values, they are longer than expected. From the normalized and smoothed duration contour, we can identify phrasal accents and demarcate stress groups in speech recordings.

Figure 2.6: V-to-V normalized (z-scores) and smoothed (5-points moving average) duration contour of the same utterance of Figure 2.5.



2.3.3 Chunks

As already said, this research analyzes 23 acoustic measures (see Section 2.4). Some of these measures have a scope of application fixed by definition. For instance, the scope of the number of syllables in a stress group is, evidently, the stress group; that is, this measure is, by definition, always calculated within the limits of a stress group. On the other hand, some prosodic measures do not have a fixed scope of application. For example, the speech rate is the number of syllables produced per second. We can calculate the speech rate within a stress group, but also within an utterance, or even within larger chunks of speech. Actually, chunks of any length can be used to calculate speech rate. The same occurs with the median of the fundamental frequency (f_0). We can calculate it from f_0 values within

one second or 30 minutes of speech recording. I name these variables *free scope* variables, in contrast with *fixed scope* ones.

In this research, the chunk is the name of the annotation unit that I use to compute free scope prosodic variables. There are some reasons for not trying to extract free scope variables from too small or too large intervals. If the measurement is extracted from too small an excerpt, it can be the case that the data is not enough to accurately represent the speech. On the other hand, a too large scope makes the measure more insensitive to variations in time. For example, one single measurement of speech rate can be taken in a speech recording of 20 seconds. Another possibility is to divide the recording into four intervals of five seconds and compute the speech rate four times, one for each interval. The point here is that possible differences in speech rate between these 5-seconds intervals are lost when only a single measurement of 20 seconds is taken.

There is no infallible criterion to establish the length of the chunk. It is important, however, to ensure consistency between measures. That is, to establish the range of the scope and use it for all measurements. In this research, all chunks are intervals ranging from 4 to 8 seconds.

2.4 Prosodic measures

Meta-comentário:
Esta seção ainda precisa ser melhorada. Do modo como está, ela traz apenas uma descrição de como cada medida é calculada. Pretendo discutir também como as medidas têm sido interpretadas na literatura fonética.

This research analyzes 23 acoustic measures concerning rhythm and intonation. In this section, I will describe these measures, but adopting a more succinct exposition and leaving a more in-depth discussion of each of them for Chapter 3, where the results are reported.

2.4.1 Rhythm measures

Speech rate It measures how many times a unit occurs per second. Usually, the unit is the syllable, but it can also be another linguistic unit, as words or stress groups. In this research, the speech rate is the number of V-to-V units per second within a chunk. It is computed diving the number of V-to-V units by the total duration of the chunk, including silent pauses.

Stress group measures In this research, I analyze two measures related to stress groups: (i) stress group duration (in milliseconds); and (ii) the number of V-to-V units in stress groups.

Rhythm metrics

The following seven measures join the group of the so-called rhythm metrics. As explained in Section 1.3, all these metrics are alternative ways of calculating the degree of variation in the duration of a phonetic unit. Even though some of them were first proposed for a specific phonetic unit (e.g., the Rhythm Ratio was created with syllables in mind), they can all be applied to the duration of intervals associated with different units, like vowels, consonants, and syllables. Since I analyze these metrics only in application to syllable-sized intervals (the V-to-V), I will describe them using the syllable as a reference, even though not all were originally proposed by their authors with the syllable in mind.

Fuchs (2016) divides the rhythm metrics into two groups: global and local. The author explains that global metrics compute the variability in duration "by taking into account how all the intervals in an utterance or a sentence differ from each other" (Fuchs, 2016, p. 40). Out of the metrics analyzed in this research, Δ and Varco are global metrics. On the other hand, the local metrics still compute the variability considering all the intervals in an utterance, but not all the differences between each other. Instead, they consider only the differences of adjacent intervals. rPVI, nPVI, RR, VI, and YARD are the five local metrics analyzed in this research.

These metrics can also be divided between raw and normalized measures. The normalization aims to minimize the effects of speech rate. As explained by Dellwo (2006), intervals associated with vowels, consonants, and syllables tend to be longer in lower speech rates than in higher ones, and longer intervals cause higher absolute variation in duration. In other words, the syllable duration in slow speech will probably be more variable than in fast speech. However, many researchers use rhythm metrics to compare the rhythm between languages or varieties *independently* of speech rate. For this reason, normalized versions of raw metrics were proposed. The normalized metrics basically convert the raw duration into a proportion. For instance, instead of saying that the syllable duration varies on average 30 milliseconds, with a normalized measure, it would be said that the syllable duration varies 70% of the mean syllable duration of the utterance.

Delta (Δ) Proposed by Ramus, Nespor, and Mehler (1999), this rhythm measure is basically the standard deviation applied to the duration of intervals (here, V-to-V units) within a chunk. Higher values of Δ are interpreted as indicating a more stress-timed rhythm. The formula for Δ is the following:

$$\Delta = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n}} \tag{2.3}$$

Where,

d_i = Duration of syllable i

$\bar{d}$ = Mean syllable duration within the chunk

n = Number of syllables

Variability Coefficient (Varco) Proposed by Dellwo (2006), this measure is a normalized version of Δ. To attenuate speech rate effects, Varco divides Δ by the mean syllable duration so that the variability in duration

is expressed as the proportion in relation to the mean. In doing so, the value will be more reduced with lower speech rates (that tends to have a larger mean) than with higher ones. Then, the value is multiplied by 100, only to get more manageable numbers. The formula for Varco is the following:

$$\text{Varco} = \frac{\Delta \times 100}{\bar{d}} \tag{2.4}$$

Where,

$\bar{d}$ = Mean syllable duration within the chunk

Raw Pairwise Variability Index (rPVI) Proposed by Low, Grabe, and Nolan (2000), this measure first computes the differences in duration within all pairs of adjacent intervals. Then, it calculates the mean of these differences in absolute value (i.e., negative values are transformed in positive). The formula for rPVI is the following:

$$\text{rPVI} = \frac{\sum_{i=1}^{n-1} |d_i - d_{i+1}|}{n-1} \tag{2.5}$$

Where,

d_i = Duration of the first syllable of the pair

d_{i+1} = Duration of the second (adjacent) syllable of the pair

n = Number of syllables

Normalized Pairwise Variability Index (nPVI) Proposed by Grabe and Low (2002), this measure is the normalized version of rPVI. The only difference from rPVI is that nPVI divides each difference within each pair of adjacent syllables by the mean syllable duration of the respective pair. As in Varco, the final value is multiplied by 100 to get more manageable numbers. The formula for nPVI is the following:

$$\text{nPVI} = 100 \times \frac{\sum_{i=1}^{n-1} \left| \frac{d_i - d_{i+1}}{(d_i + d_{i+1})/2} \right|}{n-1} \tag{2.6}$$

Where,

d_i = Duration of the first syllable of the pair

d_{i+1} = Duration of the second (adjacent) syllable of the pair

n = Number of syllables

Rhythm Ratio (RR) Proposed by Gibbon and Gut (2001), RR is essentially the PVI, but using the quotient in the place of the difference. If the duration of the first syllable of the pair is shorter than the duration of the second, it divides the second by the first. On the contrary, if the duration of the second is shorter, it divides the first by the second. After computing the quotients of all adjacent intervals, it calculates the mean of the absolute values of these quotients and multiplies the result by 100. This is the only measure among the ones analyzed in this research in which higher values indicate more syllable-timing. The higher the RR, the more proportional the syllable duration. An RR of 100 indicates perfectly isochronous syllables. The formula for RR is the following:

$$\text{RR} = 100 \times \frac{\sum_{k=1}^{n-1} \left| \frac{d_i}{d_j} \right|}{n-1} \tag{2.7}$$

If $d_k < d_{k+1}$, $d_i = d_k$ and $d_j = d_{k+1}$

Else, $d_i = d_{k+1}$ and $d_j = d_k$

Where,

d_k = Duration of the first syllable of the pair

d_{k+1} = Duration of the second (adjacent) syllable of the pair

n = Number of syllables

Variability Index (VI) Proposed by Deterding (1994), VI differs from nPVI only by using the global mean syllable duration (instead of the local) as reference for normalization. Before calculating the difference within pairs of adjacent syllables, it divides the duration of each syllable of the pair by the mean duration of all syllables of the utterance. The formula for VI is the following:

$$\text{VI} = \frac{\sum_{i=1}^{n-2} \left| \frac{d_i}{\text{mean}(d)} - \frac{d_{i+1}}{\text{mean}(d)} \right|}{n-1} \tag{2.8}$$

Where,

d_i = Duration of the first syllable of the pair

d_{i+1} = Duration of the second (adjacent) syllable of the pair

$\text{mean}(d)$ = Mean syllable duration within the chunk

n = Number of syllables

Yet Another Rhythm Determinant (YARD) Proposed by P. Wagner and Dellwo (2004), YARD is similar to PVI, but using z-scores for normalization. It essentially consists of normalizing the duration of syllables transforming their values in z-scores and then applying the rPVI to the z-scores. Therefore, the variability is expressed in terms of units of standard deviation. The formula for transforming syllable duration in z-scores is the following:

$$z_i = \frac{d_i - \bar{d}}{s} \tag{2.9}$$

Where,

z_i = z-score of syllable i

d_i = Duration of syllable i

$\bar{d}$ = Mean syllable duration within the chunk

s = Standard deviation of syllable duration within the chunk

2.4.2 Intonation measures

I analyze 13 acoustic measures concerning the contour of the fundamental frequency. They can be arranged in three groups: (i) measures applied globally to all f_0 values throughout the chunk; (ii) measures concerned only with peaks of f_0; and (iii) measures that indicate the rate of change in the fundamental frequency (i.e., derivatives of f_0).

Global f_0 measures This group comprises five common statistics applied to all f_0 values within a chunk, that, in conjunction, give a general description of the distribution of these values. The measures are the following: (i) median; (ii) standard deviation of f_0; (iii) semi-amplitude between quartiles (also known as interquartile range); (iv) minimum value; and (v) maximum value.

Peaks of f_0 The second group comprises four measures that describe the behavior of f_0 peaks within a chunk: (i) standard deviation of f_0 peaks; (ii) width of f_0 peaks; (iii) f_0 peak rate; and (iv) standard deviation of interval between adjacent f_0 peaks.

Derivatives of f_0 The last group refers to rates of change in the fundamental frequency. In mathematics, the rate of change of a function is computed by derivatives. Thus, a derivative of f_0 simply indicates how fast the values of f_0 change in time. If the derivative is positive, then f_0 is rising. If negative, f_0 is falling. In the graphical representation of the f_0 contour, in which the x-axis is time and the y-axis is the frequency in Hz, the derivatives correspond to the inclination (usually called *slope*) of the contour. In this research, I analyze the mean and the standard deviation of positive and negative derivatives of f_0 within chunks.

2.5 Automation

Meta-comentário:
Para a versão final da dissertação, vou expandir esta seção, explicando passo-a-passo como automatizei a segmentação e a etiquetagem das gravações.

As said in Section 2.3, to compute the measures described above from many speech recordings, I used computational tools of automation. For the speech annotation of ALCP and SP2010's recordings, I developed a script in Python that implements techniques of speech recognition based on hidden Markov models. These techniques allow for the automatic phonetic alignment of speech recordings. From the user's point of view, the script is simple. It requires audio files of speech recordings accompanied by their orthographic transcription in EAF format (i.e., the format of annotation files in ELAN). Then, it automatically extracts chunks of the speech of the participant (ignoring the interviewer's parts) ranging from 4 to 8 seconds from the speech recordings. The script uses the chunks to train a model of speech recognition using the Hidden Markov Models Toolkit (HTK) (Young et al., 2015; Gorman, Howell, and M. Wagner, 2011). The trained model, then, is used to automatically segment and label the chunks in vowels, consonants, and V-to-V units. The segmentation and labeling are written in TextGrid files (i.e., the format of annotation files in Praat), which can be read in Praat. It is important to mention that, for the development of this script, I relied on the source code of the AlinhaPB, a phonetic aligner for Brazilian Portuguese developed by Kruse and Barbosa (2020), and the ProsodyLab-Aligner (Gorman, Howell, and M. Wagner, 2011). The annotation of the chunks in stress groups was made by using the script SGDetector, developed by Barbosa (2013), to be executed in Praat.

The extraction of the acoustic measures was also automated. For the extraction of rhythm metrics, I use the Praat script *Metrics and Acoustics Extractor* by Silva Jr and Barbosa (2019). All the other measures were extracted using Barbosa's (2013) Praat script *Prosody Descriptor Extractor*.

Chapter 3

Results

In this chapter, I report the results of mixed-effects linear regression analyses of 23 acoustic variables concerning speech prosody. I first discuss the results about rhythm, in Section 3.1, and then I move on to intonation in Section 3.2. In both sections, I adopt the same order of exposition. First, I analyze prosodic differences between Alagoan migrants and lifelong native Paulistanos, with models having sample (ALCP, SP2010) and gender (women, men) as predictor variables. Then, I move to the second type of analysis, focused on the effects of migration on the migrants' prosody. The models of this second analysis were constructed only from the ALCP data. Besides gender, they also include the age of arrival (before 19 years old, 20 years old or older) and length of residence (less than 10 years, 10 years or more) as predictor variables. In all models reported in this chapter, speakers are included as random effects.

The first analysis aims to determine to which degree the Alagoan migrants' prosody differs from that of the native Paulistanos. The aim is to use the results as references to guide subsequent investigations about the effects of dialect contact on the prosody of Alagoan migrants. Depending on these results about the prosodic differences and similarities between the two samples, different perspectives of analysis can be adopted. For instance, if the results show that the migrants do not differ in any way from the Paulistanos' when it comes to the prosody of speech, then the following analyses should focus on discovering if the Alagoan migrants

have completely accommodated to the prosody of São Paulo or if their prosody were already similar to that of the Paulistanos when they arrived in São Paulo — i.e., if there is no significant difference between the prosody in both states of São Paulo and Alagoas. If, on the contrary, the two samples differ in the prosodic patterns, the following analyses should investigate to what degree they have accommodated to the Paulista prosody and, in certain cases, why some migrants have not accommodated to the prosody of São Paulo.

In the intermediate cases, in which the prosody of the Alagoan migrants variably differ from the Paulistanos' in some aspects, but not in others, and in some speakers, but not in all of them, the goals that should guide further investigations should be to discover if the variation can be explained in terms of factors related to dialect contact; that is, factors that could potentially inhibit or catalyze the acquisition of new prosodic patterns by migrants. Based on the results of the comparison between ALCP and SP2010, the second type of analysis reported here adopts this perspective of searching for conditioning factors of the prosodic accommodation in a dialect contact situation. More specifically, I explore if the migrants' age when they arrived in the state of São Paulo and the number of years they have lived in this state can explain at least partially the prosodic variation in their speech.

Since the analyses reported here involve many variables, I use the same strategies of exposition for all of them, as a way of facilitating the reading and the comparison between results. Throughout the chapter, I discuss the prosodic variables in groups. For each group, I bring a graphical representation of the results in boxplots and a table reporting regression models. As for the tables, it is especially important to notice that each of them brings the information of more than one linear regression model. Each column shows estimates and coefficients of determination (R^2) of a mixed-effects model for one of the variables of the group. However, all models of each table share the same predictor variables.

3.1 Rhythm

I present rhythm variables in two groups, as shown in Table 3.1. First, I discuss the results of speech rate, duration of stress groups, and the number of syllables in stress groups. Then, I analyze in conjunction the group composed of seven metrics of rhythm.

Table 3.1: Groups of rhythm variables

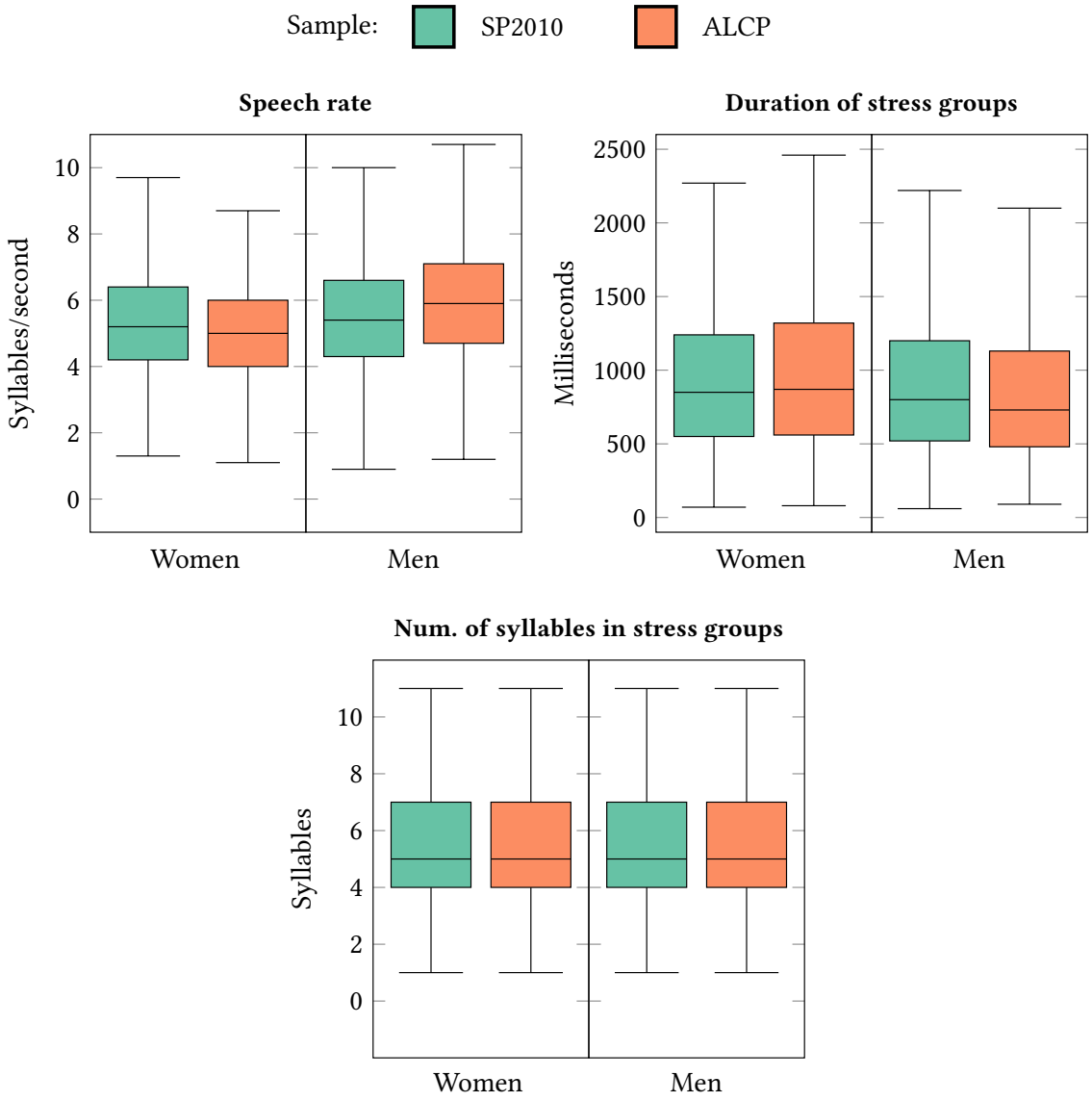
Group 1	Group 2
Speech rate	ΔS
Stress group duration	Variability Coefficient (Varco)
N° of syllables in stress group	Raw Pairwise Variability Index (rPVI)
	Normalized Pairwise Variability Index (nPVI)
	Rhythm Ratio (RR)
	Variability Index (VI)
	Yet Another Rhythm Determinant (YARD)

3.1.1 Rhythmic differences between ALCP and SP2010

Figure 3.1 shows the distribution of speech rate (in syllables per second), duration of stress groups (in milliseconds), and the number of syllables in stress groups, in the speech of males and females from the ALCP and SP2010 samples. In the three variables in both samples, there is great dispersion. The speech rate has values as low as one and a half syllables per second (syl/sec) and as high as ten syl/sec. The duration of stress groups ranges approximately from 50 to more than 2,000 msec, and the extreme values of the number of syllables in stress groups are separated by ten syllables.

Table 3.2 reports the results of three mixed-effects linear regression models, one for each rhythmic variable. The Intercept refers to the *absolute* estimate of the rhythmic variable in Paulistano women's speech. It is a reference group to which other levels of the predictor variables are compared. In all tables reporting linear models in this chapter, the Intercept is specified just below the table.

Figure 3.1: Distribution of speech rate (N=4,459), duration of stress groups (N=18,931) and number of syllables in stress groups (N=18,931) in the SP2010 (in green) and ALCP (in orange) samples.



Paulistano women speak on average at a rate of 5.1 syl/sec. The rows below the Intercept show the *relative* estimate of other levels of the predictor variables. The estimate is relative in the sense that the value is the difference from the Intercept. On average, there is no difference in speech rate between SP2010 and ALCP. Alagoan migrants and native Paulistanos have a 5.1 syl/sec estimated speech rate. On the other hand, in both samples, men speak on average at a rate of 0.4 syl/sec faster than women. The asterisk beside the estimate indicates that the difference from the Intercept is statistically significant under an alpha level (α) of 0.05. Therefore, the difference between men and women in speech rate is statistically significant. Descriptively, we see in the box plots that this difference is greater within ALCP sample, but there is no significant interaction between gender and sample (p = 0.09) in another linear regression model.

Table 3.2: Mixed-effects linear regression models of speech rate (N=4,459), stress group duration (N=18,931) and number of syllables in stress groups (N=18,931).

	Speech rate (syl/sec)	Stress group duration (msec)	N° of syllables in stress group (syl)
marginal R^2	1%	0%	0%
conditional R^2	18%	3%	0%
Intercept	5.1	981	5.2
Sample: ALCP	0	-13	-0.1
Gender: Men	+0.4 ⋆	-56 ⋆	0

⋆: statistically significant (α = 0.05)
Intercept: SP2010 (Sample) + Women (Gender)
Random effects variable: Speaker

The duration of stress groups for men is 56 msec shorter than for women, whose stress groups last on average 981 msec. As for the number of syllables in the stress group, it is clear from the box plots and reinforced by the linear model that there is no difference at all in this variable between samples and genders. In the four groups of participants, the stress groups have a median of approximately 5 syllables.

The p-value under the significance level should not be the only criterion to

evaluate the importance of a predictor in explaining variation in prosodic variables. In particular, statistically significant differences alone do not mean that they are noticeable to hearers, nor that the model fits the observed data well.

Significant differences can be small enough for native hearers not to be able to perceive them. Thus, we have to ask if native Brazilians can notice a difference of 0.4 syl/sec in speech rate and 56 msec in stress group duration. Quené (2007) examined the Just Noticeable Difference (JND) for speech rate variation in read speech in Dutch. He argues that the JND is around 5% of variation for inter-speakers differences (for intra-speaker, the value is higher). If we adopt this value for Brazilian Portuguese, then the difference of 0.4 syl/sec could be considered noticeable, insofar as it amounts to an increase of approximately 8% in relation to women's estimated speech rate of 5.1 syl/sec.

Among studies on JND of segmental duration, the threshold varies from 10 to 40 msec, but more studies are necessary to examine if these values can be applied to stress group duration. It can be the case that larger prosodic units have larger JND, requiring more than 40 msec to be noticeable. For this reason, I do not consider the 51 msec difference between men and women in stress group duration as prosodically relevant.

Statistically significant differences between levels of predictors are not indicators of goodness of fit. Even with significant estimates, a model can still be poorly fitted to the observed data. That is, it can explain a too-small proportion of the variation in the response variable. For this reason, the first two rows of the tables reporting linear models bring coefficients of determination (R^2), statistics that indicate the proportion of the variance in the response variable explained by the model. Marginal R^2 represents the variance explained by fixed effects. In the case of the models of Table 3.2, sample and gender are the fixed effects variables. Conditional R^2 indicates the variance predicted by both fixed and random effects. Speaker is the only random effect in this research. An R^2 of 100% means that the values expected by the model completely match the observed values; that is, the

model completely explains the variance in the response variable.

Although gender differences in speech rate and stress group duration are statistically significant, none of the models can be considered good predictors of their response variables. The best of them, the model of speech rate, can barely predict 1% of speech rate variance when only fixed effects are considered. When Speaker is included, the predictive power of the model increases to 18%, an interesting result suggesting that speech rate is more related to differences between subjects than between social groups. The predictive power of the other two models is virtually zero.

Figure 3.2 contains boxplots of the seven rhythm metrics, contrasting Alagoan migrants and native Paulistanos. Since I did not verify any gender difference in these metrics (see Table 3.3), I opt not to distinguish this variable in the plots. As explained in Section 2.4, in this research, I applied these metrics to the duration of V-to-V units. They are alternative strategies to calculate the degree of variation (in this sense, they are dispersion measures) in the duration of V-to-V units within a chunk. Except for Rhythm Ratio (RR), higher values in all of them indicate more variable syllable duration, which points to a more stress-timed rhythm. RR differs because it computes not the difference between syllables duration, but the quotient. Higher values of RR mean more proportional syllable duration, which characterizes a more syllable-timed rhythm.

Table 3.3 reports a mixed-effects regression model for each rhythm measure. In the seven models, the difference between Alagoan migrants and native Paulistanos is statistically significant, but not the one between men and women. ΔS and rPVI are the two measures whose values refer to raw syllable duration (in milliseconds), which makes interpretation easier. As discussed in Section 2.4, ΔS is a global measure of syllable duration variation, since it computes the dispersion in duration considering all syllables of the utterance together. On average, Alagoan migrants' syllables vary 28 msec more than native Paulistanos. Considering the local variation within pairs of successive syllables (rPVI), the local variation of raw duration is smaller than the global one (ΔS) in both samples, but Alagoan migrants

Figure 3.2: Distribution of ΔS, Variability Coefficient (Varco), Raw Pairwise Variability Index (rPVI), Normalized Pairwise Variability Index (nPVI), Rhythm Ratio (RR), Variability Index (VI) and Yet Another Rhythm Determinant (YARD), in the ALCP and SP2010 samples. (N=4,459)

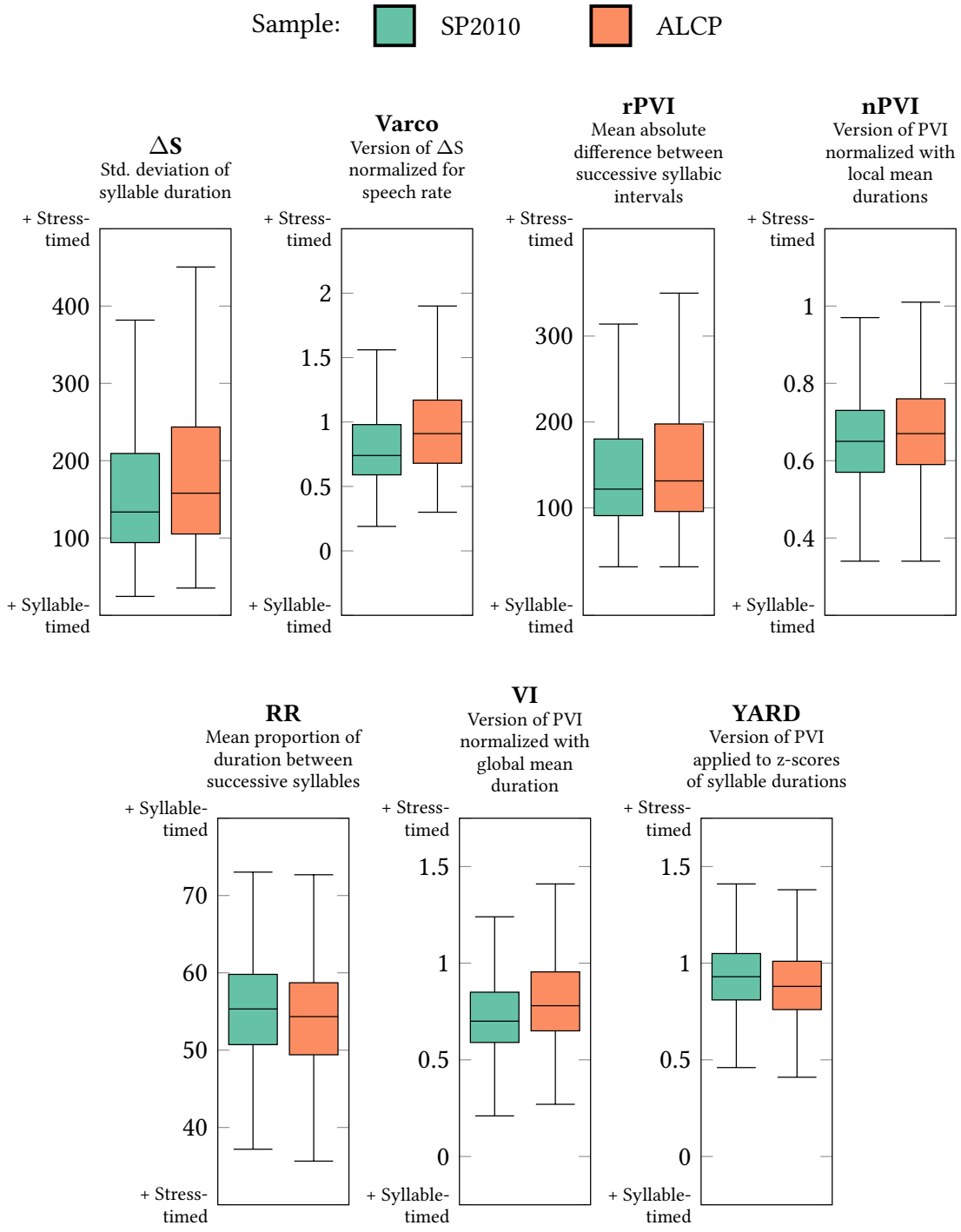


Table 3.3: Mixed-effects linear regression models of ΔS, Variability Coefficient (Varco), Raw Pairwise Variability Index (rPVI), Normalized Pairwise Variability Index (nPVI), Rhythm Ratio (RR), Variability Index (VI) and Yet Another Rhythm Determinant (YARD). (N=4,459)

	ΔS (msec)	**Varco** (%)	**rPVI** (msec)	**nPVI** (%)	**RR** (%)	**VI** (%)	**YARD** (z-score)
marginal R^2	1%	3%	1%	1%	1%	2%	0%
conditional R^2	9%	12%	7%	8%	7%	10%	3%
Intercept	174	0.8	156	0.65	55	0.73	0.94
Sample: ALCP	+28 ⋆	+0.1 ⋆	+27 ⋆	+0.03 ⋆	-2 ⋆	+0.08 ⋆	-0.02 ⋆
Gender: Men	-4	0	-5	0	-0.1	+0.01	-0.01

⋆: statistically significant (α = 0.05)
Intercept: SP2010 (Sample) + Women (Gender)
Random effect: Speaker

also show a higher value (+27 msec) than native Paulistanos.

Normalized rhythm metrics represent variation in syllable duration as a proportion. Varco is the proportion of the standard deviation of syllable duration to the mean syllable duration of the utterance. The variation in Alagoan migrants' speech is 90% of the mean syllable duration, a value higher than the 80% of native Paulistanos. The normalized correlate of local variation in syllable duration (nPVI) is also smaller than the normalized global measure (Varco). The mean variation within pairs of successive syllables is 68% of the pairs' mean syllable duration in ALCP, which amounts to 3% higher than that in SP2010.

Considering the proportion between successive syllables, RR shows that in both samples the proportion is approximately 2:1. That is, within a pair of adjacent syllables, one of them has on average twice the duration of the other. More precisely, one syllable of the pair has 53% of the duration of the other in ALCP, a value that is only 2% smaller than the value in SP2010.

VI is essentially the nPVI, but using the global mean of syllable duration as reference. Among Alagoan migrants, the variation within pairs of adjacent syllables is 81% of the global mean, while native Paulistanos

show a value of 73%. Therefore, in both samples, the local variation is higher when computed as a proportion of the global mean (VI) in the place of the local mean (nPVI). Finally, the last measure, the YARD, is the PVI applied to values of syllable duration transformed in z-scores. In the opposite direction of the other six rhythm metrics, the YARD in Alagoan migrants' speech points to a more syllable-timed rhythm compared to native Paulistanos, but the difference is small. The variation in the ALCP sample is only 0.02 z-scores less than that in SP2010.

Except for YARD, all other rhythm metrics converge. The syllable duration is significantly more variable in Alagoan migrant's speech than in native Paulistanos', a result that can be interpreted as indicating a more stress-timed rhythm among Alagoan migrants.

3.1.2 Effects of migration on speech rhythm

In this section, I analyze the same rhythm variables, but focusing on the effects of age of arrival and length of residence in the Alagoan migrants' speech. All plots and models reported in this section refer only to ALCP data. In these plots, the pair of green and orange colors is always used to distinguish the levels of the age of arrival, while the blue and pink pair refers to the length of residence. Furthermore, when the gender difference is displayed, the results for women are placed on the left, while those for men are shown on the right.

Figure 3.3 shows the distribution of speech rate, stress group duration, and the number of syllables in stress groups, concerning migrants' age of arrival, length of residence, and gender. The statistical models can be seen in Table 3.4. Concerning gender, the results are similar to those observed in the comparison between ALCP and SP2010 samples in Table 3.2. Among Alagoan migrants, men speak 0.9 syl/sec faster than women, a noticeable difference of 18%. On the other hand, the migration variables do not correlate with speech rate. Stress groups are significantly shorter (-130 ms) in men's speech than women's. I suppose that the gender differences

Figure 3.3: Boxplots representing the distribution of the speech rate (N=1,960), the duration of stress groups (N=8,510), and the number of syllables in the stress group (N=8,510), contrasting age of arrival, length of residence, and gender.

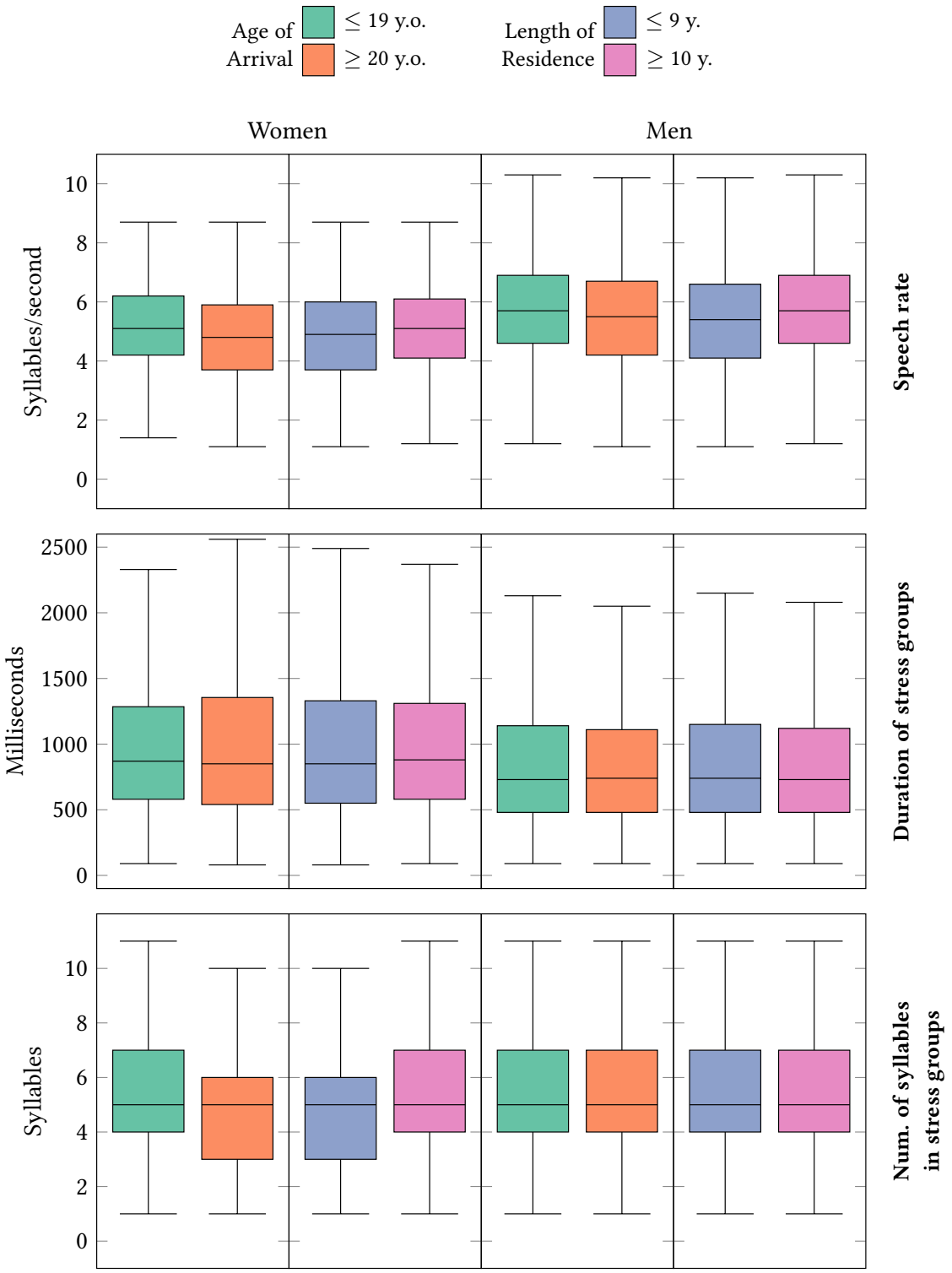


Figure 3.4: Boxplots representing the distribution of ΔS, Variability Coefficient (Varco), Raw Pairwise Variability Index (rPVI), Normalized Pairwise Variability Index (nPVI), Rhythm Ratio (RR), Variability Index (VI) and Yet Another Rhythm Determinant (YARD), in the ALCP sample (N=1,963)

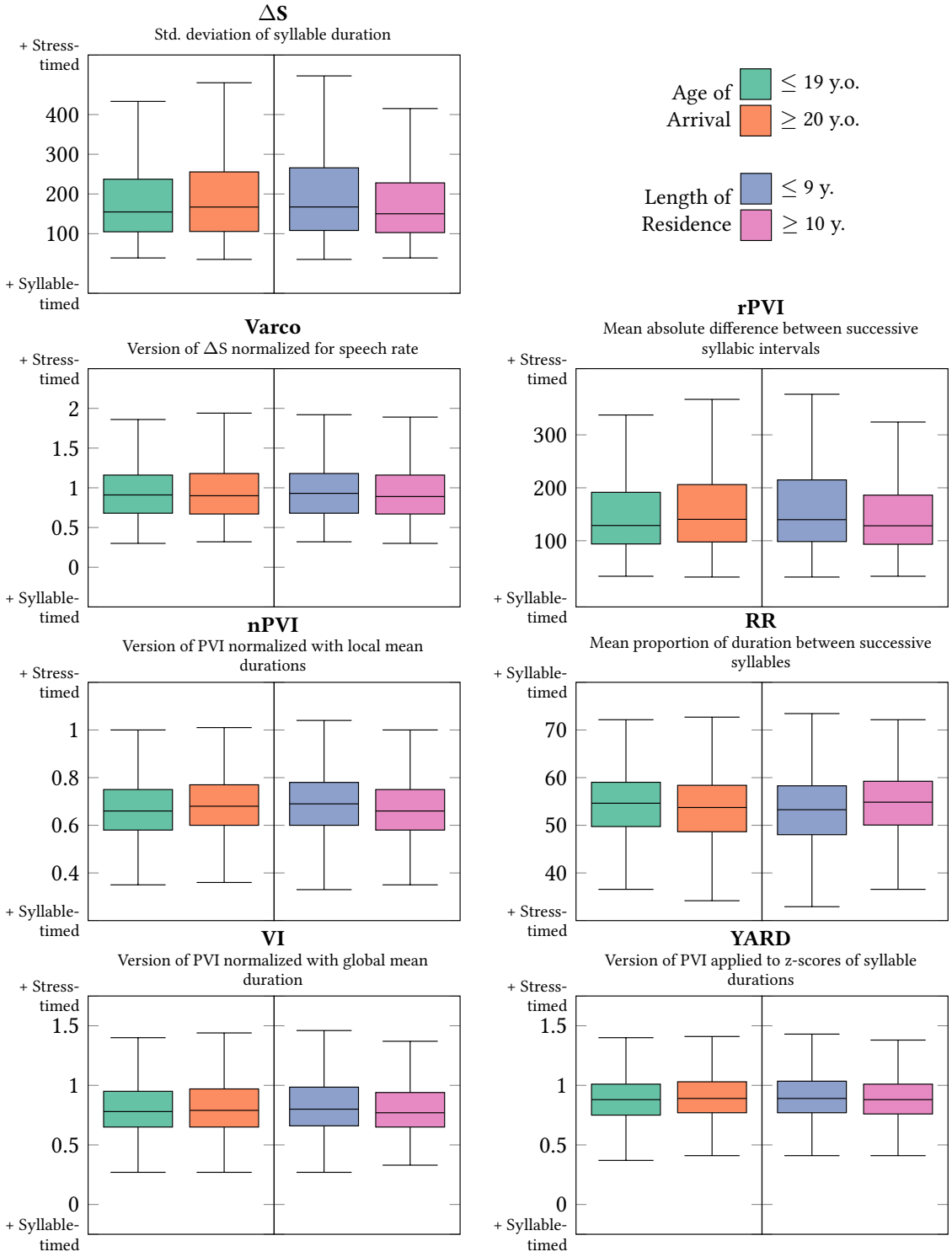


Table 3.4: Mixed-effects linear regression models of the speech rate (N=1,960), the duration of stress groups (N=8,510), and the number of syllables in the stress group (N=8,510) in the ALCP sample

	Speech rate (syl/sec)	Stress group duration (msec)	N° of syllables in stress group (syl)
marginal R^2	6%	1%	0%
conditional R^2	20%	4%	0%
Intercept	4.9	1012	5.17
Gender: Men	+0.9 ⋆	-130 ⋆	+0.09
Age of Arrival: ≥ 20 y.o.	-0.2	+14	-0.08 ⋆
Length of Residence: ≥ 10 y.	0	+7	+0.05

⋆: statistically significant (α = 0.05)
Intercept: Women (Gender) + 19 y.o or younger (Age of Arrival) + 9 y. or less (Length of Residence)
Random effect: Speaker

in speech rate and stress group duration are correlated. Shorter stress groups in men's speech can be a result of a faster speech rate.

The only significant difference concerning migration variables is verified in the number of syllables in the stress group, but it is a minuscule difference to have some phonetic or linguistic relevance. Migrants who were 20 years or older when arrived in Campinas show stress groups that have 0.08 syllables fewer than those who were younger than 20 years old. It is a meaningless difference, representing only 8% of the average duration of a syllable.

Of the three, the speech rate model has the greater predictive power, but still a low R^2. The fixed effects variables alone explain only 6% (marginal R^2) of the variance in the Alagoan migrants' speech rate. When speakers are included as random effects, the predictive power increases to 20% (conditional R^2). The other two models have lower R^2, which indicates that they do not fit well to the observed data.

Figure 3.4 illustrates the distribution of the seven rhythm metrics, contrasting age of arrival, length of residence, and gender. It is clear by the boxplots that the migration variables do not distinguish the speakers in

relation to the rhythm metrics. The regression models in Table 3.5 show no difference at all between the Alagoan migrants concerning the rhythm metrics. Furthermore, in all models, the fixed effects alone do not explain more than 2% of the variation in these metrics. That is, age of arrival and length of residence are not good predictors of the variation in syllable duration.

Table 3.5: Mixed-effects linear regression models of ΔS, Variability Coefficient (Varco), Raw Pairwise Variability Index (rPVI), Normalized Pairwise Variability Index (nPVI), Rhythm Ratio (RR), Variability Index (VI) and Yet Another Rhythm Determinant (YARD), in the ALCP sample (N=1,963)

	ΔS	Varco	rPVI	nPVI	RR	VI	YARD
marginal R^2	1%	0%	1%	2%	2%	1%	1%
conditional R^2	9%	11%	7%	8%	7%	10%	8%
Intercept	212	0.9	196	0.7	52.4	0.8	0.9
Gender: Men	-22	0.	-28	0	+1.5	0	0
Age of Arrival: ≥ 20 y.o.	+6	0	+6	0	-0.5	0	0
Length of Residence: ≥ 10 y.	-8	0	-11	0	+0.9	0	0

⋆: statistically significant (α = 0.05)
Intercept: Women (Gender) + 19 y.o or younger (Age of Arrival) + 9 y. or less (Length of Residence)
Random effect: Speaker

3.1.3 Summing up

The most important finding is the difference between the distributions of the rhythm metrics in the two analyses. The distribution of the first analysis, shown in Figure 3.2, shows that the variability in the syllable duration is greater in the Alagoan migrants' speech than in the native Paulistanos'. In Section 1.4, I discussed Abaurre-Gnerre's (1981) hypothesis that speakers from Northeast Brazil would have more syllable-timed rhythm due to higher frequency of phonological processes that inhibit vowel reduction, such as the lowering of the pretonic mid-vowels /e/ and /o/. Even though I do not analyze the relation of rhythm metrics with phonological processes, the results observed here are the opposite of Abaurre-Gnerre's hypothesis.

3.2 Intonation

Meta-comentário:
Esta seção sobre entonação é um esboço; o texto será mais bem desenvolvido para a versão final da dissertação.

In this section, I report the results about intonation dividing the variables into three groups, as shown in Table 3.6. First, I discuss five measures that summarize the central tendency and dispersion of f_0 values within chunks. Then, I report the results of four relevant variables concerning the peaks of the fundamental frequency. Finally, I analyze how fast the intonation contour rises and falls in the speech stream, describing measures about the derivatives of the fundamental frequency.

Table 3.6: Groups of intonation variables

Group 1 - Global measures of f_0
Median of f_0
Standard deviation of f_0
Semi-amplitude between quartiles of f_0 (SAQ)
Minimum f_0
Maximum f_0
Group 2 - Measures of f_0 peaks
Standard deviation of f_0 peak
f_0 peak width
f_0 peak rate
Standard deviation of intervals between f_0 peaks
Group 3 - Derivatives of f_0
Mean of positive derivatives of f_0
Mean of negative derivatives of f_0
Standard deviation of positive derivatives of f_0
Standard deviation of negative derivatives of f_0

The fundamental frequency is an acoustic parameter highly affected by anatomical differences between male and female speakers. For this reason, I decided to analyze the intonation variables separating the regression models between males and females. In other words, for each measure concerning intonation, I built a model for female speakers and another separate model for male speakers.

3.2.1 Intonation differences between ALCP and SP2010

Figure 3.5 shows boxplots of the median, standard deviation, semi-amplitude between quartiles, minimum, and maximum of the fundamental frequency in the speech of females (on the left of each graph) and males (on the right) from the ALCP and SP2010 samples. The regression models of these variables are shown in Table 3.7. The differences between males and females are striking. Male Alagoan migrants, with higher values, significantly differ from male native Paulistanos in all these measures. On average, they speak with a more high-pitched voice, and their intonation varies within a larger range. On the other hand, female Alagoan migrants do not differ from female Paulistanas in these variables, except for f_0 maximum.

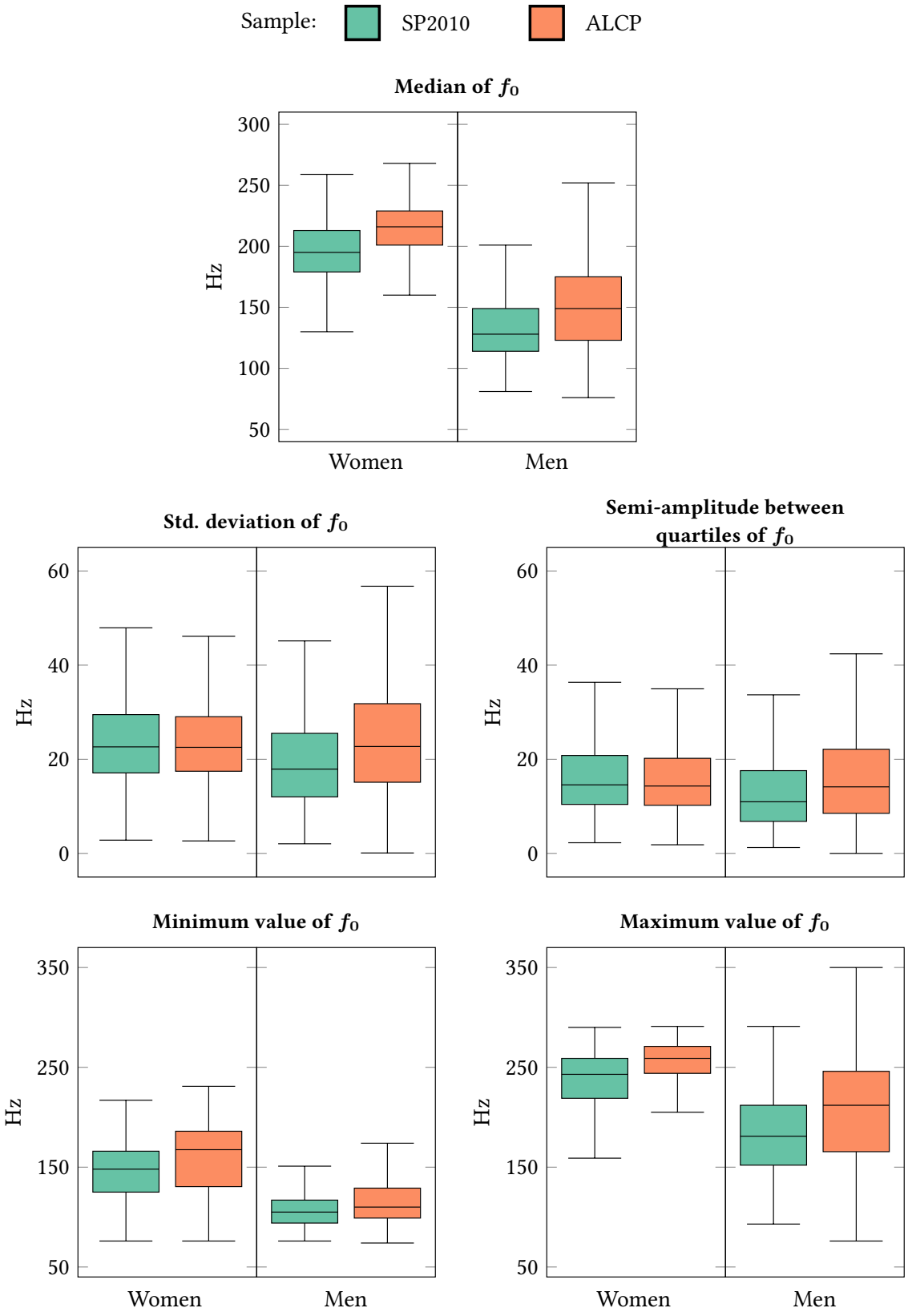
Table 3.7: Mixed-effects linear regression models of the median (Hz), standard deviation (Hz), semi-amplitude between quartiles(Hz), minimum (Hz), and maximum (Hz) of fundamental frequency (f_0) (N=4,459)

	Median (Hz)	**Std. dev.** (Hz)	**SAQ** (Hz)	**Min.** (Hz)	**Max.** (Hz)
			WOMEN		
marginal R^2	3%	0%	0%	1%	5%
conditional R^2	44%	17%	10%	32%	31%
Intercept	195	24	16.3	145	238
Sample: ALCP	+10	+2	+0.8	+7	+13 ⋆
			MEN		
marginal R^2	13%	5%	3%	5%	9%
conditional R^2	54%	19%	13%	42%	39%
Intercept	133	20	13	107	183
Sample: ALCP	+24 ⋆	+5 ⋆	+4 ⋆	+11 ⋆	+28 ⋆

⋆: statistically significant (α = 0.05)
Intercept: SP2010 (Sample)
Random effects variable: Speaker

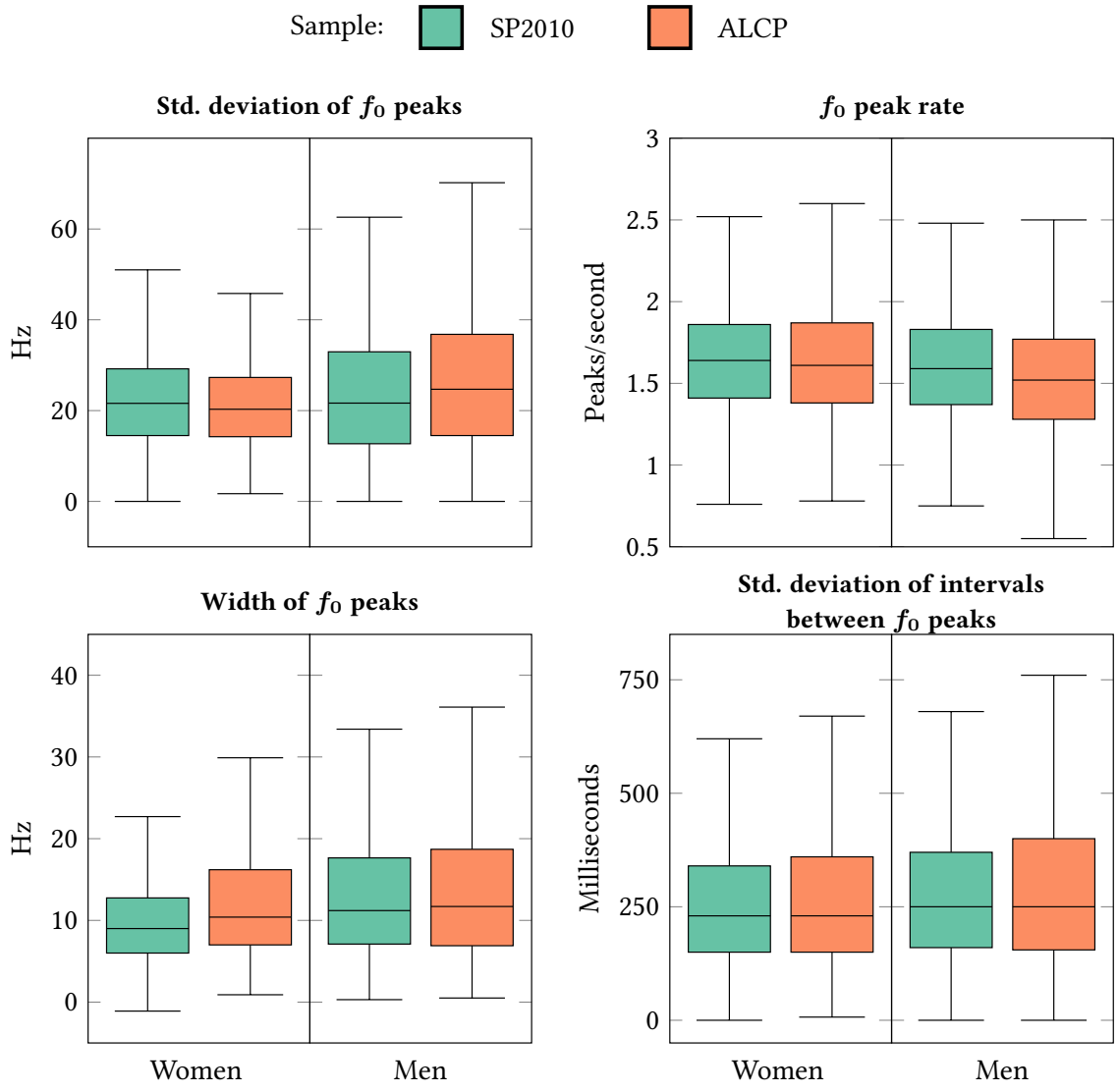
The fundamental frequency of males in the ALCP sample is on average 24 Hz higher than that of males in the SP2010. Furthermore, the variability is also greater among Alagoan migrants, as indicated by the four dispersion measures. The standard deviation of the fundamental frequency is 5 Hz

Figure 3.5: Distribution of the median (Hz), standard deviation (Hz), semi-amplitude between quartiles(Hz), minimum (Hz), and maximum (Hz) of fundamental frequency (f_0), contrasting male and female speakers from the ALCP and SP2010 samples (N=4,459)



greater in Alagoan males. The amplitude of the first and third quartiles of f_0 values is 17 Hz in the speech of male Alagoan migrants, against 13 Hz in native Paulistanos' speech. Finally, the minimum and maximum values of the fundamental frequency are also higher in Alagoan migrants' speech, with a difference of 11 and 28 Hz, respectively.

Figure 3.6: Distribution of the standard deviation of f_0 peaks (Hz), f_0 peak width (Hz), f_0 peak rate (peaks/sec), and standard deviation of intervals between f_0 peaks (ms), contrasting male and female speakers from the ALCP and SP2010 samples (N=4,459)



When we move to women's results, the situation is different. The only significant difference between ALCP and SP2010 samples is the maximum of the fundamental frequency, with female Alagoan migrants having a value of 13 Hz greater than female natives from São Paulo.

Figure 3.6 shows the distribution of variables related to peaks of the fundamental frequency. Unlike the global measures described above, the results suggest that there is not much difference between ALCP and SP2010 when it comes to peak patterns. The standard deviation of the fundamental frequency measures the variability in the height of peaks. Higher values indicate that there is a greater oscillation between taller and shorter peaks. As seen in Table 3.8, Alagoan migrants and native Paulistanos do not differ in this measure. In both groups, the estimated variation of peaks' height is around 25 Hz.

Table 3.8: Mixed-effects linear regression models of the standard deviation of f_0 peaks (Hz), f_0 peak width (Hz), f_0 peak rate (peaks/sec), and standard deviation of intervals between f_0 peaks (ms). (N=4,459)

	Peak std. dev. (Hz)	Peak width (Hz)	Peak rate (peak/sec)	Std. dev. of inter-peaks duration (msec)
WOMEN				
marginal R^2	0%	3%	0%	0%
conditional R^2	6%	11%	2%	2%
Intercept	23	11	1.7	267
Sample: ALCP	+1	+3 ⋆	-0.02	+5
MEN				
marginal R^2	1%	0%	0%	0%
conditional R^2	10%	6%	2%	1%
Intercept	24	13	1.6	268
Sample: ALCP	+3	+0.3	-0.07 ⋆	+11

⋆: statistically significant (α = 0.05)
Intercept: SP2010 (Sample)
Random effects variable: Speaker

Peak width is a measure of "peakness." In other words, higher values indicate that a peak of the fundamental frequency is "sharper." Instead, peaks with lower values of width are more flattened. Mathematically, it is the vertical distance in Hz between the maximum value and the mean of the fundamental frequency 30 ms before and after the maximum. Female Alagoan migrants' peak width is significantly higher than that of female

natives from SP2010. There is no difference in this measure among male speakers.

Although the model shows a significant difference in peak rate between males from ALCP and SP2010, it is a too small number to be considered phonetically relevant. Alagoan migrants have an estimated value of only 0.07 peak/sec less than native Paulistanos. Finally, the standard deviation of inter-peaks duration measures how the peaks of the fundamental frequency are distributed in time. No difference was observed in this variable between ALCP and SP2010. In both groups, the estimated interval that separates two adjacent f_0 peaks last on average approximately 270 milliseconds.

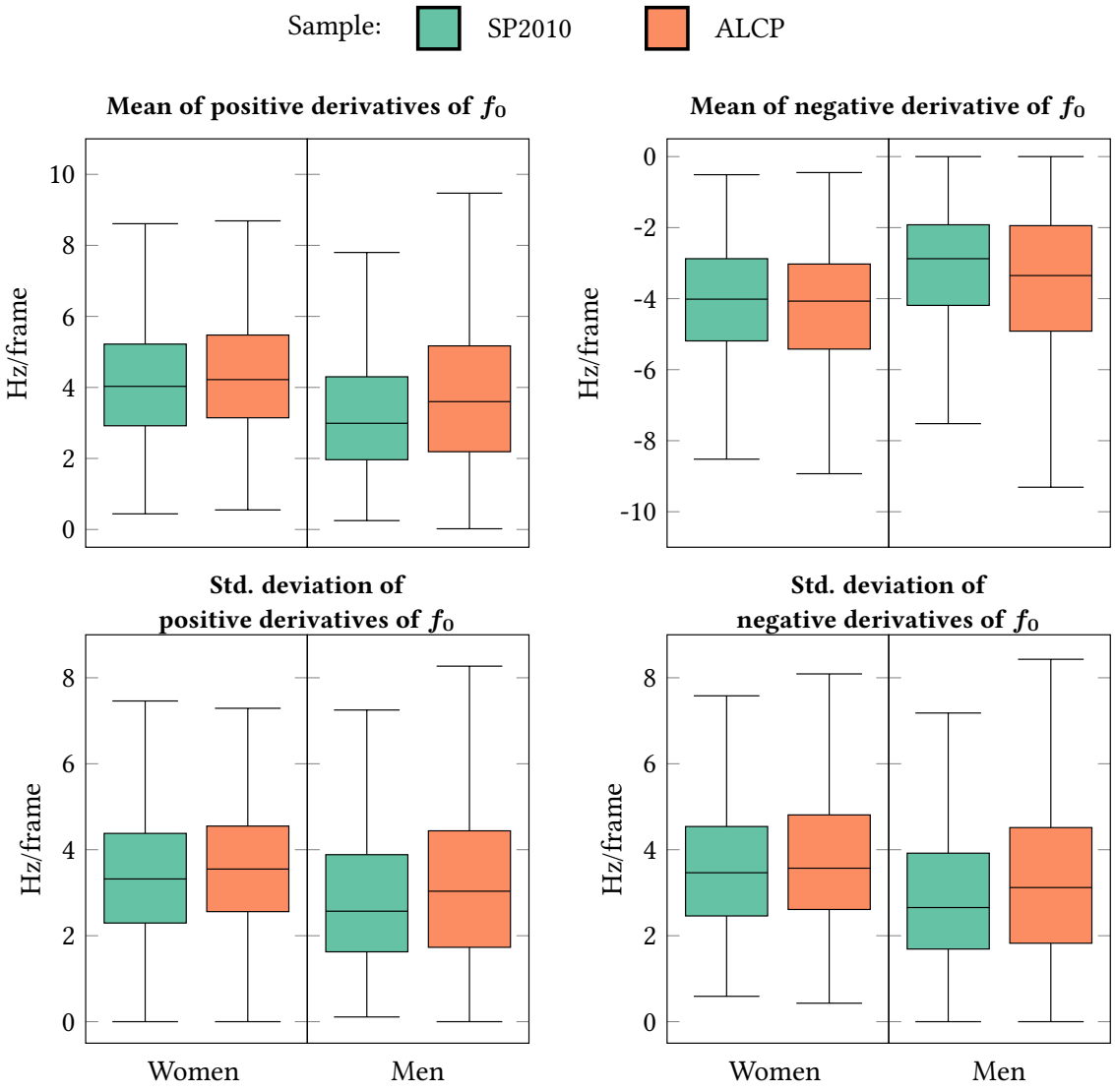
The derivative of a function measures how fast the values of the function change in time. In the case of intonation, the derivatives indicate how fast f_0 values rise or fall. In the graphical representation of the fundamental frequency, the derivatives correspond to the inclination of the intonation contour. When the fundamental frequency rises, the derivatives are positive. When it falls, they are negative. Figure 3.7 shows the distribution of the mean and standard deviation of both negative and positive derivatives of the fundamental frequency.

The fundamental frequency of Alagoan migrants rises and falls faster than that of native Paulistanos. Further, the rate of change in fundamental frequency is more variable among Alagoan migrants. However, these differences are statistically significant only in the case of male speakers. Except for the standard deviation of positive derivatives of f_0, male Alagoan migrants differ in all other variables from male Paulistanos. On the other hand, female speakers from ALCP do not differ from females from SP2010.

3.2.2 Effects of migration on intonation

I use this section to analyze the same intonation variables discussed above, but now from the perspective of migration. That is, all the following plots and models refer only to ALCP data distributed according to the age of

Figure 3.7: Distribution of the mean and standard deviation of positive and negative derivatives of f_0 (Hz/frame), contrasting male and female speakers from the ALCP and SP2010 samples (N=4,459)



arrival and length of residence. In this section, I also separate models for men and women.

Figure 3.8 shows the distribution of the median, standard deviation, semi-amplitude between quartiles, minimum, and maximum of the fundamental frequency, in both male and female speakers from the ALCP sample. Table 3.10 shows the linear regression models for each of these variables.

From a descriptive perspective, the box plots show greater differences among women. Female speakers that were younger than 20 years old when they arrived in Campinas and that have lived there more than 10 years

Figure 3.8: Distribution of the median (Hz), standard deviation (Hz), semi-amplitude between quartiles(Hz), minimum (Hz), and maximum (Hz) of fundamental frequency (f_0), contrasting gender, age of arrival and length of residence in the ALCP sample (N=1,960)

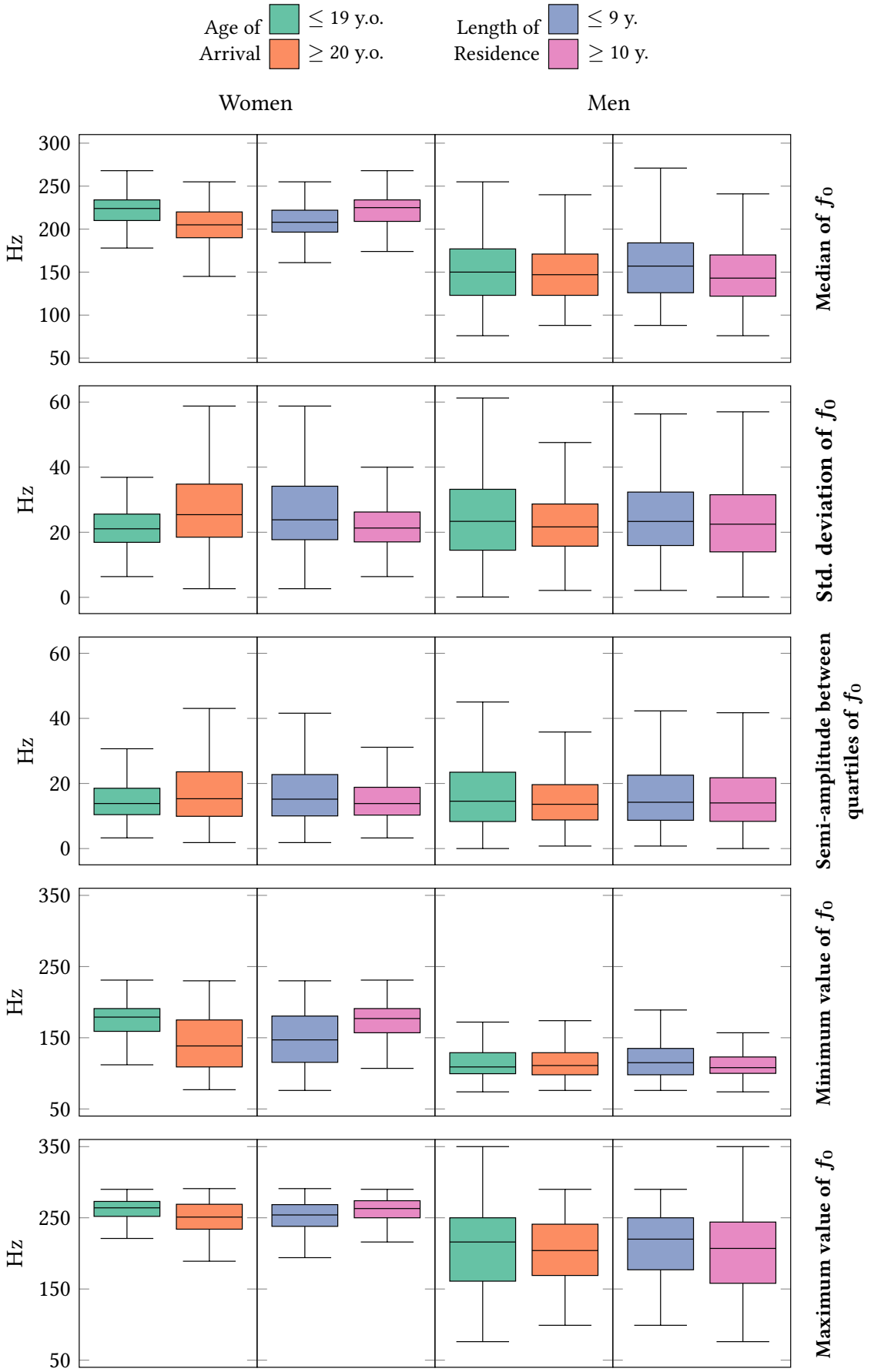


Table 3.9: Mixed-effects linear regression models of the mean and standard deviation of positive and negative derivatives of f_0 (Hz/frame). (N=4,459)

	Mean pos. deriv. (Hz/frame)	Mean neg. deriv (Hz/frame)	Std. dev. pos. deriv. (Hz/frame)	Std. dev. neg. deriv. (Hz/frame)
WOMEN				
marginal R^2	1%	0%	1%	1%
conditional R^2	13%	12%	14%	13%
Intercept	4.17	-4.14	3.43	3.56
Sample: ALCP	+0.29	-0.20	+0.27	+0.24
MEN				
marginal R^2	3%	3%	2%	2%
conditional R^2	21%	23%	18%	18%
Intercept	2.26	-3.13	2.86	2.88
Sample: ALCP	+0.66 ⋆	-0.60 ⋆	+0.43	+0.49 ⋆

⋆: statistically significant (α = 0.05)
Intercept: SP2010 (Sample)
Random effects variable: Speaker

show more high-pitched voice, with higher values in median, minimum, and maximum of f_0. They also show a slightly smaller standard deviation of f_0 when compared to female migrants who arrived at an older age and who have lived less than 10 years in the host community. However, from the inferential point of view (see Table 3.10), there is no difference between the Alagoan migrants in both genders distributed according to the age of arrival and length of residence.

In the case of the measures concerning f_0 peaks, some differences were observed among males, as shown in Table 3.11. Even though statistically significant, these differences are too small to be considered relevant. Males who were 20 years old or older and who have lived for more than 10 years in Campinas have a rate of f_0 peaks approximately 0.1 higher. That is, in one second, they produce on average one-tenth of a peak more than the other male speakers. It is unlikely that a native Brazilian can hear this difference.

Figure 3.9: Distribution of the standard deviation of f_0 peaks (Hz), f_0 peak width (Hz), f_0 peak rate (peaks/sec), and standard deviation of intervals between f_0 peaks (ms), contrasting gender, age of arrival and length of residence in the ALCP sample (N=1,960)

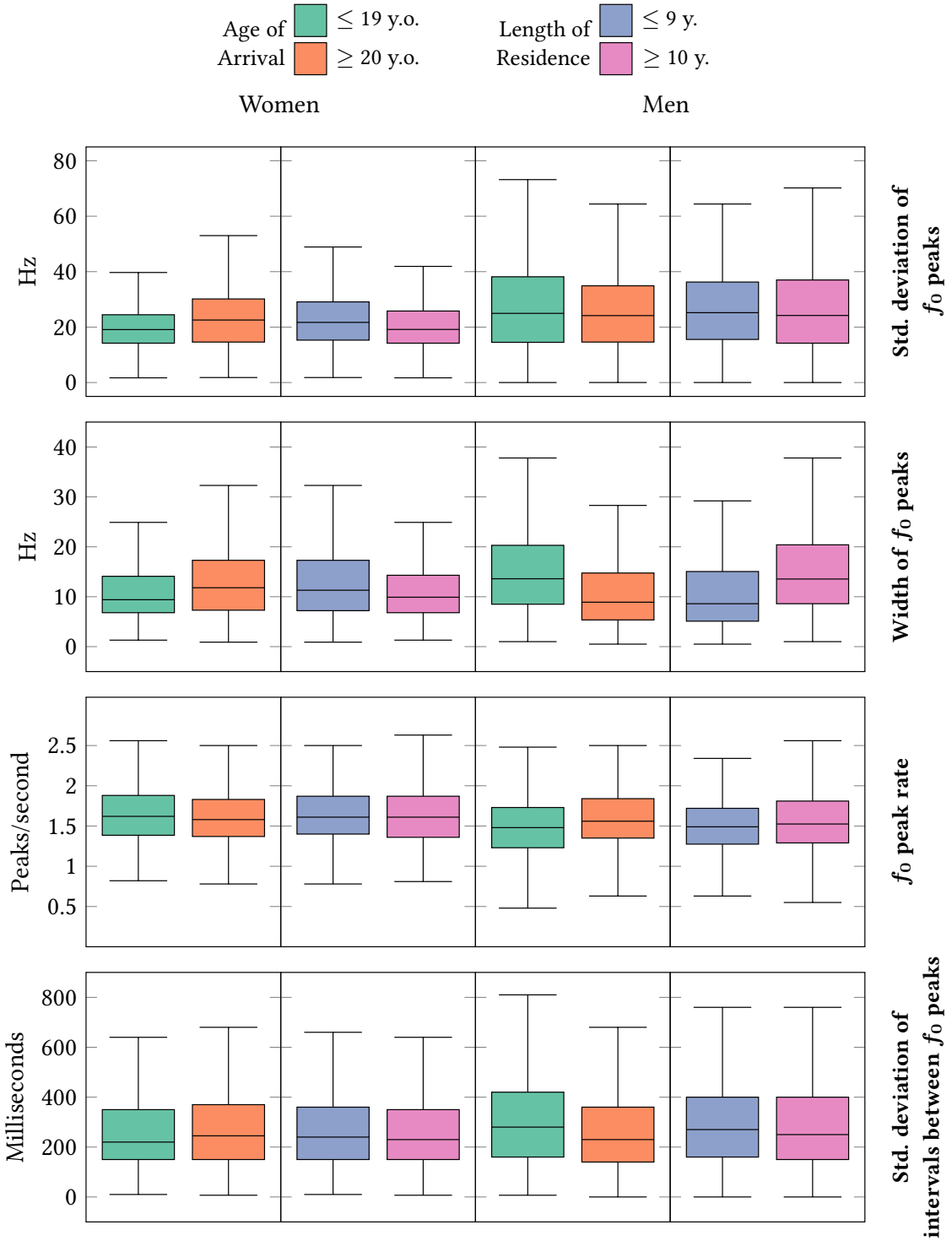


Table 3.10: Mixed-effects linear regression models of the median (Hz), standard deviation (Hz), semi-amplitude between quartiles(Hz), minimum (Hz), and maximum (Hz) of fundamental frequency (f_0), contrasting age of arrival and length of residence in the ALCP sample (N=1,960)

	Median (Hz)	Std. dev. (Hz)	SAQ (Hz)	Min. (Hz)	Max. (Hz)
WOMEN					
marginal R^2	9%	4%	3%	10%	4%
conditional R^2	48%	22%	15%	39%	32%
Intercept	206	24.7	16.9	154	251
Age of Arrival: ≥ 20 y.o.	-16	-1.2	-0.2	-15	-9
Length of Residence: ≥ 10 y.	-6	-2.3	-2.5	+4	-3
MEN					
marginal R^2	1%	0%	0%	0%	0%
conditional R^2	54%	21%	14%	32%	47%
Intercept	157	24.1	16.7	117	210
Age of Arrival: ≥ 20 y.o.	-6	-1.1	-1	-2	-5
Length of Residence: ≥ 10 y.	-4	-0.8	-0.6	-2	-7

⋆: statistically significant (α = 0.05)
Intercept: 19 y.o or younger (Age of Arrival) + 9 y. or less (Length of Residence)
Random effects variable: Speaker

Figure 3.10 illustrates the distribution of the mean and standard deviation of both positive and negative derivatives of the fundamental frequency. As can be seen in the boxplots and is reinforced by the statistical models in Table 3.12, the distribution of these measures is very similar among the Alagoan migrants. The only significant difference involves the mean of the positive derivatives. The fundamental frequency of female speakers who were 20 years old or older when they arrived in Campinas rises at a rate 0.6 Hz/frame faster than the speakers who were 19 years old or younger.

Of the 13 intonation measures, I verified statistically significant differences concerning age of arrival and length of residence in only five of them, three related to f_0 peaks and two about derivatives.

Figure 3.10: Distribution of the mean and standard deviation of positive and negative derivatives of f_0 (Hz/frame), contrasting gender, age of arrival and length of residence in the ALCP sample (N=1,960)

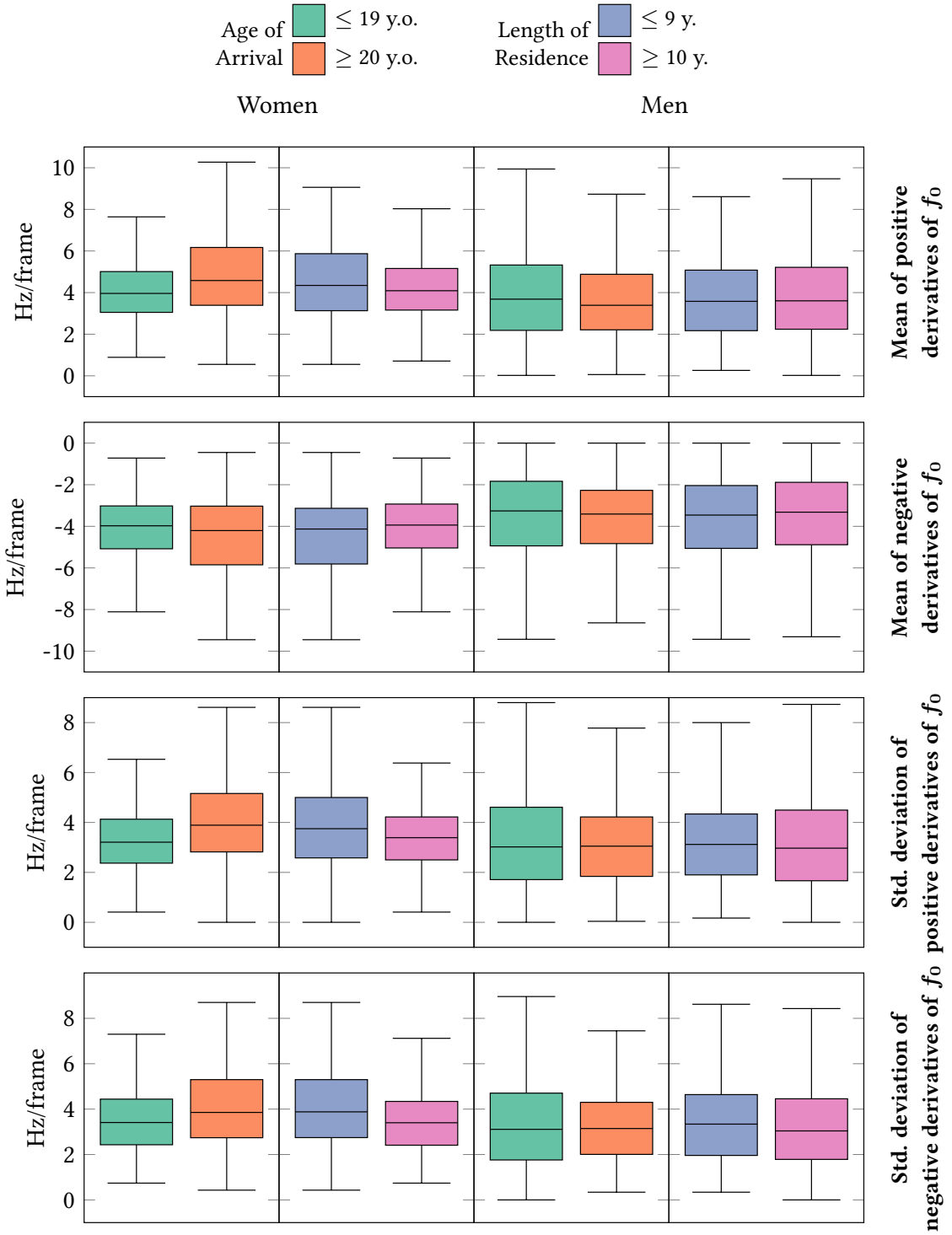


Table 3.11: Mixed-effects linear regression models of the standard deviation of f_0 peaks (Hz), f_0 peak width (Hz), f_0 peak rate (peaks/sec), and standard deviation of intervals between f_0 peaks (ms), contrasting age of arrival and length of residence in the ALCP sample (N=1,960)

	Peak std. dev. (Hz)	Peak width (Hz)	Peak rate (peak/sec)	Std. dev. of inter-peaks duration (msec)
WOMEN				
marginal R^2	2%	1%	0%	0%
conditional R^2	7%	11%	2%	2%
Intercept	23	12.8	1.64	273
Age of Arrival: ≥ 20 y.o.	+1.9	+1.3	-0.01	+11
Length of Residence: ≥ 10 y.	-0.7	+0.5	-0.04	+3
MEN				
marginal R^2	0%	2%	1%	0%
conditional R^2	11%	7%	2%	2%
Intercept	27	13.9	1.6	296
Age of Arrival: ≥ 20 y.o.	+0.04	-1.6	+0.08 ⋆	+27 ⋆
Length of Residence: ≥ 10 y.	-0.37	+2.6	+0.07 ⋆	+12

⋆: statistically significant ($\alpha = 0.05$)
Intercept: 19 y.o or younger (Age of Arrival) + 9 y. or less (Length of Residence)
Random effects variable: Speaker

3.2.3 Summing up

There are two main findings in these results concerning intonation. The first one refers to the instigating difference between male and female speakers in the comparison of samples. As we saw, male Alagoan migrants differ from male native Paulistanos in 9 of the 13 intonation parameters, while only two parameters differentiate migrants and natives among women. More than one hypothesis can be taken to explain this difference. As the second finding, the results also suggest that the age of arrival and length of residence are not relevant variables to explain the variation in the intonation of the Alagoan migrants in Campinas.

Table 3.12: Mixed-effects linear regression models of the mean and standard deviation of positive and negative derivatives of f_0 (Hz/frame), contrasting age of arrival and length of residence in the ALCP sample (N=1,960)

	Mean pos. deriv. (Hz/frame)	Mean neg. deriv (Hz/frame)	Std. dev. pos. deriv. (Hz/frame)	Std. dev. neg. deriv. (Hz/frame)
WOMEN				
marginal R^2	4%	2%	4%	3%
conditional R^2	5%	7%	8%	11%
Intercept	4.6	-4.3	3.7	3.8
Age of Arrival: ≥ 20 y.o.	+0.6 ⋆	-0.03	+0.4	-0.03
Length of Residence: ≥ 10 y.	+0.1	+0.31	-0.02	-0.42
MEN				
marginal R^2	0%	0%	0%	0%
conditional R^2	28%	34%	25%	28%
Intercept	3.9	-3.7	3.3	3.35
Age of Arrival: ≥ 20 y.o.	-0.18	-0.5	-0.05	-0.01
Length of Residence: ≥ 10 y.	+0.07	+0.14	-0.01	-0.14

⋆: statistically significant (α = 0.05)
Intercept: 19 y.o or younger (Age of Arrival) + 9 y. or less (Length of Residence)
Random effects variable: Speaker

Meta-comentário: Para a versão final, pretendo fazer uma análise de regressão em que os falantes serão incluídos como efeito fixo, a fim de investigar quais indivíduos mais se aproximam/distanciam dos padrões paulistas. Também pretendo realizar análises qualitativas com vistas a levantar hipóteses sobre que outros fatores sociais podem influenciar a acomodação prosódica.

Chapter 4

Future steps

This report has presented my academic activities in the past 16 months. In this chapter, I briefly discuss my goals for the remaining six months of research.

As discussed in Chapter 3, the findings of the analyses made so far suggest that there are significant differences in rhythm and intonation between Alagoan migrants living in Campinas and native speakers from the city of São Paulo. More specifically, Alagoan migrants differ from native Paulistanos in the seven rhythm metrics, but not in speech rate and stress group measures. In the field of intonation, generally, male Alagoans differ from male Paulistanos, but the same does not occur among female speakers. If, on the one hand, significant differences were observed in the comparison between ALCP and SP2010, on the other hand, the findings suggest that age of arrival and length of residence in the host community do not explain the prosodic variation within the speech of the Alagoan migrants. Furthermore, high values of conditional R^2 in many regression models suggest that the variation of some prosodic variables is more related to differences between subjects than between social groups. These partial results raise new questions, and the next steps of this research will focus on expanding the analyses in search of answers to these questions.

The first question concerns gender. Why do male migrants differ from Paulistanos in intonation, but female migrants do not? In Chapter 1, I discussed Bortoni-Ricardo's findings (1985) indicating that social differences

of gender, mainly related to the sexual division of labor, can condition the acquisition of new linguistic forms by migrants. In the next steps, I will conduct analyses trying to identify social differences between males and females in the ALCP and SP2010 samples that can potentially explain the gender difference observed in the partial results of this research.

Most significant prosodic differences were observed in the first type of analysis, in which the ALCP and SP2010 were contrasted, and not in the second one, concerning the effects of age of arrival and length of residence. As explained at the beginning of Chapter 3, larger differences in the comparison between samples can be evidence that the migrants did not accommodate to the prosody of the host community. This hypothesis is reinforced by the fact that too few correlations with the age of arrival and length of residence were observed. So, one important question to be addressed is the following: is prosody less susceptible to dialect contact effects than segmental features of speech? In the next steps, I will conduct more in-depth analyses to explore this question in the ALCP sample.

The values of conditional R_2 reinforce the necessity of conducting analyses focused on individual differences between subjects within the ALCP sample. It can be the case that social factors related to individual participants are more relevant for understanding the prosodic accommodation than the comparison between groups of speakers. For this reason, in the next months, I will conduct a qualitative analysis of each participant of the ALCP sample, trying to identify factors in their social contexts that can explain the prosodic variation in their speech. Furthermore, I will also examine fixed-effects regression models in which each Alagoan migrant will be included as a fixed effect, and the SP2010 as a whole will be included as if it were a single speaker among the Alagoan migrants. By these models, I can examine accommodation to the host community speaker by speaker.

Finally, I also intend to use the next months to overcome some limitations of the partial results reported here. The main limitation of the results achieved until now concerns the data used for control. I discussed in Section 2.1 that the cross-section approach to dialect acquisition, ideally,

should rely on two control samples consisting of the speech of lifelong residents, respectively, from the host and the home communities. The analyses conducted so far are based on only one control sample, and this sample is not exactly from the city where the migrants live but from a neighboring city. In the next months, I will try to obtain speech samples of lifelong residents from both the Metropolitan Area of Campinas and Alagoas.

Bibliography

Abaurre-Gnerre, Maria Bernadete Marques (1981). “Processos fonológicos segmentais como índices de padrões prosódicos diversos nos estilos formal e casual do português do Brasil”. *Cadernos de Estudos Linguísticos* (2): 23–44.

Abercrombie, David (1967). Elements of general phonetics. Edinburgh: Edinburgh University Press.

Barbosa, Plínio Almeida (2000). “"Syllable-timing in Brazilian Portuguese": uma crítica a Roy Major”. *DELTA: Documentação de Estudos em Linguística Teórica e Aplicada*, 16 (2): 369–402. doi: 10.1590/S0102-44502000000200006.

– (2006). Incursões em torno do ritmo da fala. Campinas, SP: Pontes.

– (2007). “From syntax to acoustic duration: A dynamical model of speech rhythm production”. *Speech Communication*, 49 (9): 725–742. doi: 10.1016/j.specom.2007.04.013.

– (2009). “Measuring speech rhythm variation in a model-based framework”. In: *Proceedings of the 10th Annual Conference of the International Speech Communication Association*. Brighton, UK: International Speech Communication Association, 1527–1530.

– (2012). “Panorama of experimental prosody research”. In: *Proceedings of the VIIth GSCP International Conference: Speech and Corpora*. Ed. by Heliana Mello et al. Firenze: Firenze University Press.

– (2013). “Semi-automatic and automatic tools for generating prosodic descriptors for prosody research”. In: *TRASP 2013 Proceedings*. Tools and

Resources for the Analysis of Speech Prosody. Aix-en-Provence, France, 86–89.

Barbosa, Plínio Almeida (2019). Prosódia. São Paulo: Parábola Editorial.

– (2020). Prosody Descriptor Extractor. Disponível em: https://github.com/pabarbosa/prosody-scripts.

Barbosa, Plínio Almeida; Arantes, Pablo, et al. (2005). "Abstractness in speech-metronome synchronisation: P-centres as cyclic attractors". In: *Proceedings of the 9th INTERSPEECH*. 9th European Conference on Speech Communication and Technology. Lisboa.

Barbosa, Plínio Almeida; Madureira, Sandra (2015). Manual de fonética acústica experimental: Aplicações a Dados do Português. São Paulo, SP: Cortez – 1ª edição.

Barbosa, Plínio Almeida; Papa, Paula Benassi, et al. (2019). "Harmonia vocálica e coarticulação vogal a vogal em duas variedades do português brasileiro". *DELTA: Documentação de Estudos em Linguística Teórica e Aplicada*, 35 (2): 1–32. doi: http://dx.doi.org/10.1590/1678-460X2019350202.

Barbosa, Plínio Almeida; Violaro, F. Et al. (1999). "Aiuruete: a high-quality concatenative text-to-speech system for brazilian portuguese with demisyllabic analysis-based units and a hierarchical model of rhythm production". In: *EUROSPEECH*.

Blondeau, Hélène (2013). "Studying language over time". In: Podesva, Robert; Sharma, Devyani (ed.). *Research Methods in Linguistics*. Cambridge: Cambridge University Press, 494–538.

Boersma, Paul; Weenink, David (2021). Praat: doing phonetics by computer. Version 6.1.41. Disponível em: http://www.praat.org/.

Bortoni-Ricardo, Stella Maris (1985). "The Urbanization of Rural Dialect Speakers: A Sociolinguistic Study in Brazil". Cambridge: Cambridge University Press.

Britain, David (1992). "Linguistic change in intonation: The use of high rising terminals in New Zealand English". *Language Variation and Change*, 4 (1): 77–104. doi: 10.1017/S0954394500000661.

Britain, David (2002). "Diffusion, levelling, simplification and reallocation in past tense BE in the English Fens". *Journal of Sociolinguistics*, 6 (1): 16–43. doi: 10.1111/1467-9481.00175.

– (2013). "Space, diffusion and mobility". In: Chambers, Jack; Schilling, Natalie (ed.). *The Handbook of Language Variation and Change*. Malden: Wiley-Blackwell, 472–500.

– (2018). "Dialect contact and new dialect formation". In: Boberg, Charles (ed.). *The handbook of Dialectology*. Nova Jérsei, EUA: Wiley-Blackwell, 143–158.

Cagliari, Luiz Carlos (2013). "Existem línguas de ritmo silábico?" *Estudos Linguísticos*, 42 (1): 19–32.

Cagliari, Luiz Carlos; Abaurre, Maria Bernadete (1986). "Elementos para uma investigação instrumental das relações entre padrões rítmicos e processos fonológicos no português brasileiro". *Cadernos de Estudos Lingüísticos*, 10: 39–57. doi: 10.20396/cel.v10i0.8636716.

Chambers, J. K. (1992). "Dialect acquisition". *Language*, 68 (4): 673–705. doi: 10.1353/lan.1992.0060.

Crystal, Thomas H. House, Arthur S. (1990). "Articulation rate and the duration of syllables and stress groups in connected speech". *The Journal of the Acoustical Society of America*, 88 (1): 101–112. issn: 0001-4966. doi: 10.1121/1.399955.

Cukor-Avila, Patricia; Bailey, Guy (2013). "Real time and apparent time". In: Chambers, Jack; Schilling-Estes, Natalie (ed.). *The handbook of language variation and change*. Massachusetts, EUA: Wiley-Blackwell, 237–262.

Dellwo, Volker (2006). "Rhythm and speech rate: a variation coefficient for DeltaC". In: *Language and Language-processing: Proceedings of the 38th Linguistics Colloquium*. Piliscsaba: Peter Lang Publishing Group, 231–241.

Deterding, David (1994). "The rhythm of Singapore English". In: *Proceedings of the 5th Australian International Conference on Speech Science and*

Technology. Perth: Australian Speech Science & Technology Association Inc.

Dodsworth, Robin (2017). "Migration and dialect contact". *Annual Review of Linguistics*, 3 (1): 331–346. doi: 10.1146/annurev-linguistics-011516-034108.

Fletcher, Janet (2010). "The prosody of speech: timing and rhythm". In: Hardcastle, William J. (ed.). *The Handbook of Phonetic Sciences*. Malden: Wiley-Blackwell, 523–602.

Frota, Sónia; Vigário, Marina (2000). "Aspectos de prosódia comparada: ritmo e entoação no PE e no PB". In: *Actas do XV Encontro da Associação Portuguesa de Linguística*. Braga: Associação Portuguesa de Linguística, 533–555.

Fuchs, Robert (2016). Speech Rhythm in Varieties of English. Singapura: Springer.

Gibbon, Dafydd; Gut, Ulrike (2001). "Measuring speech rhythm". In: *Proceedings of the 7th European Conference on Speech Communication and Technology*. Scandinavia: International Speech Communication Association.

Giles, Howard (1973). "Accent mobility: a model and some data". *Anthropological Linguistics*, 15 (2): 87–105. doi: en.

Gorman, Kyle et al. (2011). "Prosodylab-Aligner: A Tool for Forced Alignment of Laboratory Speech. Canadian Acoustics. 39.3. 192–193." *Canadian Acoustics.*, 39 (3): 192–193.

Grabe, Esther; Low, Ee Ling (2002). "Durational variability in speech and the rhythm class hypothesis". In: Gussenhoven, Carlos; Warner, Natasha (ed.). *Laboratory Phonology 7*. Berlin: De Gruyter Mouton, 515–546. Disponível em: https://doi.org/10.1515/9783110197105.

Gries, Stefan Thomas (2013). Statistics for Linguistics with R: A Practical Introduction. Berlin: De Gruyter Mouton.

Guy, Gregory; Zilles, Ana Maria Stahl (2007). Sociolinguística quantitativa: instrumental de análise. São Paulo: Parábola Editorial.

Hoffman, Michol F. Walker, James A. (2010). "Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English". *Language Variation and Change*, 22 (1): 37–67. ISSN: 1469-8021, 0954-3945. DOI: 10.1017/S0954394509990238.

Ide, Nancy; Pustejovsky, James (ed.). Eds. (2016). Handbook of Linguistic Annotation. New York: Springer.

IOM (2019). Glossary on Migration. Geneva: International Organization for Migration.

Johnson, Keith (2012). Acoustic and Auditory Phonetics. Malden: Wiley-Blackwell – 3rd ed.

Kerswill, Paul (1985). "A Sociolinguistic Study of Rural Immigrants in Bergen, Norway". PhD thesis. Cambridge, UK: Cambridge University.

– (1993). "Rural dialect speakers in an urban speech community: the role of dialect contact in defining a sociolinguistic concept". *International Journal of Applied Linguistics*, 3 (1): 33–56. DOI: 10.1111/j.1473-4192.1993.tb00042.x.

Kruse, João; Barbosa, Plínio Almeida (2020). AlinhaPB. Version Beta.

Labov, William (1972). Sociolinguistic Patterns. Philadelphia: University of Pennsylvania Press.

– (2006). The Social Stratification of English in New York City. Cambridge: Cambridge University Press.

Ladefoged, Peter (1996). Elements of Acoustic Phonetics. Chicago: University of Chicago Press – 2nd ed.

Lehiste, Ilse (1970). Suprasegmentals. Cambridge, MA: MIT Press.

Lenneberg, Eric H. (1967). Biological Foundations of Language. New York: John Wiley & Sons.

Low, Ee Ling et al. (2000). "Quantitative characterizations of speech rhythm: syllable-timing in Singapore English". *Language and Speech*, 43 (4): 377–401. ISSN: 0023-8309. DOI: 10.1177/00238309000430040301.

Major, Roy C. (July 1, 1981). "Stress-timing in Brazilian Portuguese". *Journal of Phonetics*, 9 (3): 343–351. ISSN: 0095-4470. DOI: 10.1016/S0095-4470(19)30977-5.

Major, Roy C. (1985). "Stress and Rhythm in Brazilian Portuguese". *Language*, 61 (2): 259–282. ISSN: 0097-8507. DOI: 10.2307/414145.

Mendes, Ronald Beline; Oushiro, Livia (2012). "O paulistano no mapa sociolinguístico brasileiro". *ALFA: Revista de Linguística*, 56 (3): 973–1001.

Mendes, Ronald Beline; Sene, Marcus Garcia (2020). "O efeito do pitch na percepção de masculinidade".

Mendonça, Clara Simone Ignácio (2003). "A sílaba em fonologia". *Working Papers em Linguística*, 7 (1): 21–40.

Migliorini, Lívia; Massini-Cagliari, Gladis (2010). "Sobre o ritmo do português brasileiro: evidências de um padrão acentual". *Revista Virtual de Estudos da Linguagem*, 8 (15): 310–328.

Milroy, Lesley (Jan. 1, 1980). Language and social networks. Oxford: B. Blackwell – Illustrated edition.

Milroy, Lesley; Gordon, Matthew (2003). Sociolinguistics: Method and Interpretation. Malden: Blackwell Publishing.

Moraes, João Antônio; Leite, Yonne (2002). "Ritmo e velocidade da fala na estratégia do discursos: uma proposta de trabalho". In: Ilari, Rodolfo (ed.). *Gramática do português falado II: Níveis de análise linguística.* 4th ed. Campinas, SP: Editora da Unicamp.

Nokes, Jacqui; Hay, Jennifer (2012). "Acoustic correlates of rhythm in New Zealand English: a diachronic study". *Language Variation and Change*, 24 (1): 1–31. ISSN: 1469-802. DOI: 10.1017/S0954394512000051.

Nycz, Jennifer (2013). "New contrast acquisition: methodological issues and theoretical implications". *English Language and Linguistics*, 17 (2): 325–357. ISSN: 1360-6743. DOI: 10.1017/S1360674313000051.

– (2015). "Second dialect acquisition: a sociophonetic perspective". *Language and Linguistics Compass*, 9 (11): 469–482. DOI: 10.1111/lnc3.12163.

Oliveira, Marcelo Augusto (2019). Dialetos em contato: acomodação dialetal por migrantes baianos habitantes da cidade de Bauru, São Paulo.

Relatório de qualificação de mestrado. Araraquara: Universidade Estadual Paulista.

Oushiro, Livia (2016). "Social and structural constraints in lectal cohesion". *Lingua*. Coherence, covariation and bricolage. Various approaches to the systematicity of language variation 172-173: 116–130. ISSN: 0024-3841. DOI: 10.1016/j.lingua.2015.10.015.

– (2018). Processos de acomodação dialetal na fala de nordestinos residentes em São Paulo. Relatório científico final - FAPESP (Processo nº 2016/04960-7). Campinas: Universidade Estadual de Campinas.

– (2019a). "A fala de migrantes internos: uma agenda de estudos". Pôster. XI Congresso Internacional da Associação Brasileira de Linguística (Alagoas). Disponível em: https://www.abralin.org/abralin50/inicio/.

– (2019b). "Linguistic uniformity in the speech of Brazilian internal migrants in a dialect contact situation". In: *Proceedings of the 19th International Congress of Phonetic Sciences*. Ed. by Sasha Calhoun et al. Melbourne, Australia: International Phonetic Association.

– (2019c). Resumo "A fala de migrantes internos: uma agenda de estudos".

– (2020). "Contrasting age of arrival and length of residence in dialect contact". *Selected Papers from New Ways of Analyzing Variation 47*, 25 (2): 79–88. Disponível em: https://repository.upenn.edu/pwpl/vol25/iss2/.

Payne, Arvilla (1976). "The acquisition of the phonological system of a second dialect". Tese de doutorado. Universidade da Pennsylvania.

– (1980). "Factors controlling the acquisition of the Philadelphia dialect by out-of-state children". In: Labov, William (ed.). *Locating Language in Time and Space*. New York: Academic Press, 143–178.

Pike, Kenneth L. (1945). The Intonation of American English. Ann Arbor: University of Michigan Press.

Podesva, Robert (2006). "Intonational variation and social meaning: categorical and phonetic aspects". *University of Pennsylvania Working Papers in Linguistics*, 12 (2): 189–202.

Pompino-Marschall, Bernd (1989). "On the psychoacoustic nature of the P-center phenomenon". *Journal of Phonetics*, 17 (3): 175–192. ISSN: 0095-4470. DOI: 10.1016/S0095-4470(19)30428-0.

Quené, Hugo (2007). "On the just noticeable difference for tempo in speech". *Journal of Phonetics*, 35 (3): 353–362. ISSN: 0095-4470. DOI: 10.1016/j.wocn.2006.09.001.

Ramus, Franck et al. (1999). "Correlates of linguistic rhythm in the speech signal". *Cognition*, 73 (3): 265–292. DOI: 10.1016/S0010-0277(00)00101-3.

Santana, Amanda de Lima (2018). "As vogais médias pretônicas na fala de sergipanos em São Paulo". Dissertação de mestrado. São Paulo: Universidade de São Paulo.

Siegel, Jeff (2010). Second Dialect Acquisition. Cambridge: Cambridge University Press.

Silva Jr, Leônidas; Barbosa, Plínio Almeida (2019). Metrics and acoustics extractor. Version 1.7.7.

Souza, Emerson Santos (2019). "O uso variável do imperativo de migrantes baianos em São Paulo". *Domínios de Lingu@gem*, 13 (4): 1433–1464. ISSN: 1980-5799. DOI: 10.14393/DL40-v13n4a2019-5.

Szakay, Anita (2006). "Rhythm and pitch as markers of ethnicity in New Zealand English". In: *Proceedings of the 11th Australian International Conference on Speech Science & Technology*. Ed. by Paul Warren and Catherine Watson. New Zealand: Australian Speech Science & Technology Association Inc., 421–426.

– (2008). Ethnic Dialect Identification in New Zealand - The Role of Prosodic Cues. Saarbrücken, Germany: VDM Verlag Dr. Mueller E.K.

Tagliamonte, Sali A. Molfenter, Sonja (Nov. 2007). "How'd you get that accent?: Acquiring a second dialect of the same language". *Language in Society*, 36 (5): 649–675. ISSN: 1469-8013, 0047-4045. DOI: 10.1017/S0047404507070911.

Thomas, Erik (2013). "Sociophonetics". In: Chambers, Jack; Schilling, Natalie (ed.). *The Handbook of Language Variation and Change*. Malden: Wiley-Blackwell, 108–127.

Thomas, Erik; Carter, Phillip (2006). "Rhythm and African American English". *English World-Wide*, 27 (3): 331–355. doi: 10.1075/eww.27.3.06tho.

Torgersen, Eivind Nessa; Szakay, Anita (2012). "An investigation of speech rhythm in London English". *Lingua*, 122 (7): 822–840. doi: 10.1016/j.lingua.2012.01.004.

Troncoso-Ruiz, Aurora; Elordieta, Gorka (2018). "Prosodic accommodation and salience: The nuclear contours of Andalusian Spanish speakers in Asturias". *Loquens*, 4 (2): 043. doi: 10.3989/loquens.2017.043.

Trudgill, Peter (1986). Dialects in Contact. Language in Society 10. Oxford: Basil Blackwell.

Wagner, Petra; Dellwo, Volker (2004). "Introducing YARD (Yet Another Rhythm Determination) and re-introducing isochrony to rhythm research". In: *Proceedings of Speech Prosody*. Nara, Japan: International Speech Communication Association, 227–230.

Walker, Abby Jewel (2014). "Crossing oceans with voices and ears: second dialect acquisition and topic-based shifting in production and perception". PhD thesis. Ohio: The Ohio State University Press.

Young, Steve et al. (2015). The HTK Book. Cambridge: Cambridge University Engineering Department – Version 3.5.

Zec, Draga (2007). "The syllable". In: De Lacy, Paul V. (ed.). *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press, 161–194.