Universidade Estadual de Campinas
Instituto de Estudos da Linguagem

Gustavo de Campos Pinheiro da Silveira

The prosody of speech in a dialect contact situation:

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

A prosódia da fala em situação de contato dialetal:
Um estudo sociofonético da fala de migrantes alagoanos em São Paulo

Campinas
2022

Gustavo de Campos Pinheiro da Silveira

The prosody of speech in a dialect contact situation:

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

A prosódia da fala em situação de contato dialetal:
Um estudo sociofonético da fala de migrantes alagoanos em São Paulo

Dissertation presented to the Institute of Language Studies of the University of Campinas in partial fulfillment of the requirements for the degree of Master, in the area of Linguistics.

Dissertação de mestrado apresentada ao Instituto de Estudos da Linguagem da Universidade Estadual de Campinas para obtenção do título de Mestre em Linguística.

Adviser: Prof. PhD. Livia Oushiro

Este arquivo digital corresponde à versão final da dissertação de mestrado defendida pelo aluno Gustavo de Campos Pinheiro da Silveira sob orientação da Prof.ª Dr.ª Livia Oushiro.

Campinas
2022

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Estudos da Linguagem
Tiago Pereira Nocera — CRB 8/10468

Si39p Silveira, Gustavo de Campos Pinheiro da, 1993-
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. — Campinas, SP : [s.n.], 2022.

Orientador: Livia Oushiro.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Estudos da Linguagem.

1. Prosódia (Linguística). 2. Contato dialetal. 3. Entoação. 4. Ritmo. 5. Migração. I. Oushiro, Livia, 1980-. II. Universidade Estadual de Campinas. Instituto de Estudos da Linguagem. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: A prosódia da fala em situação de contato dialetal : um estudo sociofonético da fala de migrantes alagoanos em São Paulo
Palavras-chave em inglês:
Prosody (Linguistics)
Dialect contact
Intonation
Rhythm
Emigration and immigration
Área de concentração: Linguística
Titulação: Mestre em Linguística
Banca examinadora:
Livia Oushiro [Orientador]
Plínio Almeida Barbosa
Cláudia Regina Brescancini
Data de defesa: 11−05−2022
Programa de Pós-Graduação: Linguística

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-2869-2024
- Currículo Lattes do autor: http://lattes.cnpq.br/9365466339986519

BANCA EXAMINADORA:

Livia Oushiro

Plínio Almeida Barbosa

Cláudia Regina Brescancini

IEL/UNICAMP
2022

Ata da defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós Graduação do IEL.

Para todos os paulistas vindos do nordeste,
e para minha mãe e meu pai

Agradecimentos

Dez anos se passaram desde que ingressei na vida acadêmica. Esse trabalho é fruto de todas as experiências que acumulei ao longo desses anos, primeiro como estudante de Filosofia, depois, de Linguística. Ao longo desse percurso, muitas pessoas contribuíram para a minha formação acadêmica, às quais agradeço profundamente.

Em primeiro lugar, agradeço à minha família, que sempre me apoiou em todos os meus projetos. À minha mãe Marcia, que, a exemplo da minha bizavó Dona Rosa, me ensinou a enfrentar os desafios da vida com comprometimento e pé no chão. Ao meu pai Marco, que me apresentou as artes e as ciências, e que é minha inspiração na vida acadêmica. À minha irmã Marina, por ser minha melhor amiga e estar sempre ao meu lado para o que der e vier.

À minha companheira Cecília, que amo e com quem compartilho a vida ao longo desses anos, pelo amor e carinho cotidiano. Obrigado por me acompanhar nessa jornada.

À minha orientadora Livia Oushiro, agradeço imensamente pela enorme dedicação com que me guiou na produção deste trabalho. Ser orientado pela Livia me fez crescer como pesquisador como jamais poderia me imaginar ser capaz. Com ela, aprendi o possível e o impossível. Aprendi linguística, estatística, programação. Aprendi sobre vida acadêmica. Mas, antes de tudo, com ela, aprendi sobre generosidade e comprometimento.

Sou profundamente grato ao professor Plínio Almeida Barbosa, que, com seu carinho e sua delicadeza, me acompanhou ao longo deste trabalho, contribuindo de modo excepcional para a minha formação em fonética acústica, sem a qual este trabalho não seria possível.

Agradeço à professora Cláudia Regina Brescancini e ao professor Plínio Almeida Barbosa por aceitarem o convite para participar da banca de defesa e pelas sugestões por ocasião do exame de qualificação.

Ao Ronald Beline Mendes, meu primeiro professor de Linguística, por ter me acolhido quando ainda era um recém-formado em Filosofia e ter me introduzido ao mundo da Linguística.

Ao professor Pablo Arantes, pela leitura da dissertação e pelas sugestões por ocasião de encontros científicos.

Ao meu professor de filosofia, Anderson Beraldo de Araújo, por nunca me deixar esquecer das grandes questões.

Ao meu amigo Gabriel Catani, por ser, como eu, um curioso incansável; pelas cotidianas e empolgadas reflexões sobre tudo que se possa imaginar.

Ao meu amigo Leônidas Silva Jr., por ter me instruído a usar *scripts* no Praat; pelas agradáveis conversas e pelas dicas sobre a língua inglesa.

Aos colegas linguistas Julia Adams, Danielle Bento, Sarah Poli, Emerson Souza, Iris Dutra, Natasha Mourão e Joana Figuereido, do Laboratório Variação, Identidade, Estilo e Mudança (VARIEM). Tenho muito orgulho de fazer parte desse grupo de pesquisa. Sou especialmente grato à Natasha Mourão por ter gentilmente me cedido suas gravações sociolinguísticas.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES) — Código de Financiamento 001.

Acknowledgments

Ten years have passed since my academic life started. This work is the result of all the experiences I have accumulated over these years, first as a student of Philosophy, then of Linguistics. Throughout this journey, many people have contributed to my academic training, to whom I am deeply grateful.

First of all, I would like to thank my family, who have always supported me in all my projects. Thanks to my mother Marcia, who, like my great-grandmother Dona Rosa, has taught me to be committed and down to earth when facing life's challenges. To my father Marco, who has introduced me to the arts and sciences, and who is my inspiration in academic life. To my sister Marina, for being my best friend and always being by my side no matter what.

I thank Cecília, whom I love and with whom I have shared my life over the years, for her daily love and affection. Thank you for accompanying me on this journey.

To my supervisor Livia Oushiro, I am immensely grateful for the enormous dedication with which she guided me in the production of this work. Being supervised by Livia made me grow as a researcher as I could never have imagined myself capable of. From her, I learned the possible and the impossible. I learned linguistics, statistics, programming. I learned about academic life. But first of all, from her, I learned about generosity and commitment.

I am deeply grateful to Professor Plínio Almeida Barbosa, who, with his affection and delicacy, accompanied me throughout this work, contributing in an exceptional way to my training in acoustic phonetics, without which this work would not be possible.

I thank Professor Cláudia Regina Brescancini and Professor Plínio Almeida Barbosa for accepting the invitation to join the examining committee and for their suggestions during the qualifying exam.

I thank Ronald Beline Mendes, my first professor of Linguistics, for having welcomed me when I had just graduated in Philosophy and for introducing me to the world of Linguistics.

I thanks Professor Pablo Arantes for reading this text and for suggestions during scientific meetings.

Thanks to my Philosophy professor, Anderson Beraldo de Araújo, for never letting me forget the big questions.

My thanks to my friend Gabriel Catani, for being, like me, a tireless curious person. Thanks for the daily and exciting reflections on everything imaginable.

I thank my friend Leônidas Silva Jr., for instructing me to use scripts in Praat. Thanks for our pleasant conversations and for helping me with the English language.

I thank my fellow linguists Julia Adams, Danielle Bento, Sarah Poli, Emerson Souza, Iris Dutra, Natasha Mourão, and Joana Figuereido, from the Variation, Identity, Style and Change Laboratory (VARIEM). I am very proud to be part of this research group. I am especially grateful to Natasha Mourão for kindly providing me with her sociolinguistic recordings.

The present work was carried out with the support of the Coordination for the Improvement of Higher Education Personnel-Brazil (CAPES) — Financing Code 001.

Resumo

Os efeitos do contato dialetal na fala de migrantes, um tópico atualmente em expansão, ainda carecem de estudos mais aprofundados que permitam maiores generalizações. A maior parte das pesquisas sociofonéticas sobre esse tópico se concentra em variáveis segmentais, sobretudo em dialetos do inglês, havendo poucos estudos sobre variáveis prosódicas no português brasileiro.

Este trabalho investiga os efeitos do contato dialetal na prosódia de vinte e dois alagoanos que vivem na Região Metropolitana de Campinas, em São Paulo, em comparação com uma amostra controle, com nove campineiros nativos. Dezessete parâmetros prosódicos foram analisados. A respeito de ritmo de fala, analisaram-se a taxa de elocução, a duração de grupos acentuais, a taxa de picos de duração silábica, a correlação entre a duração de grupos acentuais e o número de sílabas, e as métricas rítmicas VarcoV, VarcoC, nPVI-V e nPVI-C. Quanto à entoação, examinaram-se o desvio padrão da frequência fundamental (f_0), o desvio padrão dos picos de f_0, o grau de abertura dos picos de f_0, a taxa dos picos de f_0, o desvio padrão dos intervalos entre picos de f_0 adjacentes; e a média e o desvio padrão das taxas de mudança positivas e negativas de f_0. Este trabalho se propõe a analisar (i) quais desses parâmetros prosódicos mais diferenciam os alagoanos migrantes e os campineiros; (ii) em que medida a prosódia dos migrantes alagoanos sofreu alterações em virtude do contato com a variedade campineira; e (iii) se a idade de migração e o tempo de residência em São Paulo são capazes de prever o grau de acomodação prosódica à variedade campineira na fala dos migrantes alagoanos.

As gravações dos participantes, feitas em entrevista sociolinguística, foram analisadas com uso de técnicas computacionais de análise acústica. Desenvolveram-se técnicas de alinhamento forçado para automatizar a segmentação fonética, e os pârametros prosódicos foram extraídos com uso de scripts no Praat. Analisaram-se os dados com técnicas estatísticas, à luz da literatura variacionista sobre contato dialetal e da fonética acústica da prosódia da fala.

Os resultados mostram que os migrantes alagoanos e os campineiros nativos são similares na maioria dos parâmetros prosódicos. Diferenças significativas entre os dois grupos foram encontradas em oito das dezessete variáveis analisadas, duas rítmicas (VarcoV e VarcoC) e seis entoacionais (desvio padrão de f_0 e de picos de f_0, média e desvio padrão das taxas de mudança positivas e negativas de f_0). No entanto, no caso das seis variáveis entoacionais, observou-se interação com o gênero dos participantes; diferenças significativas foram encontradas apenas entre migrantes e campineiros homens. A idade de migração correlacionou-se com as seis variáveis entoacionais na fala dos migrantes homens: quanto menor a idade do participante ao migrar para São Paulo, maior é o grau de acomodação prosódica à fala paulista. Nenhuma correlação com tempo de residência foi observada. Os resultados também mostram que fatores individuais são importantes para explicar a variação prosódica na fala dos migrantes, sobretudo no caso de variáveis entoacionais.

Palavras-chave: contato dialetal; prosódia; ritmo; entoação; migração.

Abstract

The effects of dialect contact on the speech of migrants, a topic currently in expansion, still lack further studies to allow for broader generalizations. Most sociophonetic research on this topic focuses on segmental variables, especially in English dialects, with few studies on prosodic variables in Brazilian Portuguese.

This research investigates the effects of dialect contact on the prosody of twenty-two migrants from the state of Alagoas living in the Campinas Metropolitan Region, in São Paulo, in comparison with a control sample with nine native Campineiros. Seventeen prosodic parameters were analyzed. Regarding speech rhythm, I analyzed speech rate, stress group duration, the rate of syllabic duration peaks, the correlation between the duration of stress groups and the number of syllables, and the rhythm metrics VarcoV, VarcoC, nPVI-V, and nPVI-C. Concerning intonation, I examined the standard deviation of fundamental frequency (f_0), f_0 peaks standard deviation, f_0 peak width, f_0 peak rate, the standard deviation of the intervals between adjacent f_0 peaks, and the mean and the standard deviation of positive and negative f_0 change rates. This research aims to analyze (i) which of these prosodic parameters most differentiate Alagoan migrants and native Campineiros; (ii) to what extent Alagoan migrants' prosody suffered modifications due to the contact with Campinas variety; and (iii) whether age of arrival and length of residence in São Paulo are able to predict the degree of prosodic accommodation to the Campinas variety in Alagoan migrants' speech.

The recordings of the participants, in sociolinguistic interviews, were analyzed using computational techniques of acoustic analysis. Forced alignment techniques were developed to automate phonetic segmentation, and prosodic parameters were extracted using Praat scripts. Data were analyzed using statistical techniques, in light of dialect contact variationist studies and acoustic phonetics of speech prosody.

The results show that Alagoan migrants and native Campineiros are similar in most prosodic parameters. Significant differences between the two groups were found in eight of the seventeen variables, two regarding rhythm (VarcoV and VarcoC) and six regarding intonation (standard deviation of f_0 and f_0 peaks, mean and standard deviation of positive and negative f_0 change rates). However, in the case of the six intonation variables, I observed interaction with the participants' gender; significant differences were found only between male Alagoans and male Campineiros. Age of arrival correlated with the six intonation variables in male migrants' speech: the younger the participant's age when migrating to São Paulo, the greater the degree of prosodic accommodation to São Paulo speech. No correlation concerning length of residence was observed. The results also show that individual factors are important to explain the prosodic variation in migrants' speech, especially in the case of intonation variables.

Keywords: dialect contact; prosody; rhythm; intonation; migration.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

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 linguistic variety. Sociolinguists refer to these situations as *dialect contact*, a concept defined as the influence of mutually intelligible linguistic varieties upon one another (Trudgill, 1986, p. vii). Migration-induced dialect contact settings 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 (Nycz, 2015; Oushiro, 2018) and to the effects of the population dynamics on the historical evolution of regional varieties (Britain, 2013).

One phenomenon to which sociolinguists have devoted their attention concerns the effects of dialect contact on the speech of individuals (Nycz, 2015; Siegel, 2010). 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. Much empirical research on specific cases of dialect contact in different languages and concerning different linguistic variables has been conducted in the last decades (for an overview, see Siegel 2010; Dodsworth

2017; Oushiro 2022). Many of these studies observed changes in the speech not only of children but also of adults due to dialect contact, but it is still a challenging question how to interpret these changes.

There are two main theoretical frameworks to account for contact-induced modifications in the speech of individuals. One of them is the accommodation framework (Trudgill, 1986). According to this framework, individuals who moved to another dialect region modify their speech as the result of long-term *accommodation* to the new community. The concept of accommodation comes from the accommodation theory[1] developed by Giles (1973), in the field of Social Psychology, to explain why and how speakers modify their speech depending on who they are talking to. The central idea is that, in face-to-face interactions, speakers modify their speech, making it more similar or dissimilar from their interlocutors' speech, as a way of gaining the interlocutors' social approval or marking a social distance to them.

> The essence of the theory lies in the social psychological research on similarity-attraction [. . .]. This work suggests that an individual can induce another to evaluate him more favorably by reducing certain dissimilarities between them. The process of speech accommodation of course operates on this principle and as such may be a reflection of an individual's desire for social approval. [. . .] This set is of course not insensitive to the demands of the specific occasion as can be inferred from the reference to speech divergence earlier (Giles et al., 1973, p. 179).

Although originally applied to face-to-face interactions between any given pair of speakers, the concept of *accommodation* turned out to be useful to explain dialect contact phenomena. Trudgill (1986) was responsible for introducing this concept in dialect contact research, arguing that modifications in speech of mobile speakers can be described as the result of a long-term accommodation process to the native members of the community to which they migrated.

Speech modifications induced by dialect contact are also interpreted as a process of *acquisition* (Chambers, 1992; Payne, 1976). The proponents of

[1] In a more recent version, this theory is called Communication Accommodation Theory (CAT). See Giles, Coupland, et al. (1991).

the second dialect acquisition framework argue 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. In the acquisition framework, the speech modifications are described as a process analogous to the acquisition of a second language (L2), in which speakers learn how to speak a different dialect (D2) of their native language (Siegel, 2010).

This conceptual issue is of the utmost importance for linguistic research since it concerns how dialect contact impacts the speaker's linguistic knowledge. However, one obstacle that this debate faces is the enormous difficulty in implementing methodological procedures to reliably determine whether certain changes in the speaker's linguistic knowledge are permanent or not. In the absence of such procedures, dialect accommodation and dialect acquisition, most of the time, end up referring to exactly the same empirical phenomenon: new patterns in the speech of a person that emerged after they moved to a new dialectal region. For this reason, throughout this study, I use both terms interchangeably.

Within this context of dialect contact research, this study concerns the speech of internal migrants in São Paulo's state. It is the Brazilian state with the largest population and concentrates the largest urban agglomerations in the country. In fact, the São Paulo Metropolitan Region is among the ten largest metropolitan areas in the world in terms of size of the population. However, throughout Brazil's history, São Paulo has not always had such a prominent position. It was only in the 20th century that it began to experience an outstanding urban and industrial expansion. One of the central social actors of this extraordinary expansion is the rural Northeastern migrant (Fontes, 2016). Throughout the 20th century, especially in the 1950s and 1960s, São Paulo received hundreds of thousands of internal migrants from the countryside of Northeast Brazil. The accelerated industrialization attracted these migrants to São Paulo, in search of a better life and economic condition, leaving rural life behind. As a result, nowadays 46% of the population of São Paulo Metropolitan Region between the ages of 30 and 60 was born in other Brazilian states, especially in the Northern and Northeastern regions (IPEA, 2011).

Even though the migratory flow from Northeast Brazil to São Paulo has decreased in the 21st century, it has not stopped and continues to be the largest internal displacement in Brazil (Baeninger, 2012). However, in recent years, the state of São Paulo has witnessed some changes in this migratory flow. São Paulo Metropolitan Region is not the only destination anymore, since new Northeastern migrants are also moving to other urban areas, such as Campinas (Baeninger, 2005) — the second largest urban agglomerate of São Paulo.

The massive presence of internal migrants in the urban areas of São Paulo has sociolinguistic consequences. Geographical mobility and dialect contact is part of the daily lives of the inhabitants of these areas. Thus, one consequence is that a representative description of the speech of São Paulo is not possible without including internal migrants. However, as argued by Oushiro (2022), "dialect contact has received relatively less attention than language contact, on the one hand, and the speech of non-mobile, 'prototypical' native speakers of a given geographical community, on the other." For many decades, the speech of migrants was rarely discussed in sociolinguistic studies in Brazil, but this situation is changing.

The last five years experienced a remarkable expansion of sociolinguistic studies on the speech of internal migrants in São Paulo. This expansion was motivated by the pioneering work of Oushiro (2020b) who coordinates the Accommodation Project by the *Laboratório Variação, Identidade, Estilo e Mudança* (VARIEM). This project began with the first systematic study of the speech of internal migrants in São Paulo, an investigation of two speech samples of migrants from the Northeastern states of Alagoas and Paraíba, who moved to the metropolis of São Paulo and Campinas. The project aimed at determining to what degree their speech changed due to the contact with the Paulista variety (the variety spoken in the state of São Paulo). However, the project, which began as a single study, expanded its goals and became a collaborative work between researchers with a common research agenda on the outcomes of dialect contact in São Paulo. Currently, the Accommodation Project involves studies on migrant communities from Alagoas, Paraíba, and Bahia, concerning many linguistic variables, including phonetic, morphological, and syntactical variables (Guedes, 2019; Guy et al., 2022; Oushiro, 2018, 2020a, 2020b, 2020c; Souza, 2019).

As will be discussed in Chapter 2, the studies of the Accommodation

Project carried out so far gathered significant evidence of modifications in the speech of Northeastern migrants due to the contact with natives from São Paulo. Furthermore, they demonstrated that some social variables, such as the speakers' gender, age of arrival, and length of residence, can help predict how successful these speakers are in acquiring features of Paulista variety. The results of these studies (Oushiro, 2020c), however, have been restricted so far to segmental (e.g., coda /r/ as in *porta* 'door'), morphosyntactic (e.g., subject-verb agreement as in *nós fomos* versus *nós foi* 'we went'), and syntactic variables (e.g., sentential negation as *não vi* versus *vi não* or *não vi não* 'I did not see'). Prosodic variables, on the other hand, have not yet received attention.

Prosody refers to aspects of speech that emerge when speech sounds are considered together in utterances, such as intonation, rhythm, tones, stress, and tempo (Hirst & Di Cristo, 1998). To this author's knowledge, there are no studies on Brazilian Portuguese prosody from a dialect contact perspective. However, studies concerning other languages have shown that analyzing prosody can uncover important sociolinguistic phenomena. In particular, they showed that dialect contact can induce linguistic changes in prosody (Britain, 1992; Nokes & Hay, 2012; Thomas & Carter, 2006) and that mobile speakers who move to new dialectal regions can acquire prosodic patterns of these communities (Troncoso-Ruiz & Elordieta, 2018).

Once the Accommodation Project has demonstrated that the Northeastern migrants acquire segmental features of the Paulista variety, it is fair to ask if modifications in their prosody also occur. Do their intonation and rhythmic patterns change as a result of the contact with the Paulista prosody? One Alagoan migrant interviewed for the Accommodation Project seems to think that the answer to this question is positive, at least for rhythm:

(1) Interviewer: *mas eles reconheciam que você era nordestino ou alagoano mesmo?*
Interviewee: *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* […] *só que tem muitos alagoanos*

aqui [em São Paulo] *que já já puxa pro lado paulista já…* (RicardoR)[2]

In this excerpt, RicardoR, who had just moved to São Paulo at the time of the interview, says explicitly that he notices rhythmic differences between Northeastern states, such as Bahia and Alagoas, and São Paulo in "the way of speaking". More interestingly, he seems to make a reference to a process of prosodic accommodation when he says that "many Alagoans here [in São Paulo] are already speaking like Paulistas".

Other participants of the Accommodation Project also recognize that their Northeastern states differ from São Paulo in prosody:

(2) *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… sem falar no o sotaque o jeito de falar né?* (MarisaR)[3]

(3) Interviewer: *tem algum outro sotaque do Brasil aqui que você conhece?*
Interviewee: *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… mas é só na velocidade da pronúncia falaria mais devagar… já de Alagoas já é um sotaque mais acelerado* (YasminS)[4]

[2] Translation of example (1): 'Interviewer: but did they recognize that you were from the Northeast or Alagoas? Interviewee: they said 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… the 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): 'that's because we from Paraíba we know each other because here in Paulínia almost (I think that) most are all from my city then we know each other everyone… and I know right?… a Paraiban you recognize from afar 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 accent the way of speaking right?'

[4] Translation of example (3): 'Interviewer: is there some other accent in Brazil that you know? Interviewee: another accent 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 accent is more accelerated'

Both MarisaR and YasminS evaluate their native states, Paraíba and Alagoas respectively, as regions where people speak faster than in São Paulo. It seems that MarisaR, who had been living for 26 years in São Paulo at the time of the recording, can even identify other Paraiban migrants by their rhythm.

Within this scenario, this research investigates whether the speech prosody of Alagoan migrants living in Campinas Metropolitan Region changes due to the contact with the speech of the state of São Paulo. More precisely, I seek to determine which acoustic parameters distinguish the prosody of Alagoan migrants and native Paulistas from Campinas, and in which of these differentiating parameters the speech of Alagoan migrants changes due to contact with Paulista speech. I analyze seventeen acoustic parameters concerning intonation and speech rhythm in speech recordings of sociolinguistic interviews with 22 adult Alagoan migrants, balanced for gender, age of arrival, and length of residence, in contrast to a control sample of the speech of nine lifelong native residents of the city of Campinas. I analyze eight acoustic parameters concerning rhythm, namely: speech rate, stress group duration, syllabic duration peak rate, VarcoC, VarcoV, nPVI-C, nPVI-V, and the correlation between stress group duration and the number of syllables in the stress group. Concerning intonation, nine parameters are examined: fundamental frequency (f_0) standard deviation, standard deviation of f_0 peaks, f_0 peak width, f_0 peak rate, standard deviation of intervals between f_0 peaks, and mean and standard deviation of positive and negative f_0 change rates. This research addresses the following questions:

- Which acoustic parameters concerning speech rhythm and intonation most distinguish the prosody of Alagoans and Paulistas?
- Do the rhythm and intonation of Alagoan migrants in São Paulo change due to dialect contact?
- Can age of arrival and length of residence in São Paulo predict the prosodic variation in the speech of Alagoan migrants?

This text is organized into five chapters. In Chapter 2, I review the literature on dialect contact research. I first discuss why mobile speakers have not received much attention from sociolinguistic research in the variationist tradition. I also review the main theoretical works on contact-induced modifications in the speech of individuals. Then, I discuss studies on the speech of internal migrants in São Paulo. After presenting the perspective of prosody adopted here, I review dialect contact studies that analyzed prosodic variables. I end the chapter by discussing some studies concerned with prosodic variaton in Brazilian Portuguese.

Chapter 3 is an extensive description of the materials and methodological procedures of the study. I bring information about the speech samples and the social profile of the participants. Then, I move on to the acoustic procedures and instruments used to obtain measures of prosodic variables. I also discuss the computational techniques that I developed to extract and analyze a large amount of acoustic data.

The results of the study are presented in Chapter 4, which is divided into three parts, each one corresponding to a different type of analysis. In the first, I compare the prosody of Alagoan migrants and native Paulistas in order to identify the main similarities and differences between them. The second part reports the core results of the study, which concern the effects of dialect contact on the speech of Alagoan migrants. In this part, I analyze whether the speaker's age of arrival and length of residence are good predictors of prosodic accommodation. The last part of the chapter is dedicated to the analysis of individual variation in Alagoan migrants' prosody.

The dissertation ends with Chapter 5. In this closing chapter, I summarize the main conclusions that can be drawn from the results of the study and address questions for future research.

Chapter 2

Background

This chapter reviews previous studies on dialect contact and prosody. In Section 2.1, 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 2.2, I review recent research about the speech of internal migrants in the state of São Paulo. Section 2.3 is a discussion of the concept of prosody, intonation, and rhythm. In Section 2.4, I discuss sociolinguistic investigations about prosodic variation. Finally, I end the chapter with Section 2.5, briefly discussing some works on prosodic variation in Brazilian Portuguese.

2.1 Migrants in sociolinguistic research

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). Kerswill (1993) argues that the nativeness of the speakers serves in variationist tradition as one of the criteria for defining a speech community:

> People are considered to belong to the speech community in which they live only if they are 'native' to the community. Various criteria, mainly to do with place of birth or age on arrival, are applied in order to sift

> out people who are not 'native'. In Labov's (1966) New York Lower East Side study, this leads to an extreme reduction of the target population; my calculations, based on Labov's presentation (1966), suggest that well over 50% of the original sample are excluded by various nativeness-related criteria (Kerswill, 1993, p. 35).

One motivation to exclude migrants from sociolinguistic analyses relates to Lenneberg's (1967) critical period hypothesis (Nycz, 2013; Oushiro, 2016a). According to this hypothesis, around the first years of puberty, the speaker's linguistic system stabilizes, and, from then on, it becomes more resistant to changes in face of new linguistic stimuli. For Lenneberg, puberty is the dividing line between naturally acquiring a native language and laboriously learning a second language:

> Also automatic acquisition from mere exposure to a given language seems to disappear after this age [puberty], and foreign languages have to be taught and learned through a conscious and labored effort. Foreign accents cannot be overcome easily after puberty (Lenneberg, 1967, p. 176).

The argument usually used by sociolinguists to justify the nativeness criterion relies on extending Lenneberg's critical period hypothesis to the acquisition of dialects. It 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 grew up — that is, patterns they acquired during childhood before migrating. Based on this argument, excluding from sociolinguistic samples migrants who moved to the community after a certain age became a common practice. 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.

Even though Lenneberg's theory remains influential, it has been challenged by phonetic studies on second language learning. For instance, Flege et al. (1995) found a strong linear relationship (r = 0.71) between the age of learning a second language and the degree of perceived foreign accent in a perception experiment in which native English-speaking listeners evaluated the degree of accent in English sentences produced by native Italians. The result suggests that the speaker's age is a significant predictor of linguistic acquisition and learning, but it does not cause a sharp disruption in the speaker's learning capabilities.

Lenneberg's hypothesis should not discourage sociolinguists from studying the speech of mobile speakers. Quite the opposite, it should be seen as a stimulus since analyzing variation in migrants' speech in a dialect contact situation can provide valuable data to assess the adequacy of hypotheses concerning linguistic acquisition in adulthood. For instance, the relation between speakers' age when migrating to the new community and their success in acquiring new dialect features from this community can elucidate if there is a critical period in second dialect acquisition. If there is, we would expect a sharp division among migrants, with successful acquisition of the new dialect only by the ones who migrated at a young age. Another possibility is to observe results similar to Flege et al. (1995), with age of arrival showing a linear relation with the success in acquiring dialectal features. As will be seen in Chapter 4, the findings of this present study reinforce Flege's (1995) position that the speakers' age has a continuous, instead of a discrete relationship with acquisition. In other words, the results are evidence that speakers who migrate after puberty can still acquire new dialectal features, but the older they are the less successful the acquisition tends to be.

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 speakers from different regional varieties.

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 non-lateral one) 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]); (iii) 1st person plural subject-verb number agreement (e.g., *Nós queria ir* 'we wanted to go' instead of the prestige form *Nós queríamos ir*); and (iv) 3rd person plural subject-verb number agreement (*Eles queria ir* 'they wanted to go' instead of the prestige form *Eles queriam 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. Male speakers were more advanced than females in this process of accommodation in the case of the two morphosyntactic variables (1st and 3rd person plural subject-verb number agreement). Men and women were at the same stage in relation to the delateralization of [ʎ], and women were slightly ahead males in the case of the reduction of rising glides. 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 stigmatized rural variants.

Trudgill (1986) developed a broader theoretical framework to analyze the linguistic changes resulting from dialect contact. As mentioned in Chapter 1, one of his contributions was to explain dialect contact phenomena from the perspective of accommodation theory (Giles, 1973), explaining 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 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 [tʰ], 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, which advocates for the use of the term *dialect acquisition* in place of Trudgill's *dialect accommodation*. He 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 British form "coach" in place of the Canadian variant "bus") and pronunciation variables[1] (e.g., the acquisition of the aspirated plosive [tʰ] in the place of the tap [ɾ], in words like "putting"). Based on the patterns identified in these children's speech, analyzed in the light of the results of other studies on second dialect acquisition, 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 the theoretical model and the methodological procedures of Variationist Sociolinguistics (Labov, 2006) can be employed to investigate the complex phenomenon of the acquisition of dialectal features by migrants. With quantitative analysis techniques,

[1] Chambers uses the term "pronunciation variables" to refer to allophonic variation — different phonetic forms of the same phoneme.

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. The quantification allowed for the identification of linguistic- and social-related patterns of dialect acquisition. These studies made it clear 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). The studies of Bortoni-Ricardo (1985) and Trudgill (1986) 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 (2016a) 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, 2016a, p. 10, my translation). As already pointed out above, excepting for a few studies (such as the work of Bortoni-Ricardo (1985)'s, which analyzes the morphosyntactic variable first and third-person 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.

2.2 Dialect contact in São Paulo

The present research takes place in the context of expansion of sociolinguistic studies on the speech of internal migrants in Brazil, especially the ones about communities of Northeastern migrants in the state of São Paulo. As mentioned in Chapter 1, Oushiro (2020b; 2020a; 2020c) has recently conducted one of the first systematic studies about the speech of internal migrants in Brazil. She analyzed speech samples of migrants from the states of Alagoas and Paraíba, both located in Northeast Brazil, who moved to São Paulo and to Campinas in the Southeastern state of São Paulo, with the objective of determining to what degree their speech changed due to the contact with the Paulista variety (the variety spoken in the state of São Paulo). The analyses focused on the effects of age of arrival and length of residence in the host community in relation to five linguistic variables: (i) the realization of coda /r/ (e.g., *porta* 'door' pronounced as a fricative — the Northeastern variant —, or as a tap or an approximant rhotic — the Paulista variants); (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/ (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) the sentential negation (e.g., *não vi* or *não vi não/vi não* 'I haven't seen').

The results indicate that the speakers' age when they moved to São Paulo performed an important role in the accommodation to the phonetic variables (i.e., coda /r/, /t, d/ before [i], and pretonic mid vowels /e/ and /o/ height), but not to the morphosyntactic ones (i.e., nominal number agreement and sentential negation). On the other hand, length of residence in the host community correlated only to coda /r/ (Oushiro, 2020b).

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 (2020b). The results of Santana (2018) also showed that the type of migrant Sergipanos' networks (open, with more frequent contact with Paulistas, or closed, with less contact) does not correlate with dialect accommodation for the pretonic mid vowels (see also Santana and Mendes, 2019).

Souza (2019) analyzes sociolinguistic evaluations about the variable use of imperative forms with subjunctive or indicative morphology (e.g., *traga* vs. *traz o carro!* 'bring the car!'), comparing 40 non-migrants and 32 migrants from the Northeastern state of Bahia. The migrants were living in São Paulo's state. His findings show that the use of the indicative forms, which is the predominant variant in São Paulo, is higher in migrant than in non-migrant Bahians, and that less educated speakers favor the indicative morphology. In turn, Oliveira (2020) investigates the accommodation to Paulista's variants of coda /r/ (alveolar tap and approximant) in the speech of Bahians living in Bauru, a city in the western of São Paulo state. The author observed that the accommodation to coda /r/ can be conditioned by speaker's attitude, age, length of residence, age of arrival, speech style, and contact frequency with native Paulistas.

All these studies on the speech of Northeastern migrants in São Paulo have significantly contributed to understanding the effects of dialect contact on their speech. However, none of them have analyzed speech prosody, which is the topic of the next sections.

2.3 The phonetic perspective of prosody

Before discussing sociophonetic studies on prosodic variation, I will briefly review some concepts of speech prosody, according to the main phonetic works that underlie this research.

In her comprehensive review of the literature on timing and rhythm in speech, Fletcher (2010) warns the reader about the existence of different but related meanings of speech prosody. Thus, it is wise to make it clear which of these meanings is adopted here. According to Fletcher (2010), in contemporary language research, there are two main conceptions of prosody which are directly related to two different research areas, phonetics and phonology:

> Prosody or *prosodic* features are for many phoneticians and speech scientists synonymous with variations in *suprasegmental* parameters such as duration, intensity, and f_0 that contribute in various combinations to the production and perception of stress, rhythm and tempo, lexical tone, and intonation of an utterance. [...] This leads us to the more abstract phonological use of the term prosody: as the abstract

> hierarchical phonological structure(s) of an utterance, and *prominence* relations within that structure (Fletcher, 2010, p. 521).

In this research, I analyze prosody from the phonetic perspective, as an empirical phenomenon that can be observed and examined by means of instruments for articulatory, acoustic, and auditory data collection (Barbosa, 2012; Fant, 1970; Lehiste, 1970; Rubin & Vatikiotis-Bateson, 1998; Sagisaka et al., 2012; Sudhoff, 2006). More precisely, I examine prosody at the physical level, by measuring observable and quantifiable physical parameters.

Acoustically, prosody is the simultaneous variation in time of four physical parameters: *fundamental frequency* (f_0), *duration*, *intensity*, and *spectral distribution* (Barbosa & Madureira, 2015; Fant, 1970; Hirst & Di Cristo, 1998; Johnson, 2012; Ladefoged, 1996; Speaks, 1992; Stevens, 2000). This simultaneous variation results 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 into four 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; *volume*, by which the hearers distinguish strong and weak sounds; and *quality*[2] or *timbre*, which is the perception of differences between sound sources. The various combinations of these parameters through the speech flow give rise to the perception of speech prosody (Barbosa, 2012).

In this research, I analyze intonation and rhythm. *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, duration is the primary correlate of rhythm (Barbosa, 2019; Fletcher, 2010). *Rhythm* is 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

[2] In the case of speech, quality is related to the perception of differences between sounds produced by different configurations of the same vocal tract (e.g., in the production of different vowels) or sounds produced by different vocal tracts (e.g., the difference between the same vowel being produced by a small vocal tract of a child or a large vocal tract of adult male).

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.

Sociophonetic studies on prosody usually analyze either intonation or rhythm, but not both. 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 Sene (2021), 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, such as Britain's (1992) study on linguistic change in the melodic contour of declarative clauses in New Zealand English.

As for 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's (1945) and Abercrombie's (1967) works[3]. This typology emphasizes regularity rather than alternation in rhythm, separating the languages of the world into two rhythmic categories: syllable-timed and stress-timed. 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 the term "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

[3] Even though Pike (1945) and especially Abercrombie (1967) were most responsible for popularizing rhythmic types, this typology goes back to Jones (1918) and Classe (1939).

a perfect syllable-timed or stress-timed rhythm (Barbosa, 2000; Fletcher, 2010) — actually, in the late 1930s, Classé already recognized that isochrony is very rare in spontaneous speech (Classé, 1939 *apud* Cummins and Port, 1998). 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, as is the case of Brazilian Portuguese (Barbosa, 2000; Frota & Vigário, 2000).

The rise of the continuous conception of rhythm fostered the emergence of rhythm metrics (Deterding, 1994; Grabe & Low, 2002; Ramus et al., 1999). 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 (Fuchs, 2016). 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. 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 3.4.

2.4 The social embedding of prosody

One of the first sociolinguistic studies on prosody was conducted by Britain (1992). The author analyzes a process of linguistic change in an intonation feature of New Zealand English. The rising melodic contour at the end of interrogative utterances is a common feature of many languages, including English and Portuguese. This type of contour has a particularly salient use in New Zealand English: it can also characterize declarative sentences — a phenomenon that is observed in some other languages and English varieties. In the literature about English prosody, this rising pitch with a declarative function is usually called “high rising terminals” (HRT). Britain (1992) studied this phenomenon in the speech of 75 inhabitants of a small city in New Zealand, stratified by age, sex, ethnicity, and social class. His results show that young speakers, especially Maori and *Pakeha* women — in Maoris' culture, *Pakeha* are New Zealanders of European descent — favor the use of HRTs, which suggest a linguistic change in progress.

As discussed above, the majority of the sociolinguistic studies on rhythm consist of analyses of rhythm metrics. One of the first studies to use these metrics is Szakay (2006), in which rhythmic variation is analyzed in English spoken by 36 New Zealanders, of which 24 were Maoris and the others were *Pakeha*. The author calculated the Pairwise Variability Index (PVI), a metric developed by Low et al. (2000) and explained in Section 3.4. Szakay (2006) showed that Maori English has a significantly more syllable-timed rhythm than *Pakeha* English, which tends to be more stress-timed. The author also analyzed intonation and found a higher mean fundamental frequency in Maori English (128.4 Hz), compared to *Pakeha* (109.5 Hz). 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 changes in the rhythm of 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 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 between the ethnic groups 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 through 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”.

Fagyal (2010) investigated the speech rhythm of five French speakers of European origin and five French speakers of North-African origin. All of them were born and raised in *La Courneuve*, one of the poorest areas in the periphery of Paris. The author hypothesized that the heritage languages of the speakers of African origin have an influence on their speech, making it more stress-timed. On the other hand, syllable timing was expected among the participants of European origin. To test these hypotheses, Fagyal (2010) examined two rhythm metrics, namely, the vocalic percentage (V%) and the standard deviation of consonantal (ΔC) and vocalic (ΔV) intervals. Contrary to the expectations, the results show that both groups are very similar in the degree of variability in vocalic and consonantal intervals (Δ metrics). The participants of African origin showed a slightly higher percentage of vocalic interval (V%), which can be caused by vowel lengthening or consonantal reduction.

More recently, Troncoso-Ruiz and Elordieta (2018) investigated the process of prosodic accommodation in the speech of Andalusian speakers (Southeast Spain) who 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 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 with the raising contour of the Andalusian variety — was the most acquired feature by the migrants. This result is interesting in that 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.

2.5 Prosodic variation in Brazilian Portuguese

Within Pike's (1945) and Abercrombie's (1967) framework of rhythmic typology, most studies on Portuguese describe the Brazilian variety as having a hybrid rhythm (Abaurre-Gnerre, 1981; Barbosa, 2000; Cagliari & Abaurre, 1986; Frota & Vigário, 2000; Moraes & Leite, 2002), but a few studies argue in favor of classifying this variety as strictly stress-timed, such as the works of Major (1981; 1985) and, more recently, Cagliari (2013) and Migliorini and Massini-Cagliari (2010).

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 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, to be pronounced as /ˈfɾidɐ/, with the complete deletion of the pretonic vowel. An example of pretonic vowel lowering is the pronunciation of the word /foˈfɔkɐ/ 'gossip' with 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 (Fletcher, 2010; Ramus et al., 1999). 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 in Brazil), while in Rio de Janeiro and São Paulo the frequencies are, respectively, 5% and 0% (Barbosa 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, 2019, p. 687).

There are not many studies on rhythmic variation in Brazilian Portuguese, and even fewer on intonation. Most studies on the intonation of Brazilian Portuguese concern the abstract phonological description of prosodic structures according to Autosegmental-Metrical Phonology (Nespor & Vogel, 2007) and Intonational Phonology (Ladd, 2008). Although these phonological theories contribute to understanding the relationship between prosody and linguistic structure, the studies based on them are hardly concerned with sociolinguistic variation. To this author's knowledge, the few phonological studies that have a focus on intonational variation are concerned with differences in intonation between European and Brazilian Portuguese, such as Frota and Moraes (2016), Frota and Vigário (2000), and Tenani (2006).

Chapter 3

Data and methods

In this chapter, I outline the methodological procedures used in this research. I begin by discussing two methodological approaches concerning data collection in dialect contact research. Next, in Section 3.2, I present the two speech samples that served as the raw data for this study. Then, in Section 3.3, I discuss the prosodic units that are the building blocks to acoustically measure prosodic phenomena. From acoustically 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 3.4, I explain what each of these measures informs about prosody and how they are mathematically computed. Finally, in Section 3.5, I describe in detail the computational techniques used to automate speech annotation and the extraction of these acoustic measures.

3.1 Longitudinal and cross-sectional approaches

The study of how migrants' speech changes as a result of their daily interactions with people who speak a regionally different variety presents a double 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 discover how we can disentangle the patterns in 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 (Blondeau, 2013; Cukor-Avila & Bailey, 2013).

Panel data seems to be the ideal type of data for dialect contact research (Britain, 2002; Nycz, 2015). 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 identification of patterns after the speakers come into contact with another dialect. In other words, panel data enables tracking the linguistic behavior of a speaker or group of speakers in real time, as the exposure to the new dialect increases. Unfortunately, as ideal as it is to analyze dialect accommodation, panel design brings about 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 (Chambers, 1992; Tagliamonte & Molfenter, 2007; 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 recontacting the same speakers 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 (see, e.g. Oushiro 2020b; Payne 1976; Walker 2014) have turned to cross-sectional alternative approaches. Cross-sectional studies compare the behavior of two or more different populations at the same point in time, in contrast with longitudinal studies comparing the behavior of the same population over 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.

One type of cross-sectional design 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 moved to the same host community. 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.

This type of cross-sectional design can be used 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). One possible strategy to estimate accommodation is by examining the frequency of the host community's specific linguistic features in migrants' speech. The recurrent use of these features by migrants can be taken as evidence of dialect acquisition. Another strategy is to estimate accommodation degree 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 can suggest that the migrants are favoring alternative forms, probably the forms with which they have more contact in the host community.

The cross-sectional design can also be used to assess the relevance of specific social and linguistic factors as predictors of dialect accommodation. In this research, two social factors are analyzed from a cross-sectional perspective. One of them is length of residence. We can reasonably think that the longer the speakers have lived in the host community, the more likely they will have acquired the linguistic features of this community (Siegel, 2010). Panel studies measure the effects of length of residence on dialect acquisition by tracking the same speakers over the years. 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 & 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 changes).

The other social variable analyzed in this study is age of arrival; that is, the speakers' age when they arrive in the host community. Based on considerations related to the critical period hypothesis discussed in Section 2.1, we can think that the younger the speaker when moving to the new community, the greater the chances of successful acquisition of the dialectal features of this community. In both approaches, the effects of age of arrival are usually measured by 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.

The cross-sectional approach has undeniable practical advantages over the longitudinal methods of data collection. Time issues are probably the main reason that leads researchers to opt for this approach. It enables collecting speech recordings in a short time while allowing for the control of relevant variables for the study of dialect accommodation. 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 3.2).

The practical advantages do not come without 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.

These obstacles can be lessened 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.

3.2 Corpora

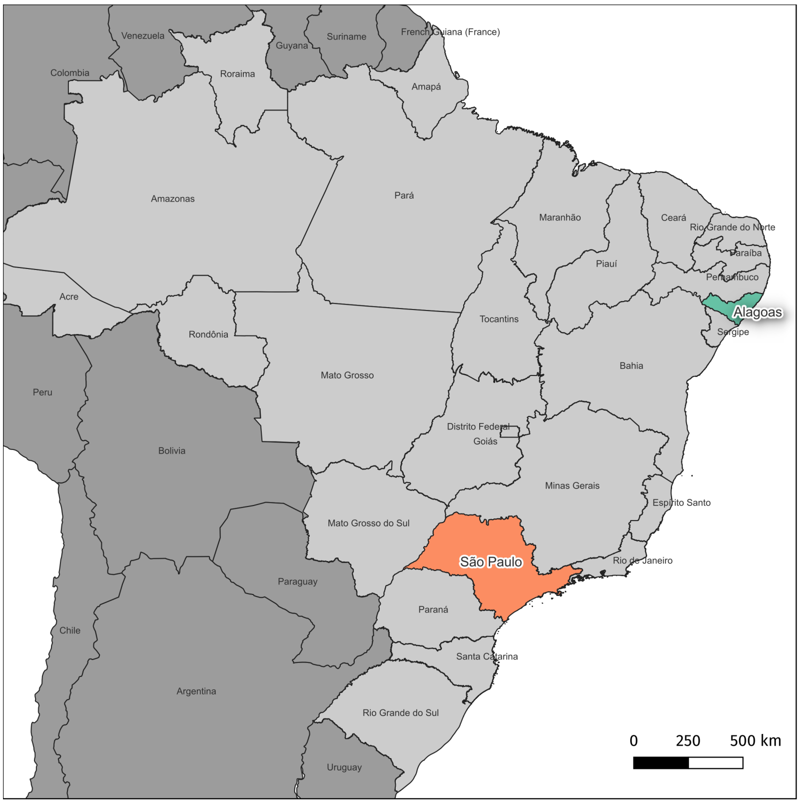
In this research, I take the cross-sectional approach for the analysis of prosodic accommodation in the speech of Alagoan migrants in São Paulo. I have analyzed two speech samples, both collected by other researchers in previous projects. One of them is the ALCP sample, collected between 2016 and 2018 by the *Laboratório Variação, Identidade, Estilo e Mudança* (VARIEM) of University of Campinas (UNICAMP), as part of the Accommodation Project (Oushiro, 2016b, 2018). The other is the CPS sample, collected in 2019 by Mourão (2018) for her master's research at UNICAMP. In the following sections, I describe the main characteristics of these samples.

3.2.1 Target sample

The ALCP sample is composed of audio recordings of the speech of 22 Brazilian speakers from the Northeastern state of Alagoas who moved to the Campinas Metropolitan Area, in the Southeastern state of São Paulo (see Figure 3.1). This sample is especially suitable for this research because it was collected 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 & Gordon, 2003). The objective of this technique is to facilitate entering a community

Figure 3.1: Location of the states of São Paulo and Alagoas in Brazil



(Tagliamonte, 2006). In this method, the researchers can use their own social network to recruit the first potential participant, by asking their friends if they know someone who fits the study's inclusion criteria. By snowballing, the researcher then asks the first participant to refer to other potential participants, 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 & Gordon, 2003).

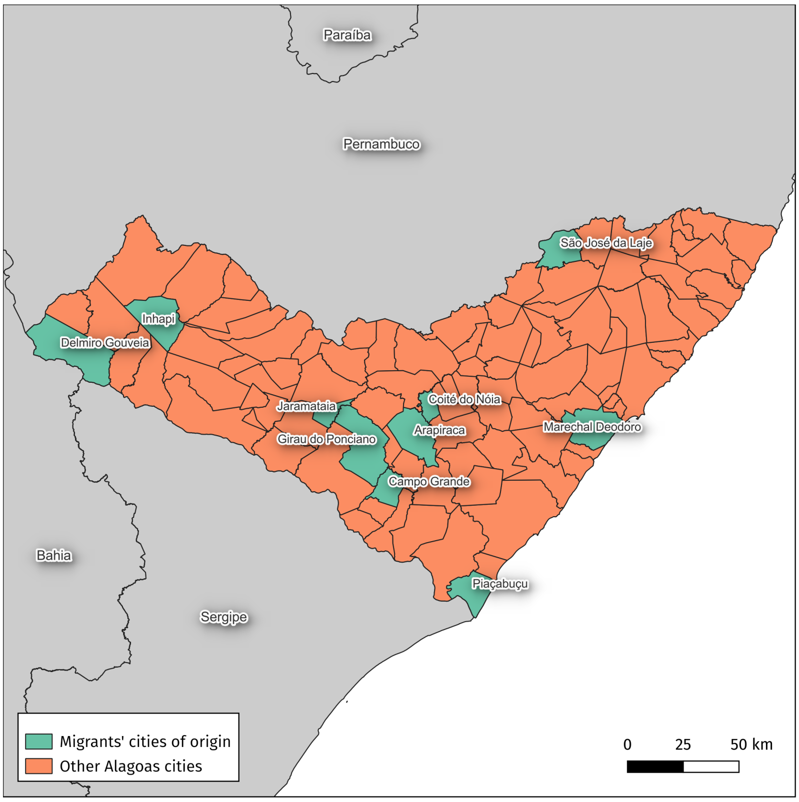
To reduce the effects of potential confounding variables (Gries, 2013), Oushiro (2018) controlled participants' profiles in relation to age, education,

Table 3.1: Social profile of the ALCP sample's participants

Participant	Gender	Age	City of origin	Job
ArthurG	Male	36	Arapiraca	Restaurant assistant
JorgeS	Male	33	Arapiraca	Security guard
JosueA	Male	23	Arapiraca	Cook
LucimaraF	Female	38	Arapiraca	Housewife
MarleneN	Female	23	Arapiraca	Manicure
RicardoR	Male	27	Arapiraca	Bartender
WellingtonF	Male	21	Arapiraca	Waiter
YasminS	Female	35	Arapiraca	Cashier
AlbertoS	Male	39	Campo Grande	Self-employed
CleoniceC	Female	27	Campo Grande	Painter's assistant
MoniqueR	Female	32	Campo Grande	Factory worker
IsabelaS	Female	60	Coité do Noia	Kitchen assistant
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

and geographical area. The research team ensured that all participants had lower levels of education (no greater than a high school diploma), had come from rural areas, and were between the ages of 19 and 65 — i.e., retirees[1] and persons in school-age were excluded. Table 3.1 lists the 22 participants of the ALCP sample by their pseudonyms, along with their gender, age, and city of origin in the state of Alagoas. The sample has a majority of male speakers (62%) with an average of 33 years of age.

Figure 3.2: State of Alagoas



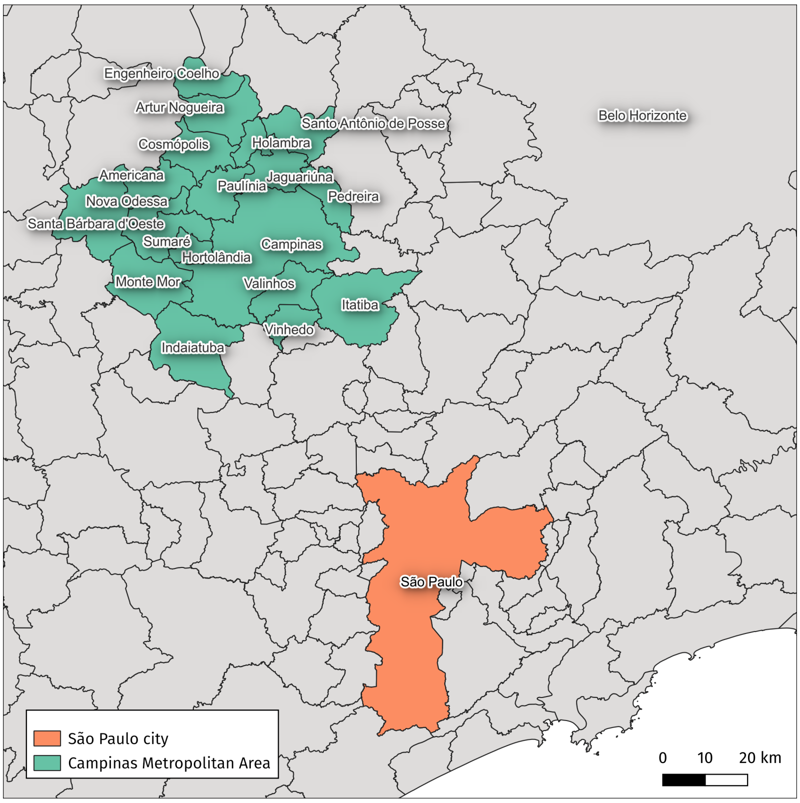
Almost all participants have come from the countryside of Alagoas, except for JosiasP from Piaçabuçu and VeronicaS from Marechal Deodoro, two coastal cities (see Figure 3.2). Compared to the Campinas Metropolitan Area, which has more than 3 million inhabitants (of which 1.2 million are residents of the city of Campinas), the participants' native cities are much smaller. The biggest is Arapiraca, with approximately 230,000 inhabitants.

[1] One exception is the participant JosiasP who was 63 years old but already retired.

All the others have no more than 50,000 residents, and Jaramataia, the native city of EraldoF, is the smallest, with only 5,000.

In addition to the city of Campinas, with the biggest population, the Campinas Metropolitan Area also comprises smaller cities (see Figure 3.3). Of the 22 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; one in Sumaré; one in Indaiatuba, and one in Vinhedo.

Figure 3.3: São Paulo city and Campinas Metropolitan Area



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 a waiter, manicure, kitchen assistant, hawker, etc., activities that make up a large part of the informal sector in Brazil.

As previously mentioned, 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, 2020b). 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 (Bortoni-Ricardo, 1985; Chambers, 1992; Payne, 1976; Siegel, 2010; Trudgill, 1986). However, the effects of these two variables can be easily confounded since, as explained by Oushiro (2020b), “it is often the case that speakers who arrived earlier are also the ones who have lived the longest in the host community”.

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

Table 3.2: Distribution of the ALCP sample’s participants by sex, age of arrival and length of residence

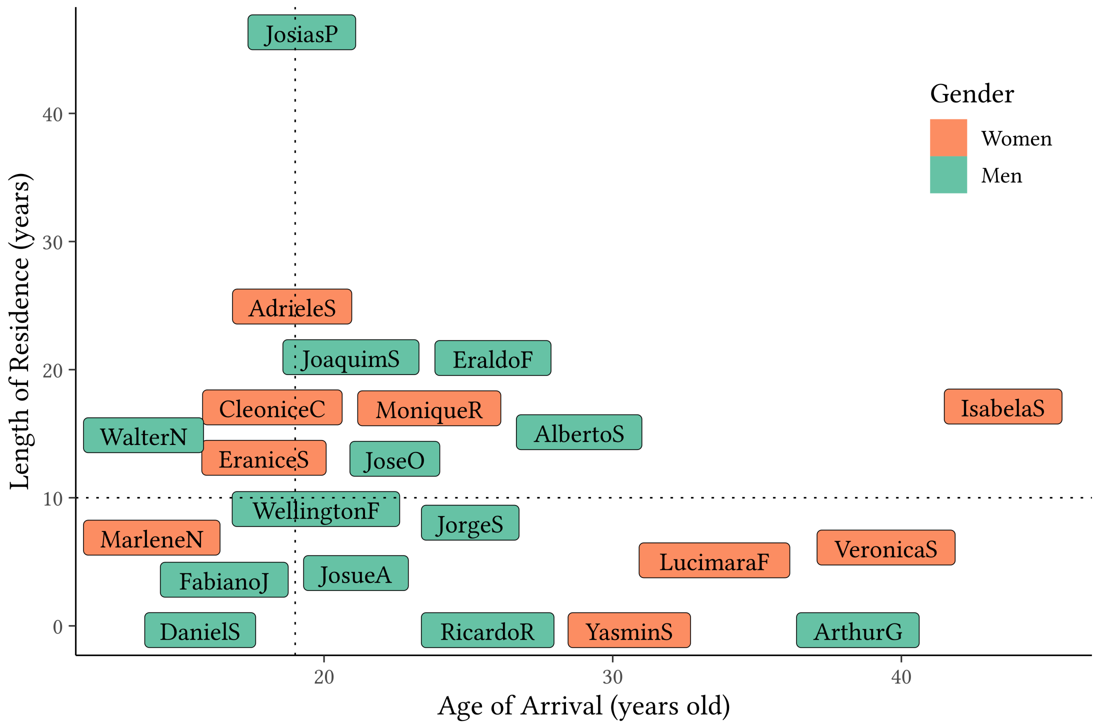
Length of Residence		Age of Arrival: Before 20 y.o.	Age of Arrival: 20 y.o. or more
Less than 10 years	Female	EraniceC MarleneN	LucimaraF VerônicaS YasminS
	Male	DanielS FabianoJ WellingtonF	ArthurG JoséO JosuéA RicardoR
10 years or more	Female	AdrieleS CleoniceC	IsabelaS MoniqueR
	Male	JoaquimS JosiasP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

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 & Zilles, 2007). Table 3.2 shows the crossed distribution of the ALCP's speakers by age of arrival, length of residence, and gender. The important fact to notice is that there is no empty cell in the table, with at least two participants filling each combination between the factors of these three variables.

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 3.4 shows the distribution of the speakers by these two variables on continuous scales in years. The difference in gender is also shown. The distribution is not as

Figure 3.4: Distribution of the ALCP sample's participants by gender, age of arrival, and length of residence



balanced as in Table 3.2, but, as noted by Oushiro (2018), the imbalance does not appear to be large enough to interfere with their statistical analyses as continuous.

Except for a married couple that was recorded together (MoniqueR and AlbertoS), all participants were individually recorded by Linguistics students native to 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 started, 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 CPS sample (see Section 3.2.2). In this text, I will refer to specific speakers from these samples only by their pseudonyms.

The interview schedule used for the collection of the ALCP sample comprised conversational modules, the reading of a word list, and two questionnaires. Each interview lasted about one hour. Each module comprised an ordered set of questions about some aspect of the participants' life: their current neighborhood in the community of destination, childhood in their home state, family, work, and leisure activities, as well as the Campinas Metropolitan Area (Oushiro, 2020b, p. 80).

Alongside the primary goal of eliciting more spontaneous speech by encouraging the participants to narrate some of their 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 with whom the participants interact the most in their daily life; their personal evaluations of the Alagoas state, the São Paulo state, and Northeast Brazil; which sociolinguistic variables are more salient to them and, by omission, which are below their level of consciousness; their migration trajectories (to which locations they have moved, how long they lived in these places, etc.); and their overt evaluations of regionally different ways of speaking (*sotaques*) in Brazil.

At the end of the interview, participants were asked to read a word list, when interviewers deemed appropriate. Since some of them were illiterate, interviewers had been explicitly instructed to sensibly assess, throughout the conversation, whether this task would potentially cause the speaker any discomfort. In this study, the word list was 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 quantifying how much the participants identify themselves with their community of origin (in the case of the ALCP's speakers, Alagoas) and also with their community of destination (São Paulo), and how strong are their ties with these communities. The second questionnaire was optional and contained questions regarding the speaker's socioeconomic status, such as the participant's average monthly income, to better evaluate possible differences in the accommodation to Paulista speech due to differences concerning the social mobility of the migrants (Oushiro, 2018).

3.2.2 Control sample

The control sample (CPS) is composed of audio recordings of the speech of native speakers from the city of Campinas. This sample is part of a larger corpus collected by Mourão (2018) with the purpose of analyzing the spatial diffusion of linguistic features (e.g., the affrication of /t/ and /d/ before [i], as in *tia* 'aunt', and the diphthongization of nasal /e/, as in *fazenda* 'farm') from São Paulo to Jundiaí and Campinas.

The recruitment of the 24 participants aimed to balance four social variables: (i) gender (males, females); (ii) age group (20–34 y.o., 35–59 y.o., 60 y.o. or more); (iii) level of education (up to high school, college); and (iv) city (Campinas, Jundiaí). In this research, I only analyze speech samples of nine speakers from Campinas who gave consent for their speech recording to be used in other studies.

The participants were recorded in sociolinguistic interviews with an average duration of 45 minutes. The structure of the interview is similar to that of the ALCP sample, described in Section 3.2.1, since both ALCP and CPS interview schedules were based on the SP2010 corpus (Mendes & Oushiro, 2012). The CPS interview has three 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 questions about linguistic topics and has the purpose of eliciting social evaluation of linguistic variables. In the third part of the interview, the speakers were recorded reading a word list, a newspaper excerpt, and an oral text. Then, they were asked to respond to a questionnaire specifically designed to gather information about the participant's social networks and mobility degree. In this research, I analyze data only from the two first parts of the interview.

3.3 Units of prosody

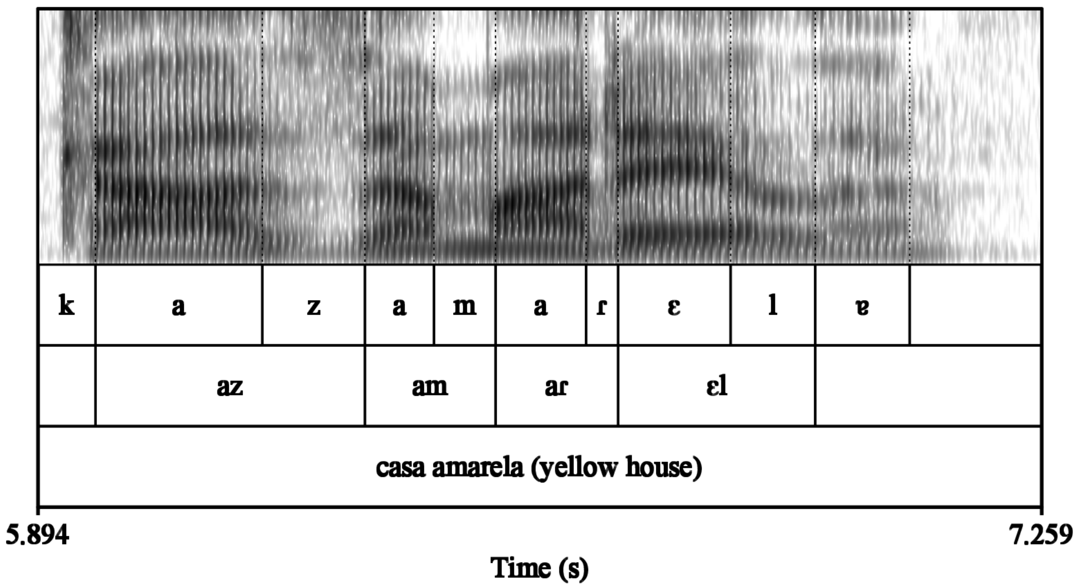
In this research, I analyze speech prosody in the ALCP and CPS samples using acoustic techniques of speech analysis. As mentioned in Chapter 1, I analyze seventeen acoustic parameters concerning intonation and speech rhythm. The following eight of them refer to speech rhythm: speech rate, stress group duration, syllabic duration peak rate, VarcoC, VarcoV, nPVI-C, nPVI-V, and the correlation between stress group duration and the number of syllables in the stress group. The remaining nine concern intonation: f_0 standard deviation, standard deviation of f_0 peaks, f_0 peak width, f_0 peak rate, standard deviation of intervals between f_0 peaks, and mean and standard deviation of positive and negative f_0 change rates. In order to compute these prosodic variables, it is necessary first to segment and annotate the speech recordings in four acoustic units, namely, phones (vowels and consonants), syllables, stress groups, and chunks. For this reason, before describing how each prosodic variable is measured (see Section 3.4), I discuss the criteria that I adopt to demarcate these units in the speech recordings.

3.3.1 Syllables

In this study, 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 (Mendonça, 2003; Zec, 2007). 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 vowel onsets (Barbosa, 2009). In other words, the V-to-V is the interval from a vowel onset until immediately before the next vowel's onset. Therefore, despite its name, this unit comprises a single vowel. Instead of a phonological syllable, the V-to-V never starts with consonants — i.e., whenever there is one or more consonants within the V-to-V unit, the consonants always follow the vowel onset.

Being a phonetic unit, the V-to-V has to be demarcated based not on the

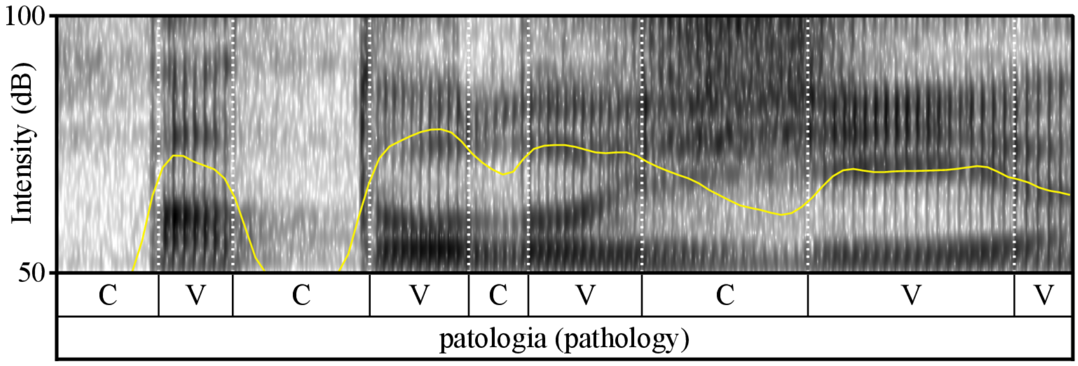
Figure 3.5: Spectrogram of the noun phrase *casa amarela* 'yellow house' exemplifying speech segmentation in vowels and consonants (first tier) and V-to-V units (second tier) using Praat.



abstract phonological representation of lexical items but on phonetic data of each particular production. Figure 3.5 shows an example of segmentation into vowels and consonants in the first tier and V-to-V units in the second one. We can see 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 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 3.6 shows the spectrogram of a recording of my own 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

Figure 3.6: 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.



becomes steeper) are precisely the transition from a consonant to a vowel.

Experimental studies on the perceptual center, such as Barbosa et al. (2005) and Pompino-Marschall (1989), support the V-to-V unit. They show that when asked to produce a sequence of syllables (e.g., "pa, pa, pa, …") in synchronicity 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 for 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 such.

3.3.2 Stress groups

Syllable groupings constitute larger units that also integrate speech prosody, such as the stress group. A stress group comprises one prominent syllable, called phrasal stress, and one or more adjacent non-prominent syllables (Barbosa, 2006; Crystal & House, 1990; Fletcher, 2010). It is noteworthy that phrasal stress here does not refer to lexical stress, but to *prosodically* stressed syllables. In other words, phrasal stress 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 stress is perceived as closing the stress group, but in some languages, like English, phrasal stress starts the stress group. A stress group ended by phrasal stress 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 stress, 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 stress. 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 capable of identifying phrasal stress with an accuracy rate higher than 80%

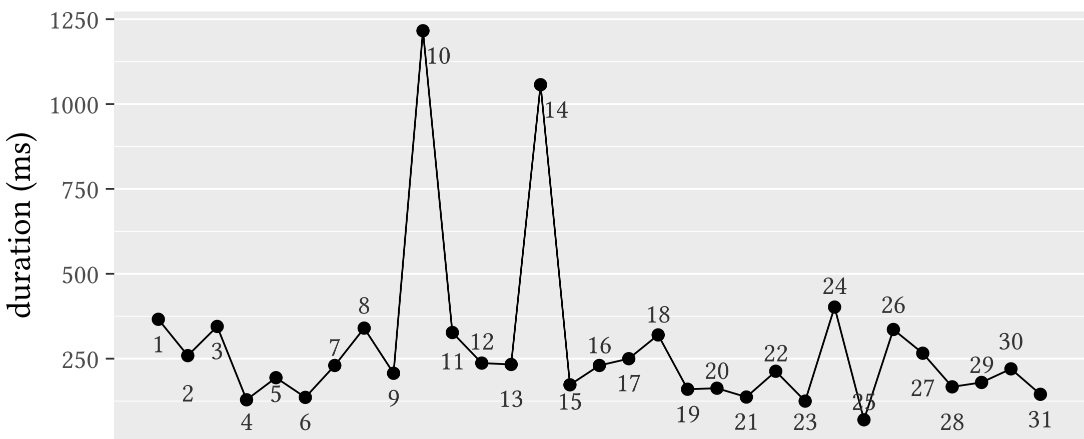
(Barbosa, 2007, p. 732). That is, more than 80% of stress groups perceived by native hearers are predicted by the normalized syllable duration.

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. A contour that indicates the absolute syllable duration (usually in milliseconds) is called *raw duration contour*. Figure 3.7 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 stresses perceived by the participants would possibly not match all duration peaks in Figure 3.7.

The reason for the mismatch is that rhythmic salience does not concern syllable duration 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

Figure 3.7: V-to-V raw duration contour of an utterance from Accommodation Project.



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 that of stressed ones since speakers already expect a stressed vowel to show 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 et al., 1999). In short, the intrinsic duration of segments must be "neutralized" by a normalization procedure.

Barbosa (2007) normalizes syllable duration by 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.

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 syllable's intrinsic duration), we can represent this distance by transforming absolute duration in z-scores, taking the intrinsic mean and standard deviation of each segment in the syllable as reference. By 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 so will have its duration value reduced and a short syllable that is longer than expected will have its value increased.

Equation 3.1 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 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

[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 et al., 1999).

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).

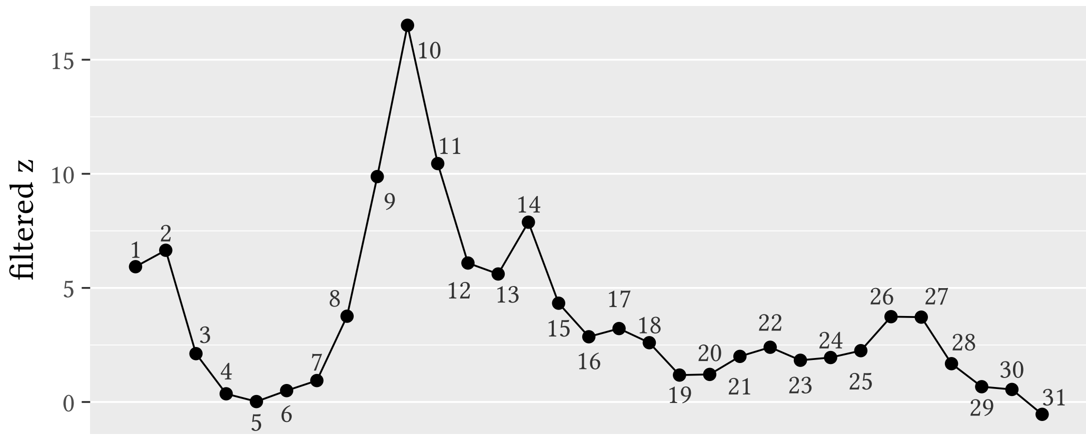
$$z_i = \frac{d_i - \sum_{j=1}^{n} \mu_j}{\sum_{j=1}^{n} \sigma_j} \tag{3.1}$$

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. The 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. The equation 3.2 represents the application of five points moving average to z-scores.

$$\text{smoothed } z_i = \frac{1 \cdot z_{i-2} + 3 \cdot z_{i-1} + 5 \cdot z_i + 3 \cdot z_{i+1} + 1 \cdot z_{i+2}}{13} \tag{3.2}$$

Figure 3.8 shows the normalized and smoothed version of the duration contour of Figure 3.7. 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 stresses and demarcate stress groups in speech recordings.

Figure 3.8: V-to-V normalized (z-scores) and smoothed (5-points moving average) duration contour of the same utterance of Figure 3.7.



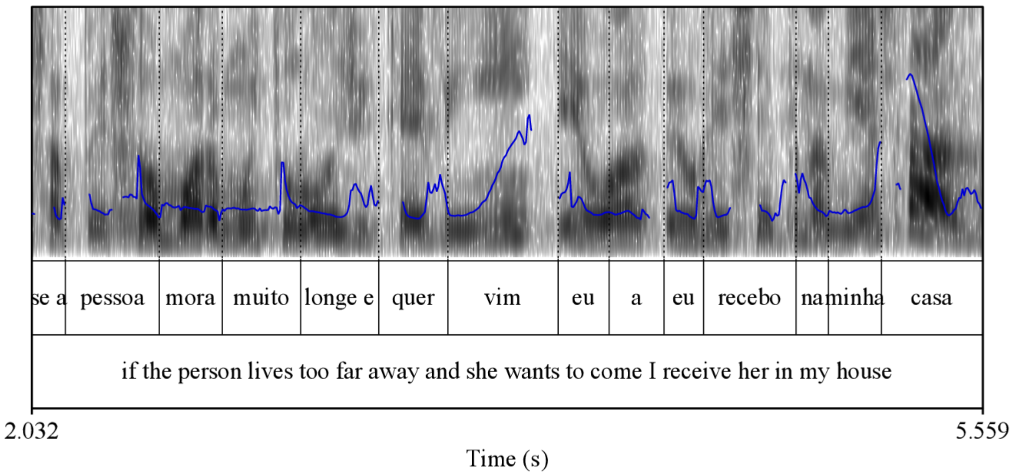
3.3.3 Chunks

Some prosodic variables can be measured from speech chunks of different lengths. 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. In fact, 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. However, there are some reasons for not measuring these variables from too small or too large intervals. If the measurement is extracted from too small an excerpt, data may not be enough to accurately represent the speech. On the other hand, a too large interval makes the measure more insensitive to variations in time. For instance, 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. To a large extent, the criterion depends on the objectives of each study. However, in all cases, it is important to ensure consistency between measures, establishing the range of the scope and using it for all measurements. In this research, I adopt three criteria to demarcate a chunk. The first is that the chunks coincide with complete utterances, i.e., utterances that a Brazilian Portuguese native listener can understand without the support of contextual information. Second, the chunks have to end in a prosodic boundary. Most of the chunks analyzed in this research end with a falling intonation contour followed by a silent pause, indicating that the speaker concluded the utterance, or a high intonation contour signaling a non-terminal boundary. The last criterion is that all chunks must be an interval ranging from 3 to 8 seconds. This range was defined to balance the sensitivity to variations in time (i.e., avoiding too long intervals) and the amount of prosodic information it comprises (i.e., avoiding too short intervals).

Figure 3.9 illustrates a speech chunk that meets the three criteria. It is a chunk by MarleneN from the ALCP sample. The blue contour over the spectrogram refers to the fundamental frequency. First, the chunk coincides with a sentence that native Brazilian Portuguese speakers can understand

Figure 3.9: Spectrogram and fundamental frequency contour (in blue) of a speech chunk



out of context. Second, the utterance clearly ends in a falling contour signaling a terminal boundary. Third, it is 3.5 seconds long, between 3 and 8 seconds. In fact, from the point of view of prosodic boundaries, this chunk could be divided in two, since the rising contour in the word *vim* is a non-terminal boundary, but, in this case, the two smaller chunks would not meet the 3-second minimum criterion and they do not sound like complete utterances for a native if isolated from each other.

3.4 Acoustic measures

The seventeen acoustic variables analyzed in this research were computed by taking measures of duration and fundamental frequency from the phonetic units discussed above. In this section, I discuss how these variables are defined from a mathematical point of view.

3.4.1 Speech rate

The definition of **speech rate** is quite straightforward: *how often a linguistic unit occurs per unit of time* (Fletcher, 2010; Kendall, 2013; Laver, 1994). The most common measurement unit for speech rate is syllables per second (syll/sec), but other units could be used as well, such as segments per second or words per milliseconds (for a review of the studies on speech rate, see Fletcher 2010, p. 568). Measuring speech rate involves counting the number of syllables in a speech chunk and dividing it by the total duration of the chunk in seconds (see Equation 3.3). In this study, instead of using the phonological syllable, the speech rate is measured in V-to-V units per second[3].

$$\text{speech rate} = \frac{\text{number of syllables}}{\text{total duration of chunk}} \tag{3.3}$$

[3] Another commonly used measure is the articulation rate, which differs from speech rate only in how silent pauses are treated. The only difference in articulation rate is that silent pauses are excluded from the calculation (for a more in-depth discussion on this distinction, see Laver 1994, p. 539).

3.4.2 Rhythm metrics

As mentioned in Section 2.3, rhythm metrics are mathematical formulas to calculate the degree of variation in the duration of vowels, consonants, and syllables. In this sense, they are not different from statistical measures of dispersion but applied to a specific phenomenon. In fact, one of the first proposed rhythm metrics is the standard deviation applied to the duration of vocalic and consonantal intervals. Despite being a well-known statistical measure, Ramus et al. (1999) decided to give a new name to the standard deviation when applied to vocalic and consonantal duration, namely, ΔV and ΔC respectively. As with any standard deviation, ΔV and ΔC summarize the average dispersion of the values from the mean. Using phonetic terms, it measures how much longer or shorter each vocalic (ΔV) and consonantal (ΔC) interval is in relation to the mean duration of these intervals in the speech chunk (see Equation 3.4). The higher the Δ, the more disperse the duration of the intervals around the mean.

$$\Delta = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n}} \tag{3.4}$$

Where,

d_i = Duration of syllable i

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

n = Number of syllables

In this research, I do not analyze Δ but a close-related metric, the **Variability Index** (Varco), proposed by Dellwo (2006) as a normalized alternative to Δ. The author explains that 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 duration of the intervals is probably more variable in slower speech rates than in faster ones. This interaction between speech rate and duration variability may be a problem because rhythm metrics are usually used to compare the rhythm between languages or varieties *independently* of speech rate. In order to minimize this speech rate influence, normalized alternatives of rhythm metrics have been suggested.

Most normalization procedures consist of converting a raw value into a proportion. In the case of Varco, the normalization is carried out by dividing the Δ by the mean duration of the interval in the speech chunk (see Equation 3.5). Thus, instead of saying that on average the duration of vocalic intervals varies 30 msec, as would be the case with ΔV, we say that the duration varies 70% of the mean vocalic duration. That is, Varco expresses the variability in duration as the proportion of Δ in relation to the mean. In doing so, the score will be more reduced in lower speech rates, in which the mean duration tends to be larger, than in faster ones. The multiplication by 100 serves only to make the numbers more manageable.

$$\text{Varco} = \frac{\Delta \cdot 100}{\bar{d}} \tag{3.5}$$

Where,

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

Fuchs (2016, p. 39) qualifies both Δ and Varco as global metrics because they are computed "taking into account how all the intervals in an utterance or a sentence differ from each other". Global metrics measure the average amount of deviation in the duration of intervals in relation to a unique and common parameter, such as the mean duration. In contrast, local metrics measure the average amount of deviation in pairs of adjacent intervals.

$$\text{PVI} = \frac{\sum_{i=1}^{n-1} |d_i - d_{i+1}|}{n-1} \tag{3.6}$$

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

The most used local metric in sociolinguistic research is the **Pairwise Variability Index** (PVI), proposed by Low et al. (2000). In order to compute this index, we first need to calculate the difference in duration within all pairs of adjacent intervals, and then to take the mean of these differences (see Equation 3.6).

The considerations about speech rate discussed above are also valid in the case of PVI. Lower speech rates tend to favor higher PVI and vice-versa. For this reason, Grabe and Low (2002) proposed the Normalized Pairwise Variability Index (nPVI). The only difference from the original PVI (usually referred to as raw PVI or rPVI) is that nPVI divides each difference within each pair of adjacent intervals by the mean duration of the respective pair (see Equation 3.7). As in Varco, the final value is multiplied by 100 to get more manageable numbers.

$$\text{nPVI} = 100 \times \frac{\sum_{i=1}^{n-1} \left| \frac{d_i - d_{i+1}}{(d_i + d_{i+1})/2} \right|}{n-1} \tag{3.7}$$

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

The idea behind PVI is that the contrast in duration is more salient for listeners when short intervals are immediately followed by long intervals and vice-versa. Thus, a rhythm metric should consider not only the variance in the duration of intervals but also how these intervals are ordered within the utterance. In particular, it should reflect the alternations of longer and shorter intervals along the speech (Low et al., 2000, p. 382).

It is important to mention that, in this study, I analyze only Varco and nPVI applied to vocalic (VarcoV, nPVI-V) and consonantal intervals (VarcoV, nPVI-C).

3.4.3 Measures of stress groups

I analyze three parameters involving stress groups. The first one is simply the **stress group duration** in milliseconds. The computation of this parameter consists of delimiting the stress groups' boundaries in Praat and measuring their acoustic duration.

I also investigate the **rate of syllabic duration peaks** (Barbosa, 2013). In Section 3.3.2, we saw that stress groups are delimited by salient peaks of syllable duration and that these salient peaks can be identified by means of the normalized and smoothed duration contour. Thus, the same procedure adopted for speech rate can be used to compute the rate of duration peaks, that is, by counting the number of peaks in the speech chunk and dividing it by the total duration of the chunk.

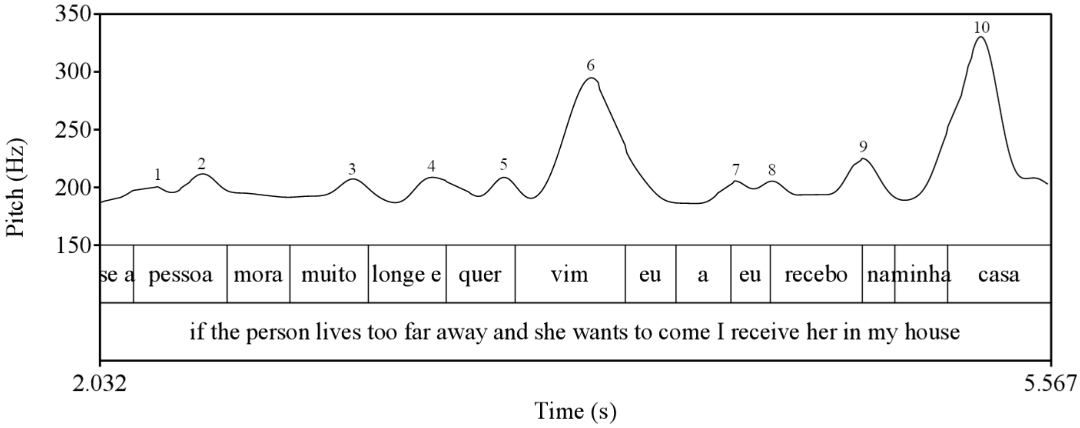
The last parameter involving stress groups is the **degree of correlation** between the duration of stress groups and the number of syllables in them. This parameter, which was proposed by Barbosa (2000), has the purpose of being a rhythmic index similar to Varco and PVI but with the advantage of being more transparent than these rhythm metrics.

In an ideal syllable-timed rhythm, syllables have exactly the same duration. Thus, each syllable that is added to a stress group causes an increase of constant magnitude in this stress group duration. On the other hand, in an ideal stress-timed rhythm, the stress groups have exactly the same duration *independently* of the number of syllables they comprehend. That is, adding or subtracting syllables from a stress group should not influence its duration, which is supposed to be constant. The central point here is that the relation between the duration of stress groups and the number of syllables clearly distinguishes the two rhythmic types, and this relation can be represented by a correlation coefficient like the Pearson's r. The closer the coefficient is to zero, which means no correlation, the more stress-timed the rhythm. In turn, the closer it is to one, meaning complete positive correlation, the more syllable-timed the rhythm. A more in-depth and contextualized discussion of this parameter is found in Section 4.1.1, where the results are reported.

3.4.4 Melodic peaks

I examine four parameters related to peaks of the fundamental frequency. Figure 3.10 shows the same speech chunk of Figure 3.9 but with a smoothed and interpolated[4] f_0 contour. There are ten peaks in this chunk. The sixth and tenth are significantly higher than the others because they act as a non-terminal and a terminal boundary respectively. Some peaks may have no perceptual relevance, as the first one, and others, like the fourth and the ninth, are probably prominence marks, drawing the listener's attention to specific words.

Figure 3.10: Smoothed and interpolated fundamental frequency contour illustrating melodic peaks.



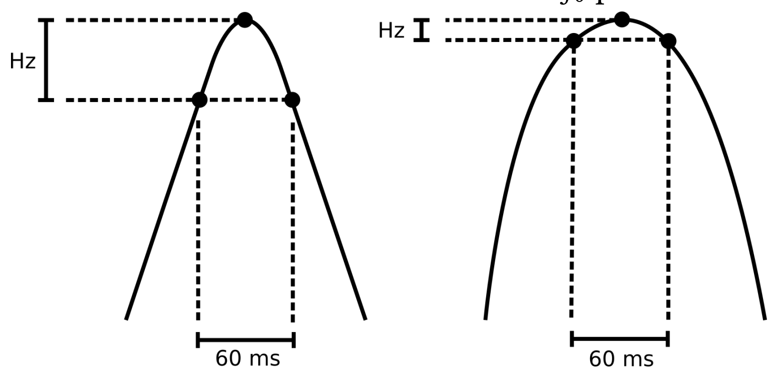
One variable I analyze is the **standard deviation of f_0 peaks**. In order to compute it, we measure the maximum f_0 of each peak and then calculate the standard deviation of these local maxima. This parameter tells us how variable is the height of f_0 peaks. In Figure 3.10, we see a large range of variation, from 200 Hz, as in the first peak, to almost 350 Hz, as in the tenth.

The second variable is the **rate of f_0 peaks**, which is computed using exactly the same procedure described above for speech rate and rate of duration peaks.

The third parameter is the **f_0 peak width**, which is a measure of "peakness". In order to compute it, we first need to identify the maximum f_0 of the peak (see Figure 3.11). Then, we measure the f_0 30 msec before

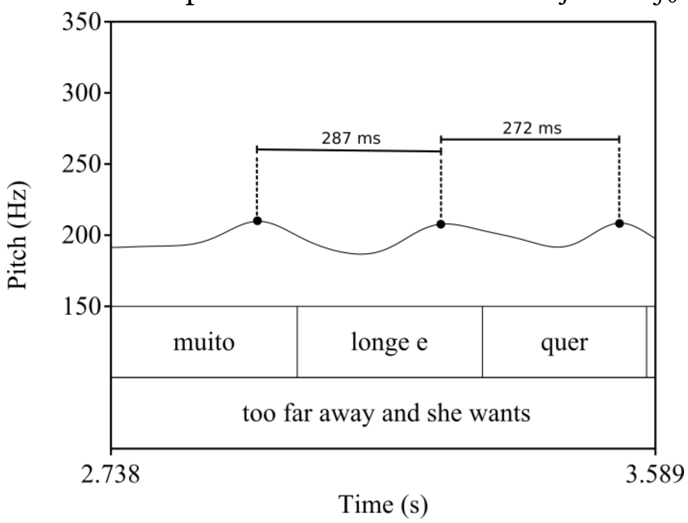
[4] In an f_0 contour, silent pauses and voiceless sounds have no f_0 values. Interpolation is the estimation of new points in order to fill the gaps in the graph of a function. It allows filling the gaps in the f_0 contour, making it easier for automatic methods to detect the peaks.

Figure 3.11: Illustration of how the width of f_0 peaks is measured



and after this maximum. Since most peaks are not symmetrical, these two values probably will not coincide, and thus the mean between these values is computed. Finally, we compute the difference between this mean and the f_0 maximum of the peak. The longer the vertical distance in the melodic contour, the sharper and more pointed the peak.

Figure 3.12: Example of intervals between adjacent f_0 peaks



Finally, the fourth parameter concerning f_0 peaks is the **standard deviation of the duration of inter-peaks intervals**. This parameter is useful to analyze how the f_0 peaks are distributed in time. An inter-peaks interval goes from one local f_0 maximum (i.e., the maximum of a peak) to the immediately following local f_0 maximum (see Figure 3.12). Therefore, this parameter involves measuring the duration of all these intervals in the speech chunk and then computing the standard deviation of the values. In a hypothetical speech chunk in which all f_0 peaks are evenly distributed in time, with a constant interval separating each other, the inter-peaks standard deviation would be zero.

3.4.5 Fundamental frequency change rates

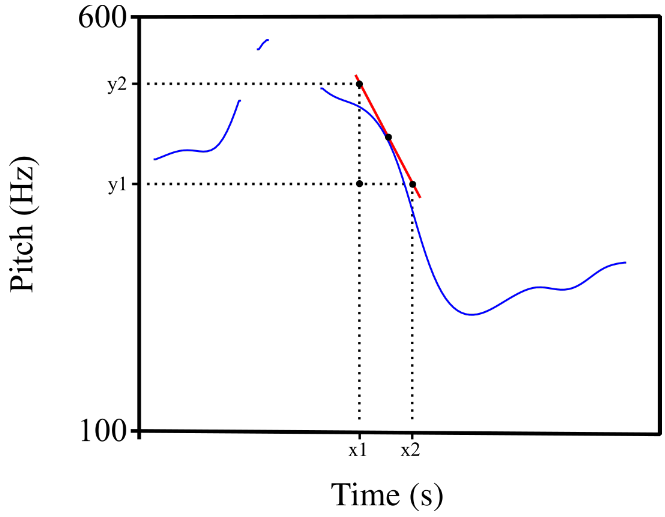
In this research, I also analyze the **fundamental frequency change rate**, a measure of how fast the f_0 contour rises and falls (Barbosa, 2013). The change rate at a point of a function's graph is the slope of the tangent line to that point. Figure 3.13 illustrates the slope of the tangent line, in red, to a point of the f_0 contour, in blue. Mathematically, the slope m of the tangent line is $m = \frac{y2-y1}{x2-x1}$.

The slope of only one point in time is not enough to analyze the change rate of a contour. The function that maps all or most points of the domain (i.e., the values in the x-axis) to the slope of y is called the first[5] derivative function. Thus, to compute f_0 change rates, we need to measure the first derivatives of the f_0 contour[6]. The first derivative indicates if the f_0, in the y-axis, is increasing (when its slope is a positive number) or decreasing (when it is a negative number) along the time in the x-axis. The magnitude of the derivative in absolute value indicates by how many Hertz the fundamental frequency changes per unit of time. In this research, I analyze the mean and the standard deviation of the positive (f_0 rises) and negative (f_0 falls) first derivatives of f_0. The measurement unit is Hertz per 50 milliseconds (Hz/50 msec).

[5] The derivative is qualified as the first because there can be higher-order derivatives. Velocity is the first derivative of object position with respect to time. The derivative of velocity is a second derivative since it is the derivative of a derivative — it is better known as acceleration. The derivative of acceleration is the third, and so on.

[6] For an excellent introduction to derivatives, see Lang (2002).

Figure 3.13: Fundamental frequency contour (in blue) and the tangent line (in red) to a point of this contour. The first derivative at the point is the slope of the tangent line.



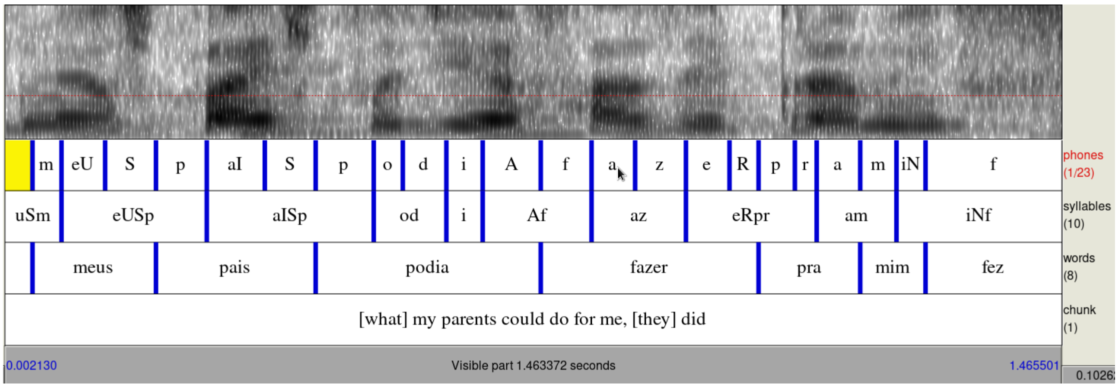
3.5 Automation

In this research, I study prosodic accommodation using acoustic techniques of speech analysis (Barbosa & Madureira, 2015; Boersma, 2013; Ladefoged, 1996), for which annotation is an essential step. Most relevant variables for phonetic and linguistic study depend on additional information to the audio signal; i.e., information involving abstract knowledge about language and speech. Since the speech recordings encode only data concerning oscillations in pressure, the researchers themselves need 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 & 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 and measuring specific linguistic units.

Time-aligned linguistic annotation of speech recordings involves computer programs like Praat (Boersma & Weenink, 2021) and ELAN (The Language Archive, 2022). Figures 3.14 and 3.15 are examples of speech annotation in Praat and ELAN respectively. In both programs, annotation consists of textual content arranged in multiple *tiers*. The researcher can use tiers to annotate different types of information. In the Praat example, the first tier is used for segmentation into phones, the second into syllables, and the third into words. In the ELAN example, the first tier, named D1, is used for transcription of the interviewer's speech and the second, named S1, of the participant's speech. Although both programs support the use of IPA symbols, in this research, I opt for a phonetic transcription system in ASCII characters to ensure compatibility between computer programs (Barbosa, 2006).

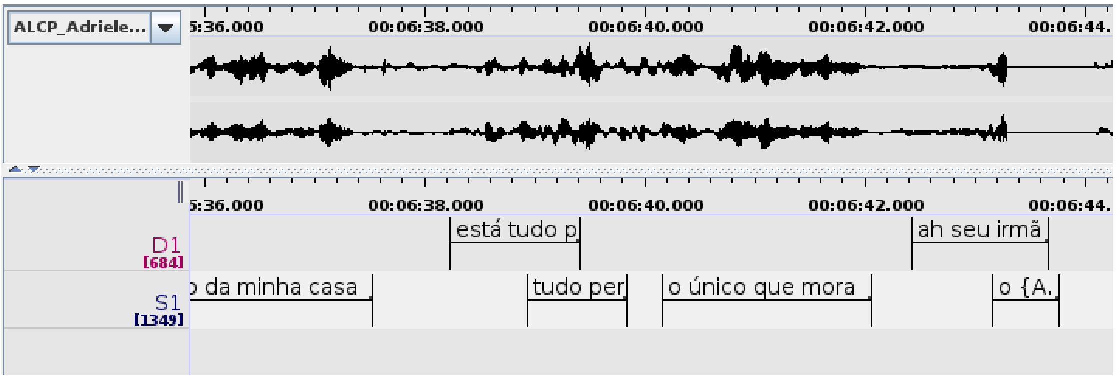
Figure 3.14: Example of segmentation and annotation of a speech chunk in Praat. From top to bottom, the first tier identifies individual phones, the second demarcates V-to-V units, while the third and fourth tiers demarcate the words and the chunk, respectively.



To extract prosodic data from the ALCP and CPS 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.

One methodological contribution of this research is the development

Figure 3.15: Example of segmentation and annotation of a speech chunk in ELAN. The first tier, named "D1", refers to the interviewer's speech, and the second tier, named "S1", concerns the interviewee's speech.



of the program Treina-PB[7], written in Python programming language, that automates phonetic segmentation of speech recordings. The tasks of selecting speech chunks from long speech recordings and manually segmenting them into vowels, consonants, and syllables in Praat are very time-consuming.[8]

When running Treina-PB on ALCP and CPS recordings, it first executes the automatic selection of speech chunks (see Figure 3.16), following criteria that were discussed in Section 3.3.3. The selection relies on time information contained in ALCP and CPS's transcription files in ELAN. The program filters out all the parts of the participant's speech that overlaps with the interviewer's because speech overlapping drastically affects the quality of the acoustic information of the audio. In particular, it is often not possible

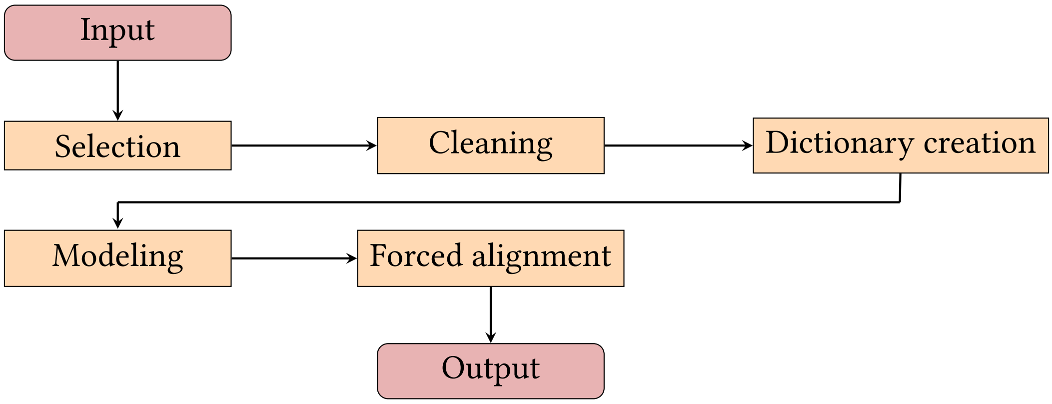
[7] TreinaPB's source code is found in the following Github repository: https://github.com/silveira7/treinaPB.

[8] The ALCP and CPS samples add up to 31 sociolinguistic interviews altogether, each with an average duration of over 45 minutes. They amount to 24 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 432,000 syllables to be manually annotated (24 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 this number is divided in half, there are still too many syllables to be manually annotated.

to completely discern which patterns in the spectrogram belong to each speaker.

After removing overlapped speech, it searches for stretches of the audio file that correspond to chunks of uninterrupted speech with a duration ranging from 3 to 8 seconds (see Section 3.3.3). Having identified the participant's eligible speech chunks, the program extracts the audio and the transcription of these chunks into new and separate files. Then, it cleans each chunk's transcription, removing special characters, such as parenthesis, and fixing possible spelling problems. This cleaning step is necessary for the execution of the automatic phonetic segmentation.

Figure 3.16: Flow diagram representing the steps executed by the Treina-PB program



The automatic phonetic segmentation of speech chunks is carried out using a technique of forced alignment based on Hidden Markov Models (HMM),[9] a type of statistical modeling largely used in learning machine systems. Nowadays, most systems of automatic speech recognition are based on HMM (Renals & King, 2012), including systems of forced alignment (Liberman, 2019).

Forced alignment consists of techniques that perform the segmentation of a speech recording into phones, syllables, and words, based on the orthographic transcription of the recordings. The idea behind the term is that, in this process of segmentation, the computer *aligns* speech segments,

[9] For a good introduction to HMM focused on applications in linguistic and speech research, see Jurafsky and Martin, 2008, p. 233; Renals and King, 2012.

inferred from the orthographic transcription, with the audio recording. In the case of Treina-PB, the forced alignment is implemented with the use of the Hidden Markov Model Toolkit (HTK) (Young et al., 2015), which is a library[10] written in C programming language specially designed to train HMMs for automatic speech recognition systems. In order to implement the HTK in Treina-PB, I largely relied on code from ProsodyLab-Aligner (Gorman et al., 2011).

For the program to know which are the phonemes in which the word needs to be segmented, a phonemic dictionary is necessary. This dictionary is a simple text file (TXT format) containing the orthographic transcription of all words that occur in the speech recordings followed by their phonemic transcription. The following is an example of three lines of the dictionary used to train the models for ALCP and CPS samples, using Barbosa's (2006) ASCII phonemic notation system for Brazilian Portuguese:

```
atento a t eN t U
atenuar a t e n u a R
atestado a t e S t a d U
```

The creation of the dictionary of all words in the ALCP and CPS sample was also automated. I used the algorithms of Alinha-PB,[11] which is a forced aligner for Brazilian Portuguese developed by Kruse and Barbosa (2021). As the name suggests, the project of Treina-PB was inspired and largely influenced by Alinha-PB. The purpose of Alinha-PB is to provide an accessible interface to phonetically align speech recordings into phones and syllables. Since the Alinha-PB already comes with a well-trained phonetic alignment model for Brazilian Portuguese, the users do not need to train their own models. Even though using an already trained model is more practical, if the user has a large spoken corpus, the quality of the alignment can be improved if the model is trained in this same corpus. For this reason, I started to develop Treina-PB, to be a complementary program to Alinha-PB, focused on training a new model on the user's data and performing the alignment with this model.

[10] In computer science, libraries are collections of programs intended to be used by other programs.

[11] Alinha-PB's source code can be found in the following Github repository: https://github.com/jkruse27/Alinha-PB

As shown in Figure 3.16, after the generation of a dictionary, Treina-PB executes the HTK library to train a speech recognition model using the audio and transcription files of the speech chunks selected from the ALCP and CPS samples. More precisely, two models were trained, one for each sample. Roughly speaking, training a model consists of using the model to repeatedly perform the intended task in a training dataset in such a way that the accuracy of each decision made by the model serves as feedback to improve it. In the case of Treina-PB, pairs of audio and transcription files are used to improve the ability of the model in detecting the boundaries of phones, syllables, and words.

As the output, for each speech chunk, Treina-PB returns an annotation file in TextGrid format, to be read in Praat, in which the boundaries of phones, V-to-V units, and words of the chunk are demarcated and annotated. Figure 3.14, discussed above, is an example of segmentation and annotation completely made by Treina-PB — except for the fourth tier at the bottom, which I manually added only to create the image.

To segment the chunks in stress groups, I used the Praat script *SGDetector*, developed by Barbosa (2013). This script uses exactly the same procedures discussed in Section 3.3.2 to obtain the normalized and smoothed duration contour of the speech recording. The detection of the boundaries of stress groups is based on finding salient duration peaks.

Finally, with all speech chunks already segmented into phones, syllables, and stress groups, the extraction of the acoustic measures was carried out with two other Praat scripts. To obtain measures of Varco and PVI, I used the *Speech Rhythm Extractor* script by Silva Jr and Barbosa (2019). In turn, all the other measures were computed with Barbosa's *Prosody Descriptor Extractor* script (2013).[12]

[12] Both scripts are open-source and freely available. All Barbosa's scripts can be found in https://github.com/pabarbosa/prosody-scripts. The *Speech Rhythm Extractor* is found in https://github.com/leonidasjr/SpeechRhythmCode.

Chapter 4

Results

In this chapter, I report quantitative analyses of seventeen prosodic variables (i.e., speech rate; stress group duration; duration peak rate; VarcoC; VarcoV; nPVI-C; nPVI-V; the correlation between stress group duration and the number of syllables in stress group; f_0 std. dev.; std. dev. of f_0 peaks; f_0 peak width; f_0 peak rate; std. dev. of intervals between f_0 peaks; and mean and std. dev. of positive and negative f_0 change rates) in the speech of Alagoan migrants in Campinas.

In Section 4.1, I analyze the differences between native Campineiros and Alagoan migrants in these prosodic parameters. Most linear regression models in this section have the sample (ALCP, CPS) as the only fixed-effect predictor variable and the speaker as random effect. The results of the comparison between ALCP and CPS have the primary purpose of being a reference to guide the analysis of the effects of migration on the speech of Alagoan migrants. Comparing these samples also contributes to the documentation of the prosody of Brazilian Portuguese's regional varieties. The prosody of most regions of Brazil, including that of the largest urban centers, such as São Paulo and Rio de Janeiro, is largely undocumented. In this sense, the prosodic comparison between Alagoan migrants and native Campineiros can yield valuable data for dialect documentation of the prosody of São Paulo and Alagoas.

In Section 4.2, I analyze the same parameters from the perspective of the variation within the group of migrants, examining the effects of dialect contact on their prosody. More specifically, Alagoan migrants are ordered according to their age of arrival in the state of São Paulo and the number of years they have lived in it, and correlations between these two social variables and the prosodic parameters are sought. The search for correlations is carried out with regression models with age of arrival and length of residence as fixed-effect predictors and the speaker as a random effect. Some models also explore the interaction between the age of arrival and gender.

Finally, in Section 4.3, I analyze prosodic variation between individual speakers. As will be reported in the following sections, for some prosodic variables, the variance between individual speakers in models in which the speaker is a random variable is significantly larger than the estimate of the fixed effects. In other words, individual differences can have a larger effect on some prosodic variables than age of arrival or length of residence. Thus, the purpose of this individual level of analysis is to explore how Alagoan speakers individually differ from each other in the light of their social profiles. The regression models in this analysis have the speaker as a fixed effect and no random variable is included.

Since the analyses conducted here involve many variables, they are reported in groups in order to better visualize and compare patterns. Most graphs display both the raw data and summary statistics (estimates and confidence intervals) from regression models.

As for the tables, it is especially important to notice that most of them report more than one linear regression model. The following rules apply throughout the chapter: (i) in a table reporting statistical models, each column corresponds to a single and different model — excepting the first column on the left, pertaining to the predictors; (ii) models reported in the same table share the same predictor variables (fixed or random) and differ from each other *only* in the response variable — the prosodic parameters. The tables bring the estimates along with their 95% confidence intervals between brackets [ ], the residual standard deviation, and the coefficients of determination (R^2). For mixed-effects models, the standard deviation of the random variable is also shown.

4.1 Prosodic differences between native and migrant speakers

4.1.1 Rhythm

Figure 4.1 shows the distribution of speech rate in syllables per second (syll/sec), the duration of stress groups in seconds (sec), and the rate of syllabic duration peaks in peaks per second (peak/sec), contrasting the CPS sample, on the left, in a lighter color, and the ALCP sample, on the right, in a darker color. The multiple dots spreading along the y-axis correspond to the raw data, each one being a single measure of the prosodic parameter[1]. The middle point of the error bar refers to the linear model estimate and the length of the line vertically crossing this point represents the 95% confidence interval. The right margin brings a density plot of the raw data to help visualize the shape of the distribution.

The three variables show considerable variability. The speech rate comprises values as low as 1.5 and as high as ten syll/sec. The duration of stress groups ranges approximately from 50 msec to more than 2 sec, and the duration peak rate ranges from low values, such as 0.3 peak/sec, to higher ones around 2 peak/sec.

Alagoan migrants are very similar to native Campineiros in the three prosodic parameters. The center of the distribution of the duration peak rate of Alagoan migrants is slightly higher than that of native Campineiros, but it is not statistically significant. The confidence intervals also overlap in the three variables, which suggests that the speech of Alagoan migrants and native Campineiros are statistically equal in these parameters.

Table 4.1 reports the mixed-effects linear regression models of the three variables. Interpreting the Intercept depends on if the predictor variable is numerical or categorical. In models involving a categorical predictor, the Intercept corresponds to the estimate of one of the levels of this predictor, arbitrarily chosen to serve as a reference. As shown at the bottom of

[1] Plotting all observations of a large data set, like the one analyzed here, demands a significant amount of computational resources. For this reason, each graph brings a smaller sample of the raw data, with 40 observations randomly selected for each speaker, adding up to a total of 1,240 plotted measurements (40 observations × 31 speakers).

Figure 4.1: Distribution of speech rate, duration peak rate, stress group duration, contrasting CPS (lighter color) and ALCP (darker color) samples. The error bars represent the 95% confidence interval of the respective angular coefficients from linear regression models.

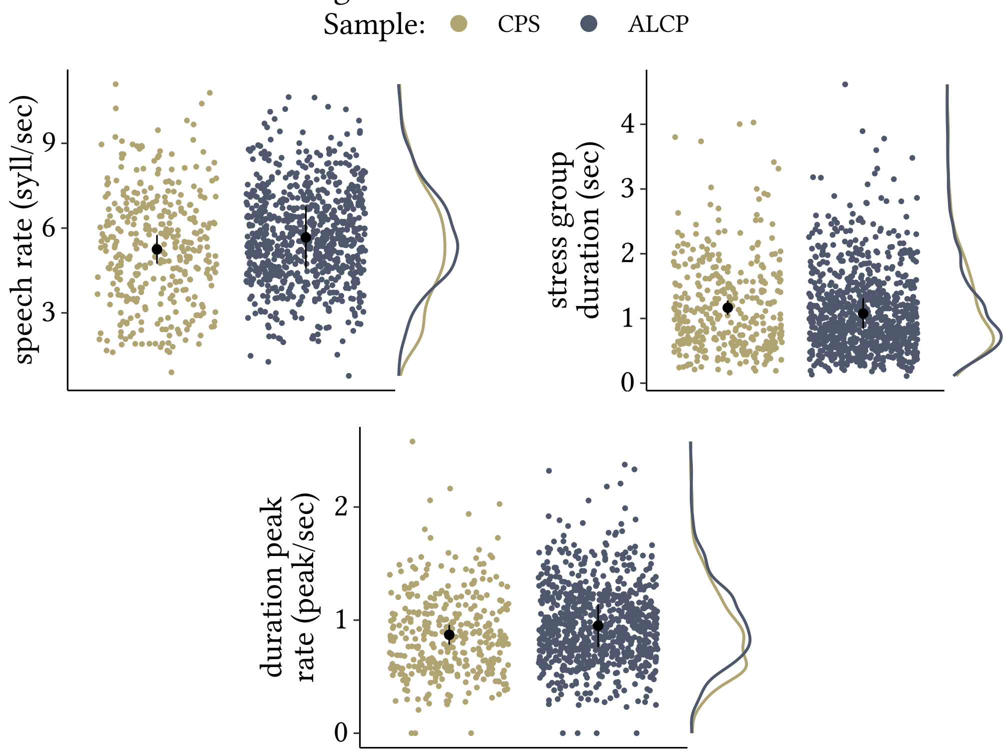


Table 4.1, the Intercept of these three models is the CPS sample. The two numbers in brackets below each estimate are the limits of the 95% confidence intervals. All the other estimates below the Intercept, and their respective confidence intervals, are *relative* values, referring to the difference of estimate from the Intercept. Values closer to zero indicate smaller differences and thus lack of correlation.

On average, native Campineiros speak at an estimated rate of 5.2 syll/sec. Alagoan migrants speak at an estimated rate of 0.4 syll/sec faster than native Campineiros (that is, 5.2 + 0.4 = 5.6 syll/sec), a difference that is not statistically significant.

The speech rate estimates in Table 4.1 agree with the results reported by Meireles and Barbosa (2009). The researchers analyzed the speech of four speakers from São Paulo and four from Minas Gerais, balanced for gender.

Table 4.1: Mixed-effects linear regression models of speech rate, stress group duration, and durational peak rate, contrasting the CPS and ALCP samples

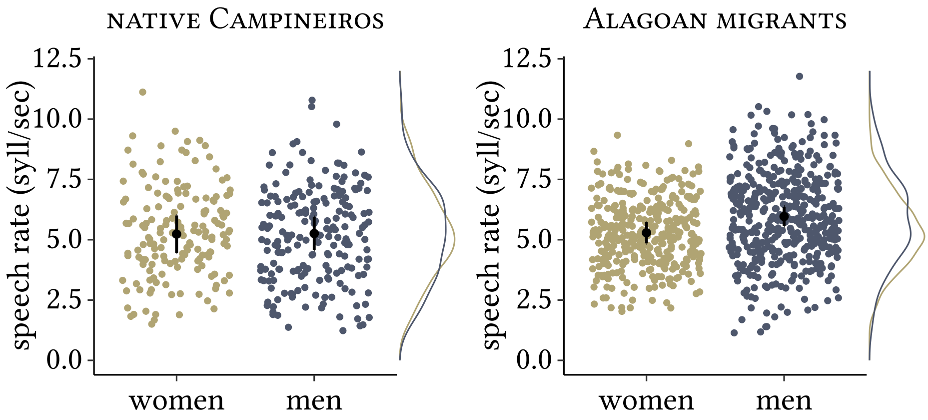
	Speech rate	Stress group duration	Duration peak rate
	(1)	(2)	(3)
Fixed effects			
(Intercept)	5.2	1.164	0.9
	[4.7, 5.7]	[1.058, 1.271]	[0.8, 1.0]
Sample: ALCP	+0.4	-0.088	+0.1
	[-0.2, 1.0]	[-0.215, 0.039]	[-0.02, 0.2]
Random effects			
Std. dev. (speaker)	0.72	0.159	0.13
Std. dev. (residual)	1.6	0.614	0.34
Marginal R^2	1%	0%	1%
Conditional R^2	18%	7%	13%
Observations	2,887	10,678	2,887

*: statistically significant ($\alpha = 0.05$)
Intercept: CPS (Sample)
R formula: `lme4::lmer(~ sample + (1|speaker))`

The participants were recorded in a controlled environment. Their results show that the participants speak at an average rate of 5.5 syll/sec, a value very similar to the one observed in ALCP and CPS samples. Meireles and Barbosa (2009) also found that female participants speak at a rate of 5.2 syll/sec, while males' speech rate is 5.8 syll/sec, a value 11% (0.6 syll/sec) faster than females.

A similar pattern of gender difference is found in the ALCP sample, but not in the CPS. Figure 4.2 shows the speech rates distribution of the CPS, on the left, and the ALCP, on the right, contrasting women and men. The density plots at the margin make it clear that there is no difference at all between males and females in the CPS sample. Native Campineiros of both genders show an average rate of 5.2 syll/sec. In its turn, the ALCP shows a pattern very similar to the one observed by Meireles and Barbosa (2009). Male Alagoans speak at an estimated rate of 6 syll/sec, a value that is 12%

Figure 4.2: Distribution of speech rate contrasting males and females in the CPS, on the left, and ALCP sample, on the right. The error bars represent the 95% confidence interval of the respective angular coefficients from linear regression models.



(0.7 syll/sec) faster than that of females (5.3 syll/sec). Table 4.2 describes the models represented in Figure 4.2. 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 for the ALCP sample, but not for CPS.

The p-value under the significance level should not be the only criterion to evaluate the importance of a predictor in explaining the variation in the response variable. Specifically for phonetic analyses, it is also necessary to consider that statistically significant differences alone do not mean that these differences are noticeable to hearers, not to mention that significant differences do not assure that the model fits the observed data well.

Statistically significant differences can be small enough for native hearers not to be able to perceive them. Ideally, a perceptual experiment should be carried out in order to assess if a difference in speech rate, such as the 0.7 syll/sec difference reported in Table 4.2, is noticeable by native Brazilian listeners. Quené (2007) examined the Just Noticeable Difference (JND) for speech rate variation in read speech in Dutch. JND is the term used to refer to the threshold of hearing for a particular phonetic variable. Thus, the

Table 4.2: Mixed-effects linear regression models of speech rate contrasting males and females in the CPS and ALCP samples

	Speech rate	
	CPS (4)	ALCP (5)
Fixed effects		
(Intercept)	5.2	5.3
	[4.6, 6.0]	[4.8, 5.7]
Gender: Men	0.0	0.7*
	[-1.0, 1.0]	[0.08, 1.3]
Random effects		
Std. dev. (speaker)	0.76	0.69
Std. dev. (residual)	1.7	1.6
Marginal R^2	0%	3%
Conditional R^2	16%	18%
Observations	882	1976

*: statistically significant ($\alpha = 0.05$)
Intercept: Women (Gender)
R formula: `lme4::lmer(~ gender + (1|speaker))`

hearer is supposed not to be sensitive to phonetic oscillations below this level (Fletcher, 2010). Quené argues that the JND of speech rate 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 0.7 syll/sec could be considered noticeable, insofar as it amounts to an increase of approximately 12% in relation to women's estimated speech rate.

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, all regression models reported here also bring coefficients of determination (R^2), statistics that indicate the proportion of the variance in the response variable explained by the model (Winter, 2019). In the case of mixed-effects models, two different R^2's are reported. Marginal R^2 represents the variance explained by fixed effects alone, and conditional R^2 indicates the variance predicted by both fixed and random effects (Winter, 2019). Speaker is the only random effect in this research. An R^2 of 100% would mean that the values expected by the model completely match the observed values; that is, the included predictors variables completely explain the variance in the response variable.

Although we observe a gender difference in speech rate that is statistically significant, none of the models discussed so far can be considered good predictors of their response variables. The best of them, Model 5 in Table 4.2, can barely predict 3% of speech rate variance when only fixed effects are considered. When the speaker is included as a random variable,

Figure 4.3: Distribution of raw data (colorful dots) of VarcoC and VarcoV contrasting CPS (lighter dots) and ALCP (darker dots) samples. The two error bars represent the 95% confidence interval of the respective angular coefficients from linear regression models.

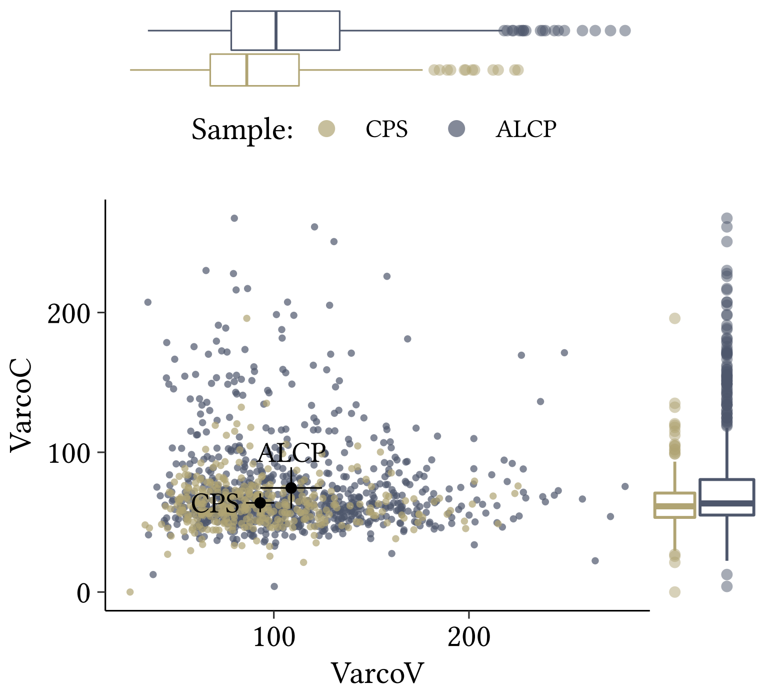


Table 4.3: Mixed-effects linear regression models of VarcoV, VarcoC, nPVI-V, and nPVI-C, contrasting the CPS and ALCP samples

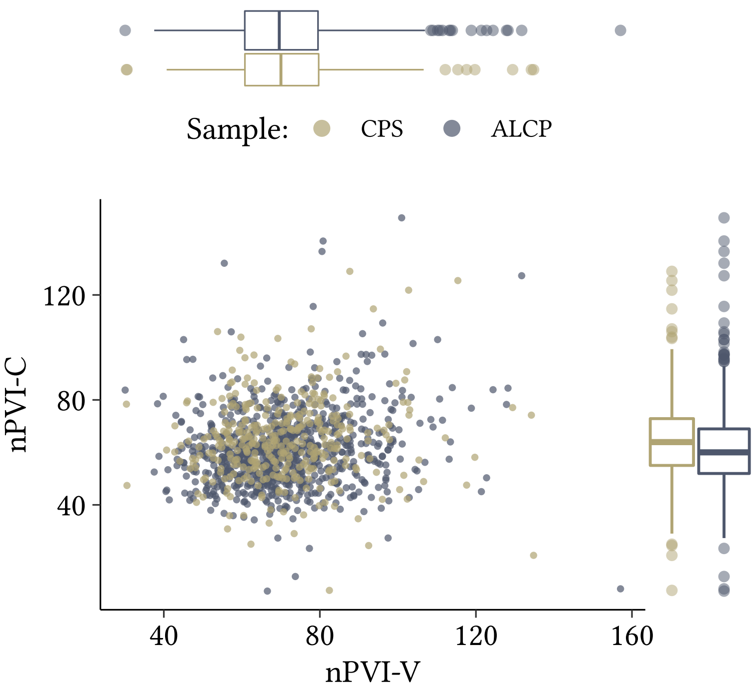
	VarcoV (6)	VarcoC (7)	nPVI-V (8)	nPVI-C (9)
Fixed effects				
(Intercept)	93.0	63.8	71.8	64.2
	[85.8, 100.2]	[57.1, 70.6]	[68.7, 74.9]	[61.1, 67.3]
Sample: ALCP	15.9*	10.7*	0.1	-2.8
	[7.3, 24.5]	[2.6, 18.7]	[-3.6, 3.8]	[-6.6, 0.9]
Random effects				
Std. dev. (speaker)	10.17	9.83	4.47	4.49
Std. dev. (residual)	38.33	29.38	15.22	14.19
Marginal R^2	3%	2%	0%	1%
Conditional R^2	10%	12%	8%	10%
Observations	2,888	2,888	2,889	2,889

*: statistically significant ($\alpha = 0.05$)
Intercept: CPS (Sample)
R formula: `lme4::lmer(~ sample + (1|speaker))`

the predictive power of this model increases to 18%. This marked difference between marginal and conditional R^2, also observed in the models of Table 4.1, suggests that differences between subjects are better predictors of speech rate than gender or sample.

In the literature of rhythm metrics (Arvaniti, 2012; Fuchs, 2016; Ramus et al., 1999; White & Mattys, 2007, among others), a conventional strategy to compare the rhythm of two or more languages is to situate these languages in a graph having a metric in the x-axis, usually concerning vowel duration, and another metric in the y-axis, usually related to consonantal duration. Following this convention, Figure 4.3 contrasts the speech rhythm of Alagoan migrants and native Campineiros in a graph having VarcoV, on the x-axis, and VarcoC, on the y-axis. The two black dots correspond to the estimate of these metrics in the models reported in Table 4.3, and the two perpendicular lines crossing each black dot refer to the 95%

Figure 4.4: Distribution of raw data (colorful dots) of nPVI-C and nPVI-V contrasting CPS (lighter dots) and ALCP (darker dots) samples.



confidence intervals. The cloud of colorful dots represents the raw data.

ALCP and CPS have very close estimated VarcoV and VarcoC, but ALCP has slightly higher values in both metrics. The duration of vocalic intervals varies in a range that is 108.9% of the mean vocalic duration in the ALCP sample and 98% in CPS (see Table 4.3). Consonantal intervals, in their turn, have a variation of 74.5% and 63.8% in Alagoans and Campineiros' speech respectively.

Even though Alagoan migrants show higher values that are statistically significant in both metrics, the distribution of the raw data (see Figure 4.3) reveals a very large dispersion, which makes the model estimates and summary statistics more imprecise. Most of the observations of both samples overlap in the same region of the graph. By observing the density plots at the margins, we can see that this overlap is especially true for VarcoC. Thus, the difference between CPS and ALCP seems to be in the outliers, with ALCP having a higher number of extreme values, and these outliers are probably responsible for an increase in the estimates of this group.

Cruz (2013), in her study of the rhythm of dialects of Portugal, reports much lower values of VarcoC and VarcoV than the ones observed in Table 4.3. In the Lisbon variety, the average score is 52 for VarcoV and 47.3 for VarcoC, values remarkably lower than the 108.9 and 98 observed here for the ALCP sample. At first, this difference can suggest that both Alagoan migrants and native Campineiros have a more stress-timed rhythm than Lisbon speakers. This suggestion contradicts the common association between vowel reduction and stress-timing (Abaurre-Gnerre, 1981; Fuchs, 2016; Ramus et al., 1999, among others), since it is a known fact that European Portuguese has more vowel reduction than Brazilian Portuguese (Frota & Vigário, 2000). By deduction, we could, thus, infer that European Portuguese should be more stress-timed, but if so, we would also expect the CPS and ALCP samples to have a Varco score smaller than that observed by Cruz (2013) for Lisbon speakers.

All these comparisons and conjunctures, however, cannot be taken as *prima facie* evidence because the data and the methods used in this study largely differ from the ones in Cruz (2013). One difference that is probably the main reason for these discrepancies is the type and the size of the samples used to compute the metrics. While Cruz (2013) computes the metrics from 5 sentences read twice by 6 speakers, a total of 60 utterances, the results reported here were computed from 2,888 utterances spoken in interviews by 31 speakers. It is expected that, in a much larger corpus of semi-spontaneous speech, a much larger dispersion will be observed in the rhythm metrics, and larger dispersions allow for the possibility of larger oscillations in the estimates.

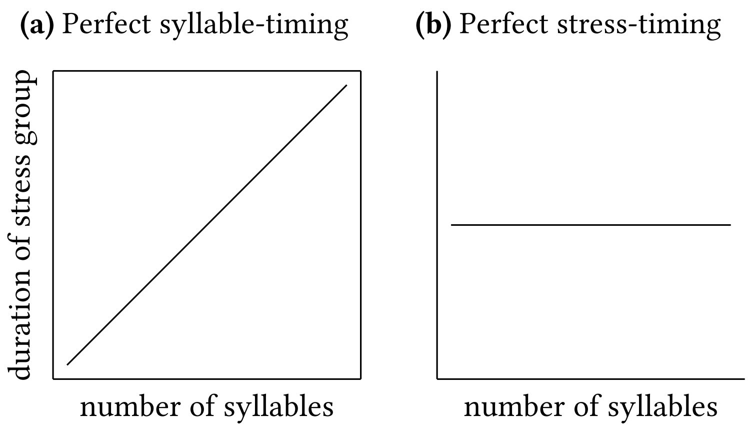
Figure 4.4 shows the same kind of graph, but now with the normalized Pairwise Variability Index applied to vocalic intervals (nPVI-V), on the x-axis, and consonantal intervals (nPVI-C), on the y-axis. Unlike Figure 4.3, this one does not bring the model estimates due to a graphical limitation: the difference between the estimates of CPS and ALCP is so small (see Table 4.3) that it is not possible to plot these estimates without completely overlapping the error bars.

As discussed in Section 3.4, Varco is a global measure of rhythm, since it computes the dispersion in duration considering all intervals of an utterance. In its turn, the Pairwise Variability Index (PVI) is a local measure because it computes the local variation within pairs of successive intervals. Instead of indicating how much consonantal and vocalic duration varies in general, PVI summarizes how much duration varies from one interval to the immediately next interval. The duration difference between a pair of intervals is normalized by dividing it by the mean duration of the intervals. Thus, like Varco, nPVI can also be taken as a percentage.

The graphical limitation mentioned above is itself an indication that the PVI scores of ALCP and CPS are very similar, and Table 4.3 just confirms this point. The ALCP score in nPVI-V is 71.9, only 0.1 higher than CPS. A slightly bigger difference is observed in nPVI-C, with ALCP having a score 2.8 smaller than CPS, a difference that can be graphically observed in the boxplots at the margins of Figure 4.4. Like Varco, PVI metrics also show great dispersion. nPVI-C has values as low as 25 and as high as 125, but the interquartile range (the 50% of in the middle) is between 52 and 70. Similarly, nPVI-V has values lower than 40 and higher than 150, but the interquartile range goes from 61 to 80.

One of the major deficiencies of rhythm metrics lies in their interpretation. The source of this problem is the fact that they are a very indirect and arguable way of measuring speech rhythm. In fact, calling these measures

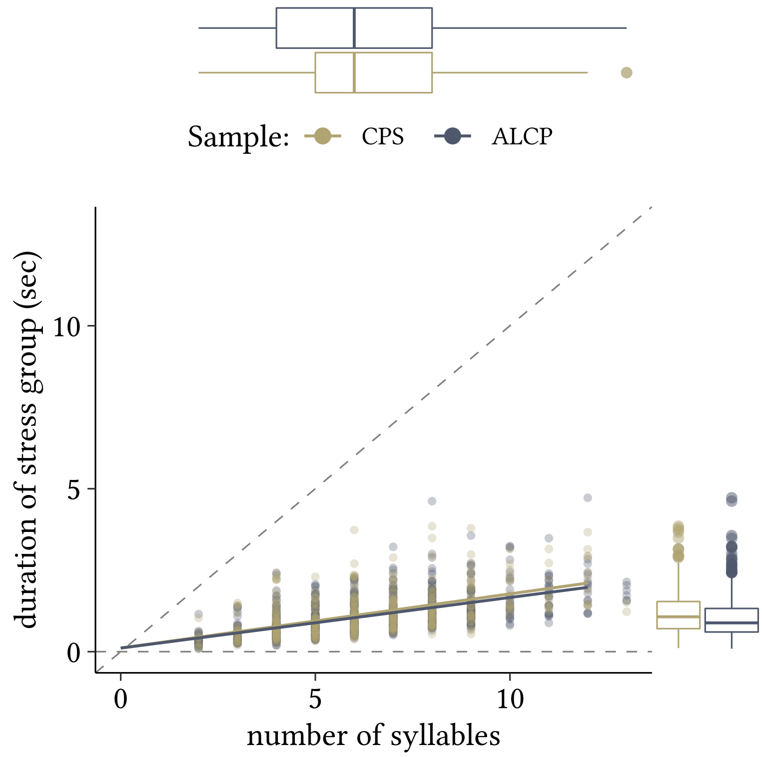
Figure 4.5: Patterns of syllable- and stress-timing in correlating stress group duration with the number of syllables in the stress group



rhythm metrics, as conventionally done in the rhythmic literature, is not precise because what they really measure is the variability in the acoustic duration of a phonetic interval, not rhythm. How do we get from summary statistics revealing the amount of dispersion in the duration of vowels, consonants, or syllables to the perception of rhythm?

Even if we rely on Abercrombie's (1967) traditional definition of stress- and syllable-timing, the connection between the metrics and this rhythmic typology is not straightforward. In a perfect syllable-timing pattern, syllables have the same duration, and, in perfect stress-timing, it is the stress groups that have the same duration. Even though metrics like Δ, Varco, and PVI have had some success in distinguishing some languages based on segmental duration dispersion, it is not clear how, for instance, the standard deviation of vocalic duration can be a reliable criterion to

Figure 4.6: Regression line of the correlation between the duration of stress group and the number of syllables in it, contrasting the CPS (lighter color) and ALCP (darker color) samples. The dots represent the raw data.



assess the regularity in syllable or stress groups duration.

As discussed in Section 3.4, Barbosa (2000) proposes a much more straightforward way of measuring stress- and syllable-timing of languages and varieties. If a hypothetical language has perfectly isochronous syllables, being prototypically syllable-timed, then the duration of a stress group will be directly proportional to the number of syllables in this stress group. On the other hand, if this language's stress groups have all the same duration, then that duration remains constant independently of the number of syllables in each group. Therefore, a straightforward way of assessing the rhythmic type of a language is by analyzing the correlation between stress group duration and the number of syllables in the stress

Table 4.4: Mixed-effects linear regression model of the duration of stress group

	Duration of stress group (10)
Fixed effects	
(Intercept)	111 [3, 221]
N° of syllables	166 [159, 173]
Sample: ALCP	-2 [-131, 127]
N° of syllables × ALCP	-11 [-19, -3]
Random effects	
Std. dev. (speaker)	152
Std. dev. (residual)	470
Marginal R^2	39%
Conditional R^2	45%
Observations	10,678

*: statistically significant ($\alpha = 0.05$)

R formula: `lme4::lmer(~ stress group duration + nº syllables * sample + (1|speaker))` group. A regression line can represent this correlation, with a diagonal one referring to perfect syllable-timing and a horizontal line representing perfect stress-timing (see Figure 4.5). This diagonal line must divide the graph into two equal parts. The less inclined the line, the more stress-timed the language.

Figure 4.6 shows an effect plot of a regression model having the stress group duration as the response variable and the number of syllables as the predictor variable (see Table 4.4). As a way of comparing the ALCP and CPS, the model also includes the interaction between the number of syllables and the sample. The two regression lines come from this interaction. For a direct comparison of the regression lines with the ideal forms illustrated in Figure 4.5, it is indispensable that the plot axes use the same scale. For this reason, in Figure 4.6, the unit used for stress group duration is second, instead of milliseconds. The dashed lines serve as a reference to assess the inclination of the regression lines.

At least three things are immediately evident from Figure 4.6. First, Alagoan migrants and native Campineiros have a very similar distribution, with their regression lines almost overlapping. One criterion to know if two variables interact in a model is to check if their regression lines are parallel or crossing each other (see Gries, 2013, chapter 5; Oushiro, 2007). It is clear that the difference between samples has no relevant effect on how the stress group duration correlates with the number of syllables. Second, we can clearly see that the increase in the number of syllables correlates with an increase in the stress group duration. More precisely, an increase of 166 msec in stress group duration is expected by adding one syllable to it (see Table 4.4). Considering that the effect of sample is irrelevant for the model, the number of syllables alone explains 39% of the variation in the stress group duration. However, although this marginal R^2 is very high for just one variable, it does not cover even half of the variation. So, the number of syllables is not the only factor that predicts stress group duration. Third, the regression lines are closer to the horizontal dashed line than to the diagonal one, which suggests that the participants have a speech rhythm with a more stress-timed tendency.

4.1.2 Intonation

The fundamental frequency is an acoustic parameter highly affected by anatomical differences between male and female speakers (Beck, 2010). Instead of trying to "neutralize" this influence of gender on the results, by transforming the data with some kind of normalization procedure (Adank et al., 2004), I include gender as an independent variable of the analysis. Thus, gender influence will not be removed, but it will be explicitly reported.

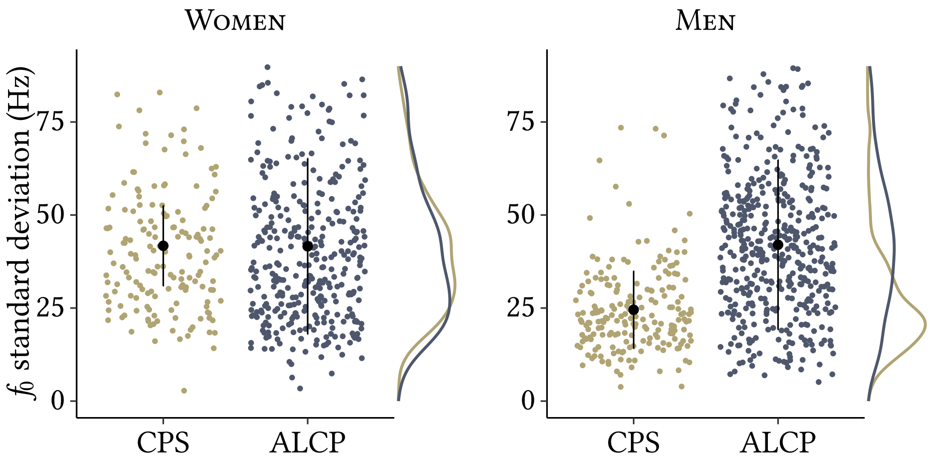
Including a predictor in the analysis in place of normalizing the response variable in relation to the predictor has the advantage of preserving the measurement unit. For instance, if we normalize the fundamental frequencies using one of the most used normalization procedures, i.e., converting the values to z-scores (Lobanov, 1971), the final data to be reported would not be in Hertz (i.e., cycles per second), but in units of standard deviation far from the mean. The problem is that discussing z-scores is not so straightforward as discussing the well-established "cycles per second."

To deal with gender influence, I first searched for interactions between gender (female, male) and the other predictor variables — in the case of this section, the sample (ALCP, CPS) is the other predictor. Since no significant interaction was found for standard deviation of f_0 and f_0 peaks, f_0 peak width, f_0 peak rate, and standard deviation of inter-peaks duration, I analyze these variables separating male and female speech in two different models. On the other hand, in the case of the mean and the standard deviation of positive and negative f_0 change rates, I analyze males and females together in the same regression model in which the interaction between gender and sample is included since this interaction is statistically significant.

Figure 4.7 shows the distribution of the f_0 standard deviation in the speech of women, on the left, and men, on the right, contrasting CPS, in a brighter color, and ALCP, in a darker color. Both female and male Alagoan migrants have a larger dispersion than that of Campineiros, but this difference is much greater in the case of males[2]. Female Alagoans have a confidence interval 34% wider than female Campineiras. The difference is even greater among men, with male Alagoans having a confidence interval 54% larger than male Campineiros (see Table 4.5).

This larger dispersion in the ALCP data may be related to a factor that has been observed by other dialect contact studies (Dodsworth, 2017; Oushiro, 2020c; Siegel, 2010): greater inter-speaker variability, a topic that will be further explored in Section 4.3. In fact, in most variables discussed in this section, migrants exhibit more dispersion than natives.

Figure 4.7: Distributions of the standard deviation of f_0 contrasting males and females in the CPS and ALCP samples



The difference between native and migrant speakers is more prominent in male speech. This difference can be better visualized in the density plots at the margins of Figure 4.7. Even though female Alagoans differ from Campineiras in the dispersion, the shape of their distribution is very similar, with a large area of overlap. The same does not occur with males. The shape of the distribution of male Alagoans is significantly flatter than

[2] To accurately interpret the data, we need to remember that when we refer to the dispersion of the f_0 standard deviation, we are discussing the dispersion of a dispersion measure (i.e., a second-order dispersion). Each observation in Figure 4.7 (i.e., each dot) is a measure of how variable the f_0 is in a specific utterance of the corpus. Thus, when we refer to the dispersion of this distribution, we mean how variable is the variation in f_0 from utterance to utterance.

that of male Campineiros. This flatness just reinforces that the average f_0 variation within an utterance changes more from utterance to utterance in Alagoans' speech.

Table 4.5: Mixed-effects linear regression models of the standard deviation of the fundamental frequency contrasting males and females in the CPS and ALCP samples

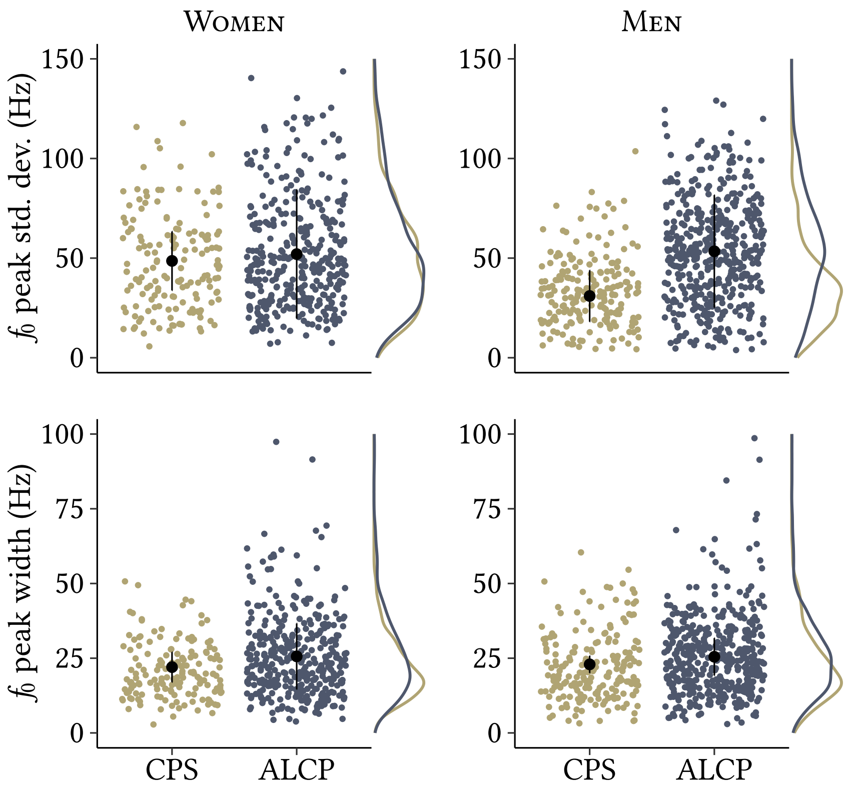
	f_0 standard deviation	
	Women (11)	Men (12)
Fixed effects		
(Intercept)	41.7	24.5
	[30.8, 52.6]	[14.1, 35.0]
Sample: ALCP	-0.1	17.5*
	[-13.1, 12.8]	[5.0, 29.9]
Random effects		
Std. dev. (speaker)	10.93	11.86
Std. dev. (residual)	16.33	13.27
Marginal R^2	0%	16%
Conditional R^2	31%	52%
Observations	1149	1739

*: statistically significant ($\alpha = 0.05$)
Intercept: CPS (Sample)
R formula: `lme4::lmer(~ sample + (1|speaker))`

Table 4.5 shows that males and females also differ in the regression estimates. While the estimated difference between female natives and female migrants is only 0.1 Hz, a value too small to be perceptually noticeable, in the male speech the difference is 17.5 Hz. The genders also differ in the explanatory power of their models, as can be seen in the marginal R^2 — which indicates the proportion of the variance in the response variable explained by the fixed effects alone. In the male speech, the difference between CPS and ALCP samples predicts 16% of the variance in f_0 standard deviation, against 0% in female speech. In both models, however, the speaker as a random effect contributes to the explanation of more than 30% of the data.

Figures 4.8 and 4.9 show the distribution of the acoustic variables related to peaks of the fundamental frequency. In the four parameters, Alagoan migrants and native Campineiros have remarkably similar distributions, which overlap almost completely. In other words, we are not able to discriminate the two groups based on the characteristics of f_0 peaks alone. One exception is the standard deviation of f_0 peaks (the two topmost plots in Figure 4.8), with male Alagoan migrants having a more dispersed distribution whose central tendency is located at a higher region of the scale compared to male Campineiros. This means that there is greater variation in the way of signalling melodic prominences in male Alagoans's speech. The shapes of the distributions of the standard deviation of f_0 peaks are very similar to the ones observed in Figure 4.7 for f_0 standard

Figure 4.8: Distributions of the standard deviation of f_0 peaks (at the top row) and f_0 peak width (at the bottom row) contrasting males and females in the CPS and ALCP samples



deviation, which is expected since the standard deviation of f_0 peaks can be seen as a subset of the f_0 standard deviation.

As shown in Table 4.6, the height of f_0 peaks in male Alagoans (Model 17) varies on average 22.4 Hz more than male Campineiros. This is one of the best models concerning f_0, with a marginal R^2 of 14%. The other models do not go beyond the 2% of variance explained by fixed effects. The peak width of migrants and natives are similar in both genders, with migrants having values slightly higher (i.e., slightly sharper f_0 peaks). The same applies to f_0 peak rate and standard deviation of inter-peaks duration. On average, the participants produce 2.5 f_0 peaks per second, and the intervals separating adjacent peaks have a duration that varies on average 200 msec.

The second group of intonation parameters concerns the derivatives

Figure 4.9: Distributions of f_0 peak rate (at the top row) and the standard deviation of intervals between adjacent f_0 peaks (at the bottom row), contrasting females (on the left) and males (on the right) in the CPS (lighter color) and ALCP (darker color) samples

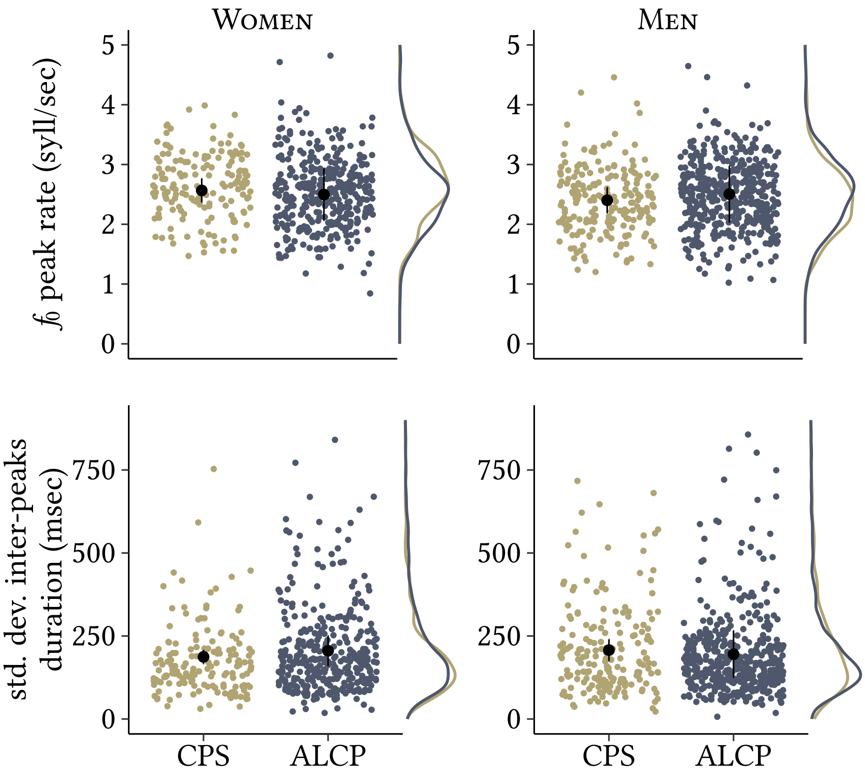


Table 4.6: Mixed-effects linear regression models of the standard deviation of $_0$ peaks, f_0 peak width, f_0 peak rate, and standard deviation of the intervals between adjacent f_0 peaks, comparing the CPS and ALCP samples in male and female speech

	Peaks of fundamental frequency			
	Std. dev.	Width	Rate	Std. dev. intervals
	Women			
Fixed effects	(13)	(14)	(15)	(16)
(Intercept)	48.6	22.0	2.6	0.2
	[33.7, 63.5]	[17.0, 27.1]	[2.4, 2.9]	[0.2, 0.2]
Sample: ALCP	3.4	3.6	-0.1	0.02
	[-14.3, 21.1]	[-2.4, 9.6]	[-0.4, 0.1]	[-0.01, 0.04]
Random effects				
Std. dev. (speaker)	14.97	4.95	0.2	0.02
Std. dev. (residual)	21.49	12.29	0.79	0.13
Marginal R^2	0%	02%	1%	0%
Conditional R^2	33%	15%	6%	2%
Observations	1147	1148	1147	1137
	Men			
Fixed effects	(17)	(18)	(19)	(20)
(Intercept)	31.0	22.9	2.4	0.2
	[18.0, 43.9]	[20.1, 25.8]	[2.2, 2.6]	[0.2, 0.2]
Sample: ALCP	22.4*	2.5	0.1	-0.01
	[7.0, 37.9]	[-0.9, 6.0]	[-0.2, 0.3]	[-0.1, 0.03]
Random effects				
Std. dev. (speaker)	14.63	2.98	0.22	0.03
Std. dev. (residual)	19.38	12.63	0.78	0.13
Marginal R^2	14%	1%	0%	0%
Conditional R^2	46%	6%	8%	7%
Observations	1731	1738	1731	1717

*: statistically significant ($\alpha = 0.05$) Intercept: CPS (Sample)

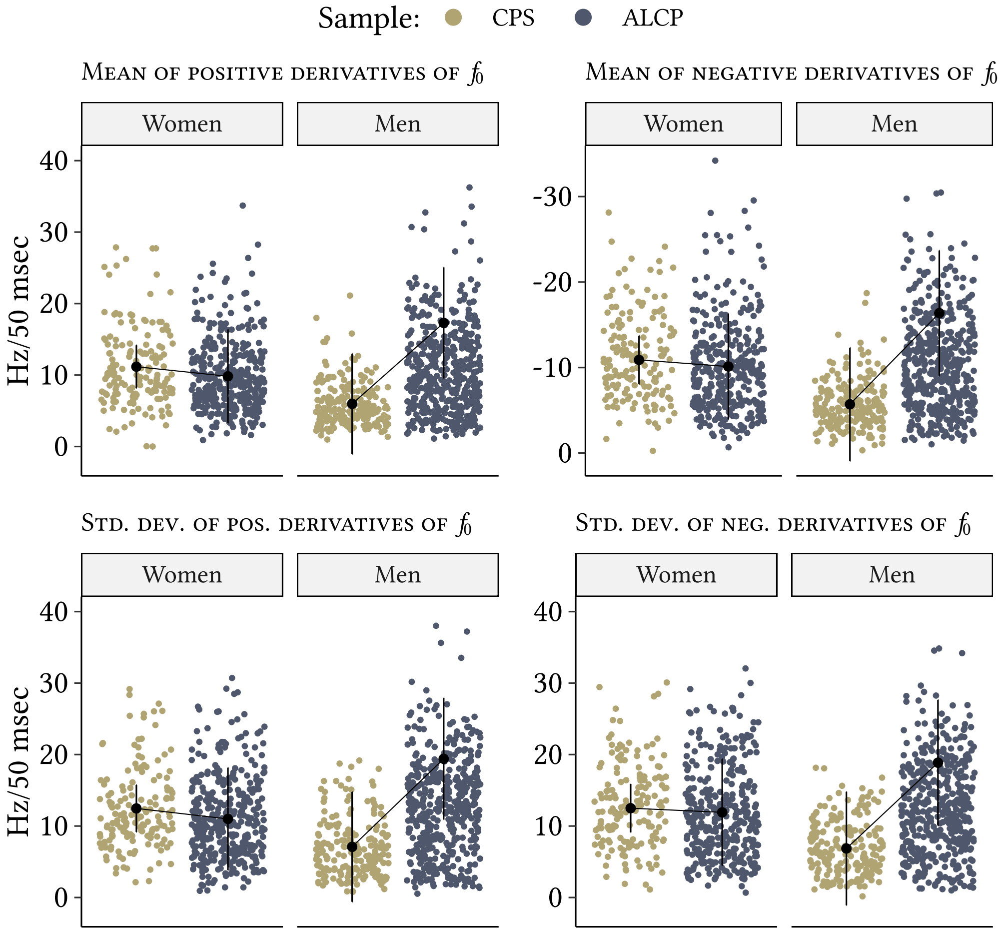
R formula: `lme4::lmer(~\ sample + (1|speaker))` of the fundamental frequency. Briefly recapitulating the discussion of Section 3.4, the first derivative of a function is a measure of how fast the values in the y-axis change according to the values in the x-axis. Usually, the x-axis represents time and the derivative is the change rate. In the case of melody, the first derivative indicates how fast f_0 values change in time. Graphically, it corresponds to the inclination of the melodic contour. When the melodic contour is rising, the first derivative is a positive number. When the contour is falling, the derivative is negative. It is important to have in mind that the amount of change rate is informed by the magnitude of the number only. For instance, in both rates of 15 Hz/50 msec and -15 Hz/50 msec, the melodic contour changes at the same "speed", but one is a rising contour and the other is a falling contour.

In the case of the derivative parameters, male and female speakers are analyzed in the same model, in which the interaction between gender and sample is considered. Figure 4.10 shows the distribution of the mean and standard deviation of both negative and positive f_0 derivatives.

Of the intonation parameters analyzed in this study, f_0 change rate is the one that most differentiates Alagoan migrants from native Campineiros. As mentioned in the previous section, one way of graphically checking if two variables interact is to verify if their regression lines are parallel or crossing each other. Because the predictors of the models in Table 4.7 are not numerical, but categorical variables, it is not possible to plot a proper regression line, like the ones correlating the number of syllables and duration of stress groups in Figure 4.6 of the previous section. However, the same correlation test can be carried out with categorical variables, using lines connecting the estimates of the model, as shown in Figure 4.10.

There is a remarkable difference between males and females in the distribution of f_0 change rates. The graphs clearly show that the degree and the direction of the difference in f_0 change rates between ALCP and CPS depend on the speakers' gender. In the four parameters, the line connecting male speakers' estimates has a steep positive inclination, showing a contrast in which male Alagoans have a significantly faster and more dispersed change rate. A different pattern is observed when female speakers are considered. The line connecting their estimates is almost horizontal, with a slight negative inclination. Therefore, female Alagoans seem to

Figure 4.10: Distribution of the mean and the standard deviation of the positive and negative f_0 change rates contrasting females and males in the CPS (lighter color) and ALCP (darker color) samples



have slightly smaller and less dispersed change rates compared to female Campineiras, but the differences are too small to be considered statistically significant.

Comparing the estimates of the four models in Table 4.7, it is clear that the positive parameters are very similar to their negative counterparts. This finding is relevant since it is not a necessity that f_0 rises have change rates similar to f_0 falls. For instance, in the production of an f_0 peak, speakers can start with a very steep f_0 rising and end with a much flatter falling. In this case, the rising part of the peak would have a larger derivative than the falling part in absolute value. The data reported here, however, suggest this "unbalancing" between rising and falling contours is not predominant in both CPS and ALCP.

Models 21 and 22 show that, on average, the melodic contour of female Campineiras changes 11 Hz every 50 msec, and, according to Models 23 and 24, this change rate has an estimated variance of 12.5 Hz/50 msec. In comparison, in absolute number, female Alagoan migrants have values slightly smaller, a difference around the magnitude of 1. Evidently, the p-value of none of these differences is under the alpha level of .05. Thus,

Table 4.7: Mixed-effects linear regression models of the mean and standard deviation of the positive and negative f_0 change rates

	Fundamental frequency 1st derivatives			
	Mean positive (21)	Mean negative (22)	Std. dev. positive (23)	Std. dev. negative (24)
Fixed effects				
(Intercept)	11.2	-10.9	12.5	12.5
	[8.2, 14.2]	[-13.7, -8.1]	[9.2, 15.7]	[9.1, 15.9]
Sample: ALCP	-1.3	0.8	-1.5	-0.6
	[-4.9, 2.2]	[-2.6, 4.1]	[-5.4, 2.4]	[-4.6, 3.4]
Gender: Man	-5.2*	5.2*	-5.4*	-5.6*
	[-9.2, -1.2]	[1.4, 9.0]	[-9.8, -1.0]	[-10.2, -1.1]
ALCP × Man	6.1*	-5.5*	6.9*	6.4*
	[1.4, 10.9]	[-10.0, -1.0]	[1.7, 12.2]	[1.0, 11.8]
Random effects				
Std. Dev. (speaker)	3.14	2.96	3.43	3.55
Std. Dev. (residual)	4.19	4	4.94	4.82
Marginal R^2	10%	11%	10%	11%
Conditional R^2	42%	43%	39%	42%
Observations	2,888	2,888	2,885	2,888

*: statistically significant ($\alpha = 0.05$)

Intercept: CPS (Sample) + Women (Gender)

R formula: `lme4::lmer(~\ sample * gender + (1|speaker))` we can conclude that female Alagoans do not differ from Campineiras in f_0 change rate parameters.

A very different situation is observed when we observe the estimates for male participants. The models in Table 4.7 show that, besides the interaction between sample and gender, the gender variable itself is statistically significant. In other words, the models suggest that there is a gender difference in f_0 change rate that also applies to the CPS sample. The intonational contour of male Campineiros changes at a rate 5.2 Hz/50 msec slower than female Campineiras, and 5.5 Hz/50 msec more disperse. In the opposite direction, male Alagoans exhibit the highest values of change rate, in absolute number. Rounding the estimates, we can say that in the four parameters, male Alagoans differ from female Campineiras in 6 Hz/50 msec. In relation to male Campineiros, this difference is even larger, ranging from 10 to 12 Hz/ 50 msec.

4.1.3 Summary

Concerning speech rhythm, Alagoan migrants and native Campineiros do not differ in speech rate, stress group duration, and rate of duration peaks. The speech rate of male migrants is significantly faster (+0.7 syll/sec) than that of female migrants, a gender difference that is in agreement with other studies on speech rate (Meireles & Barbosa, 2009). However, this difference is not observed among native Campineiros. Alagoan migrants show higher values of both VarcoV and VarcoC compared to native Campineiros, which suggests that the migrants' speech rhythm is more stress-timed. However, this rhythmic difference is not supported by another rhythmic index, i.e., the degree of correlation between stress group duration and the number of syllables in stress groups. In both samples, virtually the same degree of correlation is observed. Alagoan migrants and native Campineiros also do not differ in both nPVI-V and nPVI-C.

In relation to intonation, no significant difference is observed between female Alagoan migrants and female native Campineiras. On the other hand, male Alagoan migrants show a significantly more disperse f_0 contour in comparison to male native Campineiros. The standard deviation of f_0 and f_0 peaks is higher among male Alagoans. Furthermore, male Alagoans' f_0

has also faster and more variable change rates than Campineiros. Alagoan migrants do not differ from native Campineiros in f_0 peak width, peak rate, and standard deviation of inter-peaks duration.

4.2 Effects of migration on speech prosody

In this section, I analyze the same prosodic variables but with a focus on the effects of age of arrival and length of residence in the Alagoan migrants' speech. Since these two variables are specific to the ALCP sample, most models (and their respective plots) reported in this section refer only to ALCP data. The only exceptions are four models concerning f_0 change rates discussed at the end of Section 4.2.2, which also include the CPS data for reference. Therefore, in this section, except when explicitly mentioned, the reader can assume that the model represents only the Alagoan migrants.

Throughout the section, age of arrival and length of residence are analyzed as numeric variables in a positive integer scale whose effects on prosodic parameters, which are also numeric but in a real scale, are investigated. Different from the previous section, in which the predictors were categorical, the numerical nature of the variables examined here favors the use of the classical regression line as a graphical representation of the relationship between the prosodic and the social parameters.

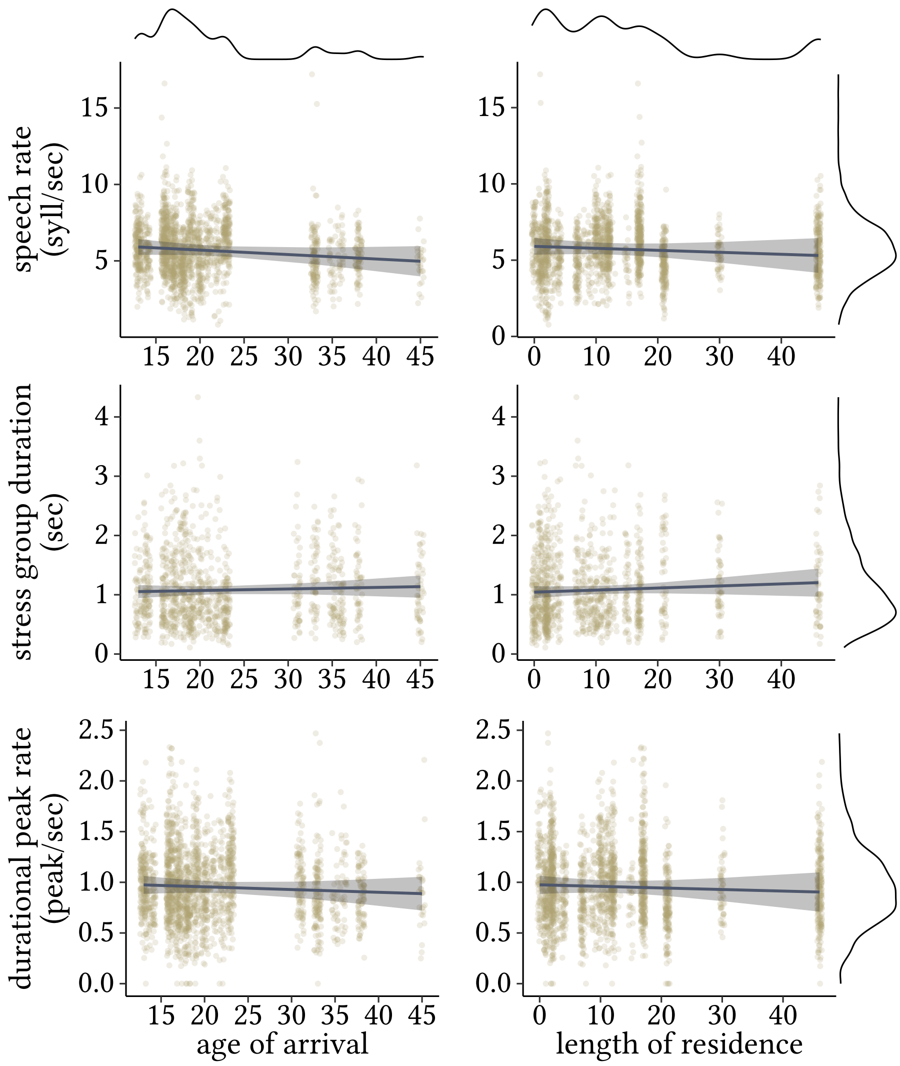
4.2.1 Rhythm

Figure 4.11 shows the distribution of speech rate, stress group duration, and duration peak rate, concerning speakers' age of arrival, on the left, and length of residence, on the right. The straight line refers to the estimate of each regression model (see Table 4.8) and the shaded area around the line corresponds to the 95% confidence intervals. The raw data is represented in the multiple dots. Similar to the plots discussed in the previous section, the shape of the distributions is illustrated in the density plots at the margins.

When analyzing correlations, the steeper the inclination of the line, determined by the angular coefficient of the linear model, the stronger the correlation. The absence of correlation has the form of a horizontal line, and the graphs in Figure 4.11 clearly show that the six regression lines are almost horizontal. In Table 4.8, the angular coefficient is represented by estimates of age of arrival and length of residence, and, in the three models, these estimates are virtually zero. In other words, varying the age of arrival and the length of residence does not have any effect on these duration parameters.

In Section 4.1.1, we discussed the correlation between gender and speech

Figure 4.11: Distribution of speech rate, stress group duration, and durational peak rate in function of the age of arrival and length of residence in the speech of Alagoan migrants (ALCP sample)



rate, showing that, in accordance with other sociophonetic studies (Meireles & Barbosa, 2009), a significant difference between male and female speakers was observed in the ALCP sample. Figure 4.12 shows that, when only Alagoan migrants are considered, significant differences between males and females are also observed in the stress group duration and the rate of duration peaks. Female Alagoan migrants have stress groups that on average last 134 msec more than male Alagoans. On the other hand, the duration peak rate is 14% faster (a difference of 0.12 peak/sec) in male speech. These two differences, however, are probably a consequence of the gender difference in speech rate. Male Alagoans have a faster speech rate, and it is expected that at faster rates the duration of the stress group will

Table 4.8: Mixed-effects linear regression models of speech rate, stress group duration, and durational peak rate, contrasting age of arrival and length of residence in the ALCP sample

	Speech rate	Stress group duration	Durational peak rate
	(25)	(26)	(27)
Fixed effects			
(Intercept)	5.5	968	0.9
	[5.0, 5.9]	[768, 1169]	[0.8, 1.0]
Age of arrival	0.003	4	0.002
	[−0.02, 0.03]	[−3.13, 11.18]	[−0.002, 0.01]
Length of residence	0.003	2	0.001
	[−0.03, 0.03]	[−3.94, 7.22]	[−0.004, 0.01]
Random effects			
Std. Dev. (speaker)	0.79	137	0.13
Std. Dev. (residual)	1.63	598.9	0.34
Marginal R^2	0%	0%	0%
Conditional R^2	19%	5%	13%
Observations	2,758	10,678	2,887

*: statistically significant ($\alpha = 0.05$)

R formula: `lme4::lmer(age_arrival + length_residence + (1|speaker)` be shorter and the rate of duration peaks will be faster.

Figures 4.13 and 4.15 represent rhythm metrics using the same strategy adopted in Section 4.1.1, i.e., crossing two metrics in the same coordinate system. Here we aim to analyze if there are significant differences in the rhythm metrics according to the speakers' age of arrival and length of residence. Visually representing three or more numerical variables in the same graph without making it too complex and difficult to understand is a challenging task, and, in our case, we are dealing with four different variables. In these two figures, I represent age of arrival and length of residence in a more categorical way. In both figures, the graph on the right shows the raw data in full scale. Each point in light gray corresponds to a single measure. Additionally, the graph brings the median of each metric for each age of migration, in green, and length of residence, in purple. This means that these figures do not show estimates of regression models, but only descriptive statistics. To facilitate the identification of each median point with the respective age of arrival or length of residence, the graph on the left is a zoomed-in area of the full-scale panel on the right. In the zoomed-in panel, the points are replaced by numbers that inform the respective age of arrival or length of residence.

If there were a clear correlation between the migration variables and Varco metrics, the values of age of arrival and length of residence would

Figure 4.12: Distribution of stress group duration and durational peak rate contrasting females and males in the ALCP sample

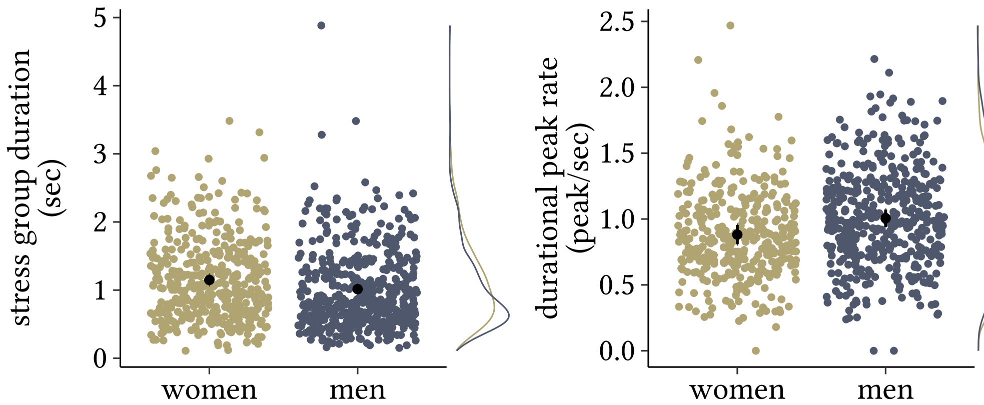
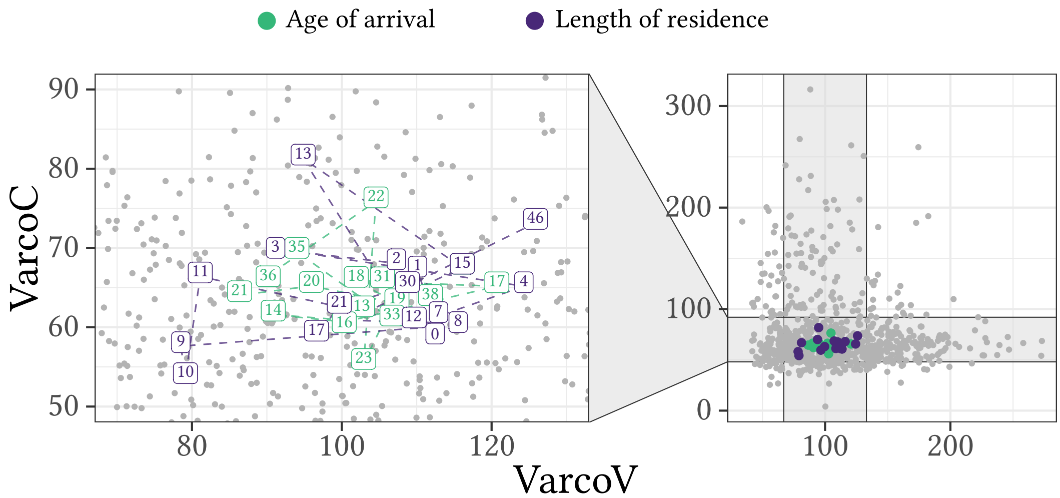
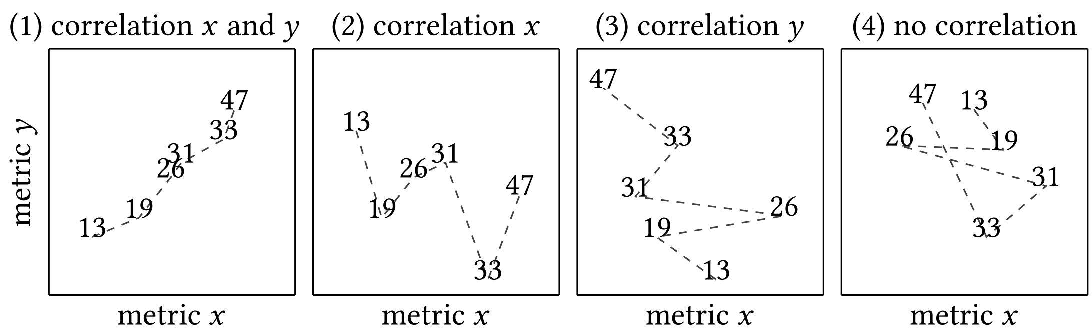


Figure 4.13: Distribution of VarcoC in function of VarcoV in the ALCP sample, and the medians according to age of arrival (lighter green color) and length of residence (darker purple color)



reveal some kind of order in one or both of the scales. For instance, if age of arrival showed a strong positive correlation with both VarcoC and VarcoV, the ages would be linearly distributed in the Cartesian plane, forming the pattern of a diagonal line that connects the origin with the opposite corner. This pattern is illustrated in the Panel 1 from Figure 4.14.

Figure 4.14: Examples of different distributions correlating two rhythm metrics and a third numerical variable



There is also the possibility of age of arrival correlating with only one metric. If age of arrival correlates only with VarcoV, in the x-axis, we would notice that, independently of their position in the y-axis, the ages would be sequentially arranged starting from the left with the lowest age and ending with the highest age in the rightmost position (see Panel 2) — in negative correlation, the order is inverted. The same applies to correlation with the y-axis, but instead of a horizontal sequencing, we would observe a vertical one (see Panel 3). There is still a fourth possibility: age of arrival may not correlate with any of the metrics. If so, the ages will be randomly distributed without order in both axes, as shown in Panel 4.

It is clear from Figures 4.13 and 4.15 that, at least from this descriptive point of view, no correlation can be detected between any metrics and the migration variables. For instance, participants who migrated at 36 years old share a similar median VarcoV with participants who migrated at the age of 14. In the case of length of residence, we observe that migrants who have lived for 46 years in Campinas have a median VarcoV very close to migrants who have lived for only 4 years. As for the VarcoC, we observe that younger ages of arrival, such as 13 and 16 years old, are very close to older ages, such as 33 years old. Examples like these ones can be easily

Figure 4.15: Distribution of nPVI-C in function of nPVI-V in the ALCP sample, and the medians according to age of arrival (lighter green color) and length of residence (darker purple color)

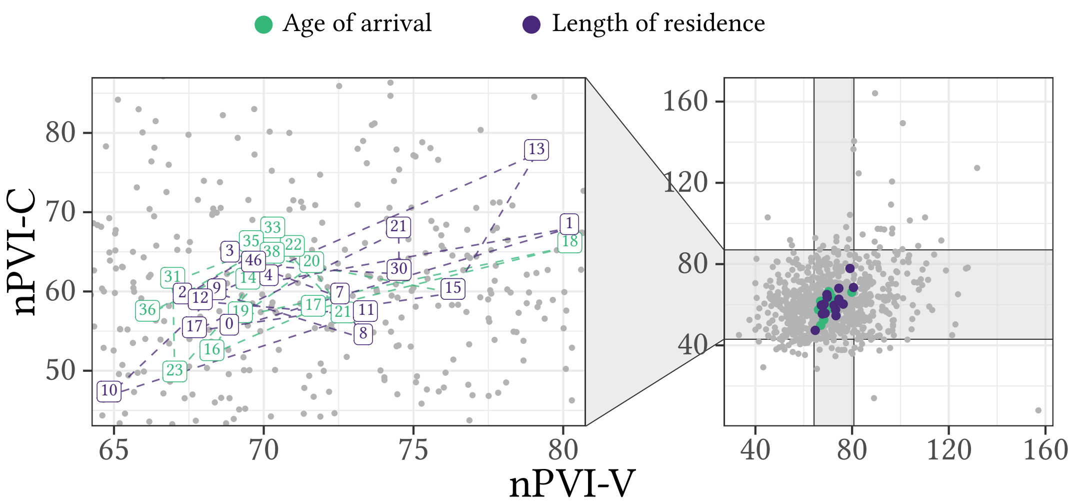


Table 4.9: Mixed-effects linear regression models of VarcoV, VarcoC, nPVI-V, and nPVI-C, contrasting age of arrival and length of residence in the ALCP sample

	VarcoV (28)	VarcoC (29)	nPVI-V (30)	nPVI-C (31)
Fixed effects				
(Intercept)	93.3	65.4	72.1	63.0
	[87.0, 99.7]	[59.0, 71.8]	[69.2, 75.0]	[60.1, 66.0]
Age of arrival	0.6***	0.3	−0.001	−0.01
	[0.2, 0.9]	[−0.02, 0.6]	[−0.1, 0.1]	[−0.2, 0.1]
Length of residence	0.3	0.2	−0.02	−0.1
	[−0.1, 0.6]	[−0.2, 0.6]	[−0.2, 0.1]	[−0.3, 0.1]
Random effects				
Std. Dev. (speaker)	9.69	10.19	4.56	4.72
Std. Dev. (residual)	38.33	29.38	15.22	14.19
Marginal R^2	4%	2%	0%	0%
Conditional R^2	10%	13%	8%	10%
Observations	2,888	2,888	2,889	2,889

*: statistically significant ($\alpha = 0.05$)
R formula: `lme4::lmer(age_arrival + length_residence + (1|speaker))`

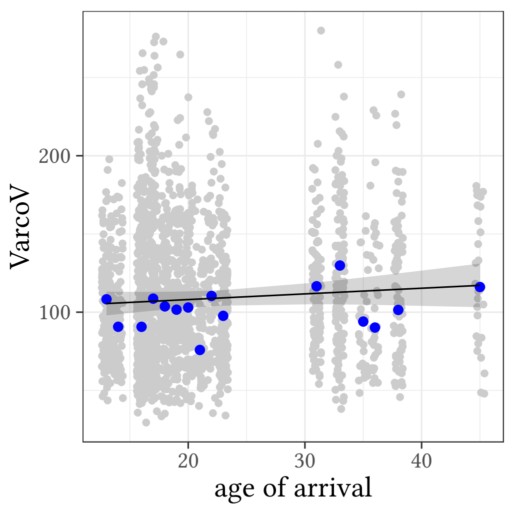
found in Figure 4.15 as well, concerning PVI metrics.

Until now, we have discussed the rhythm metrics relying only on the patterns of raw data and descriptive statistics, i.e., the medians. Table 4.9 shows inferential information from regression models concerning the four metrics. For three of them, the results are consistent with what was already discussed: age of arrival and length of residence are not good predictors of the variation in VarcoC, nPVI-V, and nPVI-C. In the three, the increase in age of arrival and length of residence has a very small effect on these variables, and the marginal R^2 of their models is not higher than 2%.

On the other hand, the angular coefficient of the age of arrival is statistically significant in the case of VarcoV. In the last section, we observed that Alagoan migrants have an estimated VarcoV that is significantly higher than that of native Campineiros, and Table 4.9 shows that the decrease in age of arrival correlates with a decrease in VarcoV. These two results combined seem to suggest that age of arrival may be a conditioning factor of rhythmic accommodation. In other words, Alagoan migrants who arrived at a younger age in Campinas also have a VarcoV more similar to that of Campineiros.

However, the value of the angular coefficient is quite small. As discussed in Section 3.4, Varco metrics express the dispersion in the duration of an interval as the proportion of the mean duration of this interval in the utterance. Thus, an estimate of 0.6 means that increasing the age of arrival in one unit correlates with an increase in the dispersion of vocalic duration that is proportional to 0.6% of the mean vocalic duration. Figure 4.16 helps to visualize the correlation between age of arrival and VarcoV. The points in blue correspond to the same medians of VarcoV displayed in Figure 4.13. The regression line shows a slight positive inclination, but it is still much closer to a horizontal position. Furthermore, based only on the raw data or on the median points, it seems not possible to detect an inclination pattern.

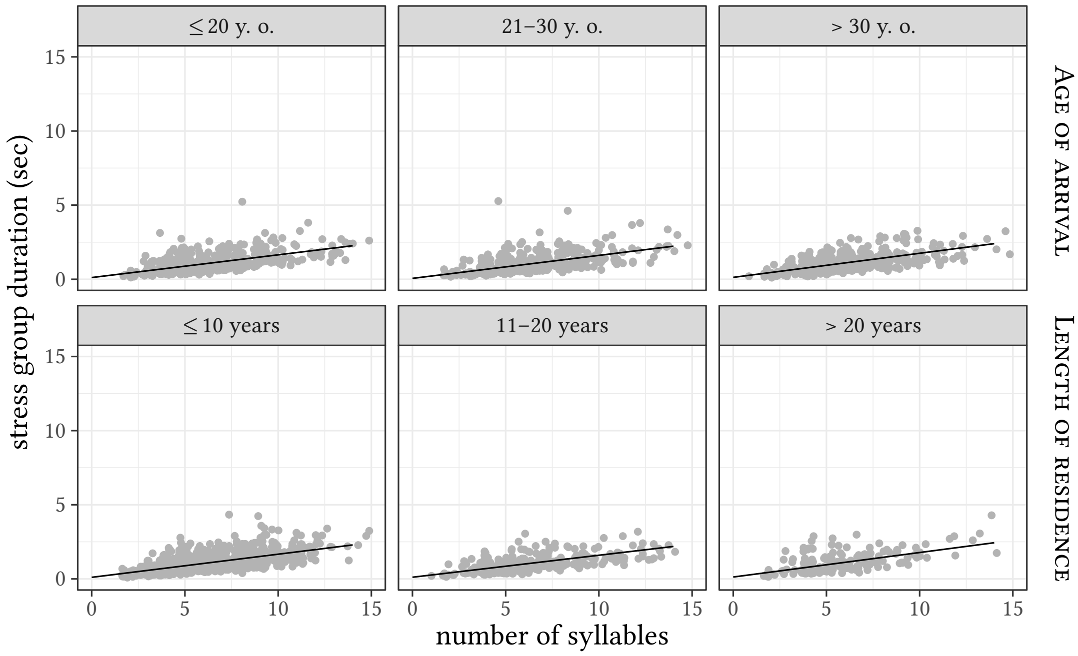
Figure 4.16: Regression line of the correlation between VarcoV and age of arrival in ALCP sample. The blue dots refer to the median of each participant.



The last rhythm parameter to be analyzed in this section is the duration of stress groups with respect to the number of syllables therein. In the case of this parameter, the plotting of a regression line is a useful technique to visualize the degree of syllable- or stress-timing, since this degree is directly related to the line inclination (see Section 3.4). To represent the migration variables in conjunction with stress group duration and the number of syllables, I converted the numerical scale of age of arrival and length of residence into a categorical scale in which the participants are divided into three groups. In the case of age of arrival, the groups are the following: participants who arrived in Campinas (i) at the age of 20 or younger; (ii) from the age of 21 to 30; and (iii) at the age of 30 or older. In relation to their length of residence in Campinas, the participants are grouped in (i) those who have lived (i) for 10 years or less; (ii) for more than 10 but less than 21 years; and (iii) for more than 20 years.

Figure 4.17 shows the regression line of each group, along with the respective raw data. More precisely, the lines correspond to the estimate of the interaction between each group and the number of syllables in the stress groups, from the models reported in Table 4.10.

Figure 4.17: Regression line of the correlation between stress group duration and the number of syllables in the stress group according to age of arrival and length of residence.



Speakers' age of arrival or length of residence is not a contrastive factor of the rhythmic tendencies of their speech. The inclination of the regression lines and the distribution of the raw data are virtually identical. The estimates in Table 4.10 show that varying the groups of age of arrival and length of residence correlates with a change in the effect of the number of syllables on the stress group duration of no more than 9 msec. To put this number in perspective, we can compare it with the effect of each individual participant. The standard deviation of the random effects summarizes the variance in the response variable due to the random variables. Since the participant is included in the model as a random variable, this parameter quantifies how much of the stress group duration is affected by changing from one participant to another. Rounding down, in both models, random differences between the individual speakers of the ALCP sample is responsible for a variance of 127 msec in the duration of the stress groups, a value 27 times higher than the better fixed-effects interaction estimates.

Table 4.10: Mixed-effects linear regression models of stress group duration

Duration of stress group			
Age of arrival (AoA) (32)		Length of residence (LoR) (33)	
Fixed effects			
(Intercept)	0.115	(Intercept)	0.106
	[0.034, 0.198]		[0.030 0.182]
N° of syllables × AoA 21–30 y. o.	0.001	N° of syllables × LoR 11–20 y.	−0.007
	[-0.008, 0.012]		[-0.017 0.002]
N° of syllables × AoA > 30 y. o.	0.009	N° of syllables × LoR > 20 y.	0.009
	[-0.002, 0.020]		[-0.002 0.020]
Random effects			
Std. dev. (speaker)	0.127	Std. dev. (speaker)	0.128
Std. dev. (residual)	0.457	Std. dev. (residual)	0.457
Observations	7,772	Observations	7,772

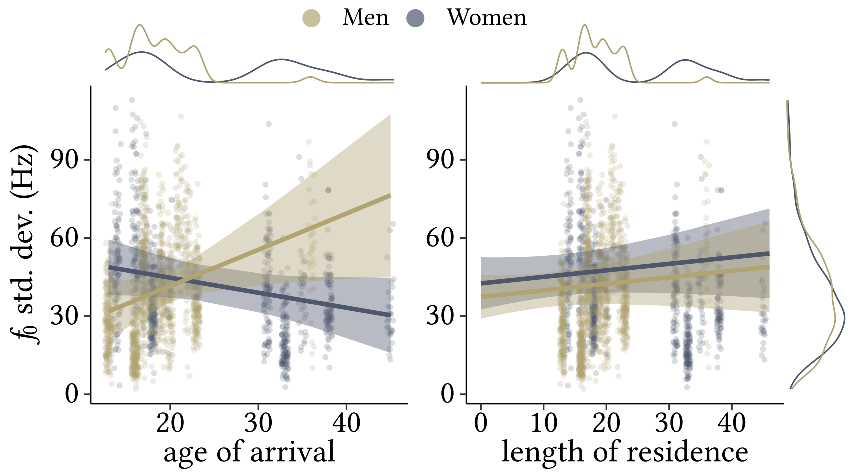
*: statistically significant ($\alpha = 0.05$)
R formula: `lme4::lmer(~ stress group duration + nº syllables * migration variable + (1|speaker))`

4.2.2 Intonation

In this section, I discuss the effects of age of arrival and length of residence on the intonation parameters in the Alagoan migrants' speech. As explained at the beginning of Section 4.1.2, instead of normalizing the fundamental frequency for gender influence, I include gender in the analyses. For most intonation variables discussed in this study, a statistically significant interaction of gender with the age of arrival is observed, but not with the length of residence. Except for the two models discussed at the end of this section, all the others include speakers' age of arrival, length of residence, and gender as predictors and the interaction between the age of arrival and gender as well.

Figure 4.18 shows the distribution of the f_0 standard deviation according to the participants' age of arrival, on the left, and length of residence in Campinas, on the right, distinguishing male, in a lighter and brownish color, from female Alagoan migrants, in a darker and bluish color. The regression lines were plotted based on the estimates for the interaction between gender and each of these migration variables (see Table 4.11).

Figure 4.18: Distribution of f_0 standard deviation in function of age of arrival and length of residence in the ALCP sample



These graphs show a pattern that will be recurrent along this section: males and females have very similar behavior when it comes to the effect of the length of residence on intonation, but the same does not hold for age of arrival. The increase in age of arrival has a different effect on most intonation variables whether the participant is male or female. In the case of f_0 standard deviation, among male Alagoans, increasing the speaker's age of arrival correlates with an increase in the f_0 dispersion. Female Alagoans show an opposite pattern, with a negative correlation between age of arrival and f_0 standard deviation.

In Table 4.11, we observe that gender is the most important predictor of f_0 variability, with male Alagoans having a f_0 contour that varies 42.6 Hz less than female Alagoans. This result is directly related to the anatomical differences between males and females. Higher pitched voice also tends to have higher f_0 dispersion.

Table 4.11: Mixed-effects linear regression models of f_0 standard deviation

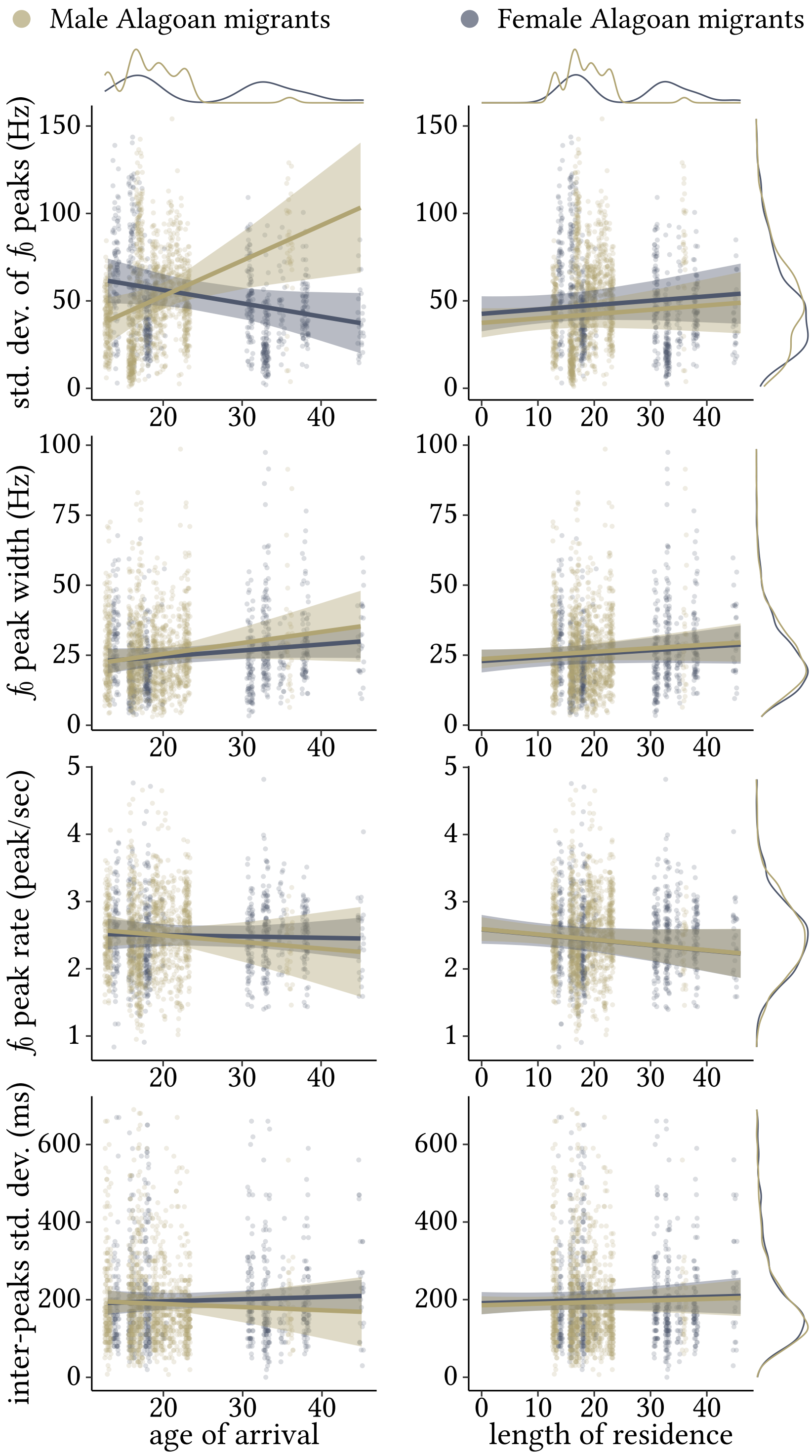
	f_0 standard deviation (34)	
Fixed effects		
(Intercept)	53.5	(33.5, 73.6)
Age of arrival	+0.6	(-1.2, 0.1)
Gender: Men	−42.6	(-73.3, -11.9)
Length of residence	+0.3	(-0.2, 0.7)
Age of arrival × Gender: Men	+2.0*	(0.6, 3.3)
Random effects		
Std. Dev. (speaker)	11.04	
Std. Dev. (residual)	14.79	
Marginal R^2	11%	
Conditional R^2	43%	
Observations	2,000	

*: statistically significant ($\alpha = 0.05$)

R formula: `lme4::lmer(length_residence + age_arrival * gender + (1|speaker))`

More interesting to the discussion is the result concerning the interaction between gender and age of arrival. In the speech of males, the younger the speaker moved to Campinas, the less variable his fundamental

Figure 4.19: Distribution of std. dev. of f_0 peaks, f_0 peak width, f_0 peak rate, and std. dev. of inter-peaks duration according to age of arrival and length of residence in the ALCP sample



frequency — a decrease of 2 Hz per year of age. On the other hand, among female Alagoan migrants, age of arrival does not correlate with f_0 standard deviation.

Even though Model 34 has statistically significant predictors, its explanatory power is not high. Fixed effects can explain 11% of the raw data (marginal R^2), and individual differences between speakers (11 Hz) are greater than those related to age of arrival, which, as we have seen, are not greater than 2 Hz.

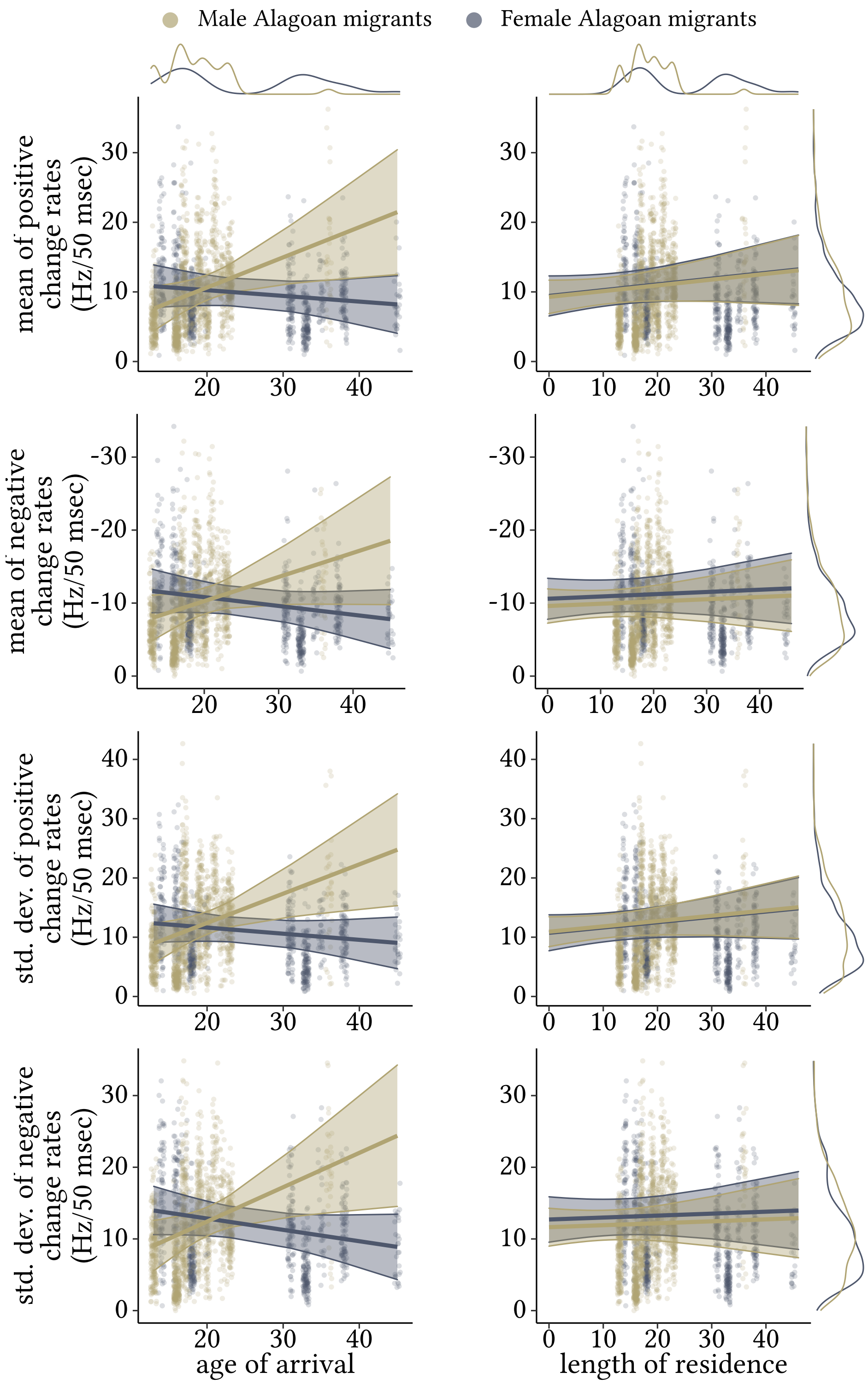
The effects of age of arrival and length of residence on the measures concerning f_0 peaks are more similar between male and female Alagoan migrants, as shown in Figure 4.19. An exception is the standard deviation of f_0 peaks, which shows a pattern that is very close to the one observed in Figure 4.18 concerning f_0 standard deviation. As discussed previously, this similarity between both prosodic parameters is due to the fact that both measure close-related phenomena.

Except for f_0 peak standard deviation, the differences due to age of arrival and length of residence in the other three parameters are not so sharp. On average, male Alagoan migrants have f_0 peaks with a width that is 2.7 Hz flatter than female Alagoans, a not-so-small difference despite not being statistically significant. Age of arrival and length of residence do not account for a difference in f_0 peak width greater than 2 Hz. In relation to f_0 peak rate and standard deviation of inter-peaks duration, the relative estimates are very small and completely irrelevant from the phonetic perspective. For instance, in male migrants' speech, increasing age of arrival by one corresponds to a decrease in the standard deviation of inter-peaks duration of only 1 msec, i.e., from 202 to 201 msec.

Figure 4.20 illustrates the distribution of the mean and standard deviation of both positive and negative derivatives of the fundamental frequency in relation to age of arrival and length of residence. In Section 4.1.2, I showed that the f_0 change rates are the melodic parameters in which Alagoan migrants and native Campineiros most differ. These parameters are also the ones correlating with speakers' age of arrival. However, this correlation is only attested in the male migrants' speech.

Table 4.13 shows the regression models of the four f_0 parameters. It is clear from the estimates that length of residence cannot predict any

Figure 4.20: Distribution of the mean and standard deviation of positive and negative f_0 change rates, according to age of arrival and length of residence in the ALCP sample



relevant portion of the variation in the response variables. In turn, age of arrival proved to be more relevant for understanding variation in f_0 change rates. In male Alagoan migrants' speech, age of arrival positively correlates to the magnitude (mean) and the dispersion (standard deviation) in the rate of both f_0 rises and falls. An increase in the age of arrival by one year corresponds to an estimated increase of 0.5 Hz/50 msec in the four response variables. In other words, the melodic contour of male Alagoans who were older when migrating to Campinas shows steeper and more variable falls and rises than male Alagoans who were younger. In contrast,

Table 4.12: Mixed-effects linear regression models of std. dev. of f_0 peaks, f_0 peak width, f_0 peak rate, and std. dev. of inter-peaks duration

	Peaks of fundamental frequency			
	Std. dev.	Width	Rate	Std. dev. intervals
Fixed effects	(35)	(36)	(37)	(38)
(Intercept)	67	19	2.6	200
	(43, 91)	(11, 27)	(2.2, 3.1)	(100, 300)
Age of arrival (Women)	−0.8	0.2	−0.002	0
	(−1.5, 0.03)	(−0.1, 0.5)	(−0.02, 0.01)	(2, 3)
Age of arrival × Men	2.8***	0.2	−0.005	−1
	(1.2, 4.4)	(−0.4, 0.7)	(−0.03, 0.02)	(−1, 3)
Length of residence	0.4	0.1	−0.01	0
	(−0.1, 1.0)	(−0.1, 0.3)	(−0.02, 0.002)	(−1, 2)
Random effects				
Std. Dev. (speaker)	13.04	4.16	0.23	0.03
Std. Dev. (residual)	20.73	12.54	0.64	0.13
Marginal R^2	12%	2%	2%	0%
Conditional R^2	37%	12%	13%	7%
Observations	1,992	1,999	1,992	1,973

*: statistically significant ($\alpha = 0.05$)

R formula: `lme4::lmer(length_residence + age_arrival * gender + (1|speaker))` female Alagoans' age of arrival seems not to influence their f_0 change rates. Even though in Figure 4.20 their regression lines fall, the respective angular coefficients are not significant, corresponding to values as small as 0.1 Hz/50 msec.

In the analyzes conducted so far, the effects of age of arrival and length of residence were examined considering only the ALCP sample. These analyzes show if the differences between Alagoan migrants in these two migration variables contribute to understanding the variation in their

Table 4.13: Mixed-effects linear regression models of the mean and standard deviation of positive and negative f_0 change rates

	Fundamental frequency 1st derivatives			
	Mean positive (39)	Mean negative (40)	Std. dev. positive (41)	Std. dev. negative (42)
Fixed effects				
(Intercept)	11	−13	13	16
	(5.2, 16.7)	(−18.5, −7.3)	(6.6, 18.8)	(9.4, 22.0)
Age of arrival (Women)	−0.1	0.1	−0.1	−0.2
	(−0.3, 0.1)	(−0.1, 0.3)	(−0.3, 0.1)	(−0.4, 0.05)
Age of arrival × Men	0.5**	−0.5*	0.6**	0.6**
	(0.1, 0.9)	(−0.8, −0.1)	(0.2, 1.0)	(0.2, 1.1)
Length of residence	0.1	−0.03	0.1	0.03
	(−0.1, 0.2)	(−0.2, 0.1)	(−0.1, 0.2)	(−0.1, 0.2)
Random effects				
Std. Dev. (speaker)	3.17	3.09	3.32	3.48
Std. Dev. (residual)	4.21	3.99	4.97	4.82
Marginal R^2	10%	7%	1%	9%
Conditional R^2	42%	42%	38%	4%
Observations	2,000	2,000	1,999	2,000

*: statistically significant ($\alpha = 0.05$)

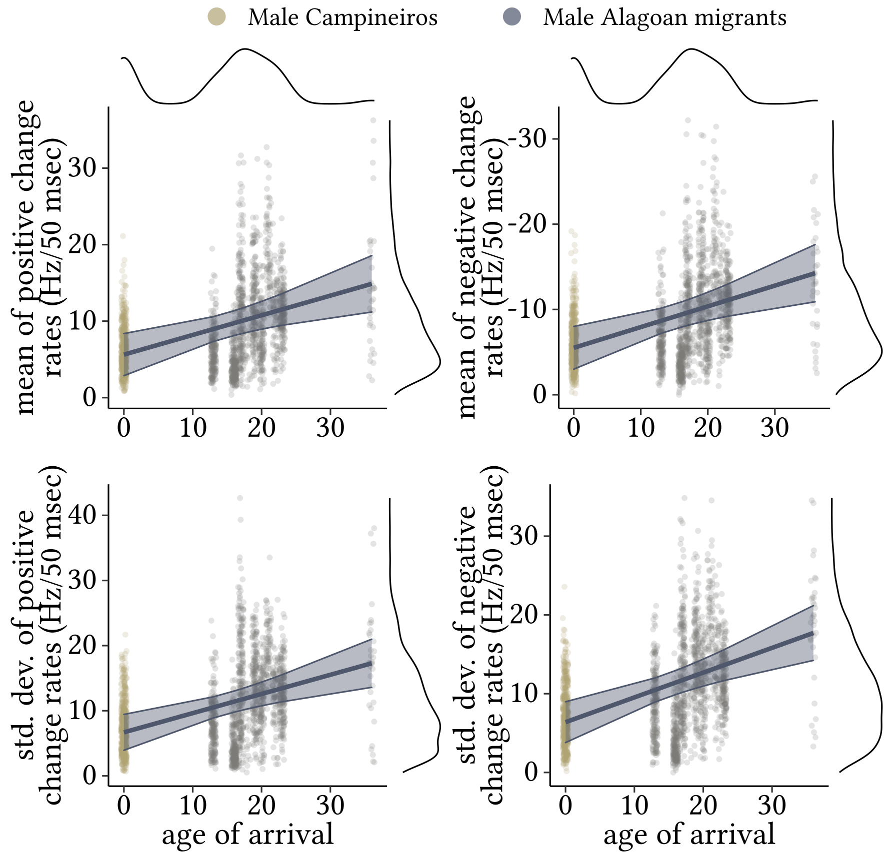
R formula: `lme4::lmer(length_residence + age_arrival * gender + (1|speaker))` speech prosody. However, they cannot inform if the observed correlations are in the expected direction of accommodation to native Campineiros' prosody. It is possible that an observed strong correlation is in a direction that is the opposite of the expected for the accommodation. For instance, the correlation with length of residence can be such that more years the migrants live in Campinas correspond to more difference in relation to the prosody of native Campineiros. Using Giles' (1991) terms, in this case, the length of residence would be increasing a divergence, instead of convergence, in migrants' prosody in relation to natives. Therefore, further analyzes are necessary to assess if the correlation with migration variables can be taken as evidence of dialect accommodation.

To better interpret the correlation between age of arrival and f_0 change rates in Alagoan migrants' speech, Table 4.14 shows linear models in which the data of Campineiros is also included. More precisely, these models include both the ALCP and CPS data and the speaker's age of arrival in Campinas as the only fixed-effect predictor. As all models discussed so far, they also include the participants as a random effect.

In these models, for all native Campineiros, age of arrival was set to zero. When the predictor is a numeric variable, the Intercept refers to the value of the response variable, in the y-axis, when this predictor, on the x-axis, is zero. Therefore, in the models of Table 4.14, the Intercept refers to the estimated f_0 change rate when the age of arrival is zero. Insofar as the native Campineiros are the only ones who have the age of arrival being zero, the Intercept can be interpreted as being the estimated f_0 change rate in native Campineiros' speech. Thus, the angular coefficient, shown just below the Intercept, expresses what occurs in the f_0 change rates when we increase the age of arrival by 1 year old, having the Intercept (i.e., the CPS estimate) as a reference.

Figure 4.21 shows the relationship between the speaker's age of arrival and each of the four f_0 change rate parameters in male participants' speech. The raw data of male Alagoan migrants are in a darker and bluish color and spread along with the age of arrival scale in the x-axis. In turn, the raw data of male native Campineiros are in a lighter and brownish color and concentrate in the zero point in the x-axis. A regression line modeling all the raw data together (both ALCP and CPS data) is also shown.

Figure 4.21: Distribution of the mean and standard deviation of positive and negative f_0 change rates, according to age of arrival in the speech of males



In the four f_0 change rate parameters, on average the Alagoan migrants have values higher than native Campineiros, but decreasing the migrants' age of arrival correlates to a decrease in the prosodic parameters, a result that is in agreement with the pattern observed in the literature for segmental variables (see Chapter 2). This result is evidence of a sociophonetic process of prosodic accommodation in which the speaker's age of arrival is an important factor to model and predict the degree of this type of dialect accommodation.

A detailed description of the respective linear models is found in Table 4.14. As mentioned above, when the predictor is a numeric variable, the Intercept refers to the value of the response variable, in the y-axis, when this predictor, in the x-axis, is zero. Therefore, in the models of Table 4.14, the Intercept refers to the estimated f_0 change rate when the age of arrival is zero — that is, f_0 change rate in native Campineiros' speech.

The f_0 of male native Campineiros rises and falls at an estimated rate of respectively 5.6 and -5.5 Hz/50 msec. Lowering male Alagoan migrants' age of arrival by one year correlates to a decrease (in absolute value) of 3 Hz/50 msec in the mean of both positive and negative first derivatives of f_0. This same amount of decrease is also observed in the standard deviation of negative and positive f_0 change rates, but in this case both native Campineiros and Alagoan migrants have values slightly higher than the means.

The importance of age of arrival to predict the data concerning f_0 change rates is confirmed by the marginal coefficient of determination. From all mixed-effects linear models reported so far, Models from 43 to 46 in Ta-

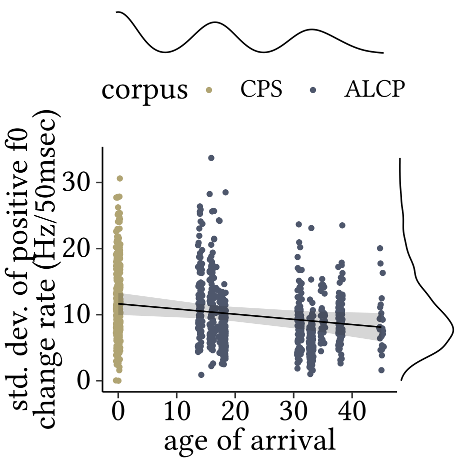
Table 4.14: Mixed-effects linear regression models of the mean and standard deviation of positive and negative f_0 change rates in the speech of males

	Fundamental frequency 1st derivatives			
	Mean positive (43)	Mean negative (44)	Std. dev. positive (45)	Std. dev. negative (46)
Fixed effects				
(Intercept)	5.6	−5.5	6.7	6.4
	[2.9, 8.4]	[−8, −3]	[3.9, 9.4]	[3.8, 9.0]
Age of arrival	0.3*	−0.3*	0.3*	0.3*
	[0.1, 0.4]	[−0.4, −0.1]	[0.1, 0.5]	[0.2, 0.5]
Random effects				
Std. Dev. (speaker)	3.35	3.05	3.34	3.14
Std. Dev. (residual)	3.89	3.65	4.81	4.51
Marginal R^2	17%	18%	17%	21%
Conditional R^2	52%	51%	44%	47%
Observations	1,739	1,739	1,739	1,739

*: statistically significant ($\alpha = 0.05$)

R formula: `lme4::lmer(age_arrival + (1|speaker))` ble 4.14 are those with the best explanatory power. The four models have marginal R^2 of 17% or higher, with the best of them, Model 46 concerning the standard deviation of negative f_0 change rates, having an R^2 of 21%. Evidently, explaining 21% of the variance in the response variable also means *not* explaining the remaining 79%, which is a very large fraction of the raw data being unexplained. However, we need also to pay attention to the fact that these models include only one predictor variable, age of arrival, and a single social predictor being capable of explaining 21% of a linguistic variable is remarkable.

Figure 4.22: Distribution of the standard deviation of positive f_0 change rates, according to age of arrival in the speech of female participants



Another interesting result is the similarity of the regression lines in the four graphs of Figure 4.21. This similarity suggests at least two types of parallelism. The first type concerns the parallelism between positive and negative f_0 change rates. In both native Campineiros and Alagoan migrants' speech, the rate of f_0 rises is very close to the rate of f_0 falls. As discussed in Section 4.1.2, there is not any physiological restriction requiring that the rate of f_0 rises matches the rate of f_0 falls. From the perspective of speech production, the speakers can produce f_0 rises independently of f_0 falls (e.g., producing a fast f_0 rise followed by a slow fall). Without this production property, the possibility of using f_0 contour to encode linguistic, paralinguistic, and sociolinguistic information would be largely restricted. To further investigate the relationship between positive and negative derivatives of f_0 in Brazilian Portuguese is an important task for future studies.

The second type of parallelism concerns the process of prosodic accommodation. The similarity in the regression lines suggests that age of arrival seems to equally condition the rate of accommodation in the four f_0 change rate parameters. Parallelism in dialect accommodation to closely related linguistic variables may be expected, but this expectation was not confirmed in two previous studies about Northeastern migrants in São Paulo. As discussed in Chapter 2.2, the height of pretonic mid vowels /e/ and /o/ is a phonetic feature that distinguishes Northeastern from Southeastern varieties of Brazilian Portuguese, and parallelism between the /e/ and /o/ is observed. For instance, a variety that has a lower pretonic /e/, such as many Northeastern varieties, usually also has a lower pretonic /o/. Thus, it may be expected that an Alagoan migrant in São Paulo who has accommodated to the raised pronunciation of the Paulista pretonic /e/ will also have accommodated to the Paulista raised pretonic /o/. Nonetheless, Oushiro (2019) and Santana (2018) have shown that this is not necessarily so. Their results reveal that many Northeastern migrants have accommodated to only one of the two vowels.

In contrast with the height of pretonic /e/ and /o/, parallelism is observed in accommodation involving f_0 change rate parameters. Male Alagoan migrants who accommodated to Paulista positive f_0 change rate parameters have also accommodated to the negative parameters, and this parallelism applies both to the mean and the standard deviation of f_0 change rates.

The patterns of accommodation involving f_0 change rates discussed so far concern male Alagoan migrants. When it comes to female Alagoan participants, age of arrival is not a relevant factor even in models in which female Campineiras are also included. The best linear model involving female speakers regards the mean of positive f_0 change rates. It is the only one in which age of arrival is a statistically significant predictor, but its angular coefficient is only -0.1 Hz/50 msec and its marginal R^2 is 5%. This model is visually represented in Figure 4.22. Different from males, the decrease in the age of arrival correlates with an increase in the mean of positive f_0 change rates. Despite the significance of the regression line, Figure 4.22 shows that the correlation is not so strong as in the case of male Alagoans. In particular, the wide dispersion of female Campineiras, encompassing the range of variation of all Alagoan migrants, makes it difficult to interpret this result as dialect accommodation.

4.2.3 Summary

Age of arrival and length of residence do not seem to be relevant predictors of rhythmic variation in the speech of Alagoan migrants. Except for VarcoV, no significant difference was observed in the rhythmic variables when the migrants are compared according to these two migration variables. Age of arrival seems to have a weak effect on VarcoV, acting as a conditioning factor of accommodation. The speakers who have moved at a younger age to Campinas are the ones with values of VarcoC that are closer to those of Campineiros.

On the other hand, regarding intonation, the results reveal different behaviors between male and female migrants. We observed a correlation between age of arrival and six of the nine intonation variables, but only in the case of male speakers. An increase in the age of arrival correlated to an increase in the standard deviation of f_0 and f_0 peaks, f_0 positive change rates, and f_0 negative change rates. It also correlated with the mean of both positive and negative f_0 change rates. Furthermore, these correlations are in the expected direction for dialect accommodation: the younger the speakers were when they moved to Campinas, the closer they are to the Campineiros in these six intonation variables. Except for the standard deviation of positive f_0 change rates, age of arrival is not a significant predictor in female participants' speech. No correlation involving the length of residence was observed in this study.

4.3 Individual variation

Along the previous sections, I discussed variation in speech prosody in terms of groups of speakers. In particular, I compared the groups of native Campineiros and Alagoan migrants in Section 4.1. I also compared Alagoan migrants according to different ages of arrival and lengths of residence in Campinas in Section 4.2. However, in the course of these comparisons, it became evident that, in addition to between-groups variation, another type of variation is important to understand our data: variation between individual participants. The results of most prosodic parameters discussed so far suggest that individuals may be more important than the samples, age of arrival, and length of residence in explaining the variance in these parameters. One evidence for the importance of individual variation is the data dispersion, with most parameters showing greater dispersion for the migrants than for the non-migrants.

Individual variation is an important factor causing this dispersion of prosodic data. However, we need to distinguish two types of individual variation: the inter-speaker and the intra-speaker (or stylistic) variation. The first concerns the variance due to differences between speakers, while the second refers to variance resulted from different values of the same speaker. Although both types simultaneously contribute to variation in speech, in this study, I will focus only on inter-speaker variability, because it is the type that can most help understand the acquisition of new prosodic features in a dialect contact situation. In particular, analyzing inter-speaker differences can help understand why some Alagoan migrants have prosodic patterns similar to native Campineiros, while others do not. Besides, in most linear models reported above, the standard deviation of the random factor, which refers to the individuals, is larger than the estimates of the fixed-effects predictors, which suggests that inter-speaker variation is relevant. For instance, in Model 1 (see Table 4.1), the difference between Alagoan migrants and native Campineiros correlates to a difference of 0.4 syll/sec in speech rate, while the variance in this prosodic parameter due to random differences between individuals is 0.72 syll/sec. In other words, individuals seem to be more relevant than the ALCP-CPS dichotomy to explain differences in speech rate.

In order to assess the influence of inter-speaker variation on the prosodic parameters, I devised an Inter-Speaker Variation Index (ISVI), which is a simple algorithm combining elementary statistical procedures. First, I compute the arithmetic mean of the prosodic parameter for each individual speaker. The result is a list of means, each one representing a single participant. Then, the standard deviation of this set of means is calculated as a way of assessing the amount of variation between participants. To compare the index of different prosodic parameters, since some parameters naturally have higher values than others and they can differ in the measurement unit, I also apply a simple normalization procedure, dividing the standard deviation by the absolute value of the arithmetic mean of the participants' means. The absolute value is necessary to prevent the standard deviation from being divided by a negative number. The normalization basically converts the unit of the standard deviation to a proportion in relation to the mean — this measure is called coefficient of variation. For instance, if the coefficient has a value of 0.25, it means that the variance is 25% of the value of the mean. So, higher values indicate higher inter-speaker variability.

Table 4.15 shows the ISVI of all prosodic parameters sorted from the highest, on the top, to the lowest value, on the bottom. The indexes were computed only from Alagoan migrants' data. The table also shows the prosodic domain of each parameter. That is, it specifies if the parameter is a measure of the fundamental frequency or of the duration of some phonetic unit. It is important to mention that, in the case of parameters concerning f_0, the index for male participants was computed separately from the index

Table 4.15: Distribution of the prosodic variables according to their Inter-Speakers Variation Index

Prosodic variable	Domain	Individual variation index
Positive f_0 change rates mean	Frequency	0.35
Negative f_0 change rates mean	Frequency	0.33
Negative f_0 change rates std. dev.	Frequency	0.32
Positive f_0 change rates std. dev.	Frequency	0.31
f_0 std. dev.	Frequency	0.29
f_0 peaks std. dev	Frequency	0.28
Inter-peaks duration std. dev	Duration	0.21
VarcoC	Duration	0.18
Width of f_0 peaks	Frequency	0.15
Durational peak rate	Duration	0.14
Stress group duration	Duration	0.13
VarcoV	Duration	0.11
Rate of f_0 peaks	Duration	0.10
nPVI-C	Duration	0.10
nPVI-V	Duration	0.08

for females. Then, the final value was obtained by calculating the mean of both indexes. Even though the rate of f_0 peaks and the standard deviation of inter-peaks duration consist of melodic variables, they do not measure frequency but duration. For this reason, they are classified in the duration domain.

There is a very explicit pattern in the distribution of the prosodic variables: except for f_0 peak width, the frequency-related measures stand at the top, with the highest values of inter-speaker variability. In other words, there is a clear division between frequency- and duration-related parameters, according to which Alagoan migrants differ from each other more in the frequency domain.

Figure 4.23 graphically displays the variation between male Alagoan participants in the two prosodic variables with the highest and lowest ISVI; respectively, the mean of positive f_0 change rates, on the left, and the nPVI-V, on the right. The error bars represent the estimates and their 95% confidence intervals from a simple linear model in which the participants are included as the only fixed-effects predictor (no random factors are included). The vertical column in light gray corresponds to the 95% confidence interval of male Campineiros considered as a group, and it serves for reference.

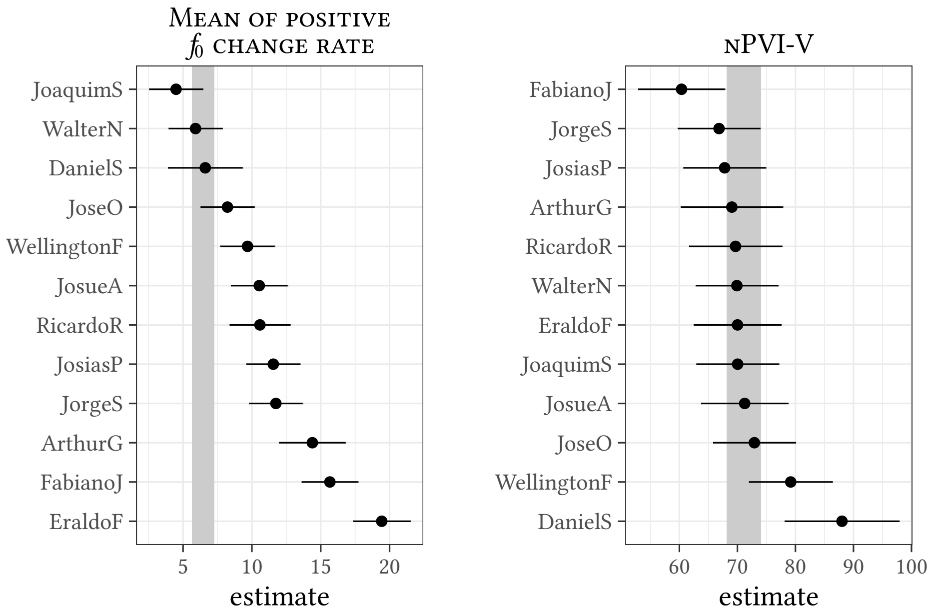
The difference of variability in both variables is remarkable. In the nPVI-V plot, except for FabianoJ, WellingtonF, and DanielS, the error bars practically form a vertical column in which all of them overlap. The case of FabianoJ is worthy of notice: of all male Alagoan migrants, FabianoJ is a participant with a remarkable high speech rate of 7.7 syll/sec, a value that distinguishes him from all the other participants. As explained in Section 3.4, faster speech rates correlate with less variability in the duration of vocalic and consonantal intervals. Even having a normalization for speech rate, it is possible that nPVI-V could not completely "neutralize" the fast rate of FabianoJ. The left plot, on the other hand, shows a very different situation, with a significantly greater dispersion of the error bars along the scale of positive f_0 change rates mean.

Another difference between the two plots is the distribution of the error bars in relation to the CPS confidence interval (the vertical light gray column). In the left plot, except for JoaquimS and WalterN, and maybe

DanielS, the estimate and the error bar of all the other participants appear to the right of the CPS confidence interval, and most of them do not overlap with it. In the right plot, the error bars are balanced around the CPS confidence interval, and most overlap with it. This difference clarifies why in Section 4.1 we observed a significant difference between Alagoan migrants and Campineiros in positive f_0 change rate but not in nPVI-V.

The primary purpose of analyzing between-speaker variation in a dialect acquisition study is to examine how close each migrant is from the patterns of the host community, as a way of assessing if an accommodation process can be observed and, if so, in which stage of this process each migrant speaker is. Instead of analyzing this accommodation process speaker by speaker and variable by variable, I devised another index with the purpose of summarizing how much each participant accommodated to Campinas' prosody in general. Since this index takes into account all prosodic parameters discussed in this study, it can be considered as a global measure of prosodic accommodation.

Figure 4.23: Distribution of linear regression angular coefficient and 95% confidence interval of each male Alagoan migrant. The vertical gray color represents the 95% confidence interval of male Campineiros



The computation of this index is based on linear regression models. First, I compute a linear model for each prosodic parameter, in which this parameter is the response variable and the participants are included as the sole predictor (for example, see Table 4.24). Native Campineiros are also included, but as if they were a single participant. Therefore, in these models, each Alagoan migrant has his own estimate, while a single estimate is computed for all native Campineiros as a group. The purpose of the CPS estimate is to serve as a reference to which we can compare the estimates of individual Alagoan migrants.

The computation of this Global Accommodation Index (GLA) starts by calculating a score for each Alagoan migrant in relation to each prosodic parameter. For instance, let's suppose we want to calculate FabianoJ's score for speech rate. First, we create a linear model in which the speech rate is the response variable and the participants are a fixed-effects predictor, in such a way that each participant will have her own estimate. As an example, this model is illustrated in Table 4.24.

As explained above, the CPS sample is included as a single participant. FabianoJ's score, then, is obtained by dividing the difference between his and CPS sample's estimate by the mean of all estimates of the model, and then getting the absolute value of the quotient (see Equation 4.1). Therefore, according to Table 4.24, FabianoJ's estimated speech rate is 7.7 syll/sec and CPS's is 5.2. Thus, the difference is 2.5 syll/sec. Then, we divide this difference by the mean of all estimates of the model, which is 5.9 syll/sec, getting a quotient of 0.42. This score is already positive, but the procedure involves obtaining the absolute value of this quotient to ensure that negative values are transformed to positive.

Figure 4.24: Linear regression model's estimates of speech rate

Speech rate	
CPS sample	5.2
FabianoJ	7.7
RicardoR	6.8
JoaquimS	6.7
JorgeS	6.7
WalterN	6.0
EraldoF	5.9
ArthurG	5.6
JosiasP	5.5
JosueA	5.5
WellingtonF	5.2
JoseO	5.0
DanielS	4.7

`lm(speech rate ~ speaker)`

The score is just a normalized estimate relative to the CPS sample and, in this example, it quantifies the difference between FabianoJ and CPS sample in speech rate. However, since we want a value that summarizes the global prosodic difference of FabianoJ in relation to CPS, one which synthesizes the differences in more than one prosodic variable, we need to repeat this procedure, calculating the score for each of the other prosodic variables. With all FabianoJ's scores already computed, we obtain the final index simply summing up the scores together. The higher the index, the greater the prosodic difference between FabianoJ and CPS sample. From a different perspective, the *smaller* the index, the greater the accommodation. Equation 4.1 formally describes the index computation.

$$\begin{aligned} \text{score}_i &= \left|\frac{e_i - e_{CPS}}{\bar{e}}\right| \\ \text{index}_j &= \sum_{i=0} \text{score}_{i+1} \end{aligned} \tag{4.1}$$

Where,

e_i = Estimate of speaker i

e_{CPS} = Estimate of CPS sample

$\bar{e}$ = Mean of all estimates of the model

These steps described above were carried out for all participants of the ALCP sample in relation to all prosodic variables analyzed here. However, the indexes of male Alagoan migrants were computed separately from those of females (i.e., from different linear models). Due to the difference between frequency- and duration-related measures discussed above, we also computed different indexes for these two groups of prosodic variables.

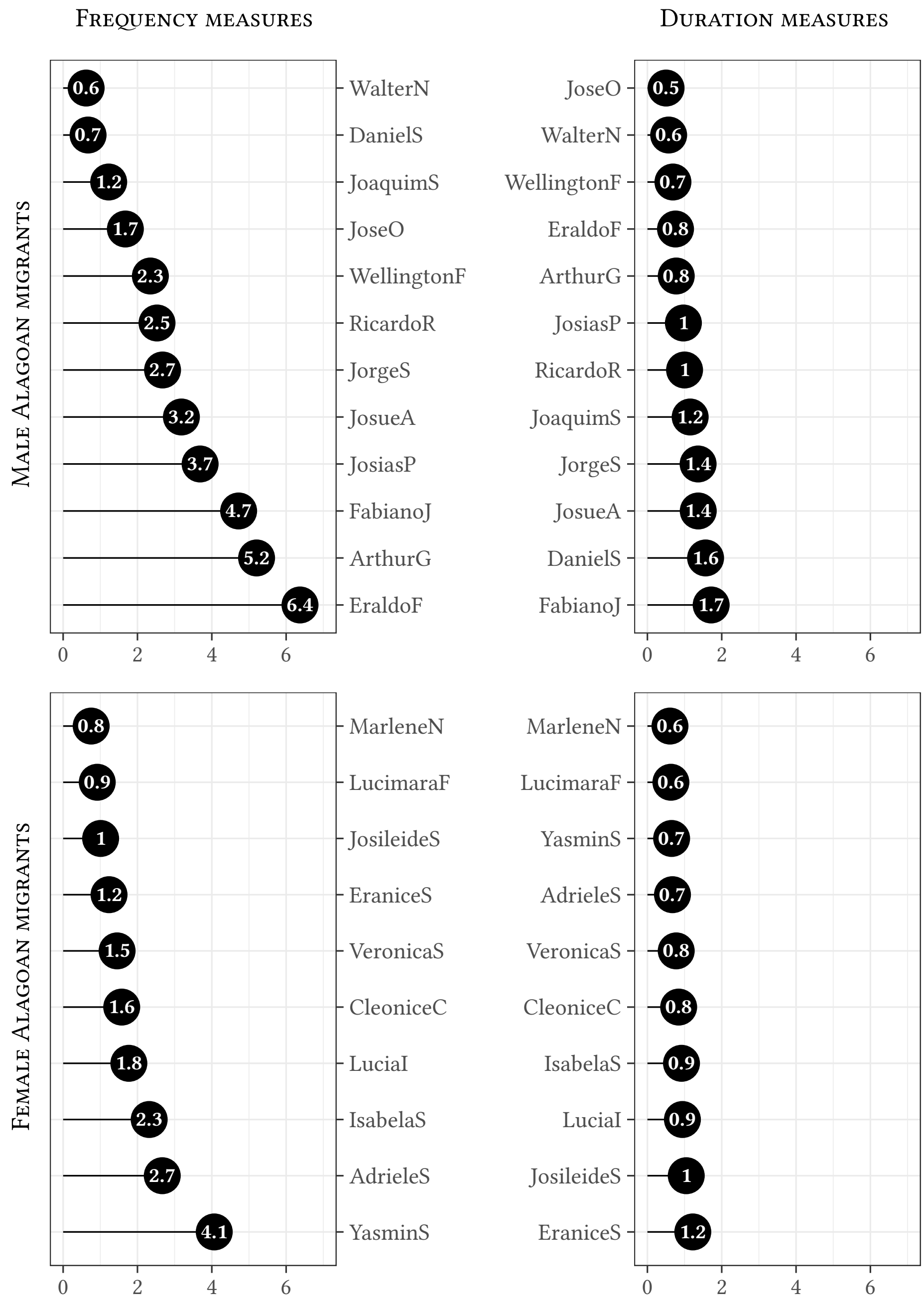
Figure 4.25 plots these indexes according to these two divisions (males, females; frequency-related, duration-related variables). It is important to notice that the participants are sorted according to their index, from the lowest at the top to the highest at the bottom. Thus, the closer the participant is to the top, the more similar her prosody is to that of Campinas. Another important point is that the participants of the CPS sample, included as the reference, also change from plot to plot. For instance, in the top left plot, the indexes of male Alagoan migrants were computed having only male Campineiros as the reference. In the index scale on the x-axis, zero means no difference in relation to the CPS sample.

These plots are consistent with the results discussed so far. In particular, they reinforce two patterns. First, in comparison to female Alagoan migrants, male Alagoans exhibit a greater global prosodic difference in relation to native Campineiros in both groups of prosodic variables. Second, in both male and female Alagoans' speech, the differences in relation to the CPS sample are greater within the group of frequency-related prosodic parameters.

One pattern that is worth mentioning is the order of the participants. We expected that the order would be similar in frequency- and duration-related measures, but Figure 4.25 show that this is not the case, especially in male speech. For instance, in frequency-related measures, EraldoF is in the last position, being the participant who most differs from Campineiros. Regarding duration-related measures, he goes up to the fourth position, being much closer to Campineiros. The opposite situation occurs with DanielS, who goes from the second to the penultimate position when comparing his frequency-related score to his duration-related one. On the other hand, female speakers seem to be much more consistent than males. The order in both types of scores is very similar, with some speakers having the same position in both, such as MarleneN, LucimaraF, VeronicaS, and CleoniceC.

This research did not aim to conduct in-depth analyses of other social variables besides the speakers' age of arrival and length of residence. To further investigate whether other social characteristics of the participants can help explain the individual variance observed in this section is an important task for future studies. This task is far from being easy since multiple factors can influence why an individual participant speaks a particular way. Additionally, it can be very challenging to give a rigorous treatment for many of these factors. For instance, social identity with the communities of origin and destination, motivations and attitudes, and social networks are all individual factors that may play an important role in dialect accommodation, and in linguistic behavior more generally. However, the challenge is to devise methodological procedures to operationalize these

Figure 4.25: Distribution of the ALCP's participants according to their Global Accommodation Index, contrasting frequency and duration measures



factors, especially those concerning identity, motivations, and attitudes, to obtain reliable results that are both comparable and reproducible.

As a preliminary and exploratory analysis, I examined the sociodemographic information about the participants of the ALCP sample collected by means of the application of two questionnaires (see Section 4.1) concerning socioeconomic profile, social network, habits, and social evaluations of São Paulo and Northeastern region of Brazil. In particular, I analyzed the following three sociodemographic factors: age, level of education, and social class. I also examined the participants' responses to two questions concerning identification with São Paulo and Alagoas. One of them is the following: *on a scale from 0 to 10, how much do you consider yourself Paulista today?* I expected that the participants who differ least from the prosody of Campinas were the ones who consider themselves to be Paulista the most, but this correlation was not observed. The two participants that self-attributed the nine and ten to this question, JosiasP and JorgeS respectively, are not the ones with the lowest GLA — i.e., they are not the ones more similar to native Campineiros in prosody (see Table 4.25). On the other hand, JoseO, who has a GLA lower than JosiasP and JorgeS, self-attributed a one.

The second question was about the participants' identification with Alagoas: *on a scale from 0 to 10, how much do you consider yourself Alagoan today?* Almost all of them self-attributed a score of eight or more for this question. Discussing this same variable in the ALCP and PBCP[3] samples, Oushiro (2020b, p. 84) argues that this result was expected "since 'denying one's origins' or 'forgetting one's roots' is generally frowned upon". Since the participants showed very similar behavior to this question about their degree of identification with Alagoas, it cannot help to understand the individual differences in prosodic accommodation.

I conducted linear regression analyses to test possible correlations between these five social variables and the seventeen prosodic parameters, and did not find any correlation.

[3] The PBCP is another sample of the Accommodation Project with recordings of the speech of migrants from the Northeastern state of Paraíba living in São Paulo (Oushiro, 2020b).

4.3.1 Summary

We observed a clear distinction between frequency- and duration-related variables in relation to individual variation. Except for f_0 peak width, the variability between speakers is significantly higher in all frequency-related variables. We also observed significant differences between the participants in how close they are to the intonation patterns of the CPS sample. These differences, however, did not correlate with social characteristics of the participants.

Chapter 5

Conclusion

This study investigated the speech prosody of Northeastern migrants from the state of Alagoas living in Campinas, a region in the Southeast state of São Paulo. The primary goal was to examine to what degree their prosody changed due to the contact with Paulista speakers. I analyzed a group of seventeen acoustic variables concerning rhythm and intonation from three different perspectives. In the absence of descriptions of the prosody of São Paulo and Alagoas, I compared Alagoan migrants and native speakers from Campinas, aiming at identifying the main similarities and differences between both groups in these acoustic variables. Then, in the second analysis, I investigated the effects of the age of arrival and length of residence in prosodic accommodation. In particular, I tried to determine if these two social variables can predict the speaker's success in acquiring Paulista prosodic patterns. Finally, an investigation on individual variation was carried out in order to verify if social differences between individual speakers could help explain the patterns of prosodic accommodation.

5.1 Major findings

Alagoan migrants and native Paulistas are similar in most prosodic variables. Contrary to my initial expectations, from the results of the first analysis comparing Alagoan migrants and native Paulistas, we can conclude that, in general, both groups are similar in most prosodic variables analyzed in this study. From the seventeen acoustic variables, significant differences between both groups were found in eight of them, namely: VarcoV, VarcoC, f_0 standard deviation, the standard deviation of f_0 peaks, standard deviation and mean of both positive and negative f_0 change rates. On the other hand, both groups are similar in speech rate, stress group duration, rate of duration peaks, nPVI-V, nPVI-C, correlation coefficient between stress group duration and number of syllables in stress groups, f_0 peak rate, f_0 peak width, and the standard deviation of intervals between f_0 peaks.

The variability in prosodic variables is greater in Alagoan migrants' speech than in native Paulistas. Alagoan migrants scored higher values than Paulistas in the eight variables in which differences between both groups were found. Except for the mean of positive and negative f_0 change rates, the other six variables consist of dispersion measures. Therefore, we can state that, having higher scores in these variables, Alagoan migrants' main difference to native Paulistas is that they show more variability in prosody.

Alagoan migrants and native Paulistas differ less in rhythm than in intonation. In Section 1, we discussed metalinguistic commentaries by Alagoan migrants on differences between São Paulo and Alagoas. In these commentaries, the participants say they recognize there are differences in the prosody of both locations. Interestingly, most of them make reference to rhythmic instead of intonation differences. For instance, RicardoR says that "each one has a rhythm…I have mine…the Baiano has a different way of speaking right? But many Alagoans here [in São Paulo] are already speaking like Paulistas." Speech rate is the rhythmic parameter most mentioned by them as a distinguishing feature between Southeast and Northeast Brazil. In general, Alagoan migrants evaluate Alagoas as a region where people speak faster. YasminS, for example, says that Alagoan accent "is more accelerated". In fact, this study observed a higher estimated speech rate (+0.4 syll/sec) in Alagoan migrants' speech. Besides, the participants of this study who have the fastest speech rates are all Alagoans (FabianoJ, RicardoR, JorgeS, and JoaquimS), not Paulistas. However, the difference in speech rate between Alagoans and Paulistas was not large enough to be considered statistically significant.

In general, we can state that, contrary to the metalinguistic commentaries discussed above, the results of this study show that the most relevant prosodic differences between Alagoan migrants and native Paulistas come from intonation, not rhythm. The only significant rhythmic differences observed concern VarcoV and VarcoC, with Alagoan migrants' having vocalic and consonantal intervals with more variable duration, which suggests a more stress-timed speech. However, the results of VarcoV and VarcoC were not supported by other three rhythm metrics: nPVI-C, nPVI-V, and the correlation coefficient between stress group duration and the number of syllables in stress groups. In these three metrics, both groups showed similar distributions.

Differences between Alagoan migrants and native Paulistas are greater among males. As mentioned in the previous conclusion, major differences were observed in intonation variables. However, an unexpected finding of this study is that these differences are only found when male participants are considered. In other words, while male Alagoan migrants differ from male Paulistas in some intonation variables, female Alagoan migrants do not differ from female Paulistas in any intonation variable. The intonation of male Alagoan migrants has faster and more variable f_0 rises and falls than male Paulistas. Besides, the overall variability of the f_0 contour is also greater among male Alagoans.

In the analysis of the effects of the age of arrival and length of residence on prosodic accommodation, a gender difference is also observed. Interaction between gender and age of arrival was observed in mixed-effects linear models concerning the following six intonation variables: f_0 standard deviation, standard deviation of f_0 peaks, mean and standard deviation of positive and negative f_0 change rates. These six variables are exactly the ones in which male Alagoan migrants significantly differ from male Paulistas. The interaction reveals that there is a significant correlation between age of arrival and each of these six prosodic variables, but only in the case of males.

Age of arrival is a better predictor of prosodic accommodation than length of residence In Section 2.2, we discussed the results of Oushiro (2020b) concerning the effects of age of arrival and length of residence on four segmental phonetic variables in the speech of Alagoan and Paraiban migrants in São Paulo, and we saw that the age of arrival correlated to three of these variables, while the length of residence correlated with only one. In this study, we observed a similar result in the case of the speech prosody of Alagoan migrants. In the six intonation variables in which differences between male Alagoans and male Paulistas were attested, the age of arrival was a statistically significant predictor and contributed to explain the variance in all of them. On the other hand, the length of arrival did not correlate with any of the seventeen prosodic variables analyzed in this study.

In order to evaluate if a social variable is a good predictor of dialect accommodation, only observing correlation between the social variable and the linguistic feature is not enough. It is also necessary to verify if the direction of the correlation is coherent with the accommodation process. In the six intonation variables correlated with age of arrival, the direction of the correlation is coherent: the younger the male Alagoans when they migrated to São Paulo, the closer they were to male Paulistas in these variables. We conclude, therefore, that the age of arrival is a relevant predictor of prosodic accommodation

Individual factors are important to explain prosody in dialect contact This study showed that individual variation in prosody is high, especially in migrants' speech. Most linear models in this study included the speakers as a random variable, and, in most of them, the explanatory power of the models was remarkably improved when the effect of the random variable was considered (marginal R^2). In other words, in predicting the values of prosodic variables, it may be more relevant to know who the speaker that "produced" these values is than knowing if these values were produced by an Alagoan migrant or a native Paulista, a migrant who migrated younger or older, a migrant who had lived few or many years in São Paulo.

It is possible that the importance of one variable comes from one or more related variables. For this reason, I analyzed the social profile of each participant to verify if the effect of the individual on prosodic variable was not due to other social variables, such as social class, level of formal education, social networks, and social identity. However, no evidence of correlation between these social variables and the prosodic parameters was found.

5.2 Suggestions for future investigation

The scientific enterprise works by a domino effect. Each scientific study draws upon the outcomes of previous investigations, addressing the questions that were left unanswered, trying to face challenges arisen in previous studies but not yet overcome, and taking unexplored paths that earlier investigations only reached the entrance. I hope that the unfilled gaps and loose ends of this research be the starting point for exciting new studies, making the row of dominoes of science advance. In what follows, I outline some questions that can be addressed by future investigations.

Social variation and prosodic functions This research examined the prosody in the speech of Alagoan migrants by analyzing the distribution of several acoustic variables in relation to social variables. This approach allows for a general prosodic-acoustic profile of a group of speakers to be obtained and to determine correlations between prosodic-acoustic patterns and social factors. However, it does not take into account the role of these prosodic-acoustic patterns in signaling communicative functions.

Melodic and rhythmic patterns can carry linguistic and pragmatic meanings. For instance, the lengthening of syllabic duration associated with a falling melodic contour can work as a terminal prosodic boundary, a way of signaling the end of an utterance. In a different context, syllable lengthening can be associated with hesitation and speech planning, i.e., a way for the speakers to gain time to better organize what they are going to say. In turn, a high peak in the melodic contour can be a strategy used by the speaker to emphasize a particular element of the speech.

An extension of this work would be to analyze the prosody of internal migrants in Brazil from a functional perspective. Does the acoustic realization of prosodic functions vary between regional varieties? If so, does dialect contact have an effect on how internal migrants use prosody to communicate linguistic and pragmatic meanings? Future investigations can delimitate specific prosodic functions and examine if internal migrants vary in how these functions are acoustically encoded.

Descriptions of regional variation in prosody One challenge that the studies concerned with the effects of dialect contact on speech prosody in Brazilian Portuguese still have to face is the lack of descriptions of prosodic differences between the regions of Brazil. The regional differences in the segmental domain are better known. For example, it is a known fact that many Northeastern varieties favor the fricative pronunciation of coda /r/, while the Paulista variety favors the tap and approximant pronunciations. Supported by this dialectological knowledge, the increase of the approximant pronunciation of coda /r/, for instance, in the speech of Northeastern migrants in São Paulo can be more reliably interpreted as the outcome of an accommodation process. In contrast, the analysis of prosodic accommodation still does not have a solid literature about regional prosodic differences to rely upon. It would be a great contribution that future studies take up the task of identifying and describing which prosodic variables are most relevant in the differentiation of Brazilian regional varieties, both from the production and the perception point of view.

Sociophonetic perception of prosodic features As discussed in Section 4.1, studies concerning speech production should be closely accompanied by studies on speech perception. Perception experiments are important not only to validate the phonetic relevance of production data, by showing if the phonetic differences can be noticed by hearers, but also to help accurately interpret these data. In this work, I have found evidence of prosodic accommodation in the speech of Alagoan migrants, especially in the fundamental frequency change rates. However, I have not examined the perceptual dimension of the acoustic variables. An exciting path for future studies to pursue is to analyze Northeastern migrants' perception of Paulista prosody and compare these perception data with the production data analyzed in this research. Which prosodic parameters in Paulista speech are more salient to the migrants? Do the migrants associate social meanings with specific prosodic patterns? Does the way Northeastern migrants perceive the prosody of their host and home communities helps to explain the production patterns observed in this study? These are some questions worth asking in future investigations.

Bibliography

Abaurre-Gnerre, M. B. M. (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. https://periodicos.sbu.unicamp.br/ojs/index.php/cel/article/view/8636610 (cit. on pp. 43, 44, 91)

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University Press. https://doi.org/10.1515/9781474463775. (Cit. on pp. 39, 43, 93)

Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, *116*(5), 3099–3107. https://doi.org/10.1121/1.1795335 (cit. on p. 96)

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, *40*(3), 351–373. https://doi.org/10.1016/j.wocn.2012.02.003 (cit. on p. 89)

Baeninger, R. (2005). São Paulo e suas migrações no final do século 20. *São Paulo em Perspectiva*, *19*(3), 84–96. https://doi.org/10.1590/s0102-88392005000300008 (cit. on p. 25)

Baeninger, R. (2012). Migrações internas no Brasil no século 21: entre o local e o global. *Anais do XVIII Encontro Nacional de Estudos Populacionais*, 1–24. http://www.abep.org.br/publicacoes/index.php/anais/article/view/1992 (cit. on p. 25)

Barbosa, P. A. (2000). "Syllable-timing in Brazilian Portuguese": uma crítica a Roy Major. *DELTA: Documentação de Estudos em Linguística Teórica e Aplicada*, *16*(2), 369–402. https://doi.org/10/cz36k7 (cit. on pp. 40, 43, 71, 94)

Barbosa, P. A. (2006). *Incursões em torno do ritmo da fala*. Pontes. (Cit. on pp. 38, 60–62, 76, 79).

Barbosa, P. A. (2007). From syntax to acoustic duration: A dynamical model of speech rhythm production. *Speech Communication*, *49*(9), 725–742. https://doi.org/10.1016/j.specom.2007.04.013 (cit. on pp. 60, 62, 63)

Barbosa, P. A. (2009). Measuring speech rhythm variation in a model-based framework. *Proceedings of the 10th Annual Conference of the International Speech Communication Association*, 1527–1530. https://doi.org/10.21437/interspeech.2009-463 (cit. on p. 59)

Barbosa, P. A. (2012). Panorama of experimental prosody research. In H. Mello, M. Pettorino, & T. Raso (Eds.), *Proceedings of the VIIth GSCP International Conference: Speech and Corpora*. Firenze University Press. https://doi.org/10.36253/978-88-6655-351-9. (Cit. on pp. 38, 61)

Barbosa, P. A. (2013). Semi-automatic and automatic tools for generating prosodic descriptors for prosody research. *TRASP 2013 Proceedings*, 86–89. http://www2.lpl-aix.fr/~trasp/ (cit. on pp. 71, 74, 80)

Barbosa, P. A. (2019). *Prosódia*. Parábola Editorial. (Cit. on pp. 38, 39, 62).

Barbosa, P. A., Arantes, P., Meireles, A., & Vieira, J. (2005). Abstractness in speech-metronome synchronisation: P-centres as cyclic attractors. *Proceedings of the 9th INTERSPEECH*. https://doi.org/10.21437/interspeech.2005-512 (cit. on pp. 60, 61)

Barbosa, P. A., & Madureira, S. (2015). *Manual de fonética acústica experimental: Aplicações a dados do português* (1ª edição). Cortez. (Cit. on pp. 38, 75).

Barbosa, P. A., Papa, P. B., Silva, B. A., & Mourão, N. (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. https://doi.org/http://dx.doi.org/10.1590/1678-460X2019350202 (cit. on p. 44)

Barbosa, P. A., Violaro, F., Albano, E., Simões, F., Aquino, P. A., Madureira, S., & Françozo, E. (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. *EUROSPEECH*, 2059–2062. https://www.isca-speech.org/archive/eurospeech_1999/index.html (cit. on p. 63)

Beck, J. M. (2010). Organic variation of the vocal apparatus. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic*

Sciences (pp. 523–602). Wiley-Blackwell. https://doi.org/10.1002/9781444317251.ch5. (Cit. on p. 96)

Blondeau, H. (2013). Studying language over time. In R. Podesva & D. Sharma (Eds.), *Research Methods in Linguistics* (pp. 494–538). Cambridge University Press. https://doi.org/10.1017/cbo9781139013734.025. (Cit. on p. 46)

Boersma, P. (2013). Acoustic analysis. In R. Podesva & D. Sharma (Eds.), *Research Methods in Linguistics* (pp. 375–396). Cambridge University Press. https://doi.org/10.1017/cbo9781139013734.020. (Cit. on p. 75)

Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer* (Version 6.1.41). http://www.praat.org/. (Cit. on p. 76)

Bortoni-Ricardo, S. M. (1985). *The Urbanization of Rural Dialect Speakers: A Sociolinguistic Study in Brazil.* Cambridge University Press. Cambridge. (Cit. on pp. 32, 33, 35, 54).

Britain, D. (1992). Linguistic change in intonation: The use of high rising terminals in New Zealand English. *Language Variation and Change*, *4*(1), 77–104. https://doi.org/10.1017/S0954394500000661 (cit. on pp. 26, 39, 40)

Britain, D. (2002). Diffusion, levelling, simplification and reallocation in past tense BE in the English Fens. *Journal of Sociolinguistics*, *6*(1), 16–43. https://doi.org/10.1111/1467-9481.00175 (cit. on p. 46)

Britain, D. (2013). Space, diffusion and mobility. In J. Chambers & N. Schilling (Eds.), *The Handbook of Language Variation and Change* (pp. 472–500). Wiley-Blackwell. https://doi.org/10.1002/9781118335598.ch22. (Cit. on p. 22)

Britain, D. (2018). Dialect contact and new dialect formation. In C. Boberg, J. Nerbonne, & D. Watt (Eds.), *The handbook of Dialectology* (pp. 143–158). Wiley-Blackwell. https://doi.org/10.1002/9781118827628.ch8. (Cit. on pp. 30, 49)

Cagliari, L. C. (2013). Existem línguas de ritmo silábico? *Estudos Linguísticos*, *42*(1), 19–32. https://revistas.gel.org.br/estudos-linguisticos/article/view/1083 (cit. on p. 43)

Cagliari, L. C., & Abaurre, M. B. (1986). Elementos para uma investigação instrumental das relações entre padrões rítmicos e processos fonológicos no português brasileiro. *Cadernos de Estudos Lingüísticos*, *10*, 39–

57. https://periodicos.sbu.unicamp.br/ojs/index.php/cel/article/view/8636716 (cit. on p. 43)

Chambers, J. K. (1992). Dialect acquisition. *Language*, *68*(4), 673–705. https://doi.org/10.1353/lan.1992.0060 (cit. on pp. 23, 32, 34, 35, 47, 54)

Classe, A. (1939). *Rhythm of English Prose*. Basil Blackwell. (Cit. on pp. 39, 40).

Cruz, M. (2013). *Prosodic variation in European Portuguese: Phrasing, intonation and rhythm in central-southern varieties* (PhD thesis). Universidade de Lisboa. Lisboa. http://hdl.handle.net/10451/9893. (Cit. on p. 91)

Crystal, T. H., & House, A. 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. https://doi.org/10.1121/1.399955 (cit. on p. 61)

Cukor-Avila, P., & Bailey, G. (2013). Real time and apparent time. In J. Chambers & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 237–262). Wiley-Blackwell. https://doi.org/10.1002/9781118335598.ch11. (Cit. on pp. 46–48)

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*(2), 145–171. https://doi.org/10.1006/jpho.1998.0070 (cit. on p. 40)

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for DeltaC. *Language and Language-processing: Proceedings of the 38th Linguistics Colloquium*, 231–241 (cit. on p. 68).

Deterding, D. (1994). The rhythm of Singapore English. *Proceedings of the 5th Australian International Conference on Speech Science and Technology*. https://www.degruyter.com/database/AUSBIB/entry/ausbib.key-1533/html (cit. on p. 40)

Dodsworth, R. (2017). Migration and dialect contact. *Annual Review of Linguistics*, *3*(1), 331–346. https://doi.org/10.1146/annurev-linguistics-011516-034108 (cit. on pp. 23, 97)

Fagyal, Z. (2010). Rhythm types and the speech of working-class youth in a banlieue of Paris: The role of vowel elision and devoicing. In D. Preston & N. Niedzielski (Eds.), *A Reader in Sociophonetics* (pp. 91–

132). De Gruyter Mouton. https://doi.org/10.1515/9781934078068.1.91. (Cit. on p. 42)

Fant, G. (1970). *Acoustic Theory of Speech Production*. Walter de Gruyter. (Cit. on p. 38).

Flege, J. (1995). Second language speech learning: Theory, findings and problems. (Cit. on p. 32).

Flege, J., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5), 3125–3134. https://doi.org/10.1121/1.413041 (cit. on pp. 31, 32)

Fletcher, J. (2010). The prosody of speech: Timing and rhythm. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (pp. 523–602). Wiley-Blackwell. https://doi.org/10.1002/9781444317251.ch15. (Cit. on pp. 37, 38, 40, 43, 61, 67, 87)

Fontes, P. R. R. (2016). *Migration and the making of industrial Sao Paulo*. Duke University Press. https://doi.org/10.1515/9780822374299. (Cit. on p. 24)

Frota, S., & Moraes, J. A. (2016). Intonation in European and Brazilian Portuguese. In L. Wetzels, J. Costa, & S. Menuzzi (Eds.), *The Handbook of Portuguese Linguistics* (pp. 141–166). Wiley-Blackwell. https://doi.org/10.1002/9781118791844.ch9. (Cit. on p. 44)

Frota, S., & Vigário, M. (2000). Aspectos de prosódia comparada: ritmo e entoação no PE e no PB. *Actas do XV Encontro da Associação Portuguesa de Linguística*, 533–555 (cit. on pp. 40, 43, 44, 91).

Fuchs, R. (2016). *Speech Rhythm in Varieties of English*. Springer. https://doi.org/10.1007/978-3-662-47818-9. (Cit. on pp. 39, 40, 69, 89, 91)

Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics*, *15*(2), 87–105. http://www.jstor.org/stable/30029508 (cit. on pp. 23, 33)

Giles, H., Coupland, J., & Coupland, N. (Eds.). (1991). *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511663673. (Cit. on p. 23)

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland,

& N. Coupland (Eds.), *Contexts of Accommodation: Developments in Applied Sociolinguistics* (pp. 1–68). Cambridge University Press. https://doi.org/10.1017/cbo9780511663673.001. (Cit. on p. 123)

Giles, H., Taylor, D. M., & Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in Society*, *2*(02), 177–192. https://doi.org/10.1017/S0047404500000701 (cit. on p. 23)

Gorman, K., Howell, J., & Wagner, M. (2011). Prosodylab-Aligner: A Tool for Forced Alignment of Laboratory Speech. Canadian Acoustics. 39.3. 192–193. *Canadian Acoustics.*, *39*(3), 192–193. (Cit. on p. 79).

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 515–546). De Gruyter Mouton. https://doi.org/10.1515/9783110197105. (Cit. on pp. 40, 70)

Gries, S. T. (2013). *Statistics for Linguistics with R: A Practical Introduction*. De Gruyter Mouton. https://doi.org/10.1515/9783110718256. (Cit. on pp. 51, 95)

Guedes, S. (2019). Emprego do artigo definido em situação de contato dialetal. *Domínios de Lingu@gem*, *13*(4), 1401–1432. https://doi.org/10.14393/DL40-v13n4a2019-4 (cit. on p. 25)

Guy, G., Oushiro, L., & Mendes, R. B. (2022). Indexicality and coherence. In *The Coherence of Linguistic Communities* (1st ed., pp. 53–68). Routledge. https://doi.org/10.4324/9781003134558-5. (Cit. on p. 25)

Guy, G., & Zilles, A. M. S. (2007). *Sociolinguística quantitativa: Instrumental de análise*. Parábola Editorial. (Cit. on p. 55).

Hirst, D., & Di Cristo, A. (1998). A survey of intonation systems. *Intonation systems: A survey of twenty languages*, *144* (cit. on pp. 26, 38).

Hoffman, M. F., & Walker, J. A. (2010). Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change*, *22*(1), 37–67. https://doi.org/10.1017/S0954394509990238 (cit. on p. 57)

Ide, N., & Pustejovsky, J. (Eds.). (2016). *Handbook of Linguistic Annotation*. Springer. https://doi.org/10.1007/978-94-024-0881-2. (Cit. on p. 75)

IOM. (2019). *Glossary on Migration*. International Organization for Migration. (Cit. on p. 53).

IPEA. (2011). *Perfil dos migrantes em São Paulo* (No. 115). Instituto de Pesquisa Econômica Aplicada. (Cit. on p. 24).

Johnson, K. (2012). *Acoustic and Auditory Phonetics* (3rd ed.). Wiley-Blackwell. (Cit. on p. 38).

Jones, D. (1918). *An outline of English Phonetics*. W. Heffner. (Cit. on p. 39).

Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (2nd ed. edição). Prentice Hall. https://home.cs.colorado.edu/~martin/slp.html. (Cit. on p. 78)

Kendall, T. (2013). *Speech Rate, Pause and Sociolinguistic Variation: Studies in Corpus Sociophonetics*. Palgrave Macmillan. https://doi.org/10.1057/9781137291448. (Cit. on p. 67)

Kerswill, P. (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. https://doi.org/10.1111/j.1473-4192.1993.tb00042.x (cit. on pp. 30, 31)

Kruse, J. S., & Barbosa, P. A. (2021). Alinha-PB: A Phonetic Aligner for Brazilian Portuguese. *Journal of Communication and Information Systems*, *36*(1), 192–199. https://doi.org/10.14209/jcis.2021.21 (cit. on p. 79)

Labov, W. (2006). *The Social Stratification of English in New York City*. Cambridge University Press. https://doi.org/10.1017/CBO9780511618208. (Cit. on pp. 31, 34)

Ladd, D. R. (2008). *Intonational Phonology*. Cambridge University Press. https://doi.org/10.1017/CBO9780511808814. (Cit. on p. 44)

Ladefoged, P. (1996). *Elements of Acoustic Phonetics* (2nd ed.). University of Chicago Press. https://doi.org/10.7208/chicago/9780226191010.001.0001. (Cit. on pp. 38, 75)

Lang, S. (2002). *Short Calculus*. Springer New York. https://doi.org/10.1007/978-1-4613-0077-9. (Cit. on p. 74)

Laver, J. (1994). *Principles of Phonetics*. Cambridge University Press. https://doi.org/10.1017/CBO9781139166621. (Cit. on p. 67)

Lehiste, I. (1970). *Suprasegmentals*. MIT Press. (Cit. on pp. 38, 60).

Lenneberg, E. H. (1967). *Biological Foundations of Language*. John Wiley & Sons. (Cit. on pp. 31, 35).

Liberman, M. Y. (2019). Corpus Phonetics. *Annual Review of Linguistics*, *5*(1), 91–107. https://doi.org/10.1146/annurev-linguistics-011516-033830 (cit. on p. 78)

Lobanov, B. M. (1971). Classification of russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, *49*, 606–608. https://doi.org/10.1121/1.1912396 (cit. on p. 96)

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, *43*(4), 377–401. https://doi.org/10.1177/00238309000430040301 (cit. on pp. 41, 69, 70)

Major, R. C. (1981). Stress-timing in Brazilian Portuguese. *Journal of Phonetics*, *9*(3), 343–351. https://doi.org/10.1016/S0095-4470(19)30977-5 (cit. on p. 43)

Major, R. C. (1985). Stress and Rhythm in Brazilian Portuguese. *Language*, *61*(2), 259–282. https://doi.org/10.2307/414145 (cit. on p. 43)

Meireles, A. R., & Barbosa, P. A. (2009). O papel da taxa de elocução nos processos dinâmicos de mudança linguística. *Revista (Con)textos Linguísticos*, *3*, 91–116. https://periodicos.ufes.br/contextoslinguisticos/article/view/5141 (cit. on pp. 84, 85, 105, 108)

Mendes, R. B., & Oushiro, L. (2012). O paulistano no mapa sociolinguístico brasileiro. *ALFA: Revista de Linguística*, *56*(3), 973–1001. https://doi.org/10.1590/S1981-57942012000300011 (cit. on p. 58)

Mendonça, C. S. I. de. (2003). A sílaba em fonologia. *Working Papers em Linguística*, *7*(1), 21–40. https://periodicos.ufsc.br/index.php/workingpapers/article/view/6165 (cit. on p. 59)

Migliorini, L., & Massini-Cagliari, G. (2010). Sobre o ritmo do português brasileiro: Evidências de um padrão acentual. *Revista Virtual de Estudos da Linguagem*, *8*(15), 310–328. http://hdl.handle.net/11449/124843 (cit. on p. 43)

Milroy, L. (1980). *Language and social networks*. Wiley-Blackwell. (Cit. on p. 32).

Milroy, L., & Gordon, M. (2003). *Sociolinguistics: Method and Interpretation*. Blackwell Publishing. https://doi.org/10.1002/9780470758359. (Cit. on pp. 50, 51)

Moraes, J. A., & Leite, Y. (2002). Ritmo e velocidade da fala na estratégia do discursos: Uma proposta de trabalho. In R. Ilari (Ed.), *Gramática do português falado II: Níveis de análise linguística* (4th ed.). Editora da Unicamp. (Cit. on p. 43).

Mourão, N. R. (2018). *Mobilidade e difusão linguística no interior de São Paulo: uma análise de Campinas e Jundiaí* (Projeto de mestrado). Instituto de Estudos da Linguagem/UNICAMP. (Cit. on pp. 50, 57).

Nespor, M. A., & Vogel, I. (2007). *Prosodic phonology*. Mouton de Gruyter. https://doi.org/10.1515/9783110977790. (Cit. on p. 44)

Nokes, J., & Hay, J. (2012). Acoustic correlates of rhythm in New Zealand English: A diachronic study. *Language Variation and Change*, *24*(1), 1–31. https://doi.org/10.1017/S0954394512000051 (cit. on pp. 26, 41)

Nycz, J. (2013). New contrast acquisition: Methodological issues and theoretical implications. *English Language and Linguistics*, *17*(2), 325–357. https://doi.org/10.1017/S1360674313000051 (cit. on pp. 31, 49)

Nycz, J. (2015). Second dialect acquisition: A sociophonetic perspective. *Language and Linguistics Compass*, *9*(11), 469–482. https://doi.org/10.1111/lnc3.12163 (cit. on pp. 22, 46)

Oliveira, M. A. (2020). *Dialetos em contato: Acomodação dialetal por migrantes baianos habitantes da cidade de Bauru, São Paulo* (Dissertação de mestrado). UNESP. http://hdl.handle.net/11449/193044. (Cit. on p. 37)

Oushiro, L. (2007). *Introdução à estatística para linguistas*. https://rpubs.com/oushiro/iel. (Cit. on p. 95)

Oushiro, L. (2016a). *Processos de acomodação dialetal na fala de nordestinos residentes no estado de São Paulo* (Projeto de pesquisa regular - FAPESP). Universidade Estadual de Campinas. Campinas. (Cit. on pp. 31, 35).

Oushiro, L. (2016b). Social and structural constraints in lectal cohesion. *Lingua*, *172-173*, 116–130. https://doi.org/10.1016/j.lingua.2015.10.015 (cit. on p. 49)

Oushiro, L. (2018). *Processos de acomodação dialetal na fala de nordestinos residentes em São Paulo* (Relatório científico final - FAPESP (Processo nº 2016/04960-7)). Universidade Estadual de Campinas. Campinas. (Cit. on pp. 22, 25, 49–51, 56, 57).

Oushiro, L. (2019). Linguistic uniformity in the speech of Brazilian internal migrants in a dialect contact situation. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*. International Phonetic Association. https://www.internationalphoneticassociation.org/icphs/icphs2019. (Cit. on pp. 44, 127)

Oushiro, L. (2020a). As variáveis sexo/gênero e indivíduo em situação de contato dialetal. In D. Carvalho & D. Brito (Eds.), *Gênero e língua(gem): formas e usos*. EDUFBA. (Cit. on pp. 25, 35).

Oushiro, L. (2020b). Contrasting age of arrival and length of residence in dialect contact. *Selected Papers from New Ways of Analyzing Variation 47*, *25*(2), 79–88. https://repository.upenn.edu/pwpl/vol25/iss2/ (cit. on pp. 25, 35, 36, 47, 54, 56, 137, 142)

Oushiro, L. (2020c). Múltiplas variáveis na fala de nordestinos residentes em São Paulo. In C. Brescancini & V. N. d. O. Monaretto (Eds.), *Sociolinguística no Brasil: Textos selecionados*. EdiPUCRS. (Cit. on pp. 25, 26, 35, 97).

Oushiro, L. (2022). *Dialect Contact in Lusophone Communities* [Manuscript submitted for publication]. (Cit. on pp. 23, 25).

Payne, A. (1976). *The acquisition of the phonological system of a second dialect* (Tese de doutorado). Universidade da Pennsylvania. (Cit. on pp. 23, 47, 54).

Pike, K. L. (1945). *The Intonation of American English*. University of Michigan Press. (Cit. on pp. 39, 43).

Podesva, R. (2006). Intonational variation and social meaning: Categorical and phonetic aspects. *University of Pennsylvania Working Papers in Linguistics*, *12*(2), 189–202. https://repository.upenn.edu/pwpl/vol12/iss2/15/ (cit. on p. 39)

Pompino-Marschall, B. (1989). On the psychoacoustic nature of the P-center phenomenon. *Journal of Phonetics*, *17*(3), 175–192. https://doi.org/10.1016/S0095-4470(19)30428-0 (cit. on pp. 60, 61)

Quené, H. (2007). On the just noticeable difference for tempo in speech. *Journal of Phonetics*, *35*(3), 353–362. https://doi.org/10.1016/j.wocn.2006.09.001 (cit. on p. 86)

Ramus, F., Nespor, M. A., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292. https://doi.org/10.1016/S0010-0277(00)00101-3 (cit. on pp. 40, 43, 68, 89, 91)

Renals, S., & King, S. (2012). Automatic speech recognition. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (2nd ed.). Wiley-Blackwell. https://doi.org/10.1002/9781444317251.ch22. (Cit. on p. 78)

Rubin, P., & Vatikiotis-Bateson, E. (1998). Measuring and Modeling Speech Production. In S. L. Hopp, M. J. Owren, & C. S. Evans (Eds.), *Animal Acoustic Communication: Sound Analysis and Research Methods* (pp. 251–290). Springer. https://doi.org/10.1007/978-3-642-76220-8_8. (Cit. on p. 38)

Sagisaka, Y., Campbell, N., & Higuchi, N. (2012). *Computing Prosody: Computational Models for Processing Spontaneous Speech*. Springer. https://doi.org/10.1007/978-1-4612-2258-3. (Cit. on p. 38)

Santana, A. d. L. (2018). *As vogais médias pretônicas na fala de sergipanos em São Paulo* (Dissertação de mestrado). Universidade de São Paulo. São Paulo. https://doi.org/10.11606/D.8.2018.tde-23112018-121948. (Cit. on pp. 36, 127)

Santana, A. d. L., & Beline Mendes, R. (2019). Sergipanos em São Paulo: Redes sociais, contato dialetal e pronúncia variável de vogais médias pretônicas. *Revista LinguíStica*, *15*(2), 54–85. https://doi.org/10.31513/linguistica.2019.v15n2a21392 (cit. on p. 36)

Sene, M. G. (2021). *Efeitos de variáveis fonéticas na percepção sociolinguística de gênero e sexualidade* (Relatório de qualificação de mestrado). Universidade Estadual Paulista - Júlio de Mesquita Filho. (Cit. on p. 39).

Siegel, J. (2010). *Second Dialect Acquisition*. Cambridge University Press. https://doi.org/10.1017/CBO9780511777820. (Cit. on pp. 22, 24, 48, 54, 97)

Silva Jr, L., & Barbosa, P. A. (2019). *Metrics and acoustics extractor* (Version 1.7.7). https://github.com/leonidasjr/SpeechRhythmCode. (Cit. on p. 80)

Souza, E. S. de. (2019). O uso variável do imperativo de migrantes baianos em São Paulo. *Domínios de Lingu@gem*, *13*(4), 1433–1464. https://doi.org/10.14393/DL40-v13n4a2019-5 (cit. on pp. 25, 36)

Speaks, C. E. (1992). *Introduction to Sound*. Springer. https://doi.org/10.1007/978-1-4899-7196-8. (Cit. on p. 38)

Stevens, K. N. (2000). *Acoustic Phonetics*. MIT Press. (Cit. on p. 38).

Sudhoff, S. (Ed.). (2006). *Methods in Empirical Prosody Research*. De Gruyter Mouton. (Cit. on p. 38).

Szakay, A. (2006). Rhythm and pitch as markers of ethnicity in New Zealand English. In P. Warren & C. Watson (Eds.), *Proceedings of the 11th Australian International Conference on Speech Science & Technology* (pp. 421–426). Australian Speech Science & Technology Association Inc. https://assta.org/proceedings/sst/2006/. (Cit. on p. 41)

Szakay, A. (2008). *Ethnic dialect identification in New Zealand - the role of prosodic cues*. VDM Verlag Dr. Mueller E.K. (Cit. on p. 41).

Tagliamonte, S. (2006). *Analysing Sociolinguistic Variation*. Cambridge University Press. https://doi.org/10.1017/CBO9780511801624. (Cit. on p. 51)

Tagliamonte, S., & Molfenter, S. (2007). How'd you get that accent?: Acquiring a second dialect of the same language. *Language in Society*, *36*(5), 649–675. https://doi.org/10.1017/S0047404507070911 (cit. on p. 47)

Tenani, L. (2006). Domínios prosódicos no português brasileiro: Evidências rítmica, entoacional e segmental. *Estudos Linguísticos*, *35*, 118–131 (cit. on p. 44).

The Language Archive. (2022). *Elan* (Version 6.3). https://archive.mpi.nl/tla/elan. (Cit. on p. 76)

Thomas, E., & Carter, P. (2006). Rhythm and African American English. *English World-Wide*, *27*(3), 331–355. https://doi.org/10.1075/eww.27.3.06tho (cit. on pp. 26, 41)

Torgersen, E. N., & Szakay, A. (2012). An investigation of speech rhythm in London English. *Lingua*, *122*(7), 822–840. https://doi.org/10.1016/j.lingua.2012.01.004 (cit. on p. 42)

Troncoso-Ruiz, A., & Elordieta, G. (2018). Prosodic accommodation and salience: The nuclear contours of Andalusian Spanish speakers in

Asturias. *Loquens*, *4*(2), 043. https://doi.org/10.3989/loquens.2017.043 (cit. on pp. 26, 42)

Trudgill, P. (1986). *Dialects in Contact*. Basil Blackwell. (Cit. on pp. 22, 23, 32–35, 42, 47, 54).

Walker, A. J. (2014). *Crossing oceans with voices and ears: Second dialect acquisition and topic-based shifting in production and perception* (PhD thesis). The Ohio State University Press. Ohio. http://rave.ohiolink.edu/etdc/view?acc_num=osu1397802092. (Cit. on p. 47)

White, L., & Mattys, S. L. (2007). Rhythmic typology and variation in first and second languages. In P. Prieto, J. Mascaró, & M.-J. Solé (Eds.), *Current Issues in Linguistic Theory* (pp. 237–257). John Benjamins Publishing Company. https://doi.org/10.1075/cilt.282.16whi. (Cit. on p. 89)

Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. Routledge. https://doi.org/10.4324/9781315165547. (Cit. on p. 88)

Young, S., Evermann, G., Gales, M., Hain, T., Kershwa, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., Ragui, A., Valtchev, V., Woodland, P., & Zhang, C. (2015). *The HTK Book* (Version 3.5). Cambridge University Engineering Department. https://htk.eng.cam.ac.uk/docs/docs.shtml. (Cit. on p. 79)

Zec, D. (2007). The syllable. In P. V. De Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 161–194). Cambridge University Press. https://doi.org/10.1017/cbo9780511486371.009. (Cit. on p. 59)