



ANA CAROLINA CONSTANTINI

**CARACTERIZAÇÃO PROSÓDICA DE SUJEITOS DE
DIFERENTES VARIEDADES DE FALA DO PORTUGUÊS
BRASILEIRO EM DIFERENTES RELAÇÕES SINAL-RUÍDO**

CAMPINAS,

2014



**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE ESTUDOS DA LINGUAGEM**

ANA CAROLINA CONSTANTINI

**CARACTERIZAÇÃO PROSÓDICA DE SUJEITOS DE
DIFERENTES VARIEDADES DE FALA DO PORTUGUÊS
BRASILEIRO EM DIFERENTES RELAÇÕES SINAL-RUÍDO**

**Tese de doutorado apresentada ao Instituto de
Estudos da Linguagem da Universidade Estadual
de Campinas para obtenção do título de Doutora
em Linguística.**

Orientador (a): Prof(a). Dr(a). Plinio Almeida Barbosa

CAMPINAS,

2014

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Estudos da Linguagem
Crisllene Queiroz Custódio - CRB 8/8624

C766c Constantini, Ana Carolina, 1985-
Caracterização prosódica de sujeitos de diferentes variedades de fala do português brasileiro em diferentes relações sinal-ruído / Ana Carolina Constantini. – Campinas, SP : [s.n.], 2014.

Orientador: Plínio Almeida Barbosa.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Estudos da Linguagem.

1. Fonética forense. 2. Razão sinal-ruído. 3. Dialeto - Fonética. I. Barbosa, Plínio Almeida, 1966-. II. Universidade Estadual de Campinas. Instituto de Estudos da Linguagem. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Prosodic characterization of subjects from different Brazilian Portuguese varieties in different signal-to-noise ratio

Palavras-chave em inglês:

Forensic phonetics

Signal-to-noise ratio

Dialects - Phonetics

Área de concentração: Linguística

Titulação: Doutora em Linguística

Banca examinadora:

Plínio Almeida Barbosa [Orientador]

Vandersi Sant'Ana Castro

Maria Filomena Spatti Sândalo

Sandra Madureira Fontes

Zuleica Antonia de Camargo

Data de defesa: 08-05-2014

Programa de Pós-Graduação: Linguística

BANCA EXAMINADORA:

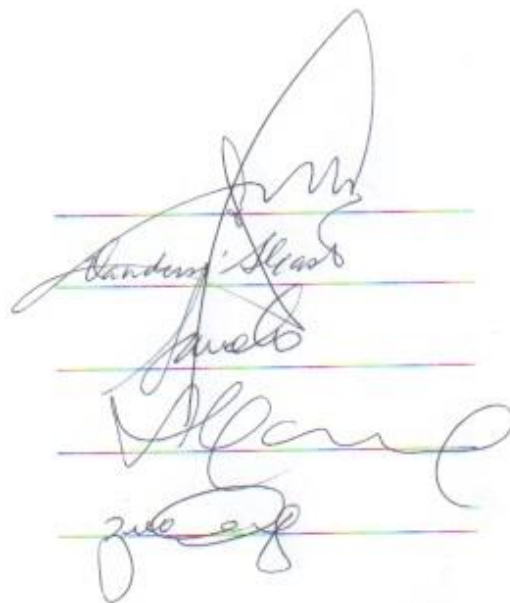
Plínio Almeida Barbosa

Vandersi Sant'Ana Castro

Maria Filomena Spatti Sândalo

Sandra Madureira Fontes

Zuleica Antonia de Camargo



Emílio Gozze Pagotto

Pablo Arantes

Cláudia Regina Brescancini

IEL/UNICAMP
2014

Ao Felipe, por estar comigo onde quer que eu esteja.

MEUS AGRADECIMENTOS

A Deus e aos santos aos quais sou devota, pela força em momentos difíceis e por todas as bênçãos recebidas.

Ao Prof. Plinio, por toda sua generosidade, por estar sempre lado a lado e defender este trabalho com todas as suas forças. Muito obrigada por todos os ensinamentos nesses seis anos de trabalho e por todo o suporte, tanto em situações da vida acadêmica, quanto em situações da vida fora da Unicamp.

A todos os professores do IEL, por contribuírem tão grandemente para a minha formação na Linguística, desde a minha graduação, até o Doutorado.

Aos meus colegas queridos do Grupo de Estudos de Prosódia da Fala, pelas experiências trocadas, angústias compartilhadas, viagens a congressos e convivência do dia-a-dia.

À Polícia Federal de Brasília, por confiar e acreditar no meu trabalho, cedendo seus dados para que este trabalho pudesse ser realizado

To Prof. Jose Hualde and Prof. Jennifer Cole, for receiving me at University of Illinois at Urbana-Champaign and to their contribution to my thesis.

Ao Felipe, por ser o meu companheiro em todos os sentidos da palavra: aquele com quem divido felicidade, sonhos, idéias, vitórias, frustrações e principalmente, a vida.

Aos meus pais, Olaerte e Soraia, à minha irmã Ana Claudia e a minha prima Célia, por sempre acreditarem em mim e me apoiarem incondicionalmente.

Às minhas amigas que me acompanham desde a faculdade: Mayla, Sabrina, Bebel, Pati, Larissa, Aline e Myrian e aos meus amigos que me acompanham a vida toda: Rodolfo, Luiza, Paulo, Dudu, Larissa, Karine, Rafael e Nara.

À FAPESP, por financiar meus estudos desde a iniciação científica e apoiar minha ida como estudante visitante para os Estados Unidos.

A banca examinadora, por aceitar participar desta etapa tão importante em minha vida acadêmica e contribuir com meu trabalho.

A todas as pessoas queridas que me rodeiam, meus familiares, sogros, cunhados e amigos, que sempre torcem por mim e ficam felizes a cada vitória.

“They did not know it was impossible, so they did it.”

(Mark Twain)

RESUMO

A prosódia é uma informação fônica que está além do nível do segmento, e é usualmente estudada a partir da análise de três parâmetros fonético-acústicos clássicos: frequência fundamental, intensidade e duração. Embora estudada para muitas finalidades, a prosódia geralmente não é a primeira opção de investigação quando se busca conhecer mais sobre diferenças entre variedades de uma mesma língua, por exemplo. Desta forma, o presente trabalho pretende preencher essa lacuna no que diz respeito aos estudos prosódicos para caracterizar e diferenciar variedades faladas no Brasil.

O objetivo desta tese de Doutorado foi estudar parâmetros prosódicos que pudessem caracterizar e posteriormente diferenciar sujeitos de diferentes variedades faladas do português brasileiro. Em um segundo momento, ruído aditivo foi incluído nas mesmas amostras de fala utilizadas para caracterizar a prosódia de diferentes variedades do português brasileiro, com o objetivo de entender melhor como os parâmetros prosódicos se comportam quando há inclusão de ruído nas amostras de fala, situação muito comum na área da Fonética Forense. O objetivo secundário da pesquisa foi aplicar testes perceptivos a ouvintes do português brasileiro com a finalidade de saber se eles seriam capazes de reconhecer e categorizar a origem dos falantes de acordo com suas falas.

Analisamos amostras de fala espontânea de 35 sujeitos, do sexo masculino, de sete regiões brasileiras: São Paulo, Minas Gerais, Rio de Janeiro, Paraná, Distrito Federal, Região Nordeste e Região Norte. Todas as amostras de fala foram segmentadas em unidades Vogal-Vogal (unidade VV), unidades do tamanho da sílaba que compreendem um segmento que vai do início de uma vogal até o início da vogal imediatamente seguinte, incluindo as consoantes entre elas. O script BeatExtractor foi utilizado para este fim. Posteriormente, outro script

(ProsodicDescriptorExtractor) foi executado para extrair oito medidas prosódico-acústicas das amostras de fala: taxa de elocução (unidades VV/s), média de z-score suavizado de duração de unidade VV, desvio-padrão de z-score suavizado de duração de unidade VV, assimetria de z-score suavizado de duração de unidade VV, taxa de proeminência (picos de z-score/s), mediana de frequência fundamental, ênfase espectral e taxa de unidades VV não proeminentes por segundo.

Após a análise estatística, os resultados mostraram que cinco dos oito parâmetros conseguiram identificar ao menos uma variedade estudada e assim, diferenciá-la de outras. A mediana de F0 e a ênfase espectral foram capazes de criar dois grandes grupos que separaram DF e Região Norte de todas as outras variedades (exceto pela não diferenciação de DF e Paraná), mostrando que DF e Norte possuem valores maiores de ênfase espectral, bem como têm valores de F0 maiores que os falantes de outras variedades. Assimetria de z-score suavizado e taxa de unidades VV não proeminentes/s foram os parâmetros que colocaram DF e Norte em grupos diferentes. O desvio-padrão de z-score apontou para uma diferença entre dialetos falados na região Norte do Brasil e da Região Sul: a região Norte se diferenciou de SP, DF e Nordeste e SP, que, por sua vez, se diferenciou do PR. Concluímos, portanto, que os parâmetros prosódicos podem revelar características próprias de variedades faladas no Brasil.

A análise das amostras de fala em diferentes relações sinal-ruído mostrou que mediana de F0 e ênfase espectral são os parâmetros que sofrem maior perturbação quando a relação sinal-ruído é baixa, sendo que os valores de ênfase espectral chegaram a sofrer mudanças de 154% em relação a seus valores originais. O resultado mostrou que a análise da estrutura rítmica é a mais robusta quando há presença de ruído nas amostras de fala.

Por fim, os testes perceptivos foram aplicados em 20 falantes do português brasileiro e a variedade mais reconhecida foi a variedade falada no Rio de Janeiro, que chegou a apresentar 90% de acerto, seguida pela variedade falada no Nordeste do Brasil. Constatamos que a proximidade da região de origem dos ouvintes com a região da variedade presente no teste facilita a identificação correta da variedade.

Palavras-chave: Fonética forense; Dialeto – Fonética; Razão sinal-ruído

ABSTRACT

Prosody is usually studied by means of three classic parameters: fundamental frequency, intensity and duration. As for as dialectology is concerned, prosody has not been the main focus of the research on different dialects. Our goal is to characterize and differentiate Brazilian Portuguese varieties using prosodic parameters. In order to do that, we analyzed the recordings of spontaneous speech from 35 male subjects from seven different Brazilian regions: São Paulo (SP), Minas Gerais (MG), Rio de Janeiro (RJ), Paraná (PR), Distrito Federal (DF), Northeast (NE) and North (N). The speech samples were segmented in Vowel-to-Vowel units (VV units) using the BeatExtractor script. Later, the ProsodicDescriptorExtractor script was used to extract eight prosodic measures which are: speech rate (VV units/s), mean, standard deviation and skewness of the normalized z-score, prominence rate (peaks of z-score/s), median of fundamental frequency, spectral emphasis and rate of non prominent VV units/s.

The statistical analysis revealed that five prosodic parameters were able to identify at least one variety and then differentiate it from the others. Fundamental frequency median and spectral emphasis created two groups which separated N and DF (DF is located at West-Central region, near North region) from all the other varieties, considering that N and DF were characterized by high values of these two parameters. On the other hand, skewness of z-score and rate of non prominent VV units/s set DF and N in different groups. Standard deviation of z-score pointed to differences between North varieties and South varieties. We concluded that prosodic parameters can be useful to differentiate Brazilian Portuguese varieties.

Another goal of the current study was to analyze the spontaneous speech ´recordings in distinct signal-to-noise ratios. The analysis has shown that Gaussian, additive noise modifies the values for median of F0 and spectral emphasis (the least has changed 154% related to the

original values). The results revealed that the rhythmic organization of the speech chain is more indicated to the analysis of acoustic parameters in the presence of noise.

Finally, 20 listeners were recruited to answer a perceptual test (free classification test) about the different varieties spoken in Brazil (we used the same spontaneous speech recordings to run the perceptual test). Rio de Janeiro was the most recognized variety, which presented 90% of correct answers, followed by the NE variety. The closeness of the listeners' own origin to the regions of the spoken varieties contributed to correct identifications.

Keywords: Forensic Phonetics; Dialects – Phonetics; Signal-to-noise ratio

LISTA DE TABELAS

Tabela 1: Números de sujeitos por região estudada _____	22
Tabela 2: Resultados obtidos para os parâmetros prosódico-acústicos estudados _____	41
Tabela 3: Valores de médias e desvio-padrão de mediana de frequência fundamental para variedades estudadas _____	45
Tabela 4: Resultados do teste Duncan para o parâmetro ênfase espectral dividindo em grupos as variedades estudadas _____	51
Tabela 5: Resultados do teste Duncan para mediana de frequência fundamental dividindo em grupos as variedades estudadas _____	52
Tabela 6: Resultados do teste Duncan para desvio-padrão de z-score suavizado de duração de unidade VV dividindo em grupos as variedades estudadas _____	52
Tabela 7: Resultados do teste Duncan para o parâmetro assimetria de z-score suavizado dividindo em grupos as variedades estudadas _____	53
Tabela 8: Resultados do teste Duncan para o parâmetro taxa de unidades VV não proeminentes/s dividindo em grupos as variedades estudadas _____	54
Tabela 9: Resultados da LDA apontando a porcentagem de acerto da técnica para cada variedade estudada _____	54
Tabela 10: Resultados da LDA utilizando cada parâmetro individualmente para medir o acerto da técnica para cada variedade estudada _____	55

Tabela 11: Resultados do Teste T para as análises das diferentes relações sinal-ruído estudadas _____	57
Tabela 12: Médias de ênfase espectral e de mediana de frequência fundamental e magnitudes de mudança após adição de ruído _____	58
Tabela 13: Resultados do teste perceptivo _____	59
Tabela 14: Porcentagem de acerto dos ouvintes no teste perceptivo para cada variedade estudada _____	60
Tabela 15: Respostas obtidas pelos participantes nos trechos de fala de sujeitos do DF ____	61
Tabela 16: Respostas obtidas pelos participantes nos trechos de fala de sujeitos do PR ____	62
Tabela 17. Respostas obtidas pelos participantes nos trechos de fala de sujeitos da região Norte _____	62

LISTA DE FIGURAS

Figura 1: Divisão dialetal do território brasileiro proposta por Nascentes	6
Figura 2. Divisão dialetal do território brasileiro proposta por Nascentes ilustrada por Noll (2008), com a presença da linha imaginária que Nascentes utilizou para dividir o território em dois grupos	7
Figura 3: Densidade demográfica do território brasileiro segundo o Censo 2010	25
Figura 4: Trecho “ <i>equipamento é de transmissão</i> ” segmentado em unidades VV	27
Figura 5: Traçado da forma da onda (janela superior) e espectrograma de banda estreita (janela inferior) da gravação original	34
Figura 6: Traçado da forma da onda (janela superior) e espectrograma de banda estreita (janela inferior) com adição de ruído de magnitude 0.01	34
Figura 7: Traçado de forma da onda (janela superior) e espectrograma de banda estreita (janela inferior) adição de ruído de magnitude 0.02	35
Figura 8. Detalhe do espectrograma de banda larga com adição de ruído de magnitude 0.02	35
Figura 9: Esquema representando a divisão das variedades estudadas de acordo com os resultados obtidos para o parâmetro ênfase espectral após utilização do teste post-hoc Tukey HSD	42

Figura 10: Boxplot mostrando os valores obtidos para o parâmetro ênfase espectral (emph) para as variedades estudadas _____ 43

Figura 11: Esquema representando a divisão das variedades estudadas de acordo com os resultados obtidos para o parâmetro mediana de frequência fundamental, após utilização do teste post-hoc Tukey HSD _____ 44

Figura 12. Boxplot mostrando os valores obtidos para o parâmetro mediana de frequência fundamental para as variedades estudadas _____ 45

Figura 13: Boxplot mostrando os valores obtidos para o parâmetro assimetria de z-score suavizado nas variedades estudadas _____ 47

Figura 14: Boxplot mostrando os valores obtidos para taxa de unidades VV não proeminentes/s para as variedades estudadas _____ 48

Figura 15: Esquema representando a divisão das variedades estudadas de acordo com os resultados obtidos para o parâmetro desvio-padrão de z-score suavizado, após utilização do teste post-hoc Tukey HSD _____ 49

Figura 16: Boxplot mostrando os valores obtidos para o desvio-padrão de z-score suavizado para as variedades estudadas _____ 50

Figura 17. Partitura da fala carioca em ilustração retirada de Marroquim (1934) _____ 69

Figura 18. Partitura da fala alagoana em ilustração retirada de Marroquim (1934) _____ 69

LISTA DE QUADROS

Quadro 1: Medidas acústicas analisadas por meio do *script* ProsodicDescriptorExtractor ____ 29

Quadro 2: Local de nascimento e idade dos sujeitos que responderam ao teste perceptivo _ 38

Quadro 3: Síntese dos resultados obtidos em função da análise dos parâmetros prosódico-acústicos _____ 71

SUMÁRIO

1. Introdução	1
1.1. Motivação para o estudo	1
1.2. Variedade e dialeto	2
1.3. Os estudos dialetológicos no Brasil	4
1.4. A prosódia na classificação dialetal	11
1.5. Impacto do ruído na extração das medidas prosódico-acústicas	14
2. Objetivo	19
2.1. Objetivo suplementar	19
3. Metodologia	21
3.1. Sujeitos	21
3.2. Tratamento dos dados	25
3.3. Parâmetros fonético-acústicos analisados	27
3.4. Análise estatística	31
3.5. Análise das diferentes relações sinal-ruído	32
3.5.1. Análise estatística	36
3.6. Testes perceptivos	37
3.6.1. Ouvintes	37
3.6.2. Amostras de fala utilizadas	38
3.6.3. Procedimento realizado	39
4. Resultados	41
4.1. Parâmetros prosódico-acústicos	41

4.2. Diferentes relações sinal-ruído	56
4.3. Testes perceptivos	58
5. Discussão	65
5.1. Parâmetros prosódico-acústicos na identificação de diferentes dialetos brasileiros	65
5.2. Análise das diferentes relações sinal-ruído	73
5.3. Testes perceptivos	75
6. Conclusão	81
7. Referências bibliográficas	83
ANEXO	87

1. Introdução

1.1. Motivação para o estudo

Esta tese de doutorado teve como objetivo estudar parâmetros prosódicos que pudessem caracterizar a fala de sujeitos de diferentes localidades brasileiras, agrupando-os de acordo com as suas regiões de origem, utilizando a análise prosódica para a diferenciação de variedades do português brasileiro. As amostras de fala dos sujeitos também foram analisadas em três relações sinal-ruído, com a finalidade de conhecermos como os parâmetros prosódicos se comportam quando há ruído na amostra.

Esta tese pretende contribuir com as pesquisas no âmbito da fonética forense no Brasil, uma vez que, em diversos casos citados na literatura, saber a região de origem de um suspeito de cometer um crime pode ser essencial para a resolução de um caso. Em uma situação de crime cometido por telefone, por exemplo, em que o dialeto usado pelo suspeito seja pertencente a uma região que não é a mesma em que a chamada telefônica foi registrada, o número de possíveis suspeitos diminui consideravelmente. Segundo Eriksson (2010), em um caso real ocorrido na Holanda, um indivíduo ameaçou sua vítima por telefone e foi preso porque foi possível identificar a cidade a qual ele pertencia, e que não era a mesma de onde foi registrada a chamada (e pelo fato de que o sujeito possuía ficha criminal), diminuindo o número de suspeitos.

A análise dos parâmetros prosódicos em diferentes relações de sinal-ruído é, certamente, uma tarefa muito comum realizada pelos peritos em fonética forense. Nestes casos, a amostra questionada (normalmente, é a amostra de fala gravada em situação adversa, e pode ser considerada a prova do crime, que pode ter ruído de fundo com energia mais elevada, alteração da qualidade vocal e/ou distorção causada pelo sinal telefônico) é comparada com a amostra de

referência (geralmente feita em um ambiente silencioso, como em uma cabine acústica, por exemplo). A comparação de duas amostras de fala obtidas em condições tão diferentes pode ser um problema quando não se leva em conta o impacto que o ruído pode causar nos parâmetros acústicos utilizados para se fazer a verificação de locutor.

Utilizar os parâmetros prosódicos na tentativa de diferenciar as variedades faladas em diferentes regiões brasileiras pretende preencher uma lacuna nessa área de conhecimento, visto que a prosódia não é muito estudada para esta finalidade. Além disso, mapear as características prosódicas das variedades faladas no país contribui para o desenvolvimento das pesquisas de fonética forense no país e também para a clínica fonoaudiológica, que é carente de dados acústicos de variedades faladas no Brasil.

Nas seções a seguir, discutiremos cada um dos tópicos que motivaram esta pesquisa: a variação dialetal encontrada no Brasil, o uso da prosódia para distinguir diferentes dialetos e o impacto das diferentes relações sinal-ruído nas medidas acústicas.

1.2. Variedade e Dialeto

Para tratarmos das questões que são os objetivos desta tese, faz-se necessário justificar a escolha da terminologia utilizada em relação a dois termos que algumas vezes são utilizados como sinônimos: variedade e dialeto.

A utilização dos termos citados acima é recorrente principalmente quando estudamos a diversidade da fala presente em comunidades, em situações reais de uso. Para Alkmin (2003), uma comunidade de fala:

“se caracteriza não pelo fato de se constituir por pessoas que falam do mesmo modo, mas por indivíduos que se relacionam, por meio de redes comunicativas diversas, e que orientam seu comportamento verbal por um mesmo conjunto de regras.” (p.31, 2003).

Alkmin (2003) afirma que as línguas não são homogêneas e a variação observável em todas elas é produto de sua história e do seu presente. Para as variedades observadas nas comunidades de fala, existem alguns tipos mais conhecidos: língua, dialeto e registro (Hudson, 1980).

O termo variedade é mais genérico, aplicando-se a qualquer manifestação da língua e, segundo Hudson, seu uso é mais flexível que o uso do termo dialeto. Segundo o autor não há uma distinção real entre língua e dialeto, a não ser quando uma possível distinção leva em consideração o prestígio: a língua seria a forma padrão, mais prestigiada, e o dialeto seria a forma menos prestigiada. Por isso, para Hudson e para outros autores, o termo variedade é considerado preferido. A noção de que dada variedade pode ser relativamente similar a outras e relativamente diferente de outras é fundamental em estudos que busquem entender melhor determinadas comunidades de fala.

Nos estudos dialetológicos e nos estudos sociolinguísticos, o termo dialeto é utilizado como termo para uma variedade que é associada à determinada área geográfica ou a diferentes segmentos da sociedade (dialeto social, por exemplo). Como veremos nas seções a seguir, a divisão dialetal do território brasileiro utilizada até hoje como referência para novos estudos é a feita por Nascentes (1953). É importante ter o cuidado de não usar o termo dialeto restringindo-o a uma área geográfica, visto que há grande dificuldade em se estabelecer fronteiras dialetais.

Além disso, muitas vezes, pensa-se que as fronteiras dialetais coincidem com fronteiras geopolíticas, fato que pode ocorrer, mas não se trata de uma relação necessária.

No entanto, a divisão geográfica e política do território brasileiro é a divisão que vem sendo utilizada para a construção dos atlas dialetológicos regionais, porém, em nenhum momento está pressuposto que a divisão utilizada implica em fronteiras dialetais. Visto que as amostras de fala dos sujeitos analisados na presente tese estão divididas de acordo com os limites geográficos dos estados brasileiros aos quais pertencem, devemos deixar claro que não acreditamos que esses limites estaduais são fronteiras dessas variedades, uma vez que é sabido que as variedades regionais se interpenetram e se influenciam principalmente em regiões de fronteira geográfica. Porém, para a realização da pesquisa, a divisão de acordo com os limites geopolíticos facilita a análise, além de nos permitir, ao final da pesquisa, repensar a delimitação desses territórios como áreas de variedades distintas. Além disso, adotaremos o termo variedade, considerado mais amplo e mais genérico, para nos referirmos à “forma de falar” dos sujeitos estudados neste trabalho.

1.3. Os estudos dialetológicos no Brasil

Para tratarmos das sobre as possíveis diferenças nos parâmetros prosódico-acústicos nos diferentes dialetos brasileiros, é interessante discorrer um pouco sobre como os estudos na área da dialetologia no Brasil começaram e como se encontram atualmente no país.

O filólogo Antenor Nascentes foi um dos pioneiros na área dos estudos dialetológicos no Brasil e foi o primeiro a propor uma divisão dialetal do território brasileiro que tomou como base fatores linguísticos em uma época em que os estudos da área não eram amplos e rigorosos. A divisão de Nascentes (1953), mesmo sendo impressionística, é tão importante para a área que

influenciou estudos posteriores e serviu como ponto de partida para outras investigações a respeito dos limites dialetais no Brasil. Para fazer sua divisão, Nascentes seguiu dois critérios:

“a cadência e a existência de pretonicas abertas em vocábulos que não sejam diminutivos nem advérbios em mente...” (Nascentes, 1953:25).

Em relação especificamente à cadência, de acordo com Nascentes, é “palpável a diferença entre a fala *cantada* do nortista e a fala *descansada* do sulista” (p.20, 1953). Utilizando os critérios mencionados, Nascentes dividiu primariamente o território brasileiro em dois grandes grupos: “os falares do Norte e os falares do Sul”, divididos então em algumas subregiões. Segundo o autor, os dois grupos

“estão separados por uma zona que ocupa uma posição mais ou menos equidistante dos extremos setentrional e meridional do país. Esta zona se estende, mais ou menos, da foz do rio Mucuri, entre Espírito Santo e Baía, até a cidade de Mato Grosso, no Estado do mesmo nome, passando cerca de Teófilo Ottoni, Minas Novas, Bocaiuva, Pirapora, rio Parnaíba, rio São Marcos, Arrepêditos, Santa Luzia, Pirenópolis, rio das Almas, Pilar, foz do rio dos Araés, Cuiabá e Mato Grosso.” (Nascentes, 1953:25).

Os falares do Norte, de acordo com Nascentes, englobam a região do Amazonas, Acre, Pará, parte de Goiás e o território do nordeste brasileiro, chegando até a Bahia, que seria uma região intermediária entre os falares do Norte e do Sul. Os falares do Sul compreendem a região que começa no Sergipe, engloba a Bahia, Minas Gerais (Norte, Nordeste e Noroeste), parte de Goiás, Rio de Janeiro, Espírito Santo, e o território sulista, que vai de São Paulo até o Rio Grande do Sul.

A figura 1 ilustra a proposta de Nascentes, em que podemos ver as seguintes divisões dos dialetos: Amazônico, Nordestino, Baiano, Mineiro, Fluminense, Sulista e Território incaracterístico, que é composto por uma parte da região Centro-Oeste do Brasil, na época, pouco habitada.



Figura 1. Divisão dialetal do território brasileiro proposta por Nascentes (1953).

A figura 2 ilustra as divisões propostas por Nascentes e especifica as regiões de acordo com as coordenadas do autor para traçar a linha horizontal no território brasileiro, que o separa em dois grupos, levando em consideração a abertura das vogais pré tônicas.



Figura 2. Divisão dialetal do território brasileiro proposta por Nascentes ilustrada por Noll (2008), com a presença da linha imaginária que Nascentes utilizou para dividir o território em dois grupos.

Apesar de o trabalho de Nascentes ser o pioneiro na área, principalmente por ser o primeiro a realizar um estudo mais rigoroso, é importante ressaltar que os estudos dialetológicos no Brasil já haviam começado antes da publicação de Nascentes. Ferreira e Cardoso (1994) fizeram uma divisão das fases dos estudos dialetológicos no Brasil, que tem sua primeira fase entre o período de 1826 a 1920 (ou seja, antes da publicação de Nascentes).

A primeira fase, segundo as autoras, é marcada por trabalhos que estudaram o léxico e as especificidades do português brasileiro, o que resultou na publicação de diversos dicionários, como por exemplo, o *Dicionário de vocábulos brasileiros*, publicado em 1889 pelo Visconde de Beaurepaire-Rohan e *Apostilas ao dicionário de vocábulos brasileiros* publicadas em 1912, de P. Carlos Teschauer.

A segunda fase da dialetologia no Brasil é marcada pela publicação de *O dialeto caipira* de Amadeu Amaral (1920). De acordo com Ferreira e Cardoso (1994), é na segunda fase que os trabalhos publicados começaram a se preocupar com a metodologia e com a realização de estudos sistemáticos, além de abordarem a morfossintaxe, fonética e fonologia de áreas específicas do Brasil.

Outro trabalho de Antenor Nascentes, *Bases para elaboração do atlas linguístico do Brasil* (1952 primeiro volume e segundo em 1961), faz parte da terceira fase dos estudos dialetológicos no Brasil, que é marcada pelo Decreto nº 30.643, de 20 de março de 1952. Tal Decreto prevê determinadas finalidades para o Centro de Pesquisas da Casa de Rui Barbosa, que foi fundada em 1928 e ainda hoje tem como objetivos o desenvolvimento da cultura, da pesquisa e do ensino, além da divulgação da vida e obra de Rui Barbosa. Dentre as finalidades previstas pelo Decreto, está a elaboração de um atlas linguístico do Brasil. Assim, o trabalho de Nascentes,

publicado em 1952, busca sistematizar ainda mais as pesquisas da área, tendo como propósito a elaboração de atlas regionais, além de um nacional.

Atualmente, o Atlas Linguístico do Brasil (ALiB) tem como meta a realização de um atlas geral no Brasil e conta com auxílio de diversas universidades brasileiras para concluir o projeto. A dificuldade encontrada até hoje para a elaboração de um atlas nacional contribuiu para a publicação de diversos atlas regionais como: *Atlas Prévio dos Falares Baianos* – AFPB (1963), *Esboço de um atlas linguístico de Minas Gerais* – EALMG (1977), *Atlas linguístico da Paraíba* – ALPB (1984), *Atlas linguístico de Sergipe* – ALS I (1987), *Atlas linguístico do Paraná* – ALPR (1990), *Atlas linguístico-etnográfico da região Sul do Brasil* – ALERS (2002), *Atlas linguístico sonoro do Pará* – ALISPA (2004), *Atlas linguístico de Sergipe II* – ALS II (2005), *Atlas linguístico do Mato Grosso do Sul* – ALMS (2007) e *Atlas linguístico do Ceará* – ALECE (2010) (informações disponíveis em <http://twiki.ufba.br/twiki/bin/view/Alib/WebHome>).

De acordo com as informações disponíveis no site do projeto ALiB, o questionário aplicado aos informantes pretende contemplar aspectos (a) fonético-fonológicos, (b) semântico-lexicais e (c) morfossintáticos. Os aspectos fonético-fonológicos são avaliados por meio de 160 questões, dentre as quais 11 relacionadas à prosódia.

As questões relacionadas à prosódia preocupam-se com a entonação. De acordo com Cunha (2006), o questionário do ALiB é sucinto e objetivo e foca nas duas modalidades de frase com valor distinto no português: padrão assertivo e padrão interrogativo, além de três padrões básicos de entonação expressiva (desagrado, contentamento e ordem). A obtenção dos dados é feita após uma instrução como a seguinte:

Instrução: “Se você / o(a) senhor(a) quer oferecer uma bebida a um amigo, e quer saber se ele toma leite ou café, como é que você / o(a) senhor(a) se dirige a ele e pergunta?”

Resposta esperada: “Oh, meu amigo, você toma leite ou café?”

Voltando à divisão dialetal do Brasil, Noll (2008) apresenta uma versão modificada da proposta de Nascentes (1953). A proposta de Noll também divide o território brasileiro em dois grandes grupos e considera o dialeto baiano como pertencente ao grupo Setentrional, diferentemente de Nascentes, que considerou a Bahia como um território intermediário. Noll propõe a seguinte modificação na divisão de Nascentes:

- Grupo Setentrional: Amazônico (Acre, Roraima, Amazonas, Pará, Amapá), Nordeste (Maranhão, Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas), Baiano (Bahia, Norte de Minas Gerais, Norte de Goiás).
- Grupo Meridional: Mineiro (centro e oeste de Minas Gerais), Fluminense (leste de Minas Gerais, Espírito Santo, Rio de Janeiro), Sulista (Sudoeste de Minas Gerais, Sul de Goiás, Mato Grosso do Sul, São Paulo, Paraná, Santa Catarina, Rio Grande do Sul).

Em sua grande maioria, os trabalhos na área da dialetologia no Brasil ocupam-se mais das diferenças fonológicas e lexicais, dando menor atenção a possíveis diferenças prosódicas. Cunha (2006) afirma que, para os estudiosos da prosódia dialetal, há dificuldade tanto na coleta, quando na análise de dados e cita como possível dificuldade, por exemplo, a dúvida no momento da elaboração do desenho do estudo: solicitar leitura ou gravar a fala espontânea?

Desta forma, na próxima seção, faremos um breve retrospecto de trabalhos que utilizaram parâmetros prosódicos para diferenciar dialetos em diferentes línguas.

1.4. A prosódia na classificação dialetal

A prosódia é uma informação fônica, que está além do nível do segmento, e é usualmente estudada a partir da análise de três parâmetros fonético-acústicos clássicos: frequência fundamental, intensidade e duração. Para alguns autores, entre eles Barbosa (2006), a prosódia resulta do acoplamento de uma informação sintática, semântica e discursiva e das restrições de um sistema de produção de fala. Se as primeiras fazem parte das escolhas linguísticas e paralinguísticas do sujeito, as segundas são resultados do modo próprio do sujeito de coordenar os gestos de fala.

Embora muito estudada para outras finalidades, a prosódia não é muito estudada quando se procura buscar diferenças entre variedades. Historicamente, a maioria dos estudos procura buscar diferenças lexicais e fonológicas, estudos que começaram a ser feitos de maneira impressionística até adotarem técnicas de análise acústica para checar os dados, como o que vem sendo feito nos Atlas Linguísticos, por exemplo, e em diversos trabalhos na área da dialetologia no Brasil.

Estudos que focam a variação prosódica nas diferentes variedades do português brasileiro ainda são escassos e por isso o presente estudo busca preencher esse espaço, analisando diferenças entre alguns parâmetros prosódico-acústicos que poderiam revelar diferenças fônicas para além de aspectos estritamente segmentais entre os dialetos do Brasil.

A mesma situação é encontrada em outras línguas, como por exemplo, no inglês norte-americano. De acordo com Clopper e Smiljanic (2011), as pesquisas que buscam encontrar diferenças entre as regiões dialetais nos Estados Unidos têm se preocupado mais com variantes segmentais, principalmente em relação ao sistema de vogais da língua. Para entender melhor o

papel da prosódia na diferenciação dos dialetos, os autores estudaram o ritmo e a entonação de falantes de dialetos da região Central e da região Sul, que caracterizariam dois grupos distintos, por meio da leitura que os sujeitos fizeram de dois textos. A taxa de elocução e a distribuição de pausas foram estudadas, bem como os movimentos de F0, utilizando como notação o sistema ToBI (Pierrehumbert, 1990). Além de tentar entender o papel da prosódia, também foi objetivo dos autores avaliar os efeitos do gênero masculino e feminino em cada um dos dialetos. Os autores concluíram que os dialetos estudados apresentam diferenças quando combinados ao tipo de padrão prosódico utilizado e ao gênero. Não houve diferença entre a taxa de elocução, o que poderia ser explicado pelo fato de a tarefa executada pelos falantes ser uma leitura. Segundo os autores, muitos outros parâmetros e fatores devem ser analisados a partir dos resultados da primeira pesquisa, incluindo a análise de dialetos mais distantes geograficamente, análise de fala espontânea ou de entrevista e uso de outros parâmetros prosódicos como a estrutura rítmica, por exemplo, que é analisada nesta tese de Doutorado.

As diferenças prosódicas entre dialetos também contribuíram para distinção de quatro dialetos do árabe: árabe iraquiano, árabe levantino, árabe egípcio e árabe do Golfo (Biadisy e Hirschberg, 2009). As diferenças encontradas estão relacionadas com a entoação, incluindo gama tonal (*pitch range*), uso de diferentes registros e em relação à estrutura rítmica, os autores encontraram diferenças na taxa de elocução e duração de vogais. O *corpus* do trabalho continha gravações de ligações telefônicas de 398 falantes no total, dos quatro dialetos estudados. Em relação à taxa de articulação, por exemplo, os autores concluíram que os falantes do árabe egípcio falam mais rapidamente, sendo seguidos pelos falantes do árabe do Golfo, pelo árabe iraquiano e árabe levantino, respectivamente em segundo, terceiro e quarto lugares.

Parâmetros prosódicos também foram analisados em um trabalho que tinha como objetivo usar técnicas para medir as distâncias linguísticas entre dialetos na Noruega e correlacionar esses achados com as percepções de ouvintes noruegueses sobre os diferentes dialetos do país (Gooskens e Heeringa, 2005). Para conduzir as análises, foram utilizadas amostras de fala do tipo leitura de texto (uma conhecida fábula do país) por 15 sujeitos (quatro homens e 11 mulheres) de 15 dialetos da Noruega. As amostras de fala foram apresentadas a 15 grupos de ouvintes (cada grupo continha de 16 a 27 ouvintes) das 15 regiões dialetais que estavam presentes nas amostras de fala gravadas previamente. A tarefa dos ouvintes era identificar de quais dialetos eram os sujeitos que tiveram suas faladas gravadas. Outra etapa do trabalho consistiu na realização de medidas objetivas que poderiam distinguir os dialetos. Os autores analisaram a pronúncia, léxico e a prosódia dos dialetos e comentam, especificamente em relação à prosódia, que é sabido que alguns dialetos noruegueses diferem quanto a dois padrões tonais no nível da palavra e, ainda, que outros dialetos não apresentam tais padrões tonais. Assim, a análise prosódica do trabalho focou as diferenças em relação a diferentes tonalidades utilizadas pelos dialetos em pares de palavras. A distância entre a prosódia de um dialeto e de outro foi calculada a partir da utilização de pesos que foram dados pelos pesquisadores para cada um dos padrões tonais encontrados. Os pesos atribuídos pelos pesquisadores aos padrões encontrados foram baseados na experiência que os autores relataram ter com os dialetos estudados. Gooskens e Heeringa (2005) concluíram que há forte correlação entre a pronúncia e o léxico de alguns dialetos, contribuindo para estes serem mais corretamente classificados pelos ouvintes. Além disso, a pronúncia pareceu prever melhor as distâncias percebidas pelos ouvintes. Apesar da afirmação de que a prosódia é um dos principais caracterizadores dos dialetos

noruegueses, os autores encontraram baixa correlação (porém significativa) entre as diferenças prosódicas dos dialetos e a distância percebida pelos ouvintes.

O projeto AMPER (Atlas Multimédia Prosodique de l'Espace Roman) reúne um grupo de pesquisadores que são responsáveis por estudar a acústica e a prosódia de línguas românicas, incluindo, assim o português, que é estudado por 15 instituições participantes do projeto. A análise prosódica é realizada a partir de seis repetições de 66 frases que fazem parte do *corpus* do projeto. As sentenças são apresentadas em forma de desenho e são afirmativas e interrogativas, como, por exemplo, “*O pássaro pateta gosta do Renato?*” e “*O pássaro gosta do Renato nadador.*” (informações disponíveis em <http://pfonetica.web.ua.pt/AMPER-POR.htm>).

Em relação à estruturação rítmica do português brasileiro, o trabalho de Meirelles et al. (2010) estudou falantes de Minas Gerais (MG) e da Bahia (BA). Para as análises, foram recrutados oito falantes (dois homens e duas mulheres de cada região) que repetiram por 10 vezes 11 sentenças em três taxas de elocução diferentes: rápida, normal e lenta. Os autores utilizaram as durações das unidades vogal-vogal ou VV (que serão descritas na seção de Metodologia da presente tese) e a duração dos grupos acentuais. Após análise estatística, utilizando a duração das unidades VV em função de cada dialeto, os resultados mostraram que os falantes de MG e BA têm a taxa de elocução considerada “normal” semelhantes. A diferença encontrada foi em relação à taxa de elocução rápida, que mostrou que os falantes de MG têm a taxa de elocução rápida mais rápida que os falantes da BA. Ao analisar também os grupos acentuais, os autores concluíram que os dados sugerem que o dialeto de MG tem ritmo mais acentual que o da BA.

1.5. Impacto do ruído na extração das medidas prosódico-acústicas

A tarefa de identificação de falantes, comum na Fonética Forense, ocorre quando há uma amostra de fala como prova de um crime, como por exemplo, uma gravação de um telefonema de um suspeito fazendo ameaças a outra pessoa. Durante o processo de identificação do falante, é preciso fazer a correlação entre a amostra de fala questionada (gravação do telefonema com a ameaça, a prova do crime) e amostras de fala de referência (gravações com amostras de fala de suspeitos de terem feito a ameaça), sendo que as últimas geralmente são feitas em situações totalmente diferentes das primeiras, podendo ser realizadas nas dependências de uma delegacia de polícia, por exemplo. A correlação entre as duas amostras de fala é importante, pois o grau de similaridade entre a amostra questionada e a amostra de referência é chamado de evidência.

Os profissionais que realizam esse tipo de atividade podem se deparar com uma série de dificuldades quando obtêm a gravação da amostra de fala questionada para análise. As dificuldades encontradas podem estar, por exemplo, relacionadas à simulação da qualidade vocal ou disfarce vocal realizado pelo suspeito durante o telefonema, distorção do sinal em que a amostra de fala foi gravada, barulho de fundo como o som do vento, barulho do trânsito e interferências no sinal de fala devido ao uso de telefone celular (Hollien, 2002; Eriksson, 2010).

Todas as situações descritas alteram a relação sinal-ruído da amostra de fala e podem comprometer a identificação correta de um indivíduo ao modificarem o sinal de fala e, conseqüentemente, as medidas acústicas que são amplamente utilizadas para auxiliar na identificação de falantes, no campo da fonética forense. Segundo Alexander et.al. (2004), o ruído de fundo presente em uma gravação pode diminuir a força da evidência quando as comparações

entre as gravações são realizadas, pois diminuem a capacidade de identificação correta do indivíduo.

Dentre os parâmetros acústicos apontados como mais robustos na identificação de falantes estão a frequência fundamental (Atal, 1972) e variáveis presentes no estilo de fala, como taxa de elocução (Kahn, 2008; Öhman et. al., 2010).

A identificação de falantes em amostras de fala com ruído foi estudada por Alexander et. al. (2004). Os autores adicionaram artificialmente ruído branco em amostras de fala gravadas por telefone e apresentaram trechos de fala com duração entre 10 e 20 segundos a ouvintes, sendo que esses trechos eram considerados como os vestígios do crime. Para o experimento, 90 ouvintes teriam que correlacionar as amostras de fala com ruído com as amostras de fala sem ruído (que eram mais longas, com 90 segundos), pertencentes aos suspeitos. Uma escala verbal com sete níveis foi estabelecida e, na escala, as respostas variavam de “eu estou certo de que os falantes das duas gravações são diferentes” até “eu estou certo de que os falantes das duas gravações são os mesmos”. Ao final do experimento, os ouvintes foram questionados sobre quais fatores tinham condicionado suas decisões. Os autores concluíram que 30% dos sujeitos levaram em conta a pronúncia e o sotaque dos sujeitos para fazer a comparação quando a amostra de fala tinha ruído; em segundo lugar, os falantes consideraram o timbre de voz, que obteve 22% das respostas; em terceiro lugar, a entonação, com 18%; e, em quarto lugar, a velocidade de fala, com 12% das respostas.

Como mostramos, o ruído é um fator comum em amostras de fala na rotina dos profissionais que trabalham com fonética forense e sua presença pode até mesmo diminuir a

força da evidência da prova de um crime. Por isso, julgamos importante conhecer mais sobre o comportamento de parâmetros prosódico-acústicos quando há ruído em uma amostra de fala.

2. Objetivo

O objetivo principal desta tese, no primeiro momento, foi verificar se os parâmetros prosódicos clássicos são capazes de diferenciar as variedades faladas em diferentes regiões brasileiras e, no segundo momento, o objetivo foi avaliar como a inclusão de ruído aditivo nas gravações de fala espontânea perturba os parâmetros prosódicos.

2.1. Objetivo suplementar

O objetivo suplementar do trabalho foi realizar testes perceptivos a fim de saber se sujeitos, falantes do português brasileiro, conseguem identificar corretamente a região de origem de outros falantes do português brasileiro.

3. Metodologia

3.1. Sujeitos

Foram analisadas amostras de fala de 35 sujeitos, do sexo masculino, de sete regiões brasileiras e com idade média de 35,4 anos. O grau de escolaridade variou de ensino superior completo até pós-graduação completa. As amostras de fala, que contêm cinco minutos de gravação cada, fazem parte de um banco de dados coletado pela Polícia Federal de Brasília, que tem como intenção promover mais estudos na área de fonética forense no Brasil com a finalidade de auxiliar na prática dos peritos da instituição. O banco de dados é composto por gravações dos próprios oficiais da Polícia Federal (PF), que foram realizadas em cabine acústica. As amostras de fala se enquadram no gênero entrevista livre, em que o entrevistador faz algumas perguntas, guiando o tema principal, e o entrevistado tem o tempo que quiser para discorrer sobre o assunto, sem ser interrompido muitas vezes pelo entrevistador. Os temas das gravações não variaram muito entre os sujeitos, que falaram sobre sua vida pessoal, profissional e viagens. O departamento da PF autorizou o uso das gravações para as análises realizadas nesta tese (ANEXO 1).

A distribuição de sujeitos por cada região se deu da seguinte maneira:

Região estudada	Nº de sujeitos
São Paulo (SP)	6
Rio de Janeiro (RJ)	6
Minas Gerais (MG)	6
Região Nordeste (NE)	6
Distrito Federal (DF)	5
Paraná (PR)	4
Região Norte (N)	2
Total	35

Tabela 1. Números de sujeitos por região estudada

A distribuição da cidade de procedência dos sujeitos estudados se dá da seguinte forma: os sujeitos da região de São Paulo são provenientes da capital (apenas um sujeito) e do interior, de cidades como: Piracicaba (dois sujeitos), Jundiaí (um sujeito), Ribeirão Preto (um sujeito) e Campinas (um sujeito). O Rio de Janeiro tem cinco representantes da capital fluminense e um de Niterói. Todos os seis sujeitos estudados de Minas Gerais são da cidade de Belo Horizonte e, os sujeitos da região Nordeste são de Natal (dois sujeitos), Fortaleza (dois sujeitos), João Pessoa (um sujeito) e Aracaju (um sujeito). Os representantes da região do Paraná são de Curitiba (dois sujeitos), Ponta Grossa (um sujeito) e São Mateus do Sul (um sujeito).

Mais especificamente em relação aos sujeitos do Distrito Federal, a escolha de analisar sujeitos nascidos nesta região (quatro deles são nascidos em Brasília entre 1980 e 1981 e um deles nasceu no interior de Goiás e mudou-se para Brasília com um mês de vida e por isso também foi incluído como pertencente à região do DF) foi proposital, pois é sabido que a região do DF passou por um intenso contato de diferentes variedades do português brasileiro após a fundação de Brasília, em 1960. Em pesquisa de 1998, a porcentagem de habitantes do DF nascidos em Brasília era de 41% (Fonte: CODEPLAN, disponibilizado por Corrêa, 1998) de um total de 1.722.190 milhão. Atualmente, com 54 anos, espera-se que a região tenha, ao mínimo, duas gerações que nasceram no DF e, por isso, seria interessante analisar as características

prosódicas do local, visto que outros trabalhos caracterizam o falar do DF como uma variedade que priorizou algumas marcas e descartou outras (Bortoni-Ricardo, 2010) e assim saber se, em relação à prosódia, a variedade falada no DF é mais semelhante a alguma outra região do Brasil.

Os sujeitos da Região Nordeste (NE) não foram rotulados pelos seus estados de origem (como as outras regiões), pois avaliamos a hipótese de que os diferentes estados da Região Nordeste (exceto a Bahia) possuem as mesmas características prosódicas, fato confirmado em nossos resultados pela baixa variação que os sujeitos apresentaram em relação aos parâmetros analisados. O desvio-padrão encontrado na análise dos parâmetros prosódicos dos sujeitos da Região Nordeste foi, muitas vezes, menor que o desvio-padrão de outras variedades estudadas como a de São Paulo, por exemplo, o que significa que os sujeitos de SP apresentaram maior variação em seus resultados que os sujeitos do Nordeste, que pertencem a diferentes estados. Consideramos que a decisão metodológica que foi tomada inicialmente (de manter sujeitos de diferentes estados da região Nordeste em um mesmo grupo) foi acertada, visto que os resultados corroboraram com nossa hipótese inicial.

Em relação à Região Norte, o número de sujeitos analisados é menor que o das outras regiões pela maior dificuldade em encontrar sujeitos que se encaixassem no critério de inclusão do estudo que previa que os sujeitos deveriam ter nascido e vivido até a idade adulta na região que representavam, além de pertencerem à organização da PF. A baixa densidade demográfica da região Norte também pode ser uma justificativa para este fato. Segundo dados do censo de 2010 do IBGE (disponíveis em <http://censo2010.ibge.gov.br/resultados>), a Região Norte do Brasil é uma das áreas com menor densidade demográfica no território brasileiro e é a maior região do país em extensão territorial. Os dados do IBGE mostram que os sete estados da região Norte estão entre os estados com menor densidade populacional do país, sendo que o estado da

região com menor densidade é Roraima, com 2,01 habitantes por km^2 (hab/ km^2) e o com maior densidade é Rondônia, com 6,56 hab/ km^2 . As justificativas para a baixa densidade demográfica podem estar relacionadas com condições naturais das regiões ou com fatos históricos ligados ao uso de territórios menos densos. A figura 3 mostra que as regiões litorâneas do Brasil apresentam a maior densidade demográfica (áreas amarelas, laranjas e vermelhas), enquanto que o interior do país, principalmente na região Norte, apresenta menor densidade demográfica (áreas em branco).

O número inferior de sujeitos da região Norte é levado em conta no momento de realizar os testes estatísticos, não interferindo nos resultados obtidos que serão apresentados nas próximas seções. Devemos ressaltar que o número de sujeitos analisados, somente dois, é pequeno para fazermos generalizações para toda a região Norte (visto que os resultados apresentados a seguir dizem respeito à população estudada nesta tese de doutorado). Ainda sim, consideramos que todos os sujeitos analisados, de todas as variedades estudadas são reconhecidos como pertencentes aos seus Estados de origem e assim, são considerados representativos de suas regiões.

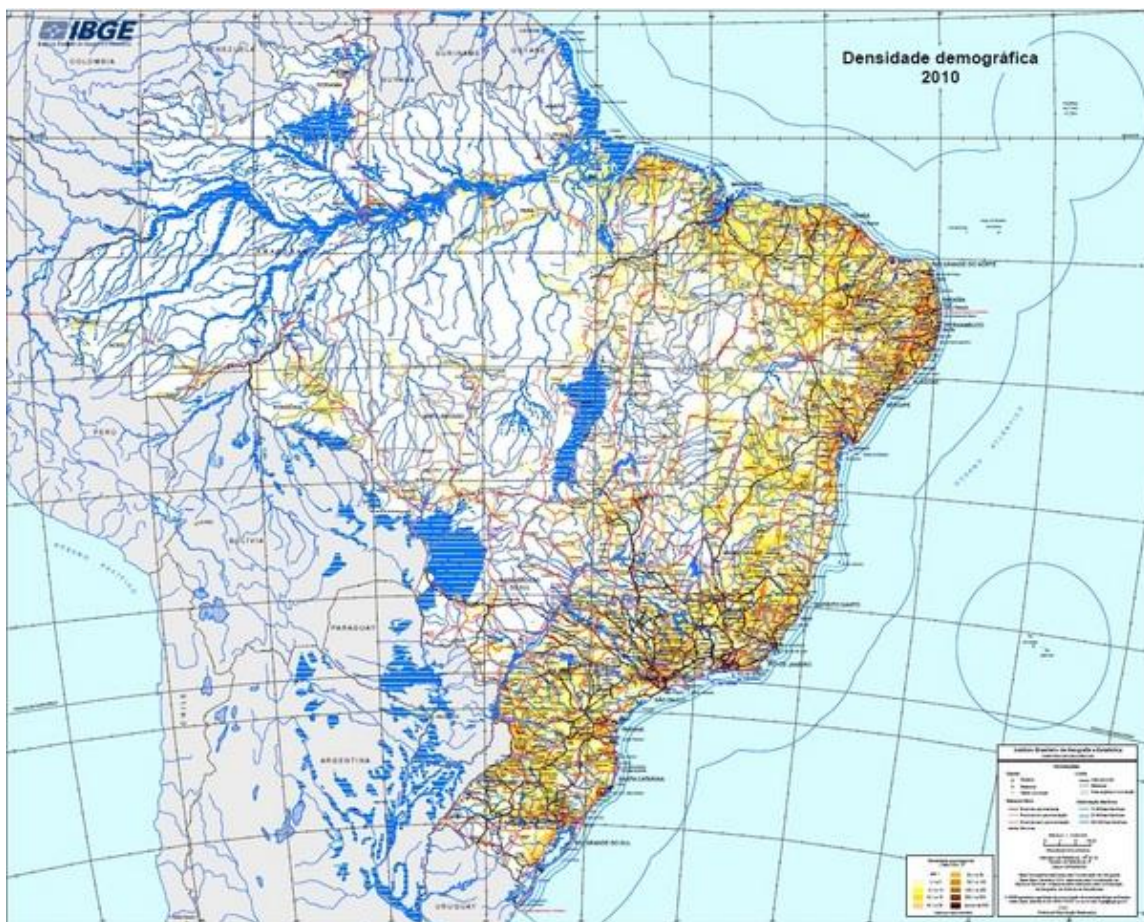


Figura 3. Densidade demográfica do território brasileiro segundo o Censo 2010. As áreas litorâneas concentram a maior parte da população: quanto mais escura a escala de amarelo, maior a densidade demográfica da área.

3.2.Tratamento dos dados

As amostras de fala (com cinco minutos de duração) foram divididas em amostras menores, de aproximadamente 100 segundos cada uma, o que permitiu que os *scripts* que foram utilizados fossem executados mais rapidamente, ao utilizarmos os arquivos de som com menor duração. Duas gravações foram divididas em quatro trechos menores e uma gravação foi dividida

em cinco trechos menores. O total de trechos, considerando todos os sujeitos analisados, foi de 109 trechos de fala.

Todas as amostras de fala foram segmentadas em unidades guiadas pelo tamanho da sílaba, as chamadas unidades VV.

A unidade VV é uma sílaba fonética compreendida entre o segmento acústico que vai do *onset* de uma vogal até o *onset* da vogal imediatamente seguinte (Barbosa, 2006). Uma vasta literatura demonstrou que os picos locais da duração normalizada dessas unidades são correlatos acústico-prosódicos do acento frasal (Barbosa, 2006) e que essas unidades são unidades mínimas de processamento do ritmo da fala por excelência (Pompino-Marschall, 1991; Dogil e Brown, 1988). Sendo assim, a duração silábica, estudada por meio da evolução das durações de cada unidade VV, poderia revelar características interessantes de cada dialeto em relação a esse parâmetro, bem como, agrupar alguns dialetos de acordo com o ritmo.

A segmentação da amostra em unidades VV foi feita de forma automática (seguida de correção manual, quando necessário) com o auxílio do *script* BeatExtractor (Barbosa, 2006:170). Após a segmentação, os trechos eram transcritos conforme uma notação ASCII. Na figura 4, encontra-se um exemplo do seguinte trecho “*equipamento é de transmissão*”, segmentado em unidades VV. Como podemos observar na figura 4, a segmentação começa com a unidade VV “ekip”. A segunda unidade VV é “am” e assim por diante. A notação utilizada permite marcar as vogais pós tônicas com letras maiúsculas, como no caso do “U”, unidade VV final da palavra “equipamento”. Além disso, há também a possibilidade de marcar as vogais abertas como em “eh”.

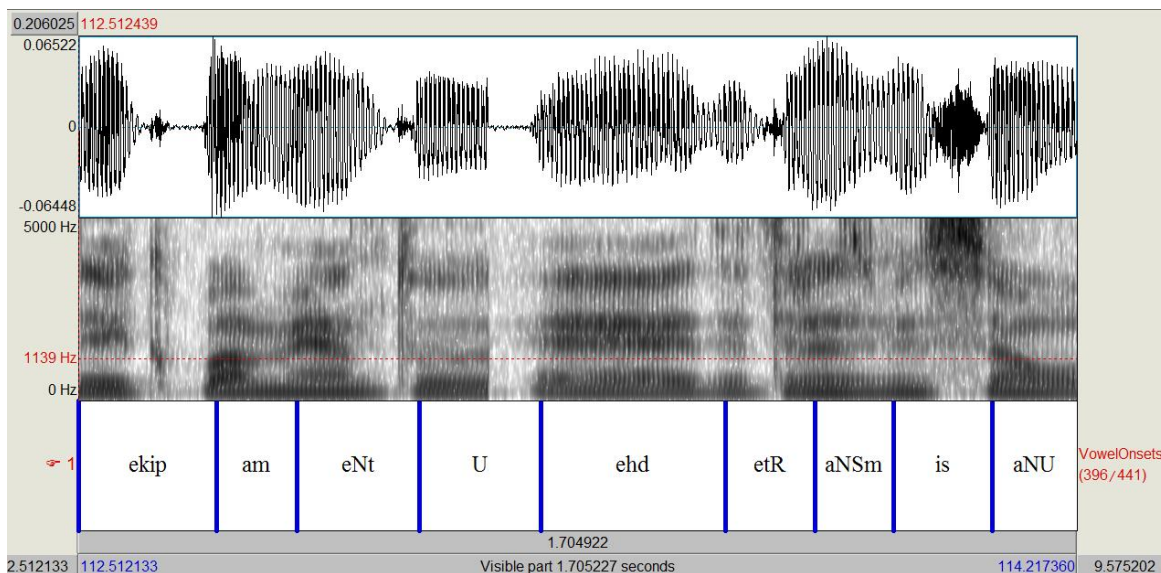


Figura 4. Trecho “*equipamento é de transmissão*” segmentado em unidades VV. Na janela superior da figura encontra-se o traçado da forma de onda, na janela intermediária visualizamos o espectrograma de banda larga e, na última janela estão as unidades VV.

Com a segmentação de todas as amostras de fala em unidades VV, o *script* ProsodicDescriptorExtractor (Barbosa, 2011) foi utilizado para extrair as medidas prosódico-acústicas selecionadas. O script ProsodicDescriptorExtractor extrai originalmente 13 medidas acústicas, porém, para este trabalho, oito medidas foram selecionadas (a serem descritas no item a seguir), pois julgamos que seriam as mais suscetíveis a revelarem características prosódicas individuais dos sujeitos. O script extrai as medidas a partir do arquivo de som, juntamente com o arquivo TextGrid, que contém as segmentações em unidades VV.

3.3. Parâmetros fonético-acústicos analisados

Oito medidas acústicas foram selecionadas para serem extraídas das amostras de fala. Estas medidas acústicas foram selecionadas com base em outro estudo que obteve resultados satisfatórios para caracterizar isoladamente a prosódia de dois tipos de enunciados: contagem de

histórias e leitura de texto (Barbosa e Silva, 2010). Os autores propuseram nova metodologia para comparar a estrutura rítmica dos dois tipos de enunciados, analisando 11 possíveis parâmetros capazes de revelar diferenças na estrutura rítmica e aplicaram um teste perceptivo para saber como os ouvintes classificariam os diferentes enunciados. Um modelo de regressão linear múltipla foi construído para que as respostas dos ouvintes pudessem ser relacionadas com os achados da análise dos 11 parâmetros extraídos de cada um dos estilos de fala. Os resultados mostraram que quatro dos 11 parâmetros poderiam explicar as respostas dadas pelos ouvintes: taxa de articulação (em unidades VV/s), taxa de unidades VV não-proeminentes/s, taxa de proeminência (picos de z-score/s) e relação entre a duração das proeminências e a força das fronteiras, sendo que este último junto com a taxa de proeminência foram capazes de explicar 71% da variância das respostas dos ouvintes.

Dessa forma, partimos da hipótese de que, caracterizando a prosódia da fala de cada indivíduo, podemos chegar a um grupo maior, que teria características próprias, baseadas na localidade em que cada indivíduo nasceu e viveu até a idade adulta.

As medidas analisadas contemplam os três parâmetros acústicos clássicos de estudo da prosódia: frequência fundamental, intensidade e duração. Os parâmetros analisados estão descritas no quadro a seguir.

Medida	Descrição
SR	Taxa de elocução (Unidades VV/s)
Mz	média de z-score suavizado de duração de unidade VV
SDz	desvio-padrão de z-score suavizado de duração de unidade VV
Skz	assimetria de z-score suavizado de duração de unidade VV
Pr	taxa de proeminência (picos de z-score/s)
F0	mediana da frequência fundamental
Emph	ênfase espectral
Ur	taxa de unidades VV não proeminentes por segundo

Quadro 1. Medidas acústicas analisadas por meio do *script* ProsodicDescriptorExtractor (Barbosa 2011).

O primeiro parâmetro analisado é a taxa de elocução (speech rate, SR) que são as unidades VV por segundo, que são extraídas utilizando o arquivo TextGrid que contém a segmentação em unidades VV, feita com auxílio do *script* BeatExtractor.

Os três parâmetros seguintes (Mz, SDz e Skz) levam em consideração o cálculo de z-score para a extração de cada uma das medidas. O cálculo de z-score assinala a duração (normalizada) das unidades VV ao longo do enunciado em termos de afastamento da média expresso em unidades de desvio-padrão. Assim, o z-score especifica o afastamento do valor medido (da duração de cada unidade VV) em relação a uma média de duração dos fones do português brasileiro (PB) que compõem a unidade. A opção por trabalhar com o z-score suavizado e seus três primeiros descritores estatísticos (média, desvio-padrão e assimetria) foi feita porque a suavização atenua os efeitos da realização local do acento lexical, assinalando apenas unidades salientes no enunciado (cf Barbosa, 2006). O z-score suavizado é calculado utilizando a fórmula abaixo:

$$z_{suav}^i = \frac{5 \cdot z^i + 3 \cdot z^{i-1} + 3 \cdot z^{i+1} + 1 \cdot z^{i-2} + 1 \cdot z^{i+2}}{13}$$

Em que z^i é o valor de z-score da duração da unidade VV, z^{i-1} é o valor de z-score da duração da unidade VV imediatamente anterior, z^{i+1} é o valor de z-score da duração da unidade VV imediatamente seguinte e assim por diante. Os valores de média, desvio-padrão e assimetria do z-score suavizado revelam como se estruturam o grau de proeminência e a força das fronteiras no enunciado e são cruciais para fazer uma descrição completa do ritmo da fala de um indivíduo (Barbosa e Silva, 2010).

O quarto parâmetro analisado (sigla PR) é referente à taxa de proeminência, que é a taxa de produção de picos de duração normalizada por segundo. A frequência fundamental, calculada a partir de sua mediana, foi o quinto parâmetro a ser analisado. A opção por extrair os valores de mediana de F0 é devido ao fato de os valores de mediana serem mais resistentes a erros de medição do que a média.

A ênfase espectral (emph) foi o sexto parâmetro analisado; a ênfase espectral e a intensidade de um trecho de fala representam duas diferentes operacionalizações da *loudness*. A medida de intensidade leva em conta todo o espectro para o cálculo final do parâmetro e a ênfase espectral é definida como a intensidade relativa que se concentra nas regiões de altas frequências do espectrograma. De acordo com uma breve revisão de Heldner (2003) sobre o assunto, a ênfase espectral é uma das medidas relacionadas com o declínio espectral, do qual também fazem parte as medidas de balanço espectral e inclinação espectral. O autor ressalta que estas medidas podem se dividir em duas classes diferentes: uma relacionada à inclinação espectral e outra relacionada à ênfase espectral. O cálculo da inclinação espectral se refere explicitamente ao declínio do espectrograma no que diz respeito à intensidade, enquanto que a ênfase espectral deve ser usada para calcular as medidas de energia relativa em altas frequências ou, em outras palavras, é a contribuição relativa das altas frequências para a intensidade total do

espectrograma. Heldner (2003) comenta que os pesquisadores da área têm utilizado diversos métodos para calcular a ênfase espectral. A definição do método utilizado para o cálculo de ênfase espectral na presente tese de Doutorado é a diferença de intensidade (em dB) do segmento a ser analisado entre toda a faixa espectral (0 a 11 kHz) e a região de frequências baixas logo acima da fundamental fixada na faixa de 0 a 350 Hz, como definido por Traunmiller e Eriksson (2000).

Por fim, o oitavo parâmetro analisado é referente à taxa de unidades VV não proeminentes por segundo (sigla ur), que é próximo à taxa de articulação, pois não leva em conta, em seu cálculo, as pausas silenciosas. As unidades VV não proeminentes são aquelas que não possuem picos de duração normalizada para as respectivas unidades VV. A taxa de unidades VV não proeminentes é calculada utilizando-se o número total de unidades VV não proeminentes em um trecho de fala, dividido pelo tempo total correspondente a essas unidades no trecho de fala, que é delimitado pelo primeiro *onset* vocálico até o último *onset* vocálico.

3.4. Análise estatística

A análise estatística dos dados foi feita com a utilização do software gratuito R, versão 3.0.2 (disponível em: www.r-project.org).

O teste estatístico One-Way Anova (com nível de significância de 0.05) e com fator VARIEDADE foi utilizado para analisar possíveis diferenças entre as medidas prosódico-acústicas estudadas nas variedades. Ao serem atestadas diferenças, um dos testes post-hoc utilizados foi o teste Tukey Honestly Significant Difference (Tukey HSD), com o mesmo nível de significância de 0.05. O teste Tukey HSD permitiu atestar entre quais variedades eram as diferenças apontadas pelo teste One-Way ANOVA. O segundo teste post-hoc utilizado foi o teste

Duncan (nível de significância de 0.05), que também permitiu que as variedades fossem agrupadas de acordo com os resultados obtidos. O teste Duncan compara os resultados obtidos a partir de suas médias e determina a criação de grupos maiores de variedades (denominamos pelas letras maiúsculas A, B, C, etc) e em alguns casos, também são criados grupos intermediários (AB, por exemplo).

O método estatístico LDA (Linear Discriminant Analysis) foi empregado com a finalidade de saber se as oito medidas analisadas poderiam, em conjunto, prever cada uma das variedades, por meio de seis funções discriminantes. A LDA encontra uma combinação linear utilizando funções discriminantes de acordo com os parâmetros que estão sendo analisados e, assim, o método consegue classificar as variáveis estudadas (em nosso caso, as sete variedades do português brasileiro). A utilização deste método é interessante para a discussão dos dados, visto que a LDA pode revelar como as combinações lineares explicam os resultados obtidos e, como mostraremos na seção dos resultados, a LDA consegue prever as variedades estudadas a partir dos resultados dos parâmetros prosódicos analisados (informações sobre o método estatístico LDA estão disponíveis no tutorial do software estatístico R: cran.r-project.org).

Os resultados obtidos serão apresentados na seção 4.1.

3.5. Análise das diferentes relações sinal-ruído.

As amostras de fala dos 35 sujeitos estudados tiveram adição de ruído gaussiano em duas magnitudes, para que possíveis mudanças nos parâmetros prosódico-acústicos, causadas pelo ruído, fossem analisadas, comparando-se os valores obtidos na gravação original (que caracteriza alta relação sinal-ruído) com os valores obtidos após adição de ruído (baixa relação sinal-ruído).

O ruído do tipo gaussiano foi adicionado às gravações originais de cada sujeito por meio do software Praat com a seguinte fórmula, em que a primeira posição entre parênteses é a média do ruído (nesse caso 0) e a segunda é o desvio-padrão em unidades arbitrárias de pressão sonora (nesse caso 0.01):

Original + randomGauss(0,0.01)

Duas magnitudes de ruído foram adicionadas, sendo que, na primeira, a magnitude do desvio-padrão da gaussiana adicionada foi 0.01 e na segunda, 0.02. O objetivo de aumentarmos ainda mais a magnitude do ruído foi testar como os parâmetros se comportariam quando diminuíssemos ainda mais a relação sinal-ruído.

Nas três figuras seguintes, pode-se observar a onda e o espectrograma original (Figura 5), o espectrograma após adição de ruído de magnitude 0.01 (Figura 6) e após adição de ruído de magnitude 0.02 (Figura 7).

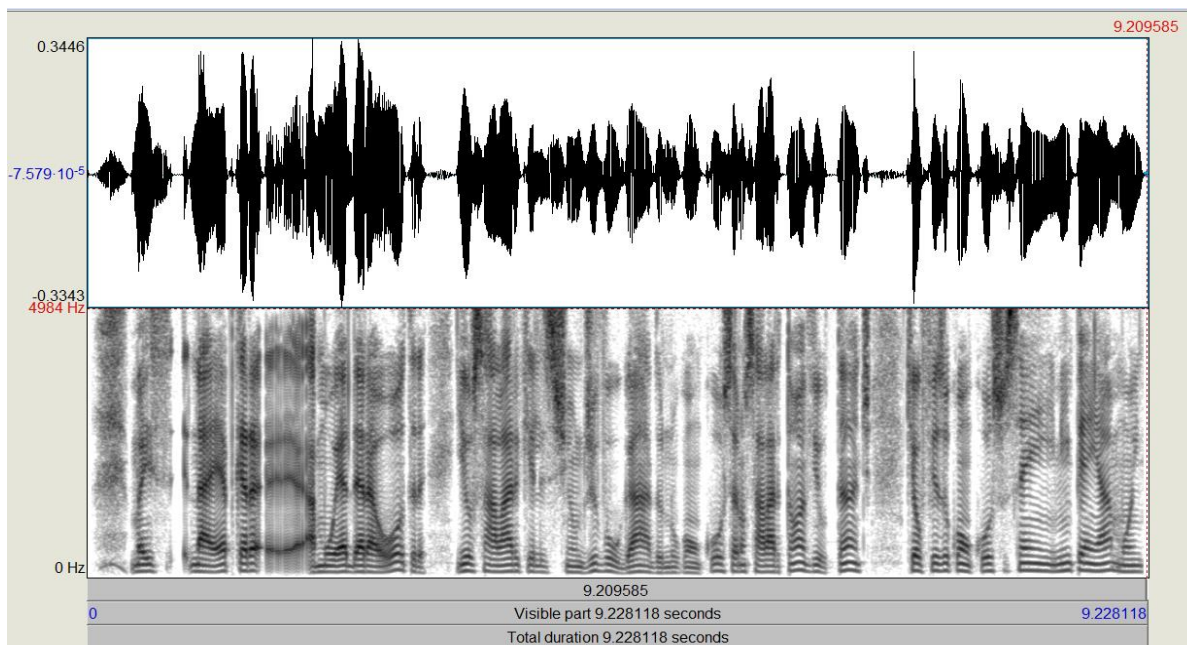


Figura 5. Traçado da forma da onda (janela superior) e espectrograma de banda estreita (janela inferior) da gravação original.

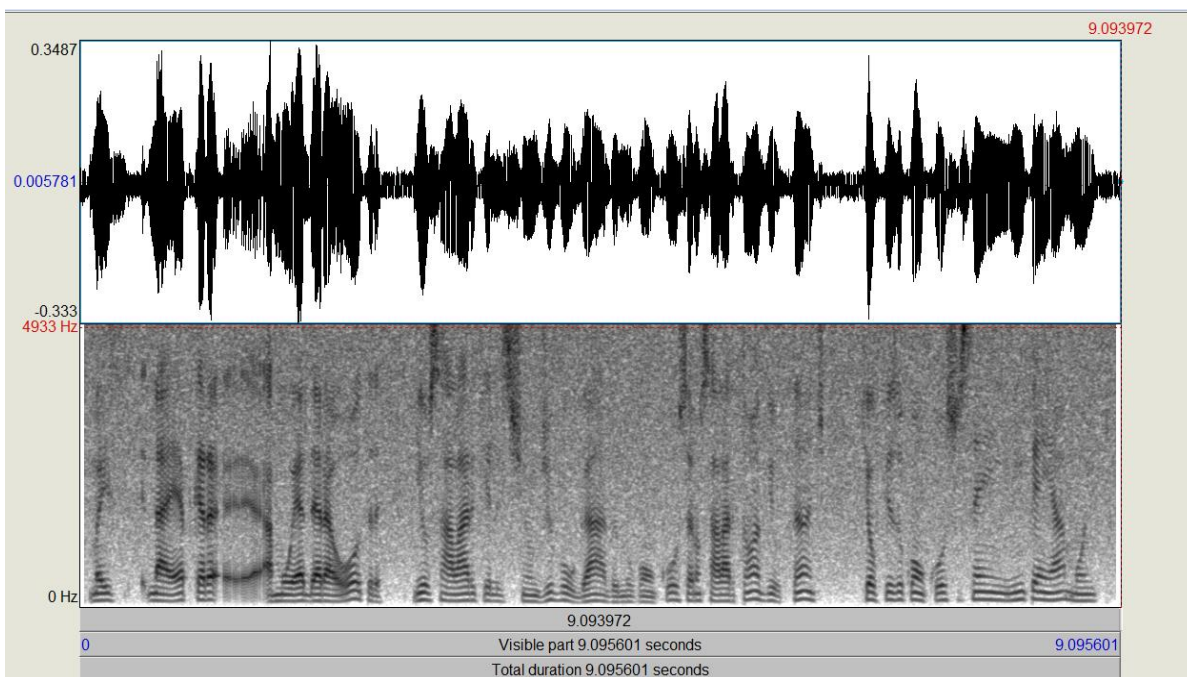


Figura 6. Traçado da forma da onda (janela superior) e espectrograma de banda estreita (janela inferior) com adição de ruído de magnitude 0.01.

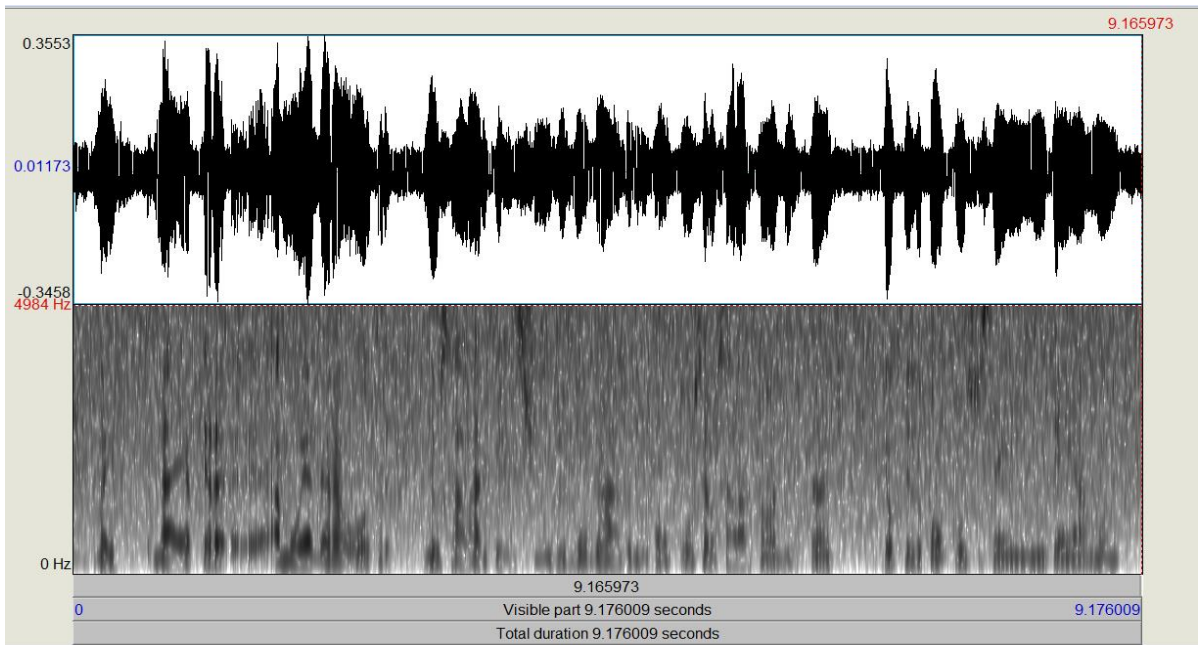


Figura 7. Traçado de forma da onda (janela superior) e espectrograma de banda estreita (janela inferior) adição de ruído de magnitude 0.02.

A figura 8 mostra um detalhe do espectrograma apresentado na figura 7, permitindo melhor visualização do impacto do ruído no espectrograma.

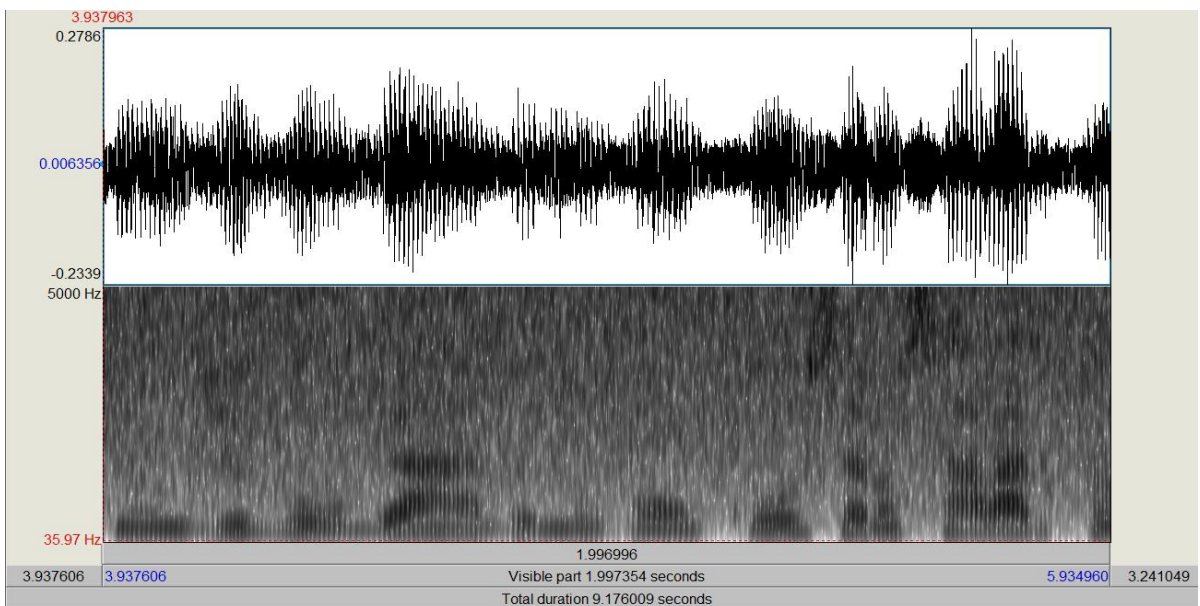


Figura 8. Detalhe do espectrograma de banda larga com adição de ruído de magnitude 0.02.

Após a adição de ruído nas duas magnitudes escolhidas (0.01 e 0.02), os scripts BeatExtractor e ProsodicDescriptorExtractor foram novamente utilizados para, respectivamente, segmentar as gravações em unidades VV e extrair as mesmas medidas prosódico-acústicas já descritas em item anterior, para posterior comparação dos valores obtidos.

3.5.1. Análise estatística

Com as medidas prosódico-acústicas extraídas, o próximo passo do estudo consistiu em fazer uma análise estatística para comparar se houve mudança estatisticamente significativa entre os parâmetros extraídos em cada situação, na comparação da gravação original com os outros dois momentos. Para isso, o teste escolhido foi o Teste T, com nível de significância de 0.016, que leva em conta as três situações que foram consideradas para análise (nível de significância 0.05/3), que são as seguintes:

- 1) Comparação dos resultados obtidos na análise da gravação original (alta relação sinal-ruído) com os resultados obtidos na análise da gravação com ruído aditivo de magnitude 0.01 (baixa relação sinal-ruído);
- 2) Comparação dos resultados obtidos na análise da gravação original (alta relação sinal-ruído) com os resultados obtidos na análise da gravação com ruído aditivo de magnitude 0.02 (relação sinal-ruído muito baixa);
- 3) Comparação dos resultados obtidos na análise da gravação com ruído aditivo de magnitude 0.01 (baixa relação sinal-ruído) com os resultados obtidos na análise da gravação com ruído aditivo de magnitude 0.02 (relação sinal-ruído muito baixa).

3.6. Testes perceptivos

Além de estudar os parâmetros prosódico-acústicos que poderiam ser capazes de individualizar os sujeitos e assim identificá-los pelas suas regiões de origem e analisar os parâmetros em diferentes relações sinal-ruído, tivemos como objetivo específico aplicar testes perceptivos, utilizando as amostras de fala disponíveis, para saber se ouvintes são capazes de identificar corretamente a região de origem dos sujeitos que tiveram suas falas analisadas.

O tipo de teste utilizado foi o de classificação livre, que foi utilizado em trabalho de Clopper e Pisoni (2007) para avaliar se os ouvintes americanos conseguiam classificar os dialetos regionais dos Estados Unidos. Nos testes tipo classificação livre, os ouvintes têm liberdade para dizer de onde acham que o falante é. Em nosso estudo, os ouvintes tiveram como auxílio um mapa geográfico do Brasil, com as legendas de cada estado, para que pudessem utilizar caso tivessem dúvidas quanto ao nome e localização de cada região.

3.6.1. Ouvintes

Para a aplicação do teste, foram recrutados 20 participantes, com idade variando entre 18 e 35 anos, que nasceram em diferentes localidades brasileiras. Os participantes tinham como língua materna o português brasileiro e os pais de todos os participantes eram brasileiros.

Houve grande cuidado durante o recrutamento para que sujeitos de diferentes regiões pudessem responder ao teste para que, assim, pudéssemos constatar se a região de origem dos ouvintes os influencia em classificar corretamente regiões que sejam vizinhas, por exemplo. Responderam ao teste sujeitos dos seguintes estados brasileiros: Goiás, São Paulo, Rio Grande do Sul, Bahia, Minas Gerais, Rio de Janeiro, Maranhão, Sergipe e Mato Grosso do Sul.

Apesar de os sujeitos terem nascido nessas regiões, alguns já moravam há algum tempo no estado de São Paulo (onde foi feita a aplicação do teste) e por isso foi feito um levantamento individual de lugares em que os sujeitos já haviam morado e por quanto tempo. O participante que mais trocou de cidade o fez por dez vezes em toda sua vida, sendo que, das 10 cidades em que morou, oito são no interior do estado de SP.

O quadro 2 mostra o local de nascimento e idade de cada um dos 20 sujeitos que responderam ao teste perceptivo.

Sujeito	Local de Nascimento	Idade
1	Salvador – BA	28
2	Salvador – BA	28
3	Ouro Preto - MG	25
4	Itajubá – MG	18
5	Itajubá – MG	21
6	Pouso Alegre – MG	18
7	São Luis – MA	31
8	Aracaju – SE	23
9	Pelotas – RS	26
10	Ibitinga – SP	28
11	Campinas – SP	27
12	Pirassununga – SP	30
13	Pirassununga – SP	32
14	Jales – SP	23
15	Campinas – SP	24
16	Campo Grande – MS	29
17	Bragança Paulista – SP	23
18	Campinas – SP	23
19	Goiânia – GO	35
20	Niterói – RJ	22

Quadro 2. Local de nascimento e idade dos sujeitos que responderam ao teste perceptivo.

3.6.2. Amostras de fala utilizadas

Para a aplicação do teste, selecionamos pequenos trechos de 10 segundos cada, de sujeitos de todas as regiões estudadas. Foram selecionados quatro sujeitos do Rio de Janeiro,

quatro de São Paulo, quatro de Minas Gerais, quatro do Paraná, quatro do Distrito Federal, três do Nordeste e dois do Norte, totalizando 25 sujeitos que tiveram suas falas analisadas. A quantidade selecionada de trechos de fala de cada sujeito variou, pois em algumas amostras de fala não era possível selecionar trechos em que os falantes discursavam sobre sua cidade natal ou sobre lugares que já haviam vivido (lembramos que esse foi um tema recorrente em todas as entrevistas realizadas) e por isso tomamos o cuidado de não selecionar trechos que pudessem dar pistas ou citar nomes de cidades. O total de trechos selecionado foi 75: 12 trechos de falantes do RJ, 12 de SP, 12 do DF e 12 do Nordeste, 11 trechos de MG, 10 do PR e 6 da região Norte.

3.6.3. Procedimento realizado

Os trechos de fala selecionados foram apresentados aos ouvintes através de *scripts* intitulados “ExperimentMFC” no Praat. Esta função do Praat permite que sejam criados testes de escolha forçada, para várias finalidades. Como o intuito do nosso teste de percepção era obter uma classificação livre das variedades faladas, por parte dos ouvintes, o script desenvolvido para o teste fazia apenas a execução dos áudios das amostras de fala e permitia que os sujeitos repetissem o áudio quando quisessem. Os participantes deveriam anotar suas respostas em uma folha. Também disponibilizamos uma segunda folha com o mapa geográfico brasileiro com as legendas de cada estado. Todos os sujeitos realizaram o teste com fones de ouvido.

Após preencher a ficha de identificação com histórico residencial e cidade e Estado de origem, os sujeitos receberam as instruções para a realização do teste. Os sujeitos foram instruídos que precisariam informar qual a região ou Estado de origem de pequenas amostras de fala de sujeitos de diferentes regiões do Brasil. Era permitido repetir apenas uma vez cada trecho e os sujeitos deveriam escrever em uma folha a resposta dada. Informamos que as amostras de

fala poderiam ser de sujeitos de todo o Brasil. Após metade dos trechos de fala ter sido tocado, os ouvintes poderiam dar uma pausa e depois continuar o teste. Os ouvintes demoraram de 20 a 30 minutos para completar o teste.

Os resultados obtidos após aplicação dos testes perceptivos serão apresentados na seção 4.3.

4. Resultados

4.1. Parâmetros prosódico-acústicos

A análise estatística das medidas obtidas apontou para os seguintes parâmetros com diferenças estatisticamente significativas ($p < 0.05$) entre as variedades: ênfase espectral, mediana de frequência fundamental, assimetria de z-score suavizado e taxa de unidades VV não proeminentes (Tabela 2). Para o parâmetro desvio-padrão de z-score suavizado, a ANOVA apontou diferenças significativas entre as variedades, porém, o teste Tukey, por ser mais conservador, não conseguiu apontar tais diferenças com o nível de significância adotado. Por isso, para esse parâmetro, adotamos um nível de significância de 0.15, que apontou diferenças marginalmente significativas entre as variedades.

Parâmetro prosódico- Acústico	Valor de p
ênfase spectral	<0,001*
mediana de F0	<0,001*
média de z-score suavizado	0.2
taxa de proeminência	0.8
Desvio-padrão de z-score suavizado	0.01*
assimetria de z-score suavizado	0.04*
taxa de elocução	0.09
taxa de unidades VV não proeminentes/s	0.02*

Tabela 2. Resultados obtidos para os parâmetros prosódico-acústicos estudados. Os valores seguidos de (*) são os estatisticamente significativos ($p < 0.05$).

Os parâmetros que melhor diferenciaram as variedades faladas nas regiões, criando grandes grupos entre eles, foram a ênfase espectral e a mediana de frequência fundamental.

A ênfase espectral conseguiu separar as variedades estudadas em dois grupos: DF e Região Norte, formaram um só grupo que se diferenciou de todas as outras variedades estudadas (exceto pela não diferenciação de DF e PR).

Primeiramente, serão apresentadas as divisões das regiões de acordo com os resultados obtidos após o teste estatístico post-hoc Tukey HSD.

O esquema abaixo mostra como as regiões se separaram em relação ao parâmetro ênfase-espectral.

$$1) DF \begin{cases} SP \\ MG \\ RJ \\ NE \end{cases} \quad 2) \text{Região Norte (N)} \begin{cases} SP \\ MG \\ RJ \\ NE \\ PR \end{cases} \quad 3) DF = PR$$

Figura 9. Esquema representando a divisão das variedades estudadas de acordo com os resultados obtidos para o parâmetro ênfase espectral após utilização do teste post-hoc Tukey HSD.

Pelos resultados, podemos concluir, ao menos inicialmente, que DF e N utilizam maior esforço vocal (correlato da ênfase espectral) que os falantes de outras regiões brasileiras, pois esse parâmetro é capaz de diferenciar os falantes dessas regiões de outras, com valores médios maiores, que podem ser observados na Figura 10 (linha horizontal em negrito no interior da caixa). O dialeto falado na Região Nordeste (NE) apresentou os menores valores médios de ênfase espectral e pequena variação desses valores. Concluimos que nesse *corpus*, a variedade falada no NE utiliza pouco esforço vocal, diferenciando-o de N e DF. Ainda em relação à ênfase espectral, assim como não se diferenciou da região Norte, o DF também não se diferenciou do PR.

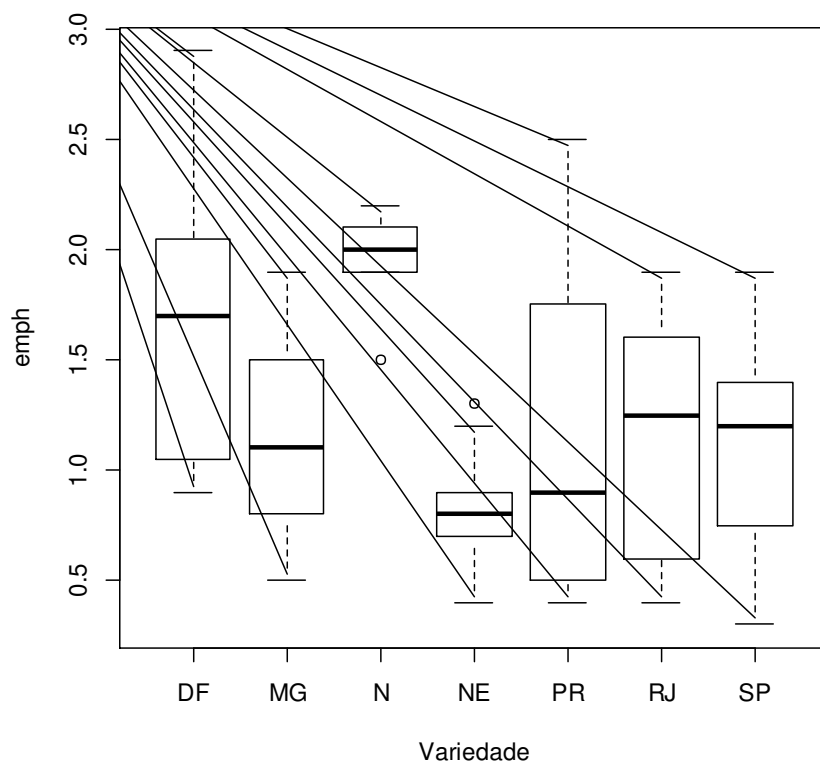


Figura 10. Boxplot mostrando os valores obtidos para o parâmetro ênfase espectral (emph) para as variedades estudadas.

A mediana da frequência fundamental permitiu que as variedades fossem separadas em dois grupos, bem semelhantes aos grupos separados pela ênfase espectral. O resultado encontrado foi que DF e N, mais uma vez, não apresentaram diferenças entre si e se diferenciaram de todas as outras regiões estudadas, como podemos ver na figura 11.

$$DF \& N \left\{ \begin{array}{l} MG \\ NE \\ PR \\ RJ \\ SP \end{array} \right.$$

Figura 11. Esquema representando a divisão das variedades estudadas de acordo com os resultados obtidos para o parâmetro mediana de frequência fundamental, após utilização do teste post-hoc Tukey HSD.

A figura 12 mostra que N e DF apresentam valores mais elevados para este parâmetro do que as outras regiões estudadas, em torno de 140 a 150 Hz. De acordo com Behlau (2001), a faixa de frequência fundamental para homens adultos se concentra entre 90 Hz a 150 Hz. O estudo de Camargo et. al (2012) verificou que a média da frequência fundamental masculina é de 127 Hz, sendo que este parâmetro foi estatisticamente significativo na comparação entre falantes do sexo feminino e do sexo masculino. A média de F0 para os falantes do sexo feminino foi 208 Hz. Assim, os falantes do N e DF estão situados entre os valores limítrofes se considerarmos a faixa de 90 a 150 Hz e estão 20 Hz acima da média dos falantes masculinos brasileiros, de acordo com Camargo et.al.(2012). Além disso, N e DF são as variedades que apresentam os maiores desvios-padrão da frequência fundamental, como mostra a tabela 3.

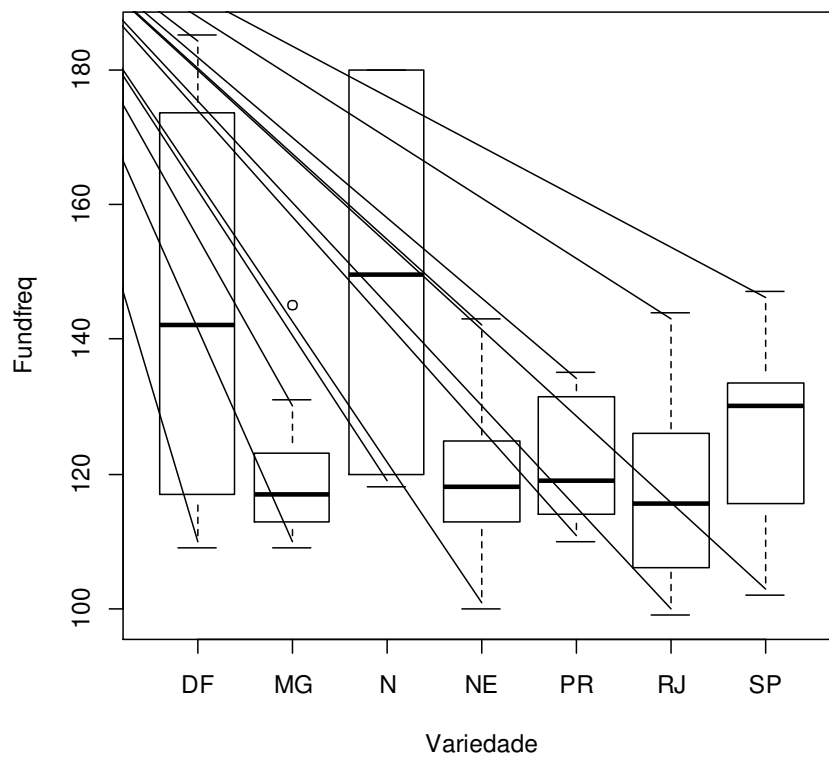


Figura 12. Boxplot mostrando os valores obtidos para o parâmetro mediana de frequência fundamental para as variedades estudadas.

Variedade	Mediana de F0	
	Média (Hz)	Desvio-Padrão (Hz)
DF	144	29
MG	119	9
N	149	32
NE	119	11
PR	121	10
RJ	117	14
SP	125	13

Tabela 3. Valores de médias e desvio-padrão de mediana de frequência fundamental para variedades estudadas.

Se por um lado a ênfase espectral e mediana de frequência fundamental são parâmetros capazes de colocar DF e N em um mesmo grupo, por outro lado, outros dois parâmetros que também apresentaram significância estatística, a taxa de unidades VV não proeminentes e a assimetria de z-score suavizado, colocaram essas duas regiões em grupos distintos.

A figura 13, que ilustra os resultados obtidos para a assimetria de z-score suavizado, mostra que N tem os maiores valores de assimetria positiva, o que pode indicar que os falantes da região aumentam mais a duração de algumas unidades VV do que DF (que tem os menores valores de assimetria positiva de todos os dialetos estudados), isto é, os falantes da região Norte fazem mais alongamentos na duração das unidades, o que resulta na assimetria de valor positivo mais elevada.

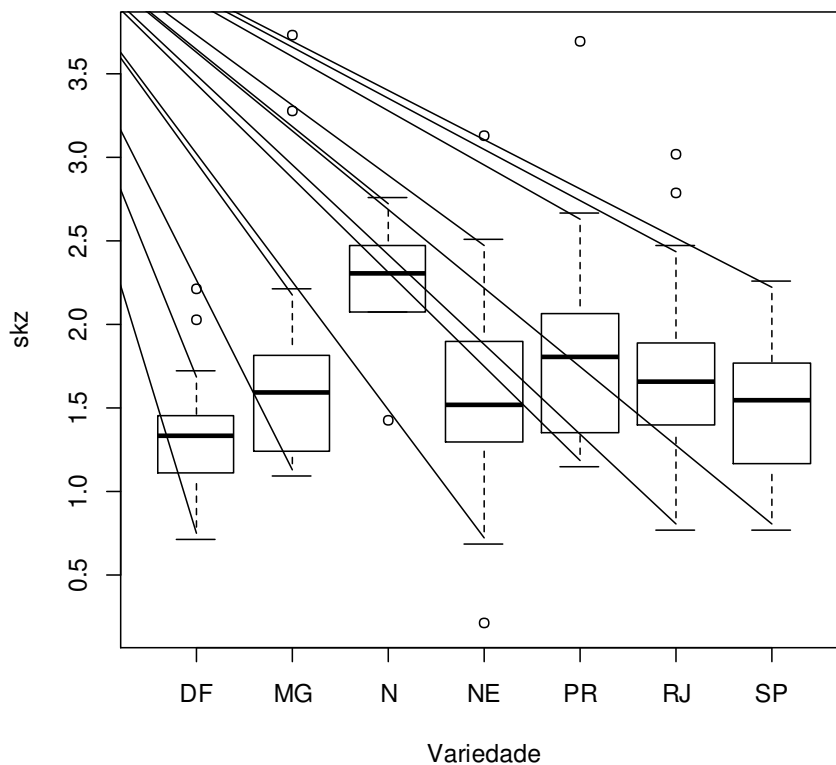


Figura 13. Boxplot mostrando os valores obtidos para o parâmetro assimetria de z-score suavizado nas variedades estudadas.

Os resultados referentes à taxa de unidades VV não proeminentes podem estar relacionados com a cadência da fala, visto que este parâmetro prosódico acústico é próximo à taxa de articulação, que por sua vez, é o número de unidades linguísticas presentes em um trecho de fala, excluindo-se as pausas deste trecho. Os resultados obtidos distinguiram a variedade falada do N das variedades de DF e SP. A Figura 14 mostra que os falantes do Norte (com maior taxa de unidades VV não proeminentes) apresentam maior cadência de fala que DF e SP.

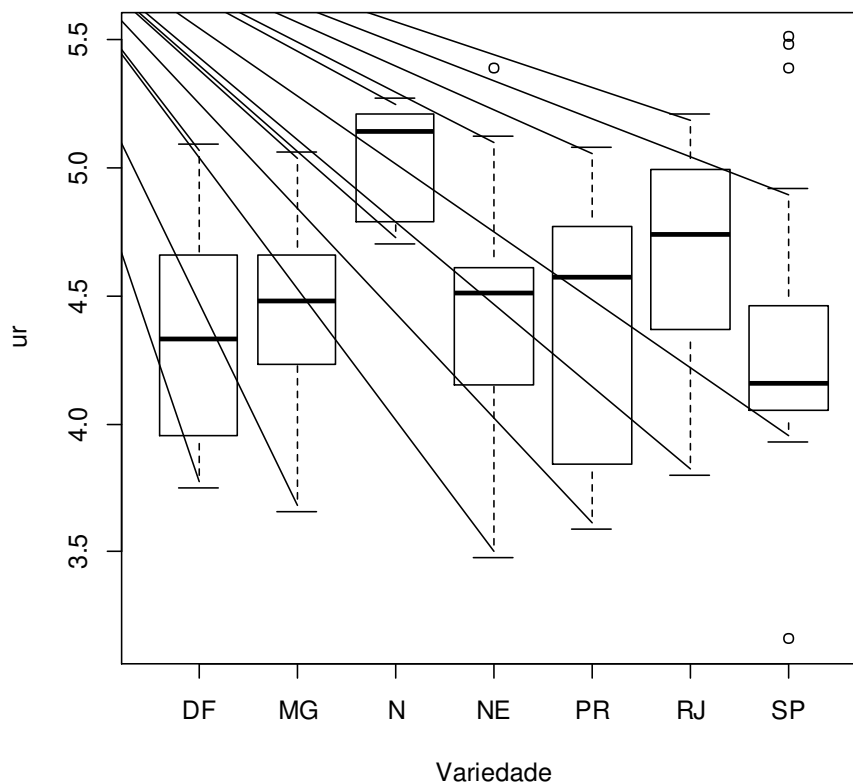


Figura 14. Boxplot mostrando os valores obtidos para taxa de unidades VV não proeminentes/s para as variedades estudadas.

O termo “cadência” utilizado no parágrafo anterior foi utilizado também por Nascentes (1953) para distinguir os falares do Norte dos falares do Sul. Porém, como na descrição apresentada pelo autor, a cadência, naquele caso, está relacionada à percepção do autor sobre as diferenças entre estes falares no que diz respeito à fala cantada (característica do Norte) *versus* a fala descansada (característica do Sul). No que diz respeito ao nosso uso do termo, a cadência está relacionada à taxa de unidades VV não proeminentes, estas últimas são conhecidas por apresentarem menor variação em suas durações, funcionando como uma espécie de “marca-passo” para o ritmo da fala. As unidades proeminentes são responsáveis, por sua vez, pelas

saliências percebidas e também apresentam maior variação de suas durações (Barbosa, 2006). Apesar da cadência utilizada como critério por Nascentes ser baseado em sua experiência e percepção, nossas medidas de taxa de unidades VV não proeminentes apresentaram resultados similares aos de Nascentes, ao separar os sujeitos da região Norte de sujeitos de SP.

O desvio-padrão de z-score suavizado apontou diferenças marginalmente significativas, tomando como nível de significância o p de 0.15. O esquema abaixo mostra como se deu a diferenciação das variedades com a análise desse parâmetro.

$$N \begin{cases} DF \\ NE \\ SP \end{cases} \qquad SP \begin{cases} PR \\ N \end{cases}$$

Figura 15. Esquema representando a divisão das variedades estudadas de acordo com os resultados obtidos para o parâmetro desvio-padrão de z-score suavizado, após utilização do teste post-hoc Tukey HSD.

A análise do desvio-padrão de z-score aponta para uma diferença entre as variedades faladas na região Norte do Brasil e da Região Sul. Vemos que o dialeto falado na Região Norte se diferenciou de SP, DF e também do Nordeste. A variedade falada em SP se diferenciou da variedade falada no PR. Os resultados em relação a esse parâmetro indicam que a região Norte apresenta maior variação da duração das unidades VV que DF, NE e SP, e estas três últimas variedades apresentaram valores baixos de desvio-padrão. Além disso, a variedade de SP apresentou menos desvio-padrão das unidades VV do que PR (Figura 16).

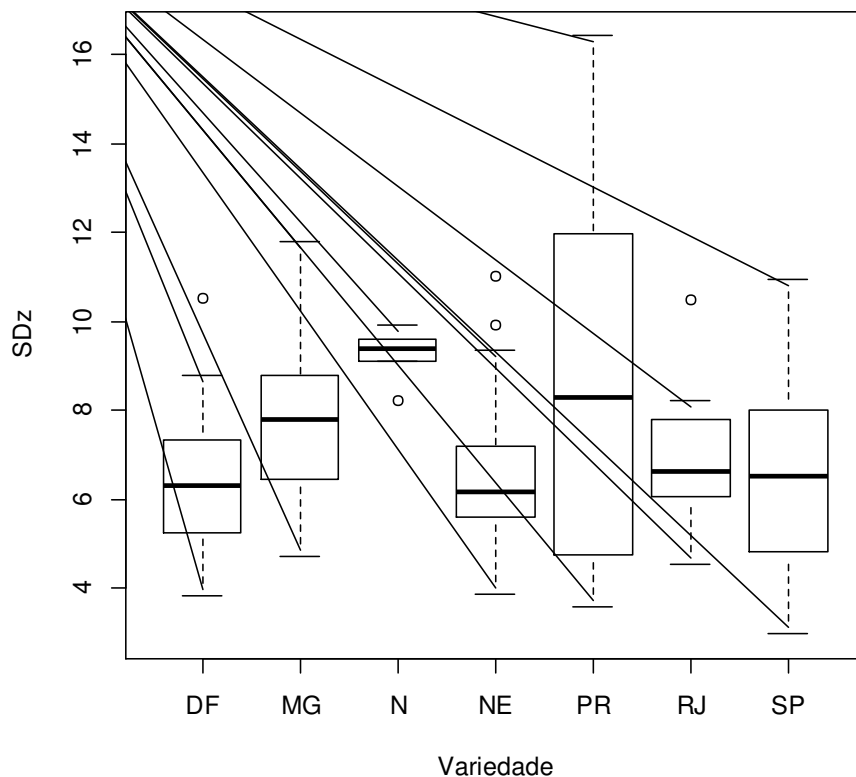


Figura 16. Boxplot mostrando os valores obtidos para o desvio-padrão de z-score suavizado para as variedades estudadas.

Finalmente, os parâmetros (i) média de z-score suavizado de duração de unidade VV, (ii) taxa de elocução e (iii) taxa de saliência duracional não apresentaram diferenças significativas entre as variedades estudadas. Assim, concluímos que, dos oito parâmetros prosódico-acústicos que selecionamos, cinco foram capazes de distinguir algumas regiões dialetais de outras (ênfase espectral, mediana de F0, assimetria de z-score suavizado, taxa de unidades VV não proeminentes e desvio-padrão de z-score suavizado), o que confirma a hipótese de que os parâmetros escolhidos são eficientes para caracterizar a fala de algumas regiões brasileiras.

A partir deste momento, serão apresentados os resultados referentes aos agrupamentos feitos pelo segundo teste post-hoc aplicado: o teste Duncan ($p < 0.05$). Mostraremos agora como este teste post-hoc separou as variedades entre si, lembrando que, para este teste, é possível que existam grupos intermediários entre dois grandes grupos, como será exemplificado a seguir.

Para a ênfase espectral, foram criados dois grupos (A e B), assim como nos resultados obtidos com o teste TukeyHSD, que separaram N e DF das outras regiões dialetais, porém, desta vez, a variedade falada no PR fica em um grupo separado do DF.

Ênfase Espectral (dB)		
Grupos	Variedades	Média
A	N	1.95
A	DF	1.69
B	PR	1.16
B	MG	1.13
B	RJ	1.11
B	SP	1.1
B	NE	0.79

Tabela 4. Resultados do teste Duncan para o parâmetro ênfase espectral dividindo em grupos as variedades estudadas.

A mediana da frequência fundamental apontou para os mesmos resultados obtidos anteriormente (tabela 5).

Mediana de Frequência Fundamental (Hz)		
Grupos	Variedades	Média
A	N	149.5
A	DF	144.6
B	PR	125.2
B	MG	121.8
B	RJ	119.9
B	SP	119.3
B	NE	117.3

Tabela 5. Resultados do teste Duncan para mediana de frequência fundamental dividindo em grupos as variedades estudadas.

O parâmetro desvio-padrão de z-score suavizado criou dois grupos (A e B) e um grupo intermediário (AB).

Desvio-padrão de z-score suavizado de duração VV		
Grupos	Variedades	Média
A	N	9.26
AB	PR	8.67
AB	MG	7.78
B	RJ	6.76
B	NE	6.59
B	DF	6.46
B	SP	6.44

Tabela 6. Resultados do teste Duncan para desvio-padrão de z-score suavizado de duração de unidade VV dividindo em grupos as variedades estudadas.

Os resultados apresentados na tabela 6 mostram que a variedade do N é a única representante do grupo A, que tem os maiores valores de desvio-padrão. Quanto maior o valor do desvio-padrão, maior a variação encontrada entre as durações das unidades VV. O grupo intermediário AB é constituído de PR e MG e o grupo B tem como representantes RJ, NE, DF e SP. Este parâmetro apresentou diferenças nos resultados obtidos entre os dois testes post-hoc aplicados. O teste Duncan apontou para a criação do grupo intermediário AB, que pode mostrar

uma tendência de PR e MG a apresentarem valores maiores de desvio-padrão, porém ainda se diferenciando de N e não chegando a valores mais baixos como os do grupo B.

A assimetria de z-score suavizado também apresentou os mesmo resultados apontados pelo TukeyHSD. O grupo A, mais uma vez, com valores de média maiores, é formado apenas pela Região Norte. O grupo intermediário AB inclui as variedades de PR, MG, RJ e NE e o grupo B, com os menores valores obtidos, inclui SP e DF.

Assimetria de z-score suavizado de duração VV		
Grupos	Variedades	Média
A	N	2.22
AB	PR	1.89
AB	MG	1.74
AB	RJ	1.73
AB	NE	1.61
B	SP	1.49
B	DF	1.36

Tabela 7. Resultados do teste Duncan para o parâmetro assimetria de z-score suavizado dividindo em grupos as variedades estudadas.

A taxa de unidades VV não proeminentes também apresentou diferença significativa no teste Duncan e novamente, no grupo A, temos a Região Norte, com a maior taxa de unidades VV não proeminentes/s. O grupo intermediário AB tem como representante apenas RJ e o grupo B é composto por todas as outras regiões: MG, NE, PR, SP e DF, que apresentam as menores taxas.

Taxa de unidades VV não proeminentes/s		
Grupos	Variedades	Média
A	N	2.22
AB	RJ	1.89
B	MG	1.74
B	NE	1.73
B	PR	1.61
B	SP	1.49
B	DF	1.36

Tabela 8. Resultados do teste Duncan para o parâmetro taxa de unidades VV não proeminentes/s dividindo em grupos as variedades estudadas.

Os seguintes parâmetros não apresentaram diferenças estatisticamente significativas tanto para o teste Tukey quanto para o teste Duncan: média de z-score suavizado de unidade VV, taxa de proeminência e taxa de saliência duracional.

O método estatístico Linear Discriminant Analysis (LDA) foi empregado com a finalidade de saber se o conjunto de parâmetros prosódicos-acústicos analisados poderia prever e discriminar corretamente cada uma das variedades estudadas. Os resultados obtidos estão na tabela 9, que mostra a porcentagem de acerto da técnica ao prever cada uma das variedades.

Variedade	Acerto
PR	75%
NE	66,60%
N	66,60%
DF	66,60%
RJ	50%
MG	38,20%
SP	31,60%

Tabela 9. Resultados da LDA apontando a porcentagem de acerto da técnica para cada variedade estudada.

A tabela 9 nos mostra que PR, com 75% de acerto após utilização do LDA, é a região que é mais corretamente discriminada ao tomarmos como base os resultados obtidos como um todo, como um conjunto, dos oito parâmetros analisados. Outras regiões bem discriminadas pelo conjunto de resultados obtidos foram Região Norte, DF e Nordeste. A região dialetal de SP foi a que teve a menor porcentagem de discriminação, com 31,6%.

A tabela 9, apresentada anteriormente, mostra os resultados da LDA que levaram em conta o conjunto dos oito parâmetros selecionados para análise dos sujeitos. Para saber como cada parâmetro se comportava individualmente ao tentar discriminar cada variedade, foi aplicada a técnica LDA para cada parâmetro. Os resultados estão na tabela 10.

Variedade	Emph	F0	Mz	Pr	SDz	Skz	SR	ur
NE	85%	71%	47%	71%	85%	52%	66%	57%
SP	31%	57%	31%	0	5%	26%	31%	15%
N	0%	0%	0	0	0	0	0	0
DF	46%	53%	0	0	0	6%	0	0
MG	16%	0%	22%	0	44%	11%	0	0
PR	0%	0%	50%	0	33%	0	0	0
RJ	0%	16%	0	44%	0	0	11%	61%

Tabela 10. Resultados da LDA utilizando cada parâmetro individualmente para medir o acerto da técnica para cada variedade estudada.

O método LDA, aplicado individualmente para cada parâmetro prosódico-acústico estudado (Tabela 10), mostrou que cada parâmetro consegue discriminar de duas a quatro variedades, embora nem sempre com muita precisão. A região Nordeste é a única região que sempre é discriminada por todos os parâmetros, sendo que a menor porcentagem de discriminação é 47% no parâmetro média de z-score suavizado de unidade VV. Para a região Nordeste, a ênfase espectral e o desvio-padrão de z-score foram os melhores identificadores, com 85% de acerto cada. A mediana da frequência fundamental conseguiu discriminar variedades

faladas em três localidades com mais de 50% de acerto, mostrando-se um parâmetro importante para contribuir com a identificação das variedades. Os falantes de SP também foram identificados por quase todos os parâmetros individualmente, porém a porcentagem de discriminação correta é pequena, variando de 5% a 57%.

Houve diferença ao compararmos os resultados da LDA aplicada individualmente e com a combinação de todos os parâmetros. A região Norte, por exemplo, não foi identificada quando a LDA foi aplicada para cada parâmetro prosódico individualmente. No entanto, a combinação de todos esses fatores aumenta a previsão da Região Norte para 66.6%. Isto mostra que, a combinação das medidas prosódico-acústicas é mais vantajosa para algumas variedades, ressaltando o impacto da intensidade relativa, frequência fundamental e medidas de duração silábica, analisados em conjunto, para a caracterização de uma área dialetal. Por outro lado, a região Nordeste apresentou bons resultados tanto da análise em conjunto (66.60%) quando da análise individual de cada parâmetro, com alto índice de acerto do método.

4.2.Diferentes relações sinal-ruído.

Após extração e análise dos parâmetros prosódico-acústicos obtidos nas gravações com ruído aditivo, observamos mudanças significativas nos valores calculados para alguns parâmetros. São eles: ênfase espectral, mediana de frequência fundamental, taxa de saliência duracional, média, desvio-padrão e assimetria de z-score de duração de unidade VV. No entanto, é importante ressaltar que os quatro últimos parâmetros citados apresentaram mudanças somente em seis trechos de fala (o total de trechos de fala analisados, considerando todos os sujeitos foi de 109). A ênfase espectral e a mediana de frequência fundamental apresentaram mudanças em

todos os trechos analisados, o que é esperado, visto que a adição do ruído interfere na captação dessas duas medidas especificamente.

As mudanças observadas nos parâmetros taxa de saliência duracional e média, desvio-padrão e assimetria de z-score suavizado de duração de unidade VV ocorreram devido ao efeito natural que o ruído causa na extração de medidas através de fórmulas do Praat. O efeito do ruído dificulta a extração das medidas, influenciando principalmente a mediana de frequência fundamental e ênfase espectral. É preciso considerar que, para os outros parâmetros que apresentaram mudanças nos valores extraídos, a dificuldade apresentada pelo *script* é mínima, já que as mudanças ocorreram em poucos trechos e com mudanças pouco significativas (da ordem de uma casa decimal).

A tabela 11 apresenta os resultados obtidos na análise estatística (Teste T, $p < 0.016$) que considerou três situações de comparação entre as gravações com e sem ruído (devidamente explicadas na seção 3.5.1.).

	Ruído Gaussiano 0,01	Ruído Gaussiano 0,02
Mediana de F0		
Alta Relação Sinal-Ruído	$p < 0,001 *$	$p < 0,001 *$
Ruído Gaussiano 0,01	-	$p < 0,001 *$
Ênfase Espectral		
Alta Relação Sinal-Ruído	$p < 0,001 *$	$p < 0,001 *$
Ruído Gaussiano 0,01	-	$p < 0,001 *$

Tabela 11. Resultados do Teste T para as análises das diferentes relações sinal-ruído estudadas.

* $p < 0.016$

A tabela 11 evidencia que, em todas as situações possíveis de análise, a estatística aponta para diferenças significativas entre os valores obtidos. Mesmo assim, é interessante saber de que

ordem são as mudanças observadas nos valores dos dois parâmetros em questão (ênfase espectral e mediana de frequência fundamental). Para isso, calculamos a média da mediana de F0 e a média dos valores de ênfase espectral nos três momentos: gravação original, gravação com ruído aditivo de magnitude 0.01 e gravação com ruído aditivo de magnitude 0.02. Os dados se encontram na tabela 12.

	Gravação original	Ruído 0,01	Magnitude de mudança	Ruído 0,02	Magnitude de mudança
M_{mediana f0}	125 Hz	126 Hz	0,8%	128 Hz	2,4%
M_{ênfase espectral}	1.18 dB	1.8 dB	55%	3 dB	154%

Tabela 12. Médias de ênfase espectral e de mediana de frequência fundamental e magnitudes de mudança após adição de ruído.

Na tabela 12 observamos que a mudança encontrada na ênfase espectral foi mais brusca que na mediana de F0, chegando a apresentar um aumento de 154% em relação à gravação original, quando há aumento na magnitude do ruído aditivo. A maior mudança observada na frequência fundamental foi de 3 Hz, que apesar de ser estatisticamente significativo, possivelmente não interfere na identificação de um sujeito, por exemplo.

4.3. Testes perceptivos

A tabela 13, abaixo, apresenta a quantidade de acerto de cada um dos estímulos apresentados aos participantes do teste perceptivo. Os estímulos sonoros marcados em vermelho e com asteriscos tiveram mais de 90% de respostas corretas. É importante lembrar que os participantes poderiam dar como resposta a região ou o estado de origem do falante da amostra.

Stml	Variedade	N acerto	%
1	Norte	2	10
2	MG	13	65
3	PR	0	0
4	RJ	16	80
5	SP	13	65
6	DF	2	10
7	MG	12	60
8	MG	16	80
9	Nordeste	16	80
10	RJ	16	60
11	SP	13	65
12	DF	1	5
13	MG	7	35
14	Norte	1	5
15	PR	3	15
16	RJ	19	95**
17	MG	4	20
18	Nordeste	17	85
19	RJ	16	80
20	Nordeste	1	5
21	DF	4	20
22	Nordeste	17	85
23	RJ	16	80
24	SP	13	65
25	Norte	0	0
26	MG	10	50
27	RJ	20	100**
28	SP	9	45
29	PR	2	10
30	Norte	1	5
31	RJ	19	95**
32	RJ	17	85
33	SP	15	75
34	MG	11	55
35	PR	4	20
36	Nordeste	13	65
37	DF	1	5
38	DF	1	5
39	SP	4	20

40	PR	0	0
41	DF	3	15
42	SP	12	60
43	DF	4	20
44	DF	3	15
45	Nordeste	8	40
46	MG	9	45
47	Nordeste	2	10
48	Nordeste	14	70
49	SP	4	20
50	RJ	10	50
51	PR	4	20
52	DF	3	15
53	SP	11	55
54	Nordeste	18	90**
55	SP	15	75
56	MG	10	50
57	DF	3	15
58	Norte	4	20
59	PR	7	35
60	MG	17	85
61	DF	1	5
62	Norte	4	20
63	Nordeste	18	90**
64	Nordeste	9	45
65	PR	0	0
66	SP	8	40
67	RJ	10	50
68	PR	8	40
69	MG	14	70
70	RJ	18	90**
71	Nordeste	0	0
72	RJ	10	50
73	PR	7	35
74	SP	12	60
75	DF	0	0

Tabela 13. Resultados do teste perceptivo.

Como é possível observar, apenas seis estímulos apresentaram porcentagem de acerto de 90% ou mais, sendo que apenas um deles chegou a 100% de identificação correta de estado/região de procedência. Quatro dos seis trechos mais corretamente identificados são de sujeitos do RJ e dois são do Nordeste.

Para produzir a tabela 14, levamos em conta o total de estímulos de cada região e o máximo de respostas possíveis por estímulo (total de 20 respostas para cada estímulo), sendo assim, a Região Nordeste, por exemplo, recebeu 240 classificações dos sujeitos que responderam ao teste (12 estímulos vezes 20 respostas possíveis por estímulo). Do total de 240 respostas possíveis, calculamos a porcentagem de repostas corretas.

	Total estímulos	% acerto
RJ	12	77.9
Nordeste	12	55.8
SP	12	53.7
DF	12	10.8
MG	11	55.9
PR	10	14.5
Norte	6	10

Tabela 14. Porcentagem de acerto dos ouvintes no teste perceptivo para cada variedade estudada.

A variedade falada no Rio de Janeiro foi a que obteve maior porcentagem de respostas corretas, com 77.9%, seja com a identificação Sudeste ou com a identificação RJ. A região Nordeste, SP e MG tiveram resultados semelhantes, em torno de 50% de identificações corretas. A variedade do PR obteve 14.5% das respostas corretas, seguida por DF e região Norte. Deve-se levar em conta o fato de que o bom índice de repostas corretas que os participantes do teste perceptivo deram para os falantes de São Paulo pode estar relacionado com a familiarização que eles tem com o falar de São Paulo, visto que todos os sujeitos que responderam ao teste, como

dito anteriormente, eram de diferentes regiões brasileiras mas na época de aplicação do teste estavam residindo no estado de São Paulo.

Com a finalidade de saber com quais variedades as três regiões menos identificadas (DF, PR e região Norte) foram confundidas, analisamos todas as respostas dadas pelos participantes que responderam ao teste. Os resultados encontram-se nas tabelas 15, 16 e 17.

Resposta obtida	DF	
	Quantidade de respostas	%
Nordeste	59	24.5%
MG	36	15.0%
Sudeste	32	13.3%
RJ	24	10.0%
Norte	19	7.9%
SP	14	5.8%
Sul	7	2.9%
SC	5	2.0%
PR	4	1.6%
ES	4	1.6%
MT	1	0.4%
RS	1	0.4%

Tabela 15. Respostas obtidas pelos participantes nos trechos de fala de sujeitos do DF.

PR		
Resposta Obtida	Quantidade de respostas	%
SP	45	22.5%
Sudeste	34	17.0%
RJ	19	9.5%
RS	8	4.0%
MG	7	3.5%
Norte	5	2.5%
Centro-Oeste	4	2.0%
SC	1	0.5%
DF	1	0.5%
Tocantins	1	0.5%
ES	1	0.5%
Nordeste	1	0.5%

Tabela 16. Respostas obtidas pelos participantes nos trechos de fala de sujeitos do PR.

Região Norte		
Resposta obtida	Quantidade de respostas	%
Nordeste	54	45.0%
Sudeste	12	10.0%
MG	11	9.1%
CO	7	5.8%
SP	4	3.3%
Goiás	2	1.6%
PR	1	0.8%
MT	1	0.8%
SC	1	0.8%

Tabela 17. Respostas obtidas pelos participantes nos trechos de fala de sujeitos da região Norte

Os falantes do DF foram mais confundidos com falantes da variedade falada no Nordeste (24.5%), porém, outros 11 tipos de respostas foram dados para classificar os falantes do Distrito Federal, mostrando que houve grande variedade nas respostas dadas pelos sujeitos. Se considerarmos que a variedade do DF foi classificada como MG, Sudeste e RJ (respectivamente

o segundo, terceiro e quarto colocados na quantidade de respostas obtidas), podemos pensar que, na maioria das vezes, o DF é confundido com as variedades faladas na região Sudeste.

Em relação ao PR, os participantes do teste perceptivo confundiram a variedade falada na região principalmente com variedades da região Sudeste, sendo que as três respostas mais frequentes foram SP, Sudeste e RJ. Este achado pode ser explicado pelo fato de que o estado do Paraná é vizinho do estado de São Paulo, e com certeza, algumas regiões do PR acabam sofrendo influência da variedade falada em SP (e vice-versa), o que faz com que as duas variedades sejam parecidas e possam até ser confundidas, visto que oito ouvintes do estado de São Paulo responderam ao teste e acabaram por classificar os falantes do PR como SP.

A tabela 17, referente às respostas obtidas quando o estímulo tocado pertencia a sujeitos da região Norte, mostra que, em 54% das vezes, os falantes da região Norte foram classificados como sendo da região Nordeste. A variação nas respostas para a região Norte foi um pouco menor que para as outras duas variedades mostradas anteriormente, com nove tipos de respostas diferentes.

Nenhum sujeito da região Norte e do Paraná respondeu ao teste perceptivo. Este fato pode ter contribuído para a quantidade inferior de acertos e maior quantidade de classificações inadequadas. Apenas um sujeito da região Centro-Oeste, mais precisamente de Goiânia, respondeu ao teste e, analisando suas respostas, observamos que o sujeito identificou corretamente três trechos de fala de sujeitos do DF, de um total de 12 trechos presentes no teste. Os trechos correspondiam à fala de três diferentes sujeitos do DF.

Um dos sujeitos que respondeu ao teste perceptivo é da cidade de São Luís, no Maranhão, estado que é muito próximo da região Norte do país. De seis estímulos de fala de

sujeitos da Região Norte, o ouvinte do Maranhão conseguiu identificar dois corretamente e, nos outros quatro estímulos faltantes, o ouvinte classificou três como sendo de sujeitos procedentes ao Maranhão e um como procedente de MG. Este fato pode mostrar que o ouvinte do Maranhão considera a variedade falada em seu estado muito próxima da variedade falada na região Norte, o que é esperado, visto que são regiões vizinhas.

5. Discussão

5.1. Parâmetros prosódico-acústicos na identificação de diferentes dialetos brasileiros.

Dos oito parâmetros analisados, cinco foram capazes de separar algumas das sete variedades estudadas. Foram eles: ênfase espectral, mediana de frequência fundamental, assimetria de z-score suavizado de duração de unidade VV, taxa de unidades VV não-proeminentes/s e desvio-padrão de z-score suavizado de duração de unidade VV. Os resultados mostram que a prosódia é uma informação fônica que é capaz de distinguir, junto com a fonologia e as diferenças de léxico, por exemplo, as diferentes variedades faladas no território brasileiro. Dos cinco parâmetros que apresentaram resultados positivos, três dizem respeito à estrutura rítmica das variedades estudadas e os outros dois referem-se à frequência fundamental e à intensidade, o que mostra que os três parâmetros clássicos de estudo da prosódia apresentam diferenças entre as variedades faladas no Brasil.

A mediana de frequência fundamental e a ênfase espectral foram responsáveis pela criação de dois grandes grupos, separando região Norte e DF das outras variedades: SP, MG, RJ, NE e PR (o caso específico da não diferenciação de DF e PR será explicado mais adiante). Podemos correlacionar os resultados encontrados para mediana de F0 com os resultados obtidos para a ênfase espectral. De acordo com Traunmiller e Eriksson (2000), o esforço vocal afeta a pressão sonora e, por consequência, as propriedades acústicas do sinal de fala também são afetadas. Os sujeitos da região Norte e do DF foram os que apresentaram maiores valores de ênfase espectral e também de mediana de F0 (valores de F0 inclusive mais elevados que a média para o sexo masculino), e, segundo a explicação de Heldner (2003), as altas frequências contribuem relativamente para a intensidade total do espectrograma e o cálculo da energia nas

altas frequências é a própria ênfase espectral, que também é o correlato acústico do esforço vocal. Além disso, é sabido que a sensação de esforço vocal está intimamente ligada a dois parâmetros relacionados aos pulsos glóticos da onda sonora: o coeficiente de velocidade das pregas vocais (razão entre o tempo de duração da fase de abertura e o tempo de duração da fase de fechamento das pregas vocais) e o declínio espectral. Quando o coeficiente de velocidade das pregas vocais é alto, grande quantidade de energia é produzida nas altas frequências, causando a sensação de qualidade vocal tensa (Holmberg et.al. 1988; Childers e Lee, 1991; Holmberg et.al. 1995), assim, novamente, mostramos que o esforço vocal (e a ênfase espectral) está relacionado com altos valores de F0, o que explicaria a relação dos resultados obtidos para N e DF.

Dois parâmetros conseguiram separar a região Norte do DF, deixando as duas variedades em grupos diferentes. São eles: taxa de unidades VV não proeminentes e assimetria de z-score suavizado. Voltando aos parâmetros clássicos de estudo da prosódia, chama atenção o fato de DF e Norte apresentarem similaridades quanto a parâmetros relacionados à intensidade e à frequência e serem distintos em relação à duração, ou mais especificamente, ao ritmo.

Durante a descrição dos resultados, mostramos que DF e PR não se diferenciaram em relação ao parâmetro ênfase espectral. Ao analisarmos a Figura 10 (página 41), com os resultados para a ênfase espectral, é possível observar que a variedade falada no PR apresenta valores médios baixos, quase próximos aos encontrados para a região Nordeste, porém há grande variação no desvio-padrão encontrado. Assim, podemos concluir que os sujeitos do PR variaram mais ao utilizarem esforço vocal durante as amostras de fala estudadas, fato que contribuiu, durante a análise estatística, para a colocação desta variedade em um mesmo grupo que a variedade do DF, sendo que esta última apresenta valores médios maiores de esforço vocal que o PR, mas também com grande variação na utilização deste recurso.

O agrupamento de sujeitos de diferentes estados do Nordeste em um mesmo grupo foi justificado pela classificação proposta por Nascentes (1953), que coloca a Bahia, junto com Sergipe, como área distinta da variedade falada do restante do Nordeste. É evidente que existem traços fonológicos e de léxico que conseguem distinguir características específicas de cada sub-região da região Nordeste, porém, tratando-se de parâmetros prosódicos, podemos concluir sim, que as diferentes localidades da região Nordeste (exceto Bahia e Sergipe, que não foram analisadas neste trabalho) pertencem a um mesmo grupo. Uma das justificativas para esta afirmação está nos valores obtidos para o desvio-padrão das medidas analisadas nos falantes do Nordeste (Figuras 9, 11, 12 e 13), que mostram que os seis sujeitos do Nordeste apresentaram desvio-padrão igual ou menor que sujeitos de um mesmo estado geográfico, como por exemplo, SP. Outra justificativa está no fato de que o método estatístico LDA, utilizado para prever cada variedade, com base nos parâmetros estudados, mostrou que, individualmente, cada parâmetro consegue prever com boa porcentagem de acerto a variedade falada no Nordeste (Tabela 10, página 53), sendo que NE foi a única variedade estudada que conseguiu ser prevista corretamente por todos os parâmetros prosódico-acústicos.

Chama a atenção o fato de que a fala dos sujeitos de SP apresentou, muitas vezes, maior desvio-padrão das medidas analisadas que os sujeitos do NE. Os sujeitos de SP, conforme relatado na seção Metodologia, são procedentes de Piracicaba, Campinas, Jundiaí, Ribeirão Preto e São Paulo (capital). Uma hipótese para o desvio-padrão ser maior para os sujeitos da região de São Paulo pode ser devido ao fato de que, prosodicamente, a fala dos sujeitos das cidades analisadas varia bastante, mesmo em cidades próximas, como Jundiaí, Campinas e Piracicaba. Novos estudos investigando a prosódia de diferentes cidades do Estado de São Paulo poderiam

trazer novos dados de como os parâmetros prosódicos se comportam nestas localidades e explicar o alto índice de desvio-padrão encontrado em algumas medidas estudadas.

Em relação aos achados das características prosódicas para a região do DF, observamos que o fato de ter havido grande contato de diferentes variedades do português brasileiro na época da fundação e início da povoação do DF fez com que o desenvolvimento dessa nova variedade falada na região fosse, em alguns momentos semelhante à variedade do Norte e em alguns momentos, semelhante a variedades faladas no Sul do Brasil, como, por exemplo, em relação à taxa de unidades VV não proeminentes, em que o DF se assemelha a SP. Assim, concluímos que os falantes estudados do DF apresentam características prosódicas próprias que caracterizam a variedade falada na região, mostrando os resultados da mistura de variedades que ocorreu durante a formação do DF.

Um achado recorrente em todos os parâmetros analisados foi a separação da variedade falada pelos dois sujeitos da região Norte de todas as outras variedades analisadas. Em todos os parâmetros analisados que obtiveram significância estatística, tanto para o teste Tukey, quando para o teste Duncan, o Norte se diferenciou das variedades faladas no Sul, mostrando que a divisão feita por Nascentes em 1953 ainda é válida para os dias de hoje, pelo menos para a população estudada neste trabalho. A região Nordeste, que na divisão de Nascentes pertence aos “falares do Norte”, acabou não se diferenciando (em relação à prosódia) das variedades faladas no sul do Brasil e diferenciou-se da região Norte. Uma hipótese pode ser a de que, prosodicamente, a região Nordeste é similar ao sul do Brasil, ficando com a fonologia o papel de distinguir essas duas regiões. Por outro lado, Marroquim, em seu livro *“A língua do Nordeste”* (1934) utiliza partituras musicais para comparar e revelar diferenças entre a entonação da fala carioca (Figura 17) e a fala de Alagoas (Figura 18).

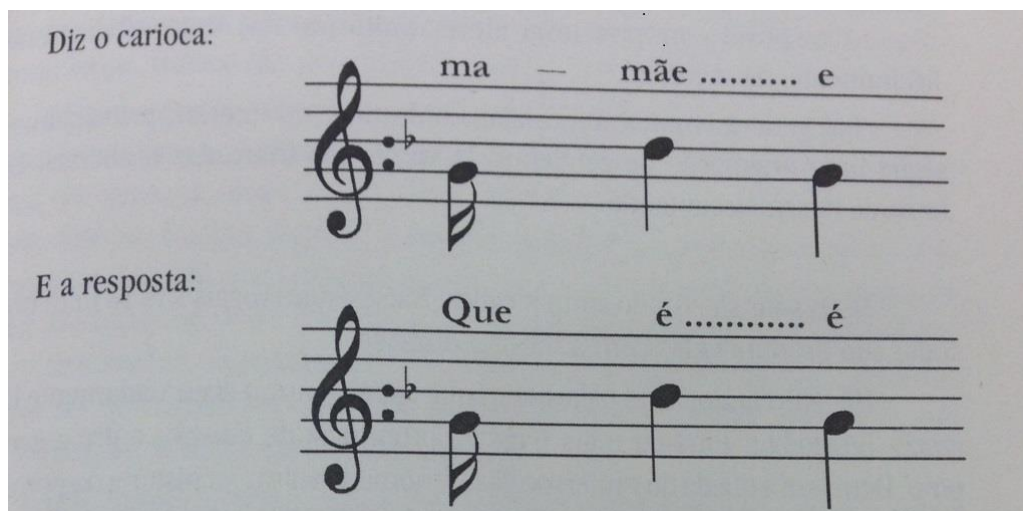


Figura 17. Partitura da fala carioca em ilustração retirada de Marroquim (1934).

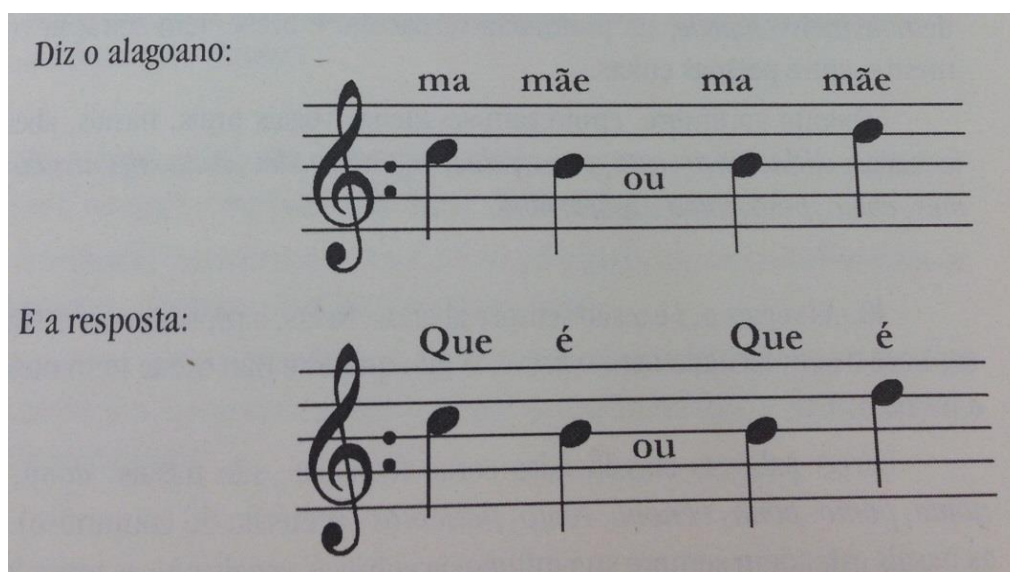


Figura 18. Partitura da fala alagoana em ilustração retirada de Marroquim (1934).

De acordo com Marroquim, a possível distinção em relação a entonação das duas falas se dá pois:

“há musicalidade na entonação carioca; a alagoana, entretanto, deixa maior impressão de fala cantada, porque as duas sílabas são pronunciadas mais vagarosamente, e têm o mesmo valor;

são duas semínimas com acento ligeiramente decrescente ou ascendente em terça com um *portamento* na voz”. (p.25; 1934)

As variedades faladas no Sul do Brasil, considerando a divisão de Nascentes, se constituíram em um grande bloco com poucas diferenciações entre as variedades faladas nesses locais. O destaque fica pela diferenciação de SP e PR em relação ao desvio-padrão de z-score suavizado, que apontou maior variação na duração das unidades VV na variedade falada no PR e menor variação na variedade falada em SP. O teste Duncan mostrou, em alguns parâmetros, a criação de grupos intermediários (grupos AB) e para três parâmetros (desvio-padrão de z-score, assimetria de z-score e taxa de unidades VV não proeminentes) foram criados grupos intermediários que eram compostos por variedades do sul do Brasil (Tabelas 6, 7 e 8), o que pode mostrar que, para este teste (que leva em conta a média de cada parâmetro), algumas variedades do sul parecem se diferenciar de outras, porém não suficientemente para criarem um grupo distinto.

O quadro 3, abaixo, sintetiza os resultados significativos encontrados para todos os parâmetros analisados, de acordo com o teste post-hoc Tukey HSD ($p < 0.05$). Após a apresentação do quadro, há uma explicação sobre os dados expostos.

Parâmetros	DF	MG	N	NE	PR	RJ	SP
ênfase spectral	+	-	+	-	+/-	-	-
frequência Fundamental	Valores límitrofes	Média	Valores límitrofes	Média	Média	Média	Média
assimetria de z-score de duração de unidade VV	unidades VV + longas		unidades VV + longas				
taxa de unidades VV não- proeminentes/s	-cadência		+ cadência				-cadência
desvio-padrão de duração de unidade VV	- Variação		+ Variação		+ Variação		- variação

Quadro 3. Síntese dos resultados obtidos em função da análise dos parâmetros prosódico-acústicos.

No quadro 3, para a ênfase espectral, correlato acústico do esforço vocal, o símbolo + significa que a variedade apresenta valores mais elevados, em média, para ênfase espectral, que, portanto, podem ser relacionados com maior esforço vocal apresentado pelos falantes das variedades. Sendo assim, nas variedades que contêm o símbolo - os falantes analisados apresentaram menor esforço vocal em suas amostras de fala. No caso do PR, o símbolo presente no quadro é +/- devido ao grande desvio padrão encontrado na variedade, o que fez com que, na análise estatística, a variedade ficasse em um mesmo grupo que a variedade do DF, porém, os falantes do PR apresentaram uma média baixa para o parâmetro em questão. Dessa forma, conclui-se que a média de ênfase espectral encontrada na variedade falada no PR tem valores baixos, porém seu desvio-padrão é alto, o que pode mostrar que alguns falantes acabam utilizando maior esforço vocal que outros nesta variedade.

Ainda no quadro 3, para a frequência fundamental, as variedades que estão marcadas com “Valores limítrofes” apresentam valores limítrofes à direita, ou seja, estão acima da média para o sexo masculino, que é de 127 Hz (Camargo et.al., 2012), e mais próximos ao limite com a frequência fundamental feminina, que tem média de 207 Hz (Camargo et.al., 2012). As outras variedades apresentam valores médios esperados para homens.

Para a assimetria de z-score, o quadro 3 destaca as variedades DF e N, que alongam mais as durações das suas unidades VV, fazendo com que a assimetria de valor positivo seja maior que nas outras variedades. Para o parâmetro taxa de proeminência de unidades VV/s, o quadro 3 aponta as três variedades que se diferenciaram em relação a esse parâmetro e mostra a variedade falada pelos sujeitos do Norte como tendo mais cadência que a variedade falada no NE e em SP. Por fim, o desvio-padrão de z-score suavizado de unidades VV diferenciou quatro variedades e colocou DF e SP como as variedades que menos variaram a duração das suas unidades VV e Norte e PR como as variedades que apresentaram maior variação.

O método estatístico LDA trouxe muita contribuição para o desenvolvimento desta pesquisa. A LDA mostrou que os resultados obtidos para cada parâmetro, quando considerados um conjunto, conseguiram prever com 75% de acerto a variedade falada pelos sujeitos do PR. Este resultado mostra que os oito parâmetros analisados em conjunto conseguem descrever bem a variedade falada no local e prever, pelos resultados de cada parâmetro, que a variedade em questão é a falada no PR. Quando a técnica foi utilizada individualmente para cada parâmetro analisado, a variedade falada pelos sujeitos do NE foi a mais corretamente identificada, com mínimo de acerto de 47% no parâmetro média de z-score de duração de unidade VV e com máximo de acerto de 85% para a ênfase espectral. O achado referente à variedade do NE reforça ainda mais a decisão metodológica tomada no início deste trabalho de colocar em um mesmo

grupo sujeitos de diferentes estados da região Nordeste, pois, de acordo com os resultados da LDA, cada parâmetro é capaz de prever, utilizando os resultados, que as medidas obtidas caracterizam a variedade do Nordeste.

5.2. Análise das diferentes relações sinal-ruído

O aumento da magnitude do ruído nas amostras de fala resultou em mudanças nos valores obtidos de dois parâmetros prosódico-acústicos: ênfase espectral e mediana de F0. Os dois parâmetros, no entanto, foram responsáveis por dividir dois grandes grupos na análise da individualização das características prosódicas de diferentes variedades faladas no Brasil (separando N e DF das outras variedades estudadas). Consideramos que os dois parâmetros são importantes ao individualizar características das variedades, porém a extração de suas medidas deve ser feita com cuidado quando há presença de ruído na amostra.

A tarefa de comparar duas amostras de fala, uma com baixa e outra com alta relação sinal-ruído é corriqueira na área da fonética forense (Rose 2002; Alexander, 2004) e por isso a análise realizada neste trabalho é de fundamental importância para contribuir para a resolução deste tipo de problema.

A frequência fundamental é apontada, juntamente com frequência dos formantes e o estilo de fala, como uma das principais pistas acústicas que os ouvintes utilizam para identificar outros falantes, segundo estudos como Atal (1972) e Kahn (2008). Na presença de ruído, o estudo de Alexander (2004) revelou que os ouvintes se baseiam no sotaque/pronúncia dos falantes para fazerem a identificação de outros falantes.

Nossos resultados mostraram que, mesmo com a análise estatística apontando para a significância das mudanças ocorridas na mediana de frequência fundamental, os valores médios

aumentaram no máximo 3 Hz (passando de 125 Hz na gravação original para 128 Hz na gravação com ruído), por isso podemos considerar que, para o tipo de ruído utilizado, mesmo com significância estatística, o aumento nos valores provavelmente não resultaria em um erro de análise, por exemplo, na identificação dos sujeitos.

A ênfase espectral vem sendo utilizada em pesquisas na área da análise prosódica por ser considerada um parâmetro mais confiável que a intensidade. Utilizar isoladamente o parâmetro intensidade pode diminuir a confiabilidade dos resultados, visto que, a intensidade pode ser altamente susceptível a efeitos relacionados à distância do sujeito ao microfone, por exemplo (Traunmiller e Eriksson 2000). Nossos resultados indicaram mudanças de até 154% de aumento dos valores de ênfase espectral quando há presença de ruído na amostra de fala e por isso seu uso com a finalidade de identificar falantes em amostras de fala em que há presença de ruído deve ser feito com cuidado ou mesmo evitado.

As figuras 6 (página 47) e 7 (página 48) mostram o espectrograma após a adição de ruído nas duas magnitudes (0.01 e 0.02) e podemos observar que muita informação é perdida. As figuras mostram redução na quantidade de harmônicos, principalmente nas regiões de altas frequências. Certamente, o impacto do ruído, que é visualizado no espectrograma, pode variar conforme o tipo de ruído presente na amostra de fala. Na figura 8 (p. 47), em que o espectrograma de banda larga é mostrado em mais detalhe, nota-se que, mesmo com a perda de informação acústica, ainda há como delimitarmos o início das vogais do trecho analisado, o que pode contribuir para as análises da identificação de um falante. Com a possibilidade de delimitação das vogais, a estrutura rítmica passa a ser uma escolha acertada para a análise de amostras de fala com ruído, sendo a duração das unidades VV um exemplo de possível análise.

De acordo com Barbosa e Silva (2011), a estrutura rítmica de um enunciado pode fazer com que eles pareçam semelhantes ou diferentes. O trabalho de Barbosa e Silva (2011) mostrou que os sujeitos levam em consideração para identificar corretamente diferenças na “forma de falar” pelo menos dois parâmetros relacionados a taxa de elocução em unidades VV/s e duração dos grupos acentuais. Como os resultados mostraram, houve mudança pontual em apenas alguns trechos analisados, e mesmo assim, tais mudanças foram de pequena magnitude. Visto que os ouvintes parecem utilizar parâmetros rítmicos para decisões corretas acerca da forma de falar de outros indivíduos, e tendo em vista os resultados obtidos neste trabalho, sugerimos que a análise de parâmetros prosódicos associados ao ritmo, como os que foram utilizados nesta tese, seja mais utilizada em pesquisas com tarefas de identificação de falantes no campo de pesquisa da fonética forense, principalmente quando há a necessidade de comparar amostras de fala com diferentes relações sinal-ruído.

5.3. Testes perceptivos

Os resultados do teste perceptivo mostraram que os sujeitos conseguem categorizar com relativa acuidade algumas variedades do português brasileiro. A variedade falada no Rio de Janeiro e a variedade falada no Nordeste foram as mais corretamente classificadas pelos participantes que responderam ao teste. Seis trechos no total receberam 90% de classificação correta, sendo que quatro deles são do RJ e dois são do Nordeste. Dos quatro trechos melhor identificados do RJ, três são trechos retirados da amostra de fala de um mesmo sujeito e os dois trechos mais identificados do Nordeste também pertencem ao mesmo sujeito, que é do estado da Paraíba. Podemos considerar que esses sujeitos sejam representantes típicos de suas regiões e, além disso, considerar que, no caso do Rio de Janeiro, a forte influência que tal variedade tem na mídia, seja em novelas e no telejornalismo, faz com que seja mais difundido e conhecido por

pessoas de diferentes localidades do Brasil. Além disso, para a variedade falada no Rio de Janeiro, há que se considerar a forte influência de outros fatores no alto índice de reconhecimento da variedade, como, por exemplo, os aspectos fonético-segmentais da variedade, que são característicos da região, como a típica realização das sibilantes e das vibrantes.

Apesar de ser muito reconhecida, a variedade falada no Rio de Janeiro é apontada como pouco aceita pelos falantes de diversas localidades brasileiras, como mostrou a pesquisa de Ramos (1997). O trabalho de Ramos buscou conhecer mais sobre a atitude linguística frente a diferentes variedades faladas no Brasil. Em relação à variedade falada no Rio de Janeiro, os resultados mostraram 8.5% de aceitação entre os falantes pesquisados. A variedade que recebeu o maior índice de aprovação nessa pesquisa foi a do Rio Grande do Sul, com 16.5%. Em outra parte da pesquisa, Ramos testou o reconhecimento de diferentes sotaques por falantes mineiros. Os resultados obtidos por Ramos corroboram os do presente trabalho: a variedade carioca foi a mais reconhecida pelos falantes mineiros, com 83.8% de reconhecimento. A variedade falada no estado de São Paulo ficou em terceiro lugar, com 70.9% de acerto, seguida pela variedade falada em Minas Gerais, com 67.7%. As diferenças entre as duas pesquisas referem-se ao fato de a pesquisa de Ramos ter sido aplicada apenas a falantes mineiros e ter opções de respostas delimitadas, sendo um teste de escolha forçada. Ramos afirma que os resultados obtidos para o Rio de Janeiro estão ligados diretamente com o fato de ser a variedade que está mais presente na mídia.

Em nossos resultados, o dialeto falado no Nordeste foi o segundo colocado, com 55.8% de identificações corretas. Ao considerarmos que o polo da mídia brasileira, tanto jornalístico quanto de entretenimento, fica na região Sudeste do Brasil, principalmente no eixo Rio-São Paulo, o fato de o Nordeste ter o segundo maior índice de reconhecimento não se explica por

essa justificativa. Uma hipótese é que os participantes que responderam ao teste perceptivo teriam mais contato com a variedade falada no Nordeste, pelo fato de que são, em sua grande maioria, universitários com contato com pessoas de diferentes localidades.

A história residencial do ouvinte e a sua proximidade com determinadas variedades contribuem para melhor acuidade na categorização dos dialetos (Clopper e Pisoni, 2007). O único ouvinte de Goiás (proveniente de Goiânia, que é próxima ao DF) conseguiu discriminar corretamente, por mais vezes, os sujeitos do DF, por exemplo. Os sujeitos do Nordeste, que eram da Bahia, Sergipe e Maranhão, identificaram corretamente os trechos de fala dos falantes nordestinos.

A proximidade com diferentes variedades ajuda no que diz respeito à construção da representação que os ouvintes têm dos dialetos falados em outras regiões. Após a aplicação do teste, muitos participantes testemunharam que haviam conseguido identificar determinados falantes, pois tinham amigos ou conhecidos daquelas regiões. Um participante do interior de SP identificou corretamente a fala de um sujeito da região Norte, categorizando corretamente inclusive o estado de procedência (Amazonas). Ao ser informado sobre o acerto, o participante relatou que durante a faculdade conviveu diariamente com um amigo do Amazonas e isso contribuiu para a identificação correta daquele trecho de fala.

Como foi mostrado anteriormente, há apenas dois sujeitos da região Norte compondo o *corpus* deste trabalho e mesmo assim os dois sujeitos analisados conseguiram se diferenciar dos sujeitos de outras variedades estudadas, o que mostra que, para a população estudada, os dois sujeitos possuem características prosódicas próprias que, provavelmente, constituem a variedade falada na região Norte do Brasil. Um fato que pode reforçar esta afirmação é a passagem descrita

no parágrafo anterior, em que um dos participantes do teste perceptivo foi capaz de dizer que um dos falantes do *corpus* era do estado do Amazonas. Podemos, então, fazer uma correlação entre os dados obtidos na análise dos parâmetros prosódicos com os resultados obtidos no teste perceptivo e reiterar que os sujeitos estudados são falantes típicos de suas regiões, podendo ter, inclusive, sua região de origem identificada por falantes que têm ou já tiveram proximidade com a variedade, fato que somente reforça positivamente os resultados obtidos nesta tese. Pode-se pontuar que a variedade falada na região Norte foi uma das menos reconhecidas pelos ouvintes. Neste sentido, precisamos levar em conta que não responderam ao teste ouvintes provenientes da região Norte e, além disso, por ter um número de sujeitos menor que nas outras variedades, somente seis trechos de fala destes sujeitos foram avaliados durante o teste perceptivo, exatamente a metade do valor total de trechos tocados dos falantes do RJ, o que pode ter contribuído para a quantidade menor de acertos que a região Norte obteve.

Ainda em relação à influência da proximidade dos sujeitos com determinadas variedades, mostramos anteriormente (Tabela 16) que a variedade falada no PR foi uma das menos corretamente discriminadas pelos sujeitos, sendo discriminada como “Sudeste” e “SP” na grande maioria das vezes. Um dos ouvintes que respondeu ao teste (sujeito 16) nasceu no MS, mudou-se aos dois anos de idade para o estado de SP, onde viveu durante 25 anos. Atualmente, o sujeito vive há dois anos em Irati, uma cidade do interior do estado do PR e, em suas respostas, todas as vezes que apareciam trechos de fala dos sujeitos do PR, o sujeito os classificou erroneamente como SP. Após a finalização do teste, o sujeito foi informado deste fato e disse que era muito difícil diferenciar a fala de SP e do PR em alguns casos, dependendo da localidade dos falantes, pois as variedades eram muito parecidas. Este fato pode mostrar que, apesar do sujeito conviver diariamente, há dois anos, com falantes do PR, a proximidade que esta região tem com SP reitera

que, em alguns casos, as duas variedades se inter-influenciam e são, muitas vezes, confundidas, corroborando com os resultados obtidos.

A história de povoamento do Estado do Paraná pode ser um forte argumento para justificar as escolhas dos ouvintes durante a participação no teste perceptivo. Segundo Aguilera (2006), linguista que se dedica há muitos anos aos estudos dialetológicos do Paraná, durante os séculos XVII, XVIII e XIX a história social do Paraná não pode ser dissociada da história social de São Paulo, visto que o Paraná foi comarca de São Paulo até 1853. Aguilera relata que, a presença do índio no território paranaense integrado à atividade econômica dos paulistas foi um fator determinante na formação da linguagem rural paranaense, evidenciada pela alta concentração de tupinismos na região litorânea e no planalto curitibano.

Apesar de a análise acústica levar em conta os parâmetros prosódicos, na aplicação do teste perceptivo não isolamos somente os aspectos suprasegmentais, como poderia ser feito utilizando-se a deslexicalização das amostras de fala, por exemplo. Alguém poderia chamar atenção para o fato de que os sujeitos podem ter levado em conta apenas aspectos fonológicos e sintáticos ao responder ao texto. Por isso, devemos ressaltar que o tipo de amostra de fala utilizada (fala espontânea) não permite que os ouvintes se limitem a um conjunto fixo de características para fazerem seu julgamento. Clopper e Pisoni (2007) afirmam que as vantagens do teste perceptivo do tipo classificação livre com uso de fala espontânea é que os ouvintes têm que julgar a similaridade das variedades na ausência de um contexto linguístico idêntico e, além disso, requer dos ouvintes o uso de representação abstrata dos falantes e de seus dialetos, baseando-se em amostras curtas e diferentes de fala. O caso do RJ, por exemplo, mostrou que, apesar de esta ser a variedade mais reconhecida, algumas vezes foi identificada apenas como Sudeste, mostrando que os ouvintes reconhecem a macro-região à qual a variedade pertence,

porém, mesmo com os aspectos segmentais presentes, não foi possível dizer especificamente qual era a variedade.

6. Conclusão

Os parâmetros prosódico-acústicos estudados são capazes de diferenciar e distinguir as variedades do português brasileiro, divididas de acordo com seus Estados. Ainda sim, existem variedades que são homogêneas em relação a alguns dos parâmetros estudados.

Oito parâmetros prosódico-acústicos foram analisados a partir de gravações de fala espontânea de 35 sujeitos. Do total de parâmetros analisados, ênfase espectral, mediana de frequência fundamental, taxa de unidades VV não proeminentes/s, assimetria de z-score e desvio-padrão de z-score foram os parâmetros que conseguiram revelar características prosódicas de algumas das variedades estudadas, o que mostra que as características prosódicas são capazes de distinguir variedades do português brasileiro e poderiam ser mais exploradas em estudos deste tipo. Os cinco parâmetros que revelaram características particulares das variedades e conseguiram diferenciá-las de outras variedades contemplam os três parâmetros acústicos clássicos de estudo da prosódia: mediana de frequência fundamental, que claramente se refere à F0, ênfase espectral, que é uma medida de intensidade relativa, referindo-se, portanto à intensidade e taxa de unidades VV não proeminentes/s, assimetria de z-score suavizado de unidade VV e desvio-padrão de z-score suavizado de unidade VV, que dizem respeito à estrutura rítmica das variedades e assim, estão relacionadas com a duração. Uma vez mais, reforçamos que as diferenças encontradas a partir desta análise não levam em conta aspectos segmentais dessas variedades e sim aspectos prosódicos, o que justifica o fato de variedades percebidas como muito distintas em relação aos aspectos segmentais não se diferenciarem em relação a determinados parâmetros prosódicos.

As variedades faladas no DF e na região Norte se diferenciaram das outras variedades formando dois grandes grupos em relação aos parâmetros mediana de frequência fundamental e ênfase espectral. Em contrapartida, outros parâmetros foram capazes de diferenciar DF da região Norte, como assimetria de z-score suavizado e taxa de unidades VV não proeminentes. Concluímos que a divisão feita por Nascentes em 1953 ainda é válida, pois todos os parâmetros estatisticamente significativos do presente estudo conseguiram distinguir as variedades faladas na região Norte das variedades faladas na região Sul.

A análise das amostras de fala espontânea dos 35 sujeitos revelou que os parâmetros que mais distinguiram a prosódia das variedades estudadas foram os parâmetros que mais sofreram perturbação em suas medidas quando há adição de ruído. A medida da ênfase espectral deve ser evitada quando há presença de ruído na amostra e o estudo da estrutura rítmica é indicado, pois as mudanças observadas foram mínimas quando a relação sinal-ruído diminuiu e, pelo espectrograma, ainda seria possível delimitar os inícios das vogais, possibilitando o estudo da duração das unidades VV, por exemplo.

Os testes perceptivos mostraram que os ouvintes reconhecem com maior precisão a variedade falada no RJ e a variedade menos reconhecida é a falada na Região Norte, fato que pode estar ligado ao menor número de trechos de sujeitos da região Norte estarem presentes no teste perceptivo. A proximidade com outras variedades mostrou-se um fator importante no reconhecimento, visto que ouvintes que têm ou tiveram contato com falantes de determinadas variedades conseguiram reconhecer mais precisamente tais variedades.

7. Referências Bibliográficas

- Aguilera, V. A. Os tupinismos na linguagem rural paranaense. In: Estudos Linguísticos. v. 7 n.12. 2006. p. 99-125.
- Alexander, A. et. al. The effect of mismatched recording conditions on human and automatic speaker recognition in forensic applications. In: Forensic Science International. 146S. 2004. p. 95–99.
- Alkmin, T. Sociolinguística. Capítulo I. In: Mussalin, F.; Bentes, A.C.(orgs) Introdução à linguística. Volume 1. 3ª edição. Cortez Editora. 2003.
- Atal, B.S. Automatic speaker recognition based on pitch contours. Journal of the Acoustical Society of America. 1972. v. 52: 1687-1697
- Barbosa, P.A. Incursões em torno de ritmo da fala. Campinas: Editora Pontes. 2006.
- Barbosa, P.A.; Silva, W. A New Methodology for Comparing Speech Rhythm Structure between Utterances: Beyond Typological Approaches. In: Computational Processing of the Portuguese Language 10th International Conference, PROPOR 2012, Coimbra, Portugal, April 17-20, 2012.
- Behlau, M. Voz: o livro do especialista. Editora Revinter. 2001.
- Bortoni-Ricardo, S.M. Contato de dialetos no Distrito Federal, Brasil. In: Bortoni-Ricardo, S.M. Freitas, V.A. e Vellasco, A.M. (orgs.) O falar candango, Brasília, Editora UnB, 2010, p.17-32.
- Camargo, Z. Et Al. Voice quality and gender: some insights on correlation between perceptual and acoustic dimensions In: 6th International Conference on Speech Prosody, 2012, Shanghai. Abstract Book Speech Prosody 2012. Shanghai: Tongji University Press, 2012. p.115 – 118.

Childers, D.G. Vocal quality factors: Analysis, synthesis, and perception. In: J. Acoust. Soc. Am 90(5), November 1991. p.2394-2410.

Clopper, C.G., Pisoni, D.B. Free classification of regional dialects of American English. Journal of Phonetics. v. 35. 2007. p. 421-438.

Clopper, C.G., Smiljanic, R. Effects of gender and regional dialect on prosodic patterns in American English. Journal of Phonetics. 2011. April 1; 39(2): 237–245.

Corrêa, C.C.: Focalização dialetal em Brasília: um estudo das vogais pretônicas e do /s/ pós-vocálico, Dissertação de Mestrado inédita, Brasília, DF: Universidade de Brasília. 1998.

Cunha, C. Atlas lingüístico do brasil: uma análise das questões de prosódia. In: Mota, J.A.; Cardoso, S.A M.(Org.). Documentos 2: projeto atlas linguistico do Brasil. Salvador: Quarteto, 2006.

Dogil, G; Braun, G. The PIVOT model of speech parsing. Verlag der Österreichischen Akademie der Wissenschaften. Viena; 1988.

Eriksson, A. The disguised voice: imitating accents or speech styles and impersonating individuals. In C. Llamas & D. Watt (Eds.), Language and Identities. Edinburgh: Edinburgh University Press. 2010. p. 86–96

Fadi B.; Hirschberg, J. Using Prosody and Phonotactics in Arabic Dialect Identification. INTERSPEECH, page 208-211. ISCA. 2009.

Ferreira,C.S., Cardoso, S.A.M. A dialetologia no Brasil. São Paulo, Cortez. 1994.

Gooskens, C.; Wilbert H. The Relative Contribution of Pronunciation, Lexical and Prosodic Differences to the Perceived Distances between Norwegian Dialects. *Literary & Linguistic Computing* . Nov 2006. Vol. 21 Issue 4, p. 477-492.

Heldner, M. On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. In: *Journal of Phonetics*. 31. 2003. 39–62.

Hollien, H. *Forensic voice identification*. San Diego, CA: Academic Press. 2002.

Holmberg, E.B. et. al. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. In: *J. Acoust. Soc. Am.* 84(2), August 1988. p. 511-529.

Holmberg, E.B et. al. Comparisons Among Aerodynamic, Electroglottographic, and Acoustic Spectral Measures of Female Voice. In: *Journal of Speech and Hearing Research*, Volume 38. December 1995. p. 1212-1223.

Hudson, R.A. *Sociolinguistics*. Cambridge University Press. 1980.

IBGE. Censo 2010. Disponível em: <http://censo2010.ibge.gov.br/resultados>

Kahn, J. *Caractéristique propres au locuteur : Traitement automatique et distance perceptive*. Université-Stendhal-Grenoble 3. Dissertação de Mestrado. 2008.

Marroquim, M. *A língua do Nordeste: Alagoas e Pernambuco*. 1ª Edição. EDUFAL. 1934.

Meireles, A.R; Tozetti, J.P.; Borges, R.R. Speech rate and rhythmic in Brazilian Portuguese. In: *Speech Prosody 2010 Conference*, 2010, Chicago. *Proceedings of the Speech Prosody 2010 Conference* . Chicago: RG. v.1. 2010. p.1-4.

Nascentes, A. *O linguajar carioca*. Rio de Janeiro: Organização Simões. 1953.

Noll, V. O português brasileiro: formação e contrastes. São Paulo: Globo, 2008.

Öhman, L. et. al. Mobile phone quality VS direct quality: how the presentation format affects earwitness identification accuracy. The European Journal of Psychology Applied to Legal Context. v. 2. 2010. p.161-182.

Pompino-Marschall, B. The syllable as a prosodic unit and the so-called P-centre effect. Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München, 29; 1991. p. 65-123.

Pierrehumbert, J. (1980) [The phonology and phonetics of English intonation](#). PhD thesis, MIT. Distributed. Indiana University Linguistics Club. 1998.

Projeto AliB. <http://twiki.ufba.br/twiki/bin/view/Alib/WebHome>

Projeto AMPER/POR: <http://pfonetica.web.ua.pt/AMPER-POR.htm>

R . Software estatístico gratuito. The R-Project for statistical computing: www.r-project.org

R Tutorial: <http://cran.r-project.org/>

Ramos, J.M. Avaliação de dialetos brasileiros: o sotaque. Rev. Est. Ling., Belo Horizonte, ano 6, n.5, v.1, jan./jun. 1997. p.103-125.

Rose, P. Forensic Speaker Identification. Forensic Science. London & New York: Taylor and Francis. 2002.

Traunmüller, H., & Eriksson, A. Acoustic effects of variation in vocal effort by men, women, and children. Journal of the Acoustical Society of America, 107(6). 2000. p. 3438–3451.



SERVIÇO PÚBLICO FEDERAL
MJ - DEPARTAMENTO DE POLÍCIA FEDERAL
DITEC - INSTITUTO NACIONAL DE CRIMINALÍSTICA

TERMO DE CONSENTIMENTO

O Serviço de Perícias em Audiovisual e Eletrônicos (SEPAEL) do Instituto Nacional de Criminalística (INC), órgão integrante da Diretoria Técnico-Científica (DITEC), tem por atribuição a execução e o desenvolvimento de exames de Comparação de Locutor. Para isso, fomenta ações de interação com o meio acadêmico-científico a fim de aprimorar e desenvolver técnicas e procedimentos.

Com o intuito de possibilitar o estudo de características linguísticas, acústicas e de processamento digital de sinais de voz, o SEPAEL tem realizado a colheita de padrões de voz que integram nosso banco de vozes de referência, intitulado: Corpus Forense do Português Brasileiro (CFPB).

Considerando que os objetivos do projeto de pesquisa da doutoranda Ana Carolina Constantini, intitulado "Individualização de características prosódicas em enunciados com baixa e alta relação sinal-ruído", orientada pelo Professor Doutor Plínio Almeida Barbosa do Instituto de Estudos da Linguagem do Departamento de Linguística da Universidade de Campinas, coadunam com os interesses de pesquisa do SEPAEL, disponibiliza-se as seguintes gravações do CFPB, em formato de áudio PCM (extensão .wav, 22050Hz de frequência de amostragem), para fins exclusivos da pesquisa em questão, resguardando o anonimato dos fornecedores:

M01AM00005232E.wav	M01MG00004142E.wav	M01RN00001330E.wav
M01AM00005232L.wav	M01MG00004142L.wav	M01RN00001330L.wav
M01AM00005333E.wav	M01MG00020436E.wav	M01RN00004337E.wav
M01AM00005333L.wav	M01MG00020436L.wav	M01RN00004337L.wav
M01AM00001728E.wav	M01MG00021743E.wav	M01SE00007333E.wav
M01AM00001728L.wav	M01MG00021743L.wav	M01SE00007333L.wav

M01AM00005054E.wav	M01MG00022041E.wav	M01RJ00000244E.wav
M01AM00005054L.wav	M01MG00022041L.wav	M01RJ00000244L.wav
M01BA00002051E.wav	M01MG00000446E.wav	M01RJ00007734E.wav
M01BA00002051L.wav	M01MG00000446L.wav	M01RJ00007734L.wav
M01CE00401138E.wav	M01MG00002428E.wav	M01SP00006131E.wav
M01CE00401138L.wav	M01MG00002428L.wav	M01SP00006131L.wav
M01CE00003733E.wav	M01MG00007629E.wav	M01SP00006428E.wav
M01CE00003733L.wav	M01MG00007629L.wav	M01SP00006428L.wav
M01DF00000629E.wav	M01MG00029239E.wav	M01RJ00000130E.wav
M01DF00000629L.wav	M01MG00029239L.wav	M01RJ00000130L.wav
M01DF00000729E.wav	M01PA00002335E.wav	M01RJ00000535E.wav
M01DF00000729L.wav	M01PA00002335L.wav	M01RJ00000535L.wav
M01DF00000827E.wav	M01PB00010038E.wav	M01RJ00002128E.wav
M01DF00000827L.wav	M01PB00010038L.wav	M01RJ00002128L.wav
M01DF00004229E.wav	M01PE00400632E.wav	M01RJ00003242E.wav
M01DF00004229L.wav	M01PE00400632L.wav	M01RJ00003242L.wav
M01DF00008938E.wav	M01PE00000935E.wav	M01RO00009333E.wav
M01DF00008938L.wav	M01PE00000935L.wav	M01RO00009333L.wav
M01DF00002728E.wav	M01PE00001543E.wav	M01RS00028336E.wav
M01DF00002728L.wav	M01PE00001543L.wav	M01RS00028336L.wav
M01DF00005619E.wav	M01PI00001235E.wav	M01SC00006932E.wav
M01DF00005619L.wav	M01PI00001235L.wav	M01SC00006932L.wav
M01DF00008130E.wav	M01PR00004644E.wav	M01SP00001841E.wav
M01DF00008130L.wav	M01PR00004644L.wav	M01SP00001841L.wav
M01DF00008732E.wav	M01PR00006839E.wav	M01SP00003344E.wav
M01DF00008732L.wav	M01PR00006839L.wav	M01SP00003344L.wav
M01MG00005834E.wav	M01PR00000341E.wav	M01SP00003529E.wav
M01MG00005834L.wav	M01PR00000341L.wav	M01SP00003529L.wav
M01MG00024439E.wav	M01PR00400036E.wav	M01SP00003633E.wav
M01MG00024439L.wav	M01PR00400036L.wav	M01SP00003633L.wav

Os pesquisadores envolvidos neste projeto estão autorizados a reproduzir as amostras em outros formatos para finalidades científicas, sendo vetado o uso e a redistribuição das amostras para outras finalidades.

Este consentimento poderá ser retirado, caso os princípios éticos que regem as pesquisas envolvendo seres humanos sejam feridos.

Em caso de dúvidas, contatar os responsáveis pelo banco de dados CFPB por meio do telefone (61) 2024 9363 ou pelo e-mail sepael.inc@dpf.gov.br.

Brasília, 3 de abril de 2014



José Jair Wermann
Diretor Técnico-Científico