



Programa de
Pós-Graduação em
Linguística

**ESTATÍSTICAS POPULACIONAIS DA FREQUÊNCIA FUNDAMENTAL DO
PORTUGUÊS BRASILEIRO PARA USO EM FONÉTICA FORENSE**

SÃO CARLOS
2023



Universidade Federal de São Carlos

Marcela Suardi da Cunha

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA

ESTATÍSTICAS POPULACIONAIS DA FREQUÊNCIA FUNDAMENTAL DO PORTUGUÊS
BRASILEIRO PARA USO EM FONÉTICA FORENSE

MARCELA SUARDI DA CUNHA

Bolsista: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)

Dissertação apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do Título de Mestre em Linguística.
Orientador: Prof(a). Dr(a). Pablo Arantes

São Carlos - São Paulo - Brasil

2023



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Linguística

Folha de Aprovação

Defesa de Dissertação de Mestrado da candidata Marcela Suardi da Cunha, realizada em 27/01/2023.

Comissão Julgadora:

Prof. Dr. Pablo Arantes (UFSCar)

Prof. Dr. Renata Regina Passetti (PUC-SP)

Prof. Dr. Julio Cesar Cavalcanti (SU)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Linguística.

Dedicado às mulheres na ciência.

AGRADECIMENTOS

Meus sinceros agradecimentos ao corpo docente da Universidade Federal de São Carlos (UFSCar) que foi minha fonte de conhecimento e inspiração para fazer ciência, bem como aos colegas que caminharam ao meu lado em muitos momentos de aprendizagem, em especial à querida Isabela de Jesus Silveira.

Agradeço o subsídio da Coordenação de Aprimoramento de Pessoal de Nível Superior (CAPES) por apoiar e possibilitar a realização deste trabalho, bem como ao Setor de Audiovisual e Eletrônicos (SEPAEL) do Instituto Nacional de Criminalística (INC) da Polícia Federal brasileira, por disponibilizarem o material de fala que foi usado como objeto da pesquisa.

Agradeço ao meu orientador Pablo Arantes que, além da maestria em seu papel, sempre se mostrou compreensível, paciente e humano diante das minhas dificuldades e pontos de melhoria. A aprendizagem que carregarei comigo será muito além de conceitos, mas sobretudo do papel ético e generoso ao compartilhar comigo seus conhecimentos, me permitindo começar a brilhar como cientista. MUITÍSSIMO obrigada por sempre acreditar em mim!

Também agradeço aos membros da banca examinadora, Renata Regina Passeti e Julio Cesar Cavalcanti, pela leitura cuidadosa e apontamentos que acrescentaram maior robustez ao texto, além de enriquecê-lo com sugestão de tópicos que complementam sua relevância para a literatura científica da área.

Meus agradecimentos ao querido Gabriel Pinheiro de Carvalho, pelo apoio emocional e técnico em todas as fases do desenvolvimento desta pesquisa, passando horas a fio me explicando conceitos matemáticos e de programação.

Por fim, a todas as mulheres que dedicam seu tempo para fazer ciência e deixar um legado ético e útil ao mundo. Vocês me inspiram a continuar, então espero inspirá-las também.

*O importante não é ser o primeiro ou primeira,
o importante é abrir caminhos.*

Conceição Evaristo

RESUMO

A presente pesquisa apresenta as medidas estatísticas da frequência fundamental (f_0) de 100 falantes do sexo masculino do português brasileiro (PB) que produziram amostras de fala semiespontânea e leitura de frases. As amostras de fala analisadas fazem parte do Corpus Forense do Português Brasileiro (CFPB), cedido pelo Setor de Audiovisual e Eletrônicos (SEPAEL), do Instituto Nacional de Criminalística (INC) da Polícia Federal brasileira. As características demográficas das 100 amostras seguem, na medida do possível, a distribuição da população brasileira nos critérios de faixa etária e distribuição entre as cinco grandes regiões brasileiras segundo a Pesquisa Nacional por Amostra de Domicílios (PNAD) de 2020. Foram analisados trechos com duração de fala contínua entre 30 e 45 segundos. Um conjunto de oito medidas estatísticas foram reportadas para cada amostra de f_0 : média, mediana, moda, valor de base, desvio-padrão, coeficiente de variação, assimetria e curtose, as quais foram apresentadas em forma de gráficos, tabelas, histogramas e diagramas de dispersão. Observou-se que, tanto o estilo de elocução quanto a conversão dos valores de f_0 da escala Hertz para a escala de semitons, aumentam a proporção de distribuições com valores que aderem à distribuição normal. A respeito da variação regional, os resultados sugerem a inexistência de uma influência importante da região geográfica sobre o valor dos estimadores estatísticos coletados. Os resultados indicam convergência com a literatura e são discutidos em relação às implicações fonético-forenses, podendo ser usados por peritos para o cálculo de valores de razão de verossimilhança em cenários reais em tarefas de comparação forense de vozes.

Palavras-chave: *Análise fonético-forense, frequência fundamental, estatística*

ABSTRACT

This research presents the fundamental frequency (f_0) statistics of 100 male speakers of Brazilian Portuguese (BP) who produced samples of semi-spontaneous speech and sentence reading. The analyzed speech samples are part of the Forensic Corpus of Brazilian Portuguese (CFPB), provided by the Audiovisual and Electronics Sector (SEPAEL), of the National Institute of Criminalistics (INC) of the Brazilian Federal Police. The demographic characteristics of the 100 samples follow, as far as possible, the distribution of the Brazilian population in terms of age group and distribution among the five major Brazilian regions according to the 2020 National Household Sample Survey (PNAD). Excerpts with duration of liquid speech between 30 and 45 seconds were analyzed. A set of eight statistical measures were reported for each f_0 sample: mean, median, mode, baseline value, standard deviation, coefficient of variation, asymmetry and kurtosis, which were presented in the form of graphs, tables, histograms and scatter plots. It was observed that both the elocution style and the conversion of f_0 values from the Hertz scale to the semitone scale increase the proportion of distributions with values that adhere to the normal distribution. Regarding the regional variation, the results suggest the absence of an important influence of the geographic region on the value of the collected statistical estimators. The results indicate convergence with the literature and are discussed in relation to the phonetic-forensic implications, and can be used by experts to calculate likelihood ratio values in real scenarios in forensic voice comparison tasks.

Keywords: *Forensic-phonetic analysis, fundamental frequency, statistics.*

LISTA DE TABELAS

Tabela I - Informações extraídas de cada pesquisa em bancos de dados.....	35
Tabela II - Tabela de distribuição do total de amostras do CFPB por idade e regiões do Brasil, incluindo o sexo masculino e feminino.....	39
Tabela III - Tabela da distribuição da população brasileira por região e idade (porcentagem)....	40
Tabela IV - Número de amostras do CFPB a serem selecionadas por região e faixa etária de acordo com as proporções da PNAD.....	41
Tabela V - Valores médios dos descritores estatísticos medidos na escala Hz em função dos dois estilos de fala.....	59
Tabela VI - Valores médios dos descritores estatísticos medidos na escala semitons (rel 1 Hz) em função dos dois estilos de fala.....	60
Tabela VII - Resultados dos testes de normalidade aplicados às distribuições amostrais dos estimadores estatísticos, separados por escala e estilo de fala (“E”: espontânea e “L”: leitura de frases”). Resultados significativos ($p < 0,05$) são marcados por “*”.....	66
Tabela VIII - Distribuição de probabilidade com melhor ajuste às distribuições de f_0 nos dois estilos de elocução (“E”: espontânea e “L”: leitura de frases”) e nas duas diferentes escalas. Quando a melhor distribuição não é a normal, indicamos entre parênteses quantas vezes melhor é o grau de ajuste daquela distribuição em comparação com a normal, segundo um teste de <i>Maximum Likelihood</i>	69
Tabela IX - Diferença mediana entre a mediana dos contornos das pausas preenchidas e as quatro medidas de valor típico dos contornos da amostra completa de f_0 . Valores nas escalas de Hertz e semitons.....	73

LISTA DE FIGURAS

Figura I - Histograma da distribuição de f_0 da fala espontânea e curva suavizada de densidade de probabilidade sobreposta. O valor da moda é indicado pela linha vertical tracejada.....	49
Figura II - Histogramas das medidas de tendência central em função dos dois estilos de elocução (“E”: espontânea e “L”: leitura de frases”), valores de f_0 em Hertz e semitons.....	62
Figura III - Histogramas das medidas de dispersão em função dos dois estilos de elocução (“E”: espontânea e “L”: leitura de frases”), valores de f_0 em Hertz e semitons.....	63
Figura IV - Histogramas de assimetria e curtose em função dos dois estilos de elocução (“E”: espontânea e “L”: leitura de frases”), valores de f_0 em Hertz e semitons.....	64
Figura V - Bloxplots das medidas de tendência central em função das regiões brasileiras, valores de f_0 em Hertz e semitons.....	69
Figura VI - Bloxplots das medidas de dispersão em função das regiões brasileiras, valores de f_0 em Hertz e semitons.....	70
Figura VII - Bloxplots das medidas de assimetria e curtose em função das regiões brasileiras, valores de f_0 em Hertz e semitons.....	71

SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO.....	13
1.1 Comparação forense de vozes.....	13
1.2 Métodos de análise em comparação de vozes.....	15
1.3 Parâmetros usados em análise acústica.....	17
1.4 Métodos para estabelecer o valor probatório da evidência.....	19
1.5 O arcabouço bayesiano ou baseado na razão de verossimilhança.....	20
1.6 Discussões sobre a população relevante.....	23
1.7 Relevância da dialetologia e sociofonética para a fonética forense.....	26
1.8 Objetivos do presente trabalho.....	27
CAPÍTULO 2 - BANCOS DE DADOS DE FALA E ESTATÍSTICAS POPULACIONAIS.....	28
2.1 Características desejáveis em bancos de dados de fala para aplicabilidade em fonética forense.....	29
2.2 Bancos de dados de fala do português brasileiro.....	29
2.3 A escolha do CFPB.....	32
2.4 Literatura prévia sobre estatísticas populacionais de f_0	33
2.5 Objetivos específicos.....	36
CAPÍTULO 3 - MATERIAIS E MÉTODOS.....	38
3.1 Descrição do Corpus Forense do Português Brasileiro.....	38
3.2 Decisões metodológicas.....	39
3.2.1. Tamanho da amostra.....	39
3.2.2. Sexo dos falantes.....	44
3.2.3. Faixa etária dos falantes.....	44
3.2.4. Variação estilística.....	45
3.2.5. Duração das amostras de fala.....	46
3.2.6. Estimadores estatísticos.....	46
3.3 Processamento das amostras de fala.....	50
3.4 Extração de f_0	51

3.5 Análise de f_0 em pausas preenchidas hesitativas.....	52
3.6 Análise estatística.....	55
CAPÍTULO 4 - RESULTADOS.....	57
4.1 Caracterização geral das distribuições de f_0 na fala encadeada.....	57
4.2 Estilo de elocução.....	59
4.3 Variação regional.....	68
4.4 f_0 em pausas preenchidas.....	72
4.5 Discussão.....	74
CAPÍTULO 5 - CONSIDERAÇÕES FINAIS.....	77
REFERÊNCIAS.....	79
APÊNDICE.....	86

CAPÍTULO 1 - INTRODUÇÃO

A fonética forense é a ciência que se utiliza de conhecimento linguístico, fonético e acústico a respeito da materialidade do som com a finalidade de investigar informações sobre o falante em contexto forense e analisar amostras de áudio gravadas durante crimes (HUGHES, 2014; p.1). Juntamente com a perícia, estudos na área podem contribuir para solucionar crimes nos quais os infratores não foram vistos, apenas ouvidos.

Tal ciência possui aplicação prática em contextos legais. Um exemplo disso é o caso Lindbergh, em 1932, em que o filho de um famoso aviador foi sequestrado e, semanas depois, encontrado morto. No processo de negociação e entrega do dinheiro do resgate, Charles Lindbergh afirmou ter ouvido a voz do sequestrador de seu filho e, em 1935, testemunhou em julgamento ter reconhecido a voz do suspeito Bruno Hauptmann.

A partir de conhecimentos linguísticos e das tecnologias disponíveis, é um dos papéis do cientista forense traçar o perfil do locutor, a priori desconhecido, a partir da observação de seus padrões de fala e assim determinar informações como contexto regional, idade, sexo etc. Para tanto, a fonética forense abrange diversas tarefas, como testemunho auricular (*earwitness testimony*), traçar o perfil de um falante a partir de amostras de sua fala (*speaker profiling*) e comparação de vozes, isto é, comparar amostras de fala para emitir um parecer técnico sobre as hipóteses das duas amostras terem ou não a mesma origem, isto é, se foram produzidas pelo mesmo trato vocal.

No presente trabalho, buscou-se fazer uma contribuição a respeito de uma dessas tarefas, a comparação forense de vozes. Para contextualizar adequadamente essa contribuição, foram apresentadas a seguir uma série de discussões a respeito dessa tarefa.

1.1 Comparação forense de vozes

Os estudos desenvolvidos no campo da fonética forense, em âmbito nacional e mundial, procuram investigar como alguns parâmetros acústicos se comportam em diferentes contextos comunicativos, paralinguísticos (por exemplo, diferentes estados emocionais de um falante) e em diferentes cenários técnicos (como gravação direta ou por telefone, diferentes níveis de ruído ambiental). Preocupa-se, ainda, em estabelecer níveis de variação para os parâmetros acústicos

devidos à variação produzida por um mesmo falante em situações distintas (variabilidade intrafalante), bem como o contraste entre vários indivíduos (variabilidade interfalante).

Segundo Rose (2002), das tarefas comuns solicitadas a peritos na área da fonética forense nas tarefas de comparação de vozes é confrontar as amostras questionadas a materiais de fala produzidos por indivíduos cuja identidade é conhecida, normalmente suspeitos de serem a fonte das amostras questionadas.

Seguindo Morrison (2009; 2010), optou-se pela expressão “comparação forense de vozes” e não termos mais consagrados como “comparação de locutor”, “verificação de locutor”, “autenticação de locutor” em função das críticas que se pode fazer à noção de que as práticas de análise usadas na ciência forense são capazes de identificar e individualizar um determinado indivíduo a partir de uma amostra, seja de fala ou de outros tipos de traços.

O uso da expressão "amostra questionada" segue essas definições gerais: são amostras de fala produzidas por um falante cuja identidade é desconhecida e que supostamente é autor ou autores de uma infração (ROSE 2002, JESSEN 2007, ERIKSSON 2011). Tal material é também chamado de “traço”, “amostra criminal” ou “disputada” e se trata, geralmente, no campo da fonética forense, de uma interceptação telefônica. Uma definição geral para o termo, que vale para toda a ciência forense, é apresentada por Bell (2004):

A amostra em questão, aquela que foi colhida como evidência. Por exemplo, uma impressão digital questionada seria aquela colhida em uma cena de crime; ela seria comparada aos padrões ou exemplares colhidos de suspeitos ou outras pessoas que as podem ter depositado. (BELL 2004, p. 213)¹ (Tradução nossa)

Já a expressão “amostra conhecida” é normalmente chamada de “amostra de referência”, “amostra padrão”, “amostra de origem conhecida” ou “amostra do suspeito”. Tal material diz respeito à amostra coletada em entrevistas policiais ou disponíveis publicamente. Portanto, a identidade dos seus autores é conhecida. Uma definição geral é proposta por Bell (2004):

Amostras com composição, identificação, tipo ou fonte conhecidas, usadas para comparação ou para garantir que uma análise está sendo realizada de maneira correta e que os resultados são confiáveis. (BELL 2004, p. 45)² (Tradução nossa)

¹ The sample in question, the one that has been collected as evidence. For example, a questioned fingerprint would be one collected at a crime scene; it would be compared to standards or examples collected from suspects or other persons who may have deposited it. (BELL 2004, p. 213)

² Samples with a known composition, identification, type, or source used for comparison or to ensure that an analysis is being properly performed and that the results are reliable. (BELL 2004, p. 45)

Outro termo encontrado na literatura é “padrão conhecido” que, segundo (BRENNER 2004, p. 151) faz referência a “um espécime de fonte identificada, adquirido para fins de comparação com uma amostra de evidência; sinônimo de exemplar”. Brenner (2004, p. 106) define o termo “exemplar” como

[...] um espécime de fonte identificada adquirido para fins de comparação com uma amostra de evidência. Um exemplo de escrita de uma pessoa, um padrão para uso em comparações, um espécime “coletado” ou “solicitado”. Amostras de provas físicas de origem conhecida (usadas para comparação com provas de cena de crimes semelhantes).³ (BRENNER, 2004, p. 106) (tradução nossa)

Apresentada essa multiplicidade de termos técnicos com significado aproximadamente idêntico, optou-se aqui, por uniformidade, por usar os termos amostra de origem questionada e de origem conhecida, bem como suas versões simplificadas “amostra questionada” e “amostra conhecida”.

1.2 Métodos de análise em comparação de vozes

Um levantamento feito pelos pesquisadores Gold e French (2011) com especialistas forenses provenientes de instituições acadêmicas e laboratórios, mostrou que há quatro abordagens metodológicas comumente utilizadas para tarefas de comparação de vozes. Esta seção faz uma breve apresentação a respeito dessas diferentes metodologias.

Análise fonética auditiva: Envolve julgamentos auditivos, geralmente qualitativos e com transcrição fonética detalhada. Nessa abordagem, não há análise acústica espectrográfica, então costuma-se observar a frequência de ocorrência de certos itens lexicais, alofonias, entre outros aspectos de qualidade de voz. Devido aos avanços tecnológicos dos quais a fonética forense se

³ A specimen of an identified source acquired for the purpose of comparison with an evidence sample. An example of a person’s writing, a standard for use in comparisons, a “collected” or a “request” specimen. Specimens of physical evidence of known origin (used for comparison with similar crime scene evidence). (BRENNER 2004, p. 106)

beneficiou ao longo dos anos, esse uso exclusivo dessa abordagem tem sido cada vez mais raro, representando apenas 9% no levantamento de Gold e French (2011).

Análise fonética acústica: Ao contrário da análise fonético-auditiva, a análise acústica envolve principalmente observações das variáveis fonético linguísticas, baseadas em medidas acústicas. Uma forma de investigação que envolve esse método é a técnica chamada *voice printing*, que consiste na comparação visual de espectrogramas. As técnicas mais modernas dessa abordagem envolvem a extração de variáveis fonético acústicas, ainda assim, no levantamento feito por Gold e French (2011), apenas um dos especialistas afirmou utilizar apenas a análise acústica em suas investigações.

Análise auditiva e fonético-acústica combinadas: Segundo Gold e French (2011), essa é a abordagem mais utilizada para apresentar as evidências da comparação forense de vozes admitidas nos tribunais de países como Reino Unido, Alemanha, Turquia, Brasil e China. Essa análise envolve a combinação das técnicas acústica e auditiva, sendo considerada ao mesmo tempo quantitativa e qualitativa. Em seu levantamento, os autores observaram que todos os especialistas usaram a análise fonética mista quando se trata de estudos sobre a frequência fundamental. Algumas variáveis fonético linguísticas também comumente observadas nessa abordagem, como variáveis segmentais e suprasegmentais (como a qualidade de voz, por exemplo), patologias da fala, variáveis linguísticas de ordem superior (como escolhas lexicais e sintaxe, por exemplo), além de variáveis comunicativas, como é o caso das hesitações ou pausas preenchidas, que será melhor desenvolvida no capítulo 3 do presente trabalho.

Reconhecimento automático de falante: Segundo Hughes (2014), os sistemas de reconhecimento automático de falantes (*automatic speaker recognition*, em inglês, ou ASR, como é mais conhecido) são softwares comerciais que funcionam de maneira independente executando processamento de sinais, seleção de falante e modelagem estatística, podendo inclusive ser construídos manualmente a partir de programas de processamento de fala ou de análise estatística. O ASR se difere das demais abordagens a partir do processamento e análise dos sinais, das variáveis que são extraídas, além dos procedimentos de modelagem estatística. Dentre seus benefícios, Hughes (2014) cita sua grande capacidade de extrair altas quantidades de

informações discriminatórias a respeito dos falantes, além do modelamento estatístico muito eficiente resultante de seus dados contínuos, sendo capaz de gerar saídas numéricas e probabilísticas. Em contrapartida, os coeficientes cepstrais que são comumente extraídos nessa abordagem são altamente sensíveis a ruídos de fundo, baixa qualidade técnica e não correspondência de canais, embora haja procedimentos capazes de minimizar esses fatores. No levantamento de Gold e French (2011), 24% dos especialistas afirmaram utilizar o ASR em suas tarefas de comparação de voz.

Na próxima seção, descreve-se alguns dos parâmetros de investigação que são comumente utilizados em casos de comparação forense de voz.

1.3 Uso da f_0 em análise acústica

Um dos principais parâmetros de investigação utilizados em análise acústica é a frequência fundamental - doravante f_0 -, medida comumente utilizada em casos de comparação forense de vozes (NOLAN, 1983; ROSE, 2002) e que recebeu atenção de muitos pesquisadores por ser um traço acústico robusto, ou seja, relativamente resistente à presença ruídos de fundo, efeitos de transmissão por canal telefônico (BRAUN, 1995), compressão de áudio e aspectos técnicos do microfone (CAVALCANTI et al, 2021).

Segundo Braun (1995), por ser um parâmetro bastante variável, a f_0 de um mesmo falante pode sofrer variações, inclusive, quando coletadas amostras de fala em momentos diferentes em um mesmo dia. A autora aponta como esse parâmetro pode ser influenciado por fatores emocionais, fisiológicos e psicológicos, tais como o uso regular de álcool; uso de drogas, esteróides, anabolizantes e testosterona; procedimentos cirúrgicos como remoção de cistos e nódulos em órgãos relacionados com a produção de fala, entre outros.

Apesar das fontes de variabilidade que afetam esse parâmetro acústico, a f_0 segue sendo extremamente utilizada por especialistas em comparação forense de vozes em todo o mundo. Gold e French (2011; 2019) conduziram dois levantamentos - com oito anos de diferença entre eles - a respeito das práticas adotadas por peritos especializados na comparação forense de vozes. O primeiro estudo contou com 36 especialistas e o segundo com 39, provenientes de 23 países e

cinco continentes, afiliados a instituições governamentais, laboratórios de pesquisa, universidades ou com atuação privada autônoma, visando identificar os métodos e os recursos de trabalho correntes na área no momento em que cada levantamento foi feito e tornar acessíveis essas informações para a comunidade científica. Os autores consideram que o conjunto das respostas documenta tanto a diversidade quanto às convergências nas práticas adotadas internacionalmente na área, além de sua evolução no período entre as duas enquetes.

O primeiro estudo aponta que 70% dos participantes reportam usar alguma forma de estatística populacional a respeito da incidência de uma ou mais características fonéticas ou acústicas para chegar a conclusões nos casos nos quais atuam e 58% declararam dispor de estatísticas populacionais previamente coletadas. Além disso, cerca de 70% utilizam a frequência fundamental em suas análises, seguida de outros parâmetros como taxa de articulação, tempo para o início do vozeamento (*voice onset time*, em inglês), frequência de formantes, entre outros. Entre os que fazem uso de estatísticas populacionais, a maioria declarou que usaria esse tipo de informação com mais frequência se houvesse disponibilidade desse recurso. Os autores sumarizam na segunda enquete as fontes dos dados de referência ou das estatísticas populacionais indicadas pelos participantes. As fontes mais comuns são publicações científicas e, em segundo lugar, bases de dados de fala públicas de diferentes línguas, organizadas em geral por consórcios de pesquisa.

De maneira geral, os dois estudos permitem chegar a duas generalizações que, por sua vez, justificam o presente trabalho: a primeira é que o uso de dados de referência ou estatísticas populacionais tem se mostrado uma tendência cada vez mais prevalente na prática da comunidade científica forense internacional em procedimentos de exame pericial; a segunda generalização diz respeito à escassez desses dados de referência acessíveis de forma pública para toda a comunidade. Em ambos os casos, os dois cenários não afetam apenas a comunidade brasileira, embora esta seja ainda mais prejudicada do que outros países e idiomas, como o inglês, o alemão e o mandarim, por exemplo.

A frequência fundamental é um dos parâmetros acústicos que ajudam a caracterizar a qualidade de voz e deve ser levado em conta a dinâmica vocal como excursões melódicas e de acentos tonais. Segundo Laver (1980, p. 1), pode-se definir qualidade de voz de uma maneira ampla como “a coloração auditiva característica da voz de um falante particular”, que resultaria da presença de traços de fala relativamente constantes na produção vocal de um falante. Ainda

segundo Laver, a qualidade de voz é devida à contribuição tanto de traços laríngeos quanto supralaríngeos. Essa visão contrasta com concepções um pouco mais restritas do termo “voz” (Ladefoged, 1971), que consideram a qualidade de voz derivada basicamente da atividade das estruturas laríngeas, descrita como diferentes estados da glote, que resultam em diferentes “modos de fonação”, que geram, por sua vez, qualidades de voz que recebem rótulos como voz modal, soprosa, crepitante, entre outras. Nessas duas concepções, a f_0 pode ser entendida como um parâmetro que ajuda a caracterizar diferentes qualidades de voz, uma vez que é o correlato acústico da frequência de vibração das pregas vocais.

No presente trabalho, olhou-se para os contornos de f_0 de uma perspectiva global e sem tocar de forma mais detida na questão da interação entre f_0 e qualidade de voz propriamente dita. Sendo assim, não investigou-se mais a fundo ocorrências pontuais de f_0 que poderiam, por exemplo, estar associadas à ocorrência de modos de fonação não modais. Essa decisão foi tomada por conta da limitação de tempo que poderia ser dedicado a esse tema durante a realização da pesquisa e não pela sua falta de importância.

1.4. Métodos para estabelecer o valor probatório da evidência

Na tarefa de comparação forense de vozes, uma vez analisadas as amostras questionada e conhecida, é preciso estabelecer se as semelhanças e diferenças entre as amostras dos dois tipos podem ser usadas para sustentar que vêm de uma mesma fonte (o suspeito é o dono da voz na amostra questionada) ou não (vozes do suspeito e da amostra questionada não vêm do mesmo trato vocal).

Na prática forense, há uma variedade de quadros de referência usados para expressar as conclusões de um laudo pericial. As principais são descritas e sumarizadas por Gold e French (2011)⁴, que quantificam o uso que seus entrevistados reportaram realizar de três diferentes metodologias: escala de probabilidade clássica, que é mencionada em 40% das respostas, seguida da chamada *UK Position Statement*, com cerca de em torno de 31% das respostas e, por último, a *likelihood ratio* (LR), expressão traduzida para o português brasileiro como razão da

⁴ Remete-se os leitores aos trabalhos de Gold e French (2011) e Eriksson (2011) para definições técnicas e detalhadas dos métodos. Por uma questão de brevidade, detalharemos apenas o chamado arcabouço bayesiano, pois é no âmbito desse método que a contribuição deste estudo se dará.

verossimilhança, que é usada por cerca de 20% dos participantes (nessa proporção estão incluídas as versões numérica e verbal da LR). Na segunda enquete, cujos resultados são apresentados em Gold e French (2019), a proporção dos participantes que declararam usar a razão de verossimilhança como base para expressar suas conclusões subiu para quase 40%, acompanhada de uma queda semelhante no uso das outras metodologias.

Na seção seguinte, descreve-se com mais detalhes a abordagem da razão de verossimilhança, bem como sua importância nesta pesquisa.

1.5. O arcabouço bayesiano ou baseado na razão de verossimilhança

As ciências forenses passaram por uma grande transformação ao longo da sua história, saindo, segundo a avaliação de Morrison (2009) de um período pré-científico para uma ciência empiricamente fundamentada, transição que o autor atribui ao predomínio atual do uso do arcabouço bayesiano, que ele (MORRISON 2009; 2010) chama de *likelihood-ratio framework*. Essa transição é resultado do que o autor caracteriza como uma “mudança de paradigma”, pois ela é vista como uma resposta às limitações de outros arcabouços de decisão, como os analisados por Champod e Mewly (2000) e que estavam presentes nas enquetes de Gold e French (2011; 2019).

Segundo Morrison (2009), o aumento na adoção do novo paradigma não é uma surpresa e também não se limita ao campo da análise da fala, sendo inicialmente implementado para a comparação de códigos de DNA, passando a ser utilizado nos tribunais e nas ciências forenses, inclusive nas tarefas de comparação de voz. Pode-se apontar como características o fato de se tratar de uma ciência empiricamente fundamentada, possuindo validade científica e podendo ser replicada por outros pesquisadores. Por se tratar de uma abordagem baseada em dados, ela permite quantificar e comunicar as limitações da comparação forense via medições das taxas de erro, conferindo confiabilidade probatória.

Nessa tarefa, o papel do perito é estabelecer a força probatória da evidência extraída dos dados e comparar duas hipóteses: a de que as amostras conhecida e questionada têm a mesma origem ou a hipótese de que ambas têm origens diferentes, ou seja, não foram produzidas pelo mesmo falante. Por sua vez, a decisão a respeito da culpabilidade ou absolvição do réu perante as

provas técnicas fornecidas pela perícia, além das demais evidências envolvidas no caso, como testemunhas oculares, entre outros, fica a cargo dos jurados ou juízes.

De maneira simplificada, pode-se dizer que, no arcabouço baseado na razão de verossimilhança, o que está em jogo é responder à pergunta:

Quão mais provável é que as diferenças observadas entre as amostras conhecidas e questionadas ocorram sob a hipótese de que a amostra questionada tenha a mesma origem que a amostra conhecida do que sob a hipótese de que ela tenha uma origem diferente? (MORRISON, 2010, p.14) (tradução nossa)⁵

A resposta a essa questão é encontrada por meio do cálculo do que se pode chamar de razão de verossimilhança. Nesse cálculo, a primeira probabilidade, representada por $p(E|H_{so})$, diz respeito à probabilidade de obter a evidência, E , dada a hipótese de que a amostra questionada e a amostra conhecida tenham a mesma origem, H_{so} (*same origin*). A segunda, $p(E|H_{do})$, refere-se à probabilidade de obter a evidência sob a hipótese de que as amostras não tenham a mesma origem, H_{do} (*different origin*). A razão entre as duas probabilidades, LR , é determinada pela expressão mostrada na fórmula 1. O numerador dessa razão pode ser entendido como um índice de *similaridade* entre as amostras de origem questionada e as amostras de origem conhecida, provenientes do(s) suspeito(s), enquanto o denominador pode ser entendido como um índice de *tipicidade* em relação à população relevante, isto é, estabelece quão prevalente na população relevante é a diferença observada nos dados analisados.

Fórmula 1: Fórmula da razão de verossimilhança.

$$LR = \frac{p(E|H_{so})}{p(E|H_{do})}$$

A estimativa quantitativa da tipicidade é derivada justamente das estatísticas populacionais que, por sua vez, são obtidas através de uma base de dados de amostras que são representativas da população. O termo “população relevante”, que discutiremos em maior

⁵ How much more likely are the observed differences between the known and questioned samples to occur under the hypothesis that the questioned sample has the same origin as the known sample than under the hypothesis that it has a different origin? (MORRISON, 2010, p.14)

detalhe na seção 1.6, diz respeito, em primeira aproximação, à população de que os falantes das amostras questionada e conhecida fazem parte.

A fórmula da LR, por sua vez, é um dos termos de outra equação matemática chamada teorema de Bayes (*Bayes' theorem*). O teorema de Bayes é uma operação matemática que multiplica dois termos: o primeiro, $p(H_{so})/p(H_{do})$, representa a crença em uma determinada proposição, que pode ser modificada conforme o surgimento ou mudança de uma nova informação; o segundo, $p(E|H_{so})/p(E|H_{do})$, representa a crença a respeito do mesmo evento que podem variar devido a diferenças nos pesos atribuídos a cada um dos componentes da informação (HUGHES, 2014), ou seja, a fórmula da LR. Por fim, a multiplicação entre as duas razões resulta em uma terceira razão, $p(H_{so}|E)/p(H_{do}|E)$, que será melhor explicada em seguida. A fórmula é apresentada a seguir:

Fórmula 2: Fórmula do teorema de Bayes.

$$\frac{p(H_{so})}{p(H_{do})} \cdot \frac{p(E|H_{so})}{p(E|H_{do})} = \frac{p(H_{so}|E)}{p(H_{do}|E)}$$

chances a priori razão da verossimilhança chances a posteriori

Na fórmula, p é a probabilidade de um determinado evento, E é a evidência, H_{so} é a proposição da acusação, H_{do} é a proposição da defesa e $|$ indica que a probabilidade em questão é condicional. As chances *a priori* introduzem um viés, visto que se trata da avaliação do julgador antes da introdução de uma evidência. Essas são expressas pela razão entre $p(H_{so})$ e $p(H_{do})$. A fórmula da razão de verossimilhança, por sua vez, fornece o cálculo do peso de cada evidência e é expresso pela razão entre $p(E|H_{so})$ e $p(E|H_{do})$. Modificando as chances *a priori*, a LR estabelece a chance *a posteriori*, podendo ser entendida como a avaliação das probabilidades das proposições concorrentes dado a força da evidência. Essa decisão final de determinar se o réu é culpado ou inocente é uma decisão da corte, não tendo qualquer interferência dos especialistas forenses, que se limitam ao cálculo da razão de verossimilhança.

A convenção para a interpretação do resultado da LR é que se ele for maior do que 1, isso suporta a proposição da acusação (H_{so}) e se o resultado for menor do que 1, isso suporta a proposição da defesa (H_{do}). A magnitude dos valores obtidos a partir da LR determina “quão

mais provável é a evidência dada uma proposição sobre a outra”⁶ (EVETT *et al*, 2000 *apud* HUGHES, 2014, p.21).

Fica estabelecida, portanto, a conexão entre o crescimento na adoção do arcabouço baseado na razão de verossimilhança e a demanda pela disponibilidade de dados de referência e estatísticas populacionais. A falta desse recurso representa um obstáculo à adoção do arcabouço baseado na razão de verossimilhança que, segundo a indicação das enquetes de Gold e French, já é a metodologia de tomada de decisão mais popular e que tem o potencial de ganhar ainda mais espaço entre os praticantes.

1.6. Discussões sobre o conceito de população relevante

A fórmula da razão de verossimilhança, explicada na seção anterior, mede a força da evidência, modificando as chances *a priori*. Enquanto seu numerador representa a similaridade entre as amostras conhecida e questionada, o denominador representa a tipicidade, isto é, quão prevalente na população relevante é o valor do traço fonético observado na amostra do suspeito. Na formulação de Hughes (2014, p.21), a tipicidade pode ser entendida com a probabilidade dos valores referentes à amostra do infrator, tendo em vista que ela tenha sido produzida não pelo suspeito, mas por outro membro da população relevante.

O conceito “população relevante” foi definido na literatura pela primeira vez por Coleman e Walls (1974, p. 276) *apud* Hughes (2014, p. 37) como

aquelas pessoas que poderiam estar envolvidas (no crime); às vezes, pode ser estabelecido que o crime deve ter sido cometido por uma determinada classe de pessoas com base na idade, sexo, ocupação ou outro subgrupo, e então não é necessário considerar o restante⁷ (tradução nossa)

Já que tal hipótese (H_{do}) não pode ser vaga ou mal definida, além de ser praticamente impossível lidar com dados em sua totalidade - no caso deste trabalho, todos os falantes

⁶ [...] how much more likely the evidence is given one proposition over the other (EVETT *et al*, 2000 *apud* HUGHES, 2014, p.21)

⁷ Those persons who could have been involved (in the crime); sometimes it can be established that the crime must have been committed by a particular class of persons on the basis of age, sex, occupation or other sub-grouping, and it is then not necessary to consider the remainder of [...] (COLEMAN E WALLS, 1994, p. 276 *apud* HUGHES, 2014, p.37).

masculinos do português brasileiros - reduz-se a população total a proporções mais administráveis, estatisticamente falando.

Há duas abordagens concorrentes adotadas para a implementação do conceito de população relevante: relevância lógica e similaridade segundo o julgamento de ouvintes leigos.

Segundo Hughes (2014), para que os critérios populacionais (como sexo, idade, entre outros) sejam considerados logicamente relevantes, eles devem afetar a distribuição de uma variável mais ampla. Dessa forma, as amostras selecionadas para compor o conjunto de dados a partir dos quais se vai estimar a distribuição populacional de um determinado parâmetro fonético-acústico devem representar subgrupos dentro da população geral, com mesmos antecedentes regionais, idade e sexo, por exemplo.

Trazendo esses conceitos para o campo da comparação forense de vozes, Rose (2004) propõe que, na ausência de uma hipótese particularizada de origem diferente para as amostras questionadas e do suspeito, a amostra questionada não deve ser entendida como sendo pertencente ao suspeito, mas a outro falante qualquer do mesmo sexo e falante da mesma língua do infrator. Duas problemáticas a respeito dessa definição são apontadas por Hughes (2014): assumir que sexo e língua de um falante são informações facilmente extraíveis de amostras de fala, e defender que esses dois parâmetros de fato apresentam uma importante fonte de variação entre os falantes. Segundo o autor, a descrição sociolinguística de uma determinada comunidade de falantes pode ajudar a identificar fatores sociodemográficos mais relevantes acerca do infrator, além do sexo e da língua.

A segunda abordagem, proposta por Morrison e colegas (2012) baseia-se na similaridade entre vozes segundo o julgamento de ouvintes leigos. Nesse procedimento, a hipótese de base colocada pela defesa do suspeito diz que o infrator na gravação questionada não é o suspeito, mas sim alguém que soa suficientemente semelhante a ele. Desse modo, um oficial da polícia ou outro indivíduo apropriado ouviria as gravações do infrator e dos suspeitos para compará-las. A população relevante, nesse caso, “consiste em falantes que soam semelhantes ao infrator com base nos julgamentos de um painel de ouvintes leigos (ou seja, não treinados)” (HUGHES, 2014, p 42)⁸.

⁸ [...] consists of speakers who sound similar to the offender based on the judgements of a panel of lay (i.e. linguistically naïve) listeners (HUGHES, 2014, p.42).

Para a implementação dessa abordagem, o painel de ouvintes leigos deve corresponder ao perfil do ouvinte inicial que submeteu as gravações para análise, ou seja, terem a mesma ocupação, histórico regional e também o mesmo nível de experiência em comparação de vozes. Além disso, o painel de ouvintes deve ser apresentado a amostras que correspondam a fatos relevantes da amostra questionada, como transmissão telefônica, se for o caso, além de corresponder às mesmas condições ambientais. Tais cuidados geram dados de referência que podem incluir falantes das mais diferentes origens sociolinguísticas.

A abordagem baseada na similaridade entre vozes pode ser mais vantajosa do que a baseada na relevância lógica, segundo Hughes (2014), na medida em que os especialistas não correm o risco de tomar decisões equivocadas a respeito de características sociolinguísticas do infrator. Além disso, na abordagem da similaridade, a população relevante é definida seguindo um único fator de agrupamento, a semelhança entre a voz do infrator e do suspeito, em vez dos diversos fatores demográficos levados em conta da perspectiva baseada na relevância lógica.

Algumas limitações do método baseado na similaridade são, segundo Hughes (2014): o perfil linguístico dos ouvintes encarregados de julgar a similaridade poderia influenciar em suas decisões que se opte por um painel de ouvintes com a mesma idade ao invés de garantir que todos sejam policiais. Além disso, tendo em vista que as bases de dados possam apresentar alta variabilidade em suas distribuições, ou seja, quanto ao sexo, origem regional e social de seus participantes, é possível que cada ouvinte se atente a características distintas do sinal de fala. Por fim, o autor discute se a decisão do ouvinte inicial em submeter a gravação do infrator a uma análise de similaridade feita por um painel de ouvintes leigos é, de fato, uma decisão objetiva ou se poderia ter sido influenciada por outras informações do caso, o que consistiria em um viés, inclusive podendo indiretamente inflar o fator tipicidade.

A partir da discussão sobre a população relevante, a depender do caminho tomado, pesquisadores podem necessitar de informações a respeito da variação linguística de uma certa língua ou comunidade de fala, segundo diversos critérios: regional, etário, estratificação por renda, ocupação profissional, entre outros. Se esse for o caso, é de extrema importância ter à disposição descrições dialetológicas e sociofonéticas da variação linguística de diversos parâmetros de interesse para a fonética forense.

Esta seção buscou apresentar um panorama geral das diferentes abordagens utilizadas e validadas pela comunidade científica para as tarefas de comparação de voz, embora não haja um consenso entre elas, ou seja, ambas possuem aspectos positivos e negativos. Do ponto de vista da abordagem da relevância lógica, além dos critérios clássicos definidos por Rose (2004) - sexo e língua - não há dados suficientes a respeito de f_0 para mostrar quais fatores sociolinguísticos afetam esse parâmetro de maneira consistente e que deveriam, portanto, ser controlados para a definição da população relevante. Nesse sentido, os dados produzidos pelo presente trabalho devem constituir uma contribuição original e relevante sobre a variabilidade de f_0 na comunidade falante do português brasileiro, com uma especial atenção à influência da variação regional e do estilo de elocução.

1.7. Relevância da dialetologia e sociofonética para a fonética forense

Quando se trata dos debates a respeito do conceito de população relevante e os critérios utilizados para defini-la, estudos sociolinguísticos, sobretudo sociofonéticos, e dialetológicos podem contribuir de maneira expressiva com o tema.

A sociofonética, subárea mais recente da sociolinguística, estuda quais fatores externos e/ ou sociais influenciam a fala, além de observar características identitárias de seus falantes, bem como seus objetivos comunicativos (FOULKES, 2010). A dialetologia, por sua vez, é uma área da linguística que busca estudar variações inerentes às línguas a partir da identificação e descrição dessas diferenças em seu uso. Cardoso (2016, p.17) aponta como principais objetivos da pesquisa dialetológica, a descrição de fatos característicos das línguas, a qualificação das ocorrências registradas a partir de um ponto de vista social e a observação dessas ocorrências sob uma perspectiva temporal.

O Brasil possui uma divisão geográfica baseada em critérios históricos e econômicos, o que não é necessariamente a melhor em termos de estudos linguísticos. Embora alguns projetos de construção de um mapa dialetológico já tenham sido propostos visando caracterizar as variedades existentes no país (CONSTANTINI, 2014; CARDOSO, 2016), a divisão estadual utilizada não pressupõe fronteiras dialetais.

Sendo assim, a colaboração de tais áreas facilitaria tarefas forenses como a comparação de vozes à medida que ajuda a mapear se e quais características de variação linguística têm efeito significativo sobre parâmetros fonéticos, incluindo a frequência fundamental. Sobretudo, ter à disposição descrições dialetológicas e sociofonéticas ajudariam a estabelecer quais fatores têm peso para a definição de uma população relevante.

1.8. Objetivos do presente trabalho

Com base no cenário desenhado nas seções anteriores, argumenta-se a favor da relevância de se produzir estatísticas de distribuição populacional de parâmetros fonéticos e linguísticos, visto que, quando se trata do português brasileiro, temos uma enorme lacuna quanto a isso. Dessa maneira, este estudo tem como objetivo geral compilar dados estatísticos a respeito da distribuição populacional de diferentes características da f_0 , tais como seu valor médio e o valor do desvio-padrão típicos em diferentes cortes demográficos da população de falantes masculinos do português brasileiro, de modo a aumentar o conhecimento a respeito deste importante traço acústico da fala e facilitar a aplicação da abordagem baseada na razão de verossimilhança no contexto de tarefas de comparação de vozes com finalidade forense.

CAPÍTULO 2 - BANCOS DE DADOS DE FALA E ESTATÍSTICAS POPULACIONAIS

As estatísticas de distribuição populacional de um determinado parâmetro acústico devem ser extraídas de material de fala. No cenário mais comum, esse material de fala é reunido em um banco de dados, que normalmente é organizado tendo como unidade uma determinada língua, quer o banco se proponha a representar toda a variabilidade regional daquela língua em um determinado território ou apenas a variedade de uma região específica. Nesse sentido, quando se trata da geração de estatísticas populacionais de parâmetros acústicos é importante discutir critérios para a escolha de um banco de dados que servirá de base para a geração dessas estatísticas. Discute-se no presente capítulo algumas características desejáveis que bancos de dados de fala devem apresentar para que seu uso seja interessante para a geração de estatísticas populacionais para uso em fonética forense.

Na seção 2.1, elencou-se características desejáveis que um banco de dados deve apresentar para que seja um bom candidato a fornecer dados de fala para a geração de estatísticas populacionais de parâmetros acústicos. Tais características foram articuladas com as discussões a respeito da definição de população relevante no contexto do arcabouço de decisão baseado na razão de verossimilhança que foram apresentadas no capítulo 1.

Na seção 2.2, foi apresentada uma seleção de bancos de dados de fala do português brasileiro existentes, além de suas principais características à luz das características desejáveis definidas na seção 2.1.

Pesando os aspectos positivos e negativos das opções disponíveis no momento, foi justificada a escolha do Corpus Forense do Português Brasileiro para o presente trabalho na seção 2.3.

Na seção 2.4, foi apresentada uma revisão de trabalhos que relatam a distribuição populacional de f_0 para auxiliar o uso da metodologia baseada na razão de verossimilhança. O objetivo desta seção foi, além de fazer um levantamento da literatura a respeito do tema até o momento, reconhecer os principais critérios usados pelos pesquisadores na área para a extração dessas estatísticas, como o número de falantes que formam a amostra utilizada, o perfil sociodemográfico dos falantes e os principais parâmetros estatísticos extraídos dos dados de f_0 . Essas informações foram relevantes para a definição da metodologia utilizadas no presente trabalho.

O capítulo apresenta os objetivos específicos da pesquisa, expandido aqueles já mencionados ao final do capítulo anterior, também discutindo o uso das estatísticas populacionais para uso em fonética forense desenvolvidos ao longo deste capítulo.

2.1 Características desejáveis em banco de dados de fala para a aplicabilidade em fonética forense

Algumas características são bastante relevantes quando se trata de bancos de dados de fala para aplicabilidade em fonética forense, visando tanto uma análise mais precisa por parte de pesquisadores e da perícia, quanto uma melhor precisão diante dos critérios que melhor definiriam uma população relevante, segundo discutido anteriormente.

Dessa maneira, é desejável que o banco tenha características como: alta cobertura geográfica que dê conta das variações linguísticas regionais; características demográficas como sexo, faixa etária e ocupação profissional dos falantes etc; material de áudio com requisitos mínimos de qualidade, visando a realização de análises para extração de parâmetros acústicos; dados contemporâneos, ou seja, relativamente atuais, embora não seja muito simples definir uma data de corte para esse critério.

Diante disso, levantou-se uma breve caracterização de alguns dos bancos de dados disponíveis para o português brasileiro, apontando a seguir os parâmetros que propiciaram a escolha da base de dados selecionada para o atual estudo.

2.2 Bancos de dados de fala do português brasileiro

Nesta seção, foi apresentado um breve levantamento a respeito de alguns dos bancos de dados disponíveis para o português brasileiro, apontando suas características principais, os tipos de materiais disponíveis, além de sua disponibilidade. Salienta-se ainda que há outros materiais semelhantes que, por limitação de tempo e recorte do trabalho, não seria viável o levantamento exaustivo dessas informações, embora reconheça-se sua importância.

Projeto da Norma Urbana Linguística Culta (NURC), disponível em <https://fale.ufal.br/projeto/nuredigital/>: Trata-se de um banco de dados de norma urbana culta que

teve seu início em 1969, contendo gravações dos anos 70 e 90 do século XX e tem o intuito de preservar a memória nacional da época. Ele conta com entrevistas gravadas em fitas magnéticas de rolo e apenas falantes de nível superior completo de cinco capitais brasileiras: Recife, Salvador, Rio de Janeiro, São Paulo e Porto Alegre. Além das amostras de fala, o corpus foi transcrito e digitalizado sendo a principal escolha entre os pesquisadores devido ao difícil acesso às gravações. Por muito tempo, um grande impedimento que dificultava ainda mais as pesquisas era a não disponibilidade das transcrições em formato digital, ou seja, conseguir os textos impressos, além de ser um processo demorado e burocrático também era passível de falhas justamente por não se poder contar com buscas automáticas dos fenômenos linguísticos.

Projeto Variação Linguística na Região Sul do Brasil (VARSUL), disponível em <https://www.varsul.org.br/>: Trata-se de uma iniciativa subsidiada pela Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES) e executada por quatro universidades brasileiras, sendo elas: Universidade Federal do Rio Grande do Sul, Pontifícia Universidade Católica do Rio Grande do Sul, Universidade Federal de Santa Catarina e Universidade Federal do Paraná. Seu objetivo é a descrição do português brasileiro falado e escrito de algumas regiões do sul do país. O banco é composto por 288 entrevistas, cada uma com duração média de sessenta minutos e originalmente gravadas em fita-cassete que começaram a ser coletadas a partir de 1988.

Segundo o site oficial, o banco vem sendo ampliado em todas as suas sedes, com acréscimo de novas amostras. Originalmente, os informantes do VARSUL não possuem curso superior e a faixa etária é acima de 25 anos, ao contrário das novas gravações coletadas, as quais contemplam outros níveis de escolaridade e faixas etárias distintas.

Projeto SP2010, disponível em <https://projetosp2010.fflch.usp.br/>: Trata-se de um banco de dados coletado pelo Grupo de Estudos e Pesquisa em Sociolinguística da USP que tem como objetivo principal contribuir para os estudos sociolinguísticos a respeito dessa variedade do português brasileiro. É composto por sessenta entrevistas com duração de aproximadamente uma hora da fala oral paulistana do sexo masculino e feminino, no entanto, o banco continua em expansão. O material possui um amplo espectro etário, com amostras de falantes de 19 a 89 anos, com diferentes níveis de escolaridade e possui amostras de residentes de bairros distintos da

cidade de São Paulo, podendo ser acessado gratuitamente por docentes, cientistas e linguistas através de cadastro no site oficial.

C-Oral Brasil, disponível em <https://www.c-oral-brasil.org/>: Iniciativa coordenada por Tommaso Raso e Heliana Mello, membros da Faculdade de Letras da Universidade Federal de Minas Gerais (UFMG), tem como objetivo promover o estudo da fala espontânea. O *corpora* é composto por amostras de fala informal do português brasileiro, em sua maioria do dialeto mineiro e com gravações de ampla variação situacional. Seus 139 textos são divididos igualmente entre diálogos, conversas e monólogos que podem ser acessados e baixados gratuitamente através do site oficial.

BrasilData: Trata-se de um banco de dados que contém amostras de fala de 80 participantes masculinos de quatro estados brasileiros (São Paulo, Minas Gerais, Alagoas e Espírito Santo) com a finalidade de simular situações forenses por meio de gravações feitas por celulares e gravadores digitais. Em sua tese de doutorado de 2018 intitulada “Análise fonético-acústica para fins forenses em cinco localidades brasileiras a partir da base Brasildata [recurso eletrônico]”, Machado (2018) estudou 42 amostras de fala provenientes das cidades de Campinas, São Paulo, Belo Horizonte, Maceió e Vitória. O corpus abrange a faixa etária entre 18 a 35 anos, sendo que o nível de escolaridade mínimo dos sujeitos foi ensino médio completo.

Amostra Linguística do Interior Paulista (ALIP), disponível em <https://www.alip.ibilce.unesp.br/>: Coletado entre os anos de 2004 a 2007, o Banco de Dados Iboruna - primeira base de dados que compõe o ALIP - é composto por materiais de fala de dois tipos: amostra censo ou comunidade, constituída por 151 entrevistas sociolinguísticas, e amostra de interação, composta por 11 interações dialógicas. Foi idealizado por pesquisadores funcionalistas da Unesp de São José do Rio Preto, noroeste do estado de São Paulo, que objetivavam construir uma base de dados própria que representasse a variedade do interior paulista do território brasileiro, ainda pouco estudada. Além disso, a dificuldade de acesso às demais bases, incluindo suas gravações originais, a necessidade de atualização das amostras disponíveis, tanto pelo fato de que algumas delas datam de até 40 anos antes do início do projeto,

quanto pela baixa qualidade acústica das gravações, foram alguns dos fatores que motivaram a criação dessa amostra.

Uma segunda base foi incorporada ao projeto, nomeada Banco de Dados de Aquisição. Ela comporta dois tipos de materiais de fala: Amostra de Aquisição do Português Brasileiro como Segunda Língua, coletada em 1996 por meio de 28 interações com seis crianças nipo-brasileiras residentes no interior paulista e Amostra de Aquisição do Português Brasileiro como Língua Materna, coletada entre os anos 2000 e 2001 a partir da interação espontânea de adultos com uma criança brasileira, filha de pais brasileiros e também residente do interior paulista. O site oficial do ALIP disponibiliza o material na íntegra mediante cadastro.

Corpus Forense do Português Brasileiro (CFPB): criado pelo Serviço de Perícias em Audiovisual e Eletrônicos (SEPAEL) do Instituto Nacional de Criminalística (INC), se trata de um material feito para uso interno, mas o acesso pode ser permitido através de acordos com pesquisadores. Possui material de fala semiespontânea além de leitura de frases, pequenos textos e dígitos. No momento em que tivemos acesso ao material, ele continha 210 gravações de falantes do sexo masculino e 69 gravações de falantes do sexo feminino das cinco regiões brasileiras e com idades que variavam entre 19 e 64 anos. Além do material de fala, seus criadores também disponibilizam alguns dados dos participantes, como escolaridade, cidades em que morou durante a vida e o tempo de permanência em cada uma delas, peso, altura e profissão em que atuava no momento das gravações.

Embora haja outras bases de dados disponíveis para o português brasileiro, além das descritas no presente capítulo, em virtude das discussões sobre população relevante e dos objetivos pretendidos para a criação das estatísticas populacionais do português brasileiro, o CFPB se mostrou a escolha mais adequada pelos critérios que serão discutidos na próxima seção.

2.3. A escolha do CFPB

O CFPB mostrou-se uma das melhores opções de base de dados de fala disponíveis no presente momento para a geração de estatísticas populacionais de f_0 entre falantes de PB pelas

seguintes razões: as amostras de áudios são relativamente recentes, no momento da realização do estudo; o material está disponível por meio de solicitação aos órgãos responsáveis; contempla amostras de fala de pessoas provenientes de todas as regiões do país assim como definidas pelo IBGE; possui amostras de falantes do sexo feminino e masculino; todas as gravações foram feitas com o mesmo equipamento, que é de boa qualidade.

Deve-se, no entanto, apontar algumas características do CFPB que podem ser entendidas como desiguais em relação às características dos falantes, por exemplo, em relação ao número de indivíduos por região (138 falantes da região Sudeste, 51 da região Nordeste, 46 da região Centro-Oeste, 23 da região Sul e 19 da região Norte); também em relação ao nível educacional dos informantes (87,28% cursaram o ensino superior com ou sem pós graduação, além de terem predominantemente a mesma profissão, ou seja, policiais); o histórico de migração desses entrevistados também não foi controlado, pois na amostra há falantes que mudaram de cidade ou estado em algum momento de sua vida.

O corpus será apresentado em detalhes no capítulo 3, referente aos materiais e métodos.

2.4. Literatura prévia sobre estatísticas populacionais de f_0

Com o intuito de observar e compreender melhor como os demais pesquisadores utilizaram bancos de dados de fala para trabalhos similares a este, sobretudo, quais critérios foram utilizados a fim de analisar a frequência fundamental, foi feito um levantamento de publicações anteriores da literatura científica que tiveram objetivos semelhantes aos da pesquisa presente. Essa busca mostrou a existência de sete publicações relevantes que foram detalhadas a seguir. A partir da análise dessas referências, foram extraídas as seguintes informações:

- Tamanho da amostra de falantes incluídos no banco de dados.
- Perfil sociodemográfico dos participantes.
- Idade e sexo dos participantes.
- Parâmetros estatísticos medidos e reportados: média, mediana, moda, desvio-padrão, coeficiente de variação, assimetria, curtose e valor de base, todos da f_0 .

A tabela I apresenta, de forma sumarizada, os principais aspectos observados em cada referência encontrada na pesquisa. O exame da tabela permite ver que a maioria dos trabalhos inclui amostras de jovens adultos do sexo masculino e o número total de amostras típico gira em torno de 100. A média e o desvio-padrão são os descritores estatísticos mais comuns. Na coluna “sexo” da tabela I, “f” significa sexo feminino e “m” sexo masculino.

Tabela I - Informações extraídas de cada pesquisa em banco de dados

Referência	Banco de dados	Idioma	Tamanho da amostra	Idade	Sexo	Parâmetros estatísticos	Estilo de elocução
Hudson et al. (2007), Grã-Bretanha	DyViS	Inglês padrão do sul britânico	100	18 - 25	m	média, mediana, moda, desvio padrão	Entrevista espontânea
Hussain et al. (2012), Paquistão	Urdu Speech Database	Urdu	82	20 - 55	f, m	média; moda; desvio padrão	Fala espontânea e leitura
Cao et al. (2017), China	PKU Mandarin Physical-Phonetic Database	Mandarim	100	Em média 24.1	m	média, mediana, moda, desvio padrão, coeficiente de variação	Vogal sustentada, palavras, sentenças, leitura de texto e fala espontânea
Lindh (2006), Suécia	Swedia Database	Sueco	109	20 - 30	m	média; mediana; desvio padrão; valor base; máximo e mínimo em Hz	Fala espontânea
Skarnitz et al. (2017), República Tcheca	Database of Common Czech	Tcheco	100	19 - 50	m	média, mediana, desvio padrão, valor de base, coeficiente de variação	entrevista policial simulada e leitura
Arantes et al. (2018), Brasil	Gravações cedidas por Anders Eriksson ⁹	Alemão, Estoniano, Francês, Inglês britânico, Italiano, Português brasileiro e Sueco.	70 (10 amostras por língua)	-	m, f	média; desvio padrão; valor de base	entrevista espontânea, leitura de frases e leitura de palavras
Costantini et al. (2015), Brasil	Corpus Forense do Português Brasileiro	Português brasileiro	35	Em média 35.4	m	média; mediana; desvio padrão; taxa de elocução; assimetria; taxa de proeminência; taxa de não proeminência; ênfase espectral	entrevista informal

⁹ O *corpus* em questão é relacionado ao projeto de Anders Eriksson intitulado "*A typology for word stress and speech rhythm based on acoustic and perceptual considerations*" e não está disponível publicamente. O projeto possui um site oficial com mais informações através do link <https://wordstress.ling.su.se/>.

De acordo com a tabela acima, os seguintes bancos de dados puderam ser encontrados: *Swedia database* (Suécia); *DyViS* (Grã-Bretanha), um dos mais conhecidos; *Urdu Speech database* (Paquistão); *PK Mandarin Physical-Phonetic database* (China); *database of Common Czech* (República Tcheca); *Canterbury Corpus of the Origins of New Zealand English database* (Nova Zelândia). É importante salientar que a pesquisa de Constantini e Barbosa (2015) utiliza o mesmo banco de dados que estamos utilizando no presente estudo. No entanto, quando foi realizado, os outros fizeram uso de apenas 35 amostras de fala.

Outras bases de dados ficaram de fora desta seleção por ainda estarem sendo construídas. É relevante citar uma proposta de criação de bancos de dados para o estado de São Paulo (SILVA, 2016), que não aparece na tabela, mas é de extrema importância para possíveis consultas.

A tabela I permite observar algumas características da amostra de participantes de cada estudo. A respeito do sexo, dentre as sete pesquisas analisadas, apenas duas incluíram tanto homens quanto mulheres, enquanto cinco delas incluíram apenas amostras de homens e nenhuma vale-se somente de mulheres. Segundo Silva (2016, p.42), o fato de “a população feminina ser constituída de um menor número de amostras justifica-se na medida em que é bem menor a participação feminina em muitos crimes, como, por exemplo, o tráfico de drogas”. Essa é uma justificativa para a não inclusão de dados da população feminina em trabalhos dessa natureza.

Observa-se também que o tamanho da amostra varia entre 35 a 109, sendo que três dos trabalhos incluem 100 indivíduos. Quanto à idade, nota-se uma gama de variação que vai de 18 a 55 anos entre os participantes. Os estilos de elocução foram: leitura de texto, frase e/ ou palavras e fala espontânea, geralmente guiada por uma entrevista informal. Algumas das decisões tomadas para o desenho do presente estudo foram baseadas nas informações colhidas nessa etapa de análise da literatura prévia. As decisões tomadas foram justificadas no capítulo 3.

2.5. Objetivos específicos

Diante do cenário apresentado a respeito das bases de dados de fala e das estatísticas populacionais disponíveis mundialmente para o parâmetros da f_0 , os objetivos específicos do atual estudo são: (1) utilizar o Corpus Forense do Português Brasileiro (CFPB) como fonte de

dados para gerar estatísticas populacionais; (2) fornecer uma visão geral da distribuição de uma série de descritores estatísticos da variação da f_0 na produção de falantes masculinos brasileiros; (3) tornar acessíveis as estatísticas populacionais relacionadas ao parâmetro f_0 , dada a escassez ou dificuldade de acesso a bancos de dados dessa natureza; (4) comparar os resultados com a literatura e discuti-los em relação às implicações para a prática forense.

CAPÍTULO 3 - MATERIAIS E MÉTODOS

3.1. Descrição do Corpus Forense do Português Brasileiro

Os dados utilizados neste trabalho foram selecionados do Corpus Forense do Português Brasileiro (CFPB), criado pelo Serviço de Perícias em Audiovisual e Eletrônicos (SEPAEL) do Instituto Nacional de Criminalística (INC), como já mencionado. É um material coletado para uso interno, mas o acesso pode ser permitido através de acordos com pesquisadores.

No momento do acesso ao banco para a realização da presente pesquisa, o corpus era composto por 210 gravações de 208 falantes do sexo masculino e 69 gravações de falantes do sexo feminino, das cinco regiões do Brasil definidas pelo Instituto Brasileiro de Geografia e Estatística (IBGE) - Norte, Nordeste, Centro Oeste, Sudeste e Sul. A idade dos falantes apresentou uma variação entre 19 e 64 anos.

A respeito da escolaridade, 3,18% dos falantes afirmaram ter cursado o ensino fundamental, 8,96% cursaram o ensino médio, 87,28% cursaram o ensino superior com ou sem pós-graduação e 0,58% não informaram sua escolaridade.

O banco de dados é composto por dois estilos de elocução, sendo eles entrevista semiespontânea, em que os entrevistados eram convidados a falar sobre tópicos diversos, como a cidade onde residiram, familiares e amigos, viagens ou assuntos de trabalho; leitura de frases, listadas no apêndice A; além de leitura de dígitos (de 0 a 9). As gravações foram feitas com equipamento de alta qualidade e em salas silenciosas.

A tabela II apresenta a distribuição por faixa etária e região geográfica do total de 277 gravações presentes no CFPB¹⁰. A notação “[[” representa um intervalo aberto, ou seja, a sequência começa a partir do número que se segue, mas sem incluí-lo; enquanto “]” representa um intervalo fechado, ou seja, o número anterior é o último incluso na sequência. Como exemplo, o intervalo representado por (1 - 5] é composto pela sequência 2, 3, 4 e 5.

¹⁰ Apesar do corpus apresentar 210 gravações, dois respondentes possuíam registros duplicados. Esses arquivos excedentes não foram contabilizados, totalizando 208 amostras de fala de 208 falantes.

Tabela II - Tabela de distribuição do total de amostras do CFPB por idade e regiões do Brasil, incluindo o sexo masculino e feminino

Região/ idade	(18 - 25]	(25 - 40]	(40 - 60]	(60 - 80]	Total
CO	2	35	9	0	46
N	0	9	10	0	19
NE	0	41	9	1	51
S	2	14	7	0	23
SE	5	81	52	0	138
Total	9	180	87	1	277

Tabela de distribuição das amostras do CFPB por idade e regiões do Brasil

A próxima seção apresenta os critérios elegidos para definir os parâmetros sociodemográficos, linguísticos e estatísticos relevantes para a coleta dos dados de f_0 a partir das amostras do CFPB.

3.2. Decisões metodológicas

3.2.1. Tamanho da amostra

Com base nos resultados analisados anteriormente na etapa de revisão da literatura, foi analisado um subconjunto de 100 amostras entre as produzidas por falantes do sexo masculino.

Na literatura prévia, como foi apresentado na seção anterior, predomina a decisão de modelar a distribuição dos falantes a serem incluídos no *corpus* a partir da distribuição demográfica da população carcerária. Foi decidido não seguir essa prática e modelar a distribuição dos falantes do CFPB a serem incluídos em nossa amostra a partir das características da população brasileira como um todo.

Elenca-se três razões para justificar a decisão de não usar a população carcerária como modelo: (1) introduziria um viés, que é considerar apenas casos em que os processos levaram a condenações e aprisionamento; (2) não leva em conta se os casos que levaram à prisão daqueles indivíduos encarcerados envolveram o uso de alguma prova técnica, em especial de natureza fonético-linguística; finalmente, (3) ignora que pode haver fatores estruturais na sociedade e no sistema de justiça que levam ao encarceramento preferencial de setores demográficos específicos da sociedade, em especial em termos etários ou proveniência geográfica, no caso em questão aqui.

Por conta das razões apresentadas anteriormente, foi decidido basear a escolha de 100 amostras (dentre o total disponível no CFPB) a serem incluídas nas estatísticas geradas pelo presente trabalho a partir da distribuição geográfica e etária da população brasileira como um todo, assim como indicada por dados da Pesquisa Nacional por Amostra de Domicílio (PNAD), feita pelo IBGE.

A tabela III mostra a distribuição da população brasileira entre as regiões geográficas e faixas etárias da PNAD de 2020¹¹.

TABELA III - TABELA DA DISTRIBUIÇÃO DA POPULAÇÃO BRASILEIRA POR REGIÃO E IDADE (PORCENTAGEM)

Região/ idade	(18 - 25]	(25 - 40]	(40 - 60]	(60 - 80]	% da população por região
CO	14,4	30,2	36,7	18,7	8
N	17,7	33,3	33,0	16,0	8
NE	14,9	30,5	33,9	20,7	26
S	12,4	28,3	35,7	23,6	15
SE	12,7	28,4	35,7	23,1	43

Pesquisa Nacional por Amostra de Domicílio (PNAD/IBGE) referente ao primeiro trimestre de 2020

¹¹ Disponível em <https://www.ibge.gov.br/estatisticas/sociais/populacao/17270-pnad-continua.html?edicao=34420&t=destaques>.

A tabela IV mostra a quantidade de amostras a serem selecionadas do CFPB por região e faixa etária seguindo as mesmas proporções da PNAD, mostradas na tabela III. Considerando apenas as regiões, o número de falantes por regiões é: CO (9), N (8), NE (24), S (15) e SE (44).

TABELA IV - NÚMERO DE AMOSTRAS DO CFPB A SEREM SELECIONADAS POR REGIÃO E FAIXA ETÁRIA DE ACORDO COM AS PROPORÇÕES DA PNAD

Região/ idade	(18 - 25]	(25 - 40]	(40 - 60]	(60 - 80]	Total
CO	2	3	4	0	9
N	0	4	4	0	8
NE	0	14	9	1	24
S	2	6	7	0	15
SE	5	17	22	0	44
Total	9	44	46	1	100

Distribuição das 100 amostras retiradas do CFPB por região geográfica brasileira e faixa etária

Pelos motivos discutidos na anteriormente, a seleção do *corpus* CFPB foi a melhor decisão para o objetivo de pesquisa. Dessa maneira, foi solicitado o uso do *corpus* através de um requerimento para os órgão responsáveis. A permissão foi concedida após as justificativas enviadas através de um projeto de pesquisa.

O material recebido consistia nas amostras de fala masculinas e femininas, além das fichas de todos os entrevistados. A partir disso, foi iniciado o processo de tabelar as informações referentes aos participantes, ou seja, sua idade, peso, altura, escolaridade e as cidades em que morou durante a vida. Dessas informações, foram dispensados peso e altura porque não são dados relevantes para essa etapa da pesquisa. A profissão também não foi tabelada por motivos que serão melhor explicados no final desta seção. A maioria dos respondentes afirmam ser policiais e muitos não colocaram seu cargo. Já se sabe que a idade influencia na frequência fundamental (BRAUN, 1995), então esse dado foi preservado inclusive para comparar com a população brasileira - como mostrado nas tabelas anteriores.

Com relação à escolaridade, ainda não se sabe se ela influencia ou não a f_0 . É importante salientar que o *corpus* é constituído de 87,28% de entrevistados que têm ensino superior completo, com ou sem pós-graduação. É também por esse motivo que não foi considerado coerente espelhar os dados na população carcerária brasileira.

Assim como a escolaridade, não se tem uma forte evidência prévia de que a profissão possa influenciar o parâmetro da f_0 , justamente por isso esses dados não foram agregados de maneira detalhada ao presente trabalho.

No entanto, segundo a sociolinguística, tanto escolaridade quanto profissão podem questões a serem levadas em conta também nos estudos fonéticos, como discute Eckert (1997) na citação a seguir:

A vida das crianças e dos adolescentes é dominada pela instituição escolar e escolaridade, a vida adulta é dominada pelo local de trabalho e a vida de muitos idosos, pela casa de repouso ou asilo. Enquanto os dois primeiros são baluartes das normas linguísticas padrão, as normas linguísticas incorporadas nas instituições para idosos são mais complicadas. (ECKERT, 1997) (tradução nossa)¹²

Para a realização do presente trabalho, foram tomadas algumas decisões metodológicas a respeito do número de falantes a serem incluídos no estudo, bem como seu sexo, faixa etária, origem geográfica e estilo de elocução. Essas decisões visam tornar a amostra final a mais representativa possível da população relevante.

Tomando como base os trabalhos citados na revisão da literatura, é possível observar que o número mais comum para a amostra da população relevante gira em torno de uma centena de falantes, embora nenhum dos estudos prévios aponte critérios explícitos e fundamentados para justificar a escolha do número de amostras de fala a ser incluído. Do ponto de vista estatístico, foi levado em conta o raciocínio básico segundo o qual quanto maior o tamanho da amostra, menores as chances de erro amostral, ou seja, há um aumento na confiabilidade das estimativas

¹² Children's and adolescents' lives are dominated by the institution of school and schooling, adult lives are dominated by the workplace, and many elderly lives by the retirement or nursing home. While the former two are bastions of standard language norms, the linguistic norms embodied in institutions for the elderly are more complicated. (ECKERT, 1997)

estatísticas derivadas da amostra. Portanto, quanto mais expressivo o número de falantes incluídos, melhor.

A partir disso, duas considerações precisam ser levadas em conta, sendo elas o custo - em recursos humanos e tempo - para o processamento de dados acústicos, bem como qual seria o grau aceitável de erro amostral para o propósito que se dispõe a investigar. Não parece haver nenhum critério objetivo, pelo menos nenhum é mencionado nos trabalhos encontrados, para se estabelecer um limiar de erro amostral aceitável no contexto da comparação de vozes com objetivo forense.

Com relação ao tamanho da amostra em trabalhos de pesquisa de cunho mais propriamente linguístico, é necessário levar em conta a heterogeneidade na manifestação do fenômeno linguístico em questão, além dos fatores que podem limitar a variação de tal fenômeno nas diferentes comunidades linguísticas dentro da população como um todo. Buchstaller e Khattab (2013), a partir de uma observação de Sankoff (1980), sugerem o teto de 150 observações, a partir do qual os ganhos com a adição de mais dados não compensariam o custo implicado na sua análise, embora esse número também não pareça ser motivado por um procedimento muito claro:

[...] mesmo para comunidades bastante complexas, amostras de mais de 150 indivíduos tendem a ser redundantes, trazendo problemas crescentes de manipulação de dados com retornos analíticos decrescentes (SANKOFF, 1980 *apud* BUCHSTALLER; KHATTAB, 2013, p. 84). (tradução nossa)¹³

Levando em consideração a dificuldade de estabelecer um número baseado em um critério metodológico de base estatística ou justificado por meio de limiares de erro aceitáveis estabelecidos a partir da experiência prática do campo, optou-se seguir uma tendência observada em outros estudos e estabelecer o número de 100 falantes como tamanho da amostra.

¹³ [...] even for quite complex communities, samples of more than 150 individuals tend to be redundant, bringing increasing data-handling problems with diminishing analytical returns (SANKOFF, 1980 *apud* BUCHSTALLER; KHATTAB, 2013, p. 84).

3.2.2. Sexo dos falantes

Ainda diante da análise de outros trabalhos semelhantes, observa-se que, na maioria dos casos, as estatísticas de distribuição populacional de f_0 baseiam-se em amostras de fala do sexo masculino. Como apontado na seção 4 do capítulo 2, essa prevalência pode ser explicada pelo fato de, na maioria dos países, pessoas do sexo masculino serem responsáveis por certos tipos de crime que chegam a ser objeto de investigação pelo sistema de justiça.

Em um primeiro momento, dispondo do tempo e recursos disponíveis, o presente trabalho analisou apenas os falantes do sexo masculino, sendo que, posteriormente, seria desejável fazer a mesma análise com falantes do sexo feminino, uma vez que o CFPB também conta com esses dados. Essa decisão se justifica pelo fato que, mesmo as mulheres sendo minoria na população carcerária, esse número ainda existe e é de grande relevância que também haja sua inclusão na caracterização sociofonética da população brasileira, visto que, na população nacional, as mulheres constituem maioria (51,1%, segundo a PNAD Contínua de 2021¹⁴). Tal estudo, além do ponto de vista forense, também seria relevante em pesquisas em linguística, em especial com relação à sociofonética.

3.2.3. Faixa etária dos falantes

Embora alguns estudos tenham procurado espelhar as características demográficas da população privada de liberdade em seus países, no presente trabalho, não se optou por semelhante escolha no quesito faixa etária dos falantes. Essa decisão leva em consideração alguns aspectos importantes, discutidos a seguir.

Como referido anteriormente, o espelhamento das características da população privada de liberdade no banco de dados introduziria um viés, que é considerar apenas casos em que os processos judiciais levaram o indivíduo à condenação e ignorar os casos em que houve absolvição. Além disso, não leva em conta se a condenação resultou do uso de alguma prova técnica, em especial de natureza fonético-linguística.

¹⁴ Ver

<https://educa.ibge.gov.br/jovens/conheca-o-brasil/populacao/18320-quantidade-de-homens-e-mulheres.html>

Tal escolha também ignora alguns aspectos estruturais da sociedade e do sistema de justiça que, infelizmente, levam ao encarceramento preferencial de certos setores demográficos específicos da sociedade, em especial em termos etários, que é o ponto abordado nesta seção.

Os dados brasileiros relativamente recentes sobre a população privada de liberdade (SANTOS, 2017) indicam que a faixa dos 18 aos 29 anos corresponde a 55% do total da população carcerária; a proporção alcança 93% de cobertura quando incluídas, cumulativamente, as faixas de 30 a 34 e 35 a 45 anos. A população carcerária brasileira concentra-se, segundo esses dados, desproporcionalmente em uma faixa etária mais jovem quando se considera a distribuição etária da população brasileira como um todo, como veremos na seção 3.4.

Dessa maneira, o presente trabalho não procura espelhar esse aspecto da população privada de liberdade, pelas questões descritas acima, além de que o banco de dados da CFPB é formado majoritariamente por falantes com idade mais avançada, como descrito anteriormente.

Para além disso, a decisão de levar em conta a distribuição da população brasileira por regiões também implica no benefício de que as estatísticas populacionais coletadas servirão também para futuras pesquisas em sociofonética, nas quais o corte etário que espelha a população privada de liberdade não faria sentido.

3.2.4. Variação estilística

Estudos como os de Arantes e Linhares (2017) apontam que o estilo de elocução exerce efeitos na f_0 no português brasileiro. Sendo assim, é importante que a tarefa de comparação de locutores utilize valores de referência apropriados para o estilo em que as falas comparadas foram produzidas.

Como o CFPB inclui tanto gravações no estilo fala semiespontânea (entrevista com perguntas abertas sobre a trajetória de vida e interesses do entrevistado) quanto no estilo fala lida (leitura de frases isoladas), tal pesquisa se utilizou de ambas as variações estilísticas, de modo que sejam geradas informações para os dois estilos.

3.2.5. Duração das amostras de fala

Uma das decisões de extrema importância para a compilação de dados é a definição da duração dos trechos das amostras de fala de cada um dos participantes do *corpus* a ser analisado. Sua relevância se justifica pelo fato de que a extração dos dados é uma tarefa automatizável, mas que precisa de algum nível de supervisão e de conferência manuais. Sendo assim, quanto mais longo for o trecho a ser analisado, maior o investimento de tempo e recursos para a execução da tarefa. Dessa maneira, o presente trabalho segue a recomendação de investigações anteriores a respeito da duração mínima necessária para a obtenção de uma estimativa confiável e robusta da f_0 (ARANTES; ERIKSSON; GUTZEIT, 2017), sendo a duração recomendada de pelo menos 30 segundos, excluindo-se trechos de disfluência ou silêncio, por exemplo.

A metodologia de pesquisa para a extração dos dados a partir do CFPB bem como a lista de parâmetros estatísticos serão apresentados a seguir.

3.2.6. Estimadores estatísticos

Com base na primeira etapa do trabalho, isto é, revisão da literatura e pesquisa em outros bancos de dados, foi possível notar que a maioria das investigações levam em conta para sua análise da f_0 medidas de tendência central (como média aritmética, moda e mediana) e também algumas medidas de dispersão ou variabilidade (mais comumente o desvio-padrão). Com base em estudos mais recentes, optou-se por ampliar o número de parâmetros a serem coletados para incluir alguns estimadores que começaram mais recentemente a ser utilizados em tarefas de comparação forense de vozes, como é o caso do valor de base, assimetria e curtose (KINOSHITA; ISHIHARA; ROSE, 2009; SILVA, 2016).

A listagem a seguir inclui os parâmetros adotados:

- Tendência central, medida de localização ou valor típico¹⁵: moda, média aritmética, mediana, valor de base.

¹⁵ Trata-se aqui os termos "tendência central", "medida de localização" e "valor típico" como sinônimos e consideramos que o valor de base pode ser considerado um estimador de valor típico, muito embora seja uma medida formulada especificamente para distribuições de f_0 , que têm características particulares, como a predominância de casos de assimetria e curtose acima do esperado para uma distribuição simétrica. Em cenários de distribuições com mais alto grau de assimetria, mesmo os estimadores genéricos, como a média e a mediana, por exemplo, não serão localizadas exatamente no "centro" da amostra.

- Dispersão: desvio-padrão, coeficiente de variação.
- Momentos estatísticos superiores¹⁶: assimetria e curtose.
- Informação dinâmica: taxa de picos.
- Distribuição teórica de probabilidade que melhor descreve a amostra de f_0 .

Como medidas de tendência central, foram extraídas média, mediana, moda e valor de base. Como medidas de dispersão, as estatísticas utilizadas foram desvio-padrão, coeficiente de variação e coeficiente de desvio de quartil. Além disso, foram medidas assimetria e curtose, bem como outras informações sobre a distribuição de f_0 , como a presença de mais do que uma moda e qual é a distribuição de probabilidade que melhor descreve cada distribuição empírica de f_0 , do modo mostrado por Arantes (2021).

A primeira medida extraída foi a média aritmética, ou seja, a soma de todos os elementos dividido pelo número de elementos. Também pode-se entendê-la como sendo a soma das medidas de frequência divididas pelo número de medidas de onda (HUSSAIN, 2012). É a medida de tendência central mais frequente na literatura, embora também seja a que mais apresenta sensibilidade em relação à assimetria. A mediana, mais robusta perante a assimetria e valores extremos, se refere ao valor que divide a amostra em duas partes iguais. Quando não há apenas um termo mediano, se calcula a média entre os dois valores centrais. Também pode ser entendida como o valor da f_0 que marca o 50º percentil da distribuição. Metade de todos os valores do conjunto são maiores que a mediana e metade são menores (HUSSAIN, 2012). A moda é o valor que mais aparece ou o mais frequente naquele conjunto de elementos. Por fim, o valor de base é “um estimador estatístico do valor típico da frequência fundamental de um falante com base na média e no desvio padrão da f_0 ” (ARANTES, 2018, p. 535). Essa medida foi proposta pelos pesquisadores Traunmüller e Eriksson (1994), especialmente para a análise de amostras de f_0 . Sua fórmula é dada pela diferença entre a média aritmética e a multiplicação da constante k - determinada empiricamente - pelo desvio padrão:

¹⁶ No jargão estatístico, o termo "momento" refere-se a diferentes medidas que ajudam a caracterizar a forma de uma distribuição de dados. A média é o primeiro momento estatístico de uma distribuição, a variância é o segundo. A assimetria e curtose são, respectivamente, o terceiro e quarto momentos e podem ser referidos genericamente como "momentos superiores".

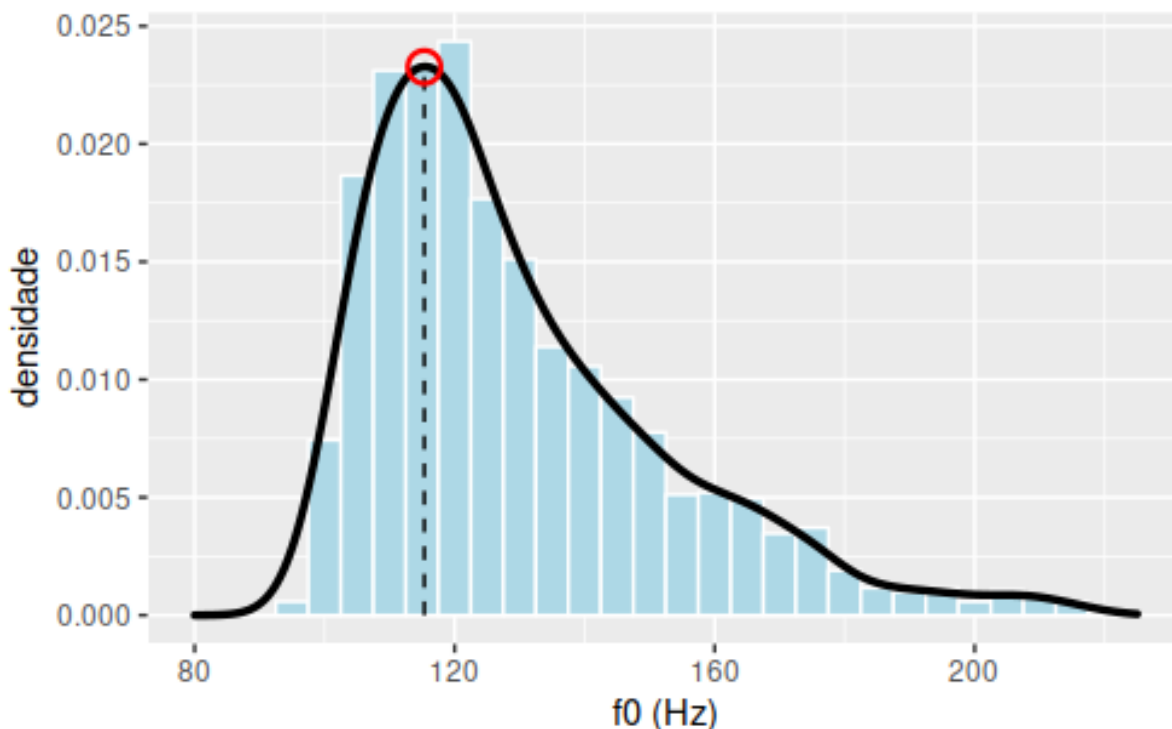
$$F_b = F_{\text{média}} - k\sigma$$

O desvio-padrão é uma das medidas de dispersão utilizadas como parâmetro estatístico. Ele mede a dispersão dos dados em comparação com a média aritmética e pode ser obtido pelo cálculo da raiz quadrada da variância. O coeficiente de variação é uma medida expressa em porcentagem que pode ser obtida através da razão entre o desvio padrão de uma população pela sua média aritmética, expressando a variabilidade dos dados em relação à média. Desse modo, quanto menor for o coeficiente de variação, mais homogêneo é o conjunto de dados.

Como medidas de dispersão relativa, utilizou-se o cálculo da assimetria e curtose, sendo que a primeira nos mostra qual o grau de desequilíbrio da curva para a direita (assimetria negativa) ou esquerda (assimetria positiva) em relação a uma distribuição normal enquanto a segunda identifica o quão achatada ou pontiaguda uma distribuição se encontra também em comparação à normal.

Seguindo a sugestão de Lennes et al. (2015), a moda amostral foi calculada a partir da aplicação da técnica *Kernel Density Estimation* (WAND; JONES, 1995) aos dados brutos de f_0 . A técnica KDE estima a densidade de probabilidade da amostra de dados em análise e pode ser entendida intuitivamente como uma versão suavizada de um histograma que fosse gerado a partir dos mesmos dados. O resultado produzido pela técnica KDE é uma curva contínua, ao contrário do histograma, que agrupa os dados brutos em intervalos. Uma das vantagens disso é que o valor no ponto máximo global dessa curva pode ser considerado o valor mais comum na amostra, portanto um candidato plausível para o valor da moda amostral. Foi usada a função *bkde* da biblioteca *KernSmooth* para o ambiente de computação estatística R (R CORE TEAM, 2020) para gerar a curva KDE de cada amostra de f_0 selecionada para compor nossa base de dados. Um parâmetro relevante passado para a função *bkde* é a largura de banda (*bandwidth*), que controla o grau de suavização do histograma alcançado pela aplicação da função. Usou-se a função *bw.nrd*, padrão na biblioteca *KernSmooth*, para determinar o valor da largura de banda para a função *bkde*, que de forma geral tende a atingir um ponto ideal entre minimizar as ondulações locais e ao mesmo tempo não eliminar movimentos importantes na distribuição. A figura I mostra o histograma (barras azuis) de uma distribuição de f_0 de fala espontânea e a curva suavizada de densidade sobreposta (traço contínuo mais grosso). O ponto máximo na curva, considerada a moda, é indicado pelo círculo vermelho ao fim da linha tracejada vertical.

Figura I: Histograma da distribuição de f_0 de fala espontânea e curva suavizada de densidade de probabilidade sobreposta. O valor da moda é indicado pela linha vertical tracejada.



A análise de ajuste da melhor distribuição teórica de probabilidade que descreve as amostras se justifica pelo fato da literatura a respeito da f_0 normalmente supor que amostras de f_0 podem ser modeladas por meio de uma distribuição normal, muito embora exista evidência mostrando que, na verdade, as distribuições de f_0 tendem a apresentar valores de assimetria e curtose incompatíveis com o esperado para a distribuição normal. Sobre esse tema, consultar Arantes (2021) e a literatura revisada no trabalho.

No contexto do presente trabalho, a questão a respeito do estabelecimento da distribuição de probabilidade adequada para modelar distribuições de f_0 é relevante porque a aplicação da razão de verossimilhança implica em adotar distribuições teóricas de probabilidade tanto para os dados do suspeito quanto para a distribuição da população relevante. Rose (2004) chama a atenção para o fato de que a maior parte das fórmulas encontradas na literatura para a derivação do valor da razão de verossimilhança assume que as amostras do suspeito e da população relevante assumem certas propriedades estatísticas, entre elas a aderência à distribuição normal, e afirma que a estimativa do valores da razão de verossimilhança pode perder acuidade quando essas propriedades não são verificadas. Em função disso, considerou-se importante analisar os

dados de f_0 quanto à melhor distribuição de probabilidade que os descrevem para que essa informação possa ser levada em conta quando esse dado for utilizado para o cálculo de razões de verossimilhança.

Para a determinação da distribuição teórica de probabilidade que melhor modela os dados de f_0 das amostras recolhidas do CFPB foram testadas as distribuições Normal, Gama, Logística, Weibull, Gumbel e Burr tipo XII. As distribuições Normal e Logística são distribuições simétricas e as demais são distribuições assimétricas. As distribuições Gama, Weibull e Burr só admitem valores positivos, portanto não foram usadas no ajuste das distribuições de dados que apresentem valores negativos. A escolha dessas distribuições teóricas baseou-se na experiência prévia em estudo prévio relatado em Arantes (2021) com o objetivo similar de determinar a melhor distribuição teórica de probabilidade para modelar a distribuição de amostras empíricas de f_0 .

As extrações e processamento dos dados foram realizados por meio de *scripts* do programa de análise acústica *Praat* (BOERSMA, 2001) e do ambiente de computação estatística R.

3.3. Processamento das amostras de fala

Do total de gravações que formam o CFPB, 100 gravações do estilo de fala da entrevista foram escolhidas aleatoriamente, mas respeitando as proporções em cada região geográfica e faixa etária observadas na população brasileira (ver tabela II). Utilizando o Praat, foram extraídos os 40-45 segundos centrais de cada amostra de áudio, respeitando os limites das frases, isto é, os cortes foram feitos em ocorrências de pausas silenciosas.

A duração média do trecho, incluindo pausas e trechos desvozeados, foi de 47,6 segundos (com desvio padrão 2,12) para o estilo espontâneo e 68 segundos (com desvio padrão 16,1) para o estilo leitura. A duração líquida média dos trechos vozeados foi de 28,3 segundos (com desvio padrão 3,42) para o estilo espontâneo e de 29,2 segundos (com desvio padrão 4,4) para o estilo leitura.

A janela de análise, em inglês *window size*, usada pelo algoritmo de extração de f_0 foi de 0,01 segundos (com desvio padrão 0,002) para o estilo espontâneo; e de 0,009 segundos (com desvio padrão 0,001) para o estilo leitura.

Por fim, o número mediano de amostras de f_0 no trecho analisado foi de 2767 para o estilo espontâneo e de 3220 para o estilo leitura.

O objetivo na seleção do trecho a ser recortado foi o de ter uma duração de fala líquida de pelo menos 30 segundos, isto é, depois de descontadas pausas e disfluências, resultando em uma amostra de fala por indivíduo. O próximo passo foi analisar as amostras de leitura, que incluíam 100 gravações dos mesmos indivíduos observados na etapa de fala espontânea. Esse processo foi feito a partir da seleção das 22 frases lidas pelos entrevistados, disponíveis no apêndice A, excluindo-se o silêncio entre elas. Quando o tempo de leitura de todas as sentenças juntas não ultrapassa pelo menos 40 segundos, foi preciso incluir na análise a repetição de alguma das frases - para os casos em que os sujeitos liam duas vezes o conjunto das frases - ou selecionar um pedaço de outro trecho lido, geralmente a leitura de pequenas fábulas.

Essa etapa também passou pela extração da f_0 a partir de dois *scripts* citados na próxima seção, além da correção manual.

3.4. Extração de f_0

A primeira etapa do processamento dos dados consistiu em extrair as informações relevantes a respeito da f_0 por meio do programa de análise acústica Praat. Em seguida, o ambiente de computação estatística R, foi usado para gerar as análises descritivas e inferenciais.

Utilizando dois *scripts* para o programa Praat, extraiu-se a frequência fundamental de cada amostra. Os resultados da extração automática foram inspecionados visualmente e erros de extração remanescentes foram manualmente corrigidos.

O primeiro *script*, *better-f0* (ARANTES, 2019), melhora a extração da f_0 a partir da otimização dos intervalos entre piso e teto do parâmetro passado pelo usuário, funcionando a partir de um algoritmo baseado na autocorrelação da f_0 .

Já o segundo, f_0 -outliers (ARANTES, 2021), sinaliza um possível erro de extração na curva da f_0 , possibilitando uma correção manual mais precisa e cuidadosa.

3.5. Análise de f_0 em pausas preenchidas hesitativas

Na fase de delimitação do trecho a ser analisado para posterior extração de f_0 , descrita na seção 3.4 do presente capítulo, notou-se a ocorrência relativamente frequente de hesitações nas amostras de fala do estilo entrevista semiespontânea. As ocorrências de pausas preenchidas hesitativas mostraram-se um ponto de atenção, sendo caracterizadas por uma duração comparativamente mais longa do que vogais em trechos fluentes, por contornos de f_0 com valores relativamente baixos de f_0 e baixa variabilidade, isto é, mantêm-se em um nível relativamente estável durante seu curso. Além disso, a qualidade dos sons que preenchem essas pausas é relativamente limitada e grande parte consiste em sons que podem ser reconhecidos como vogais tais como [ɛ] e [i], além de sons nasalizados sem qualidade vocálica aproximável a vogais específicas do português brasileiro, como [ẽ].

Adotou-se para as finalidades do presente trabalho o entendimento segundo o qual pausas preenchidas hesitativas podem ser entendidas como interrupções momentâneas na fluência da produção da fala, que seriam causadas por "dificuldades no processamento ou na verbalização da informação" (MERLO, 2006), que levam o falante a "interromper momentaneamente o texto [falado] para resolver sua dificuldade". Importante salientar que a ocorrência de hesitação é parte da "atividade de construção do texto falado" (Koch & Souza e Silva, 1996; Marcuschi, 1999 *apud* Merlo 2006), de modo que sua presença não é índice da presença de patologia ou distúrbio de fala, devendo-se esperar a sua ocorrência na fala de pessoas com desenvolvimento típico de linguagem¹⁷. É indicada a leitura de Merlo (2006) para uma discussão mais aprofundada a respeito da noção de fluência e de hesitação. Há, sobretudo, nuances de entendimento e definição do conceito em função dos interesses de diferentes autores, sendo eles pesquisadores na área de linguística ou fonoaudiologia.

¹⁷ É indicada a leitura de Merlo (2006) para uma discussão mais aprofundada a respeito da noção de fluência e de hesitação e das nuances de entendimento e definição desses conceitos em função dos interesses de diferentes autores estarem mais ligados à pesquisa na área de linguística ou fonoaudiologia. Como o foco do presente trabalho não é problematizar essas questões, adotamos a definição que nos pareceu caracterizar minimamente o fenômeno assumindo o menor número possível de compromissos teóricos.

Hesitações manifestam-se tanto no nível segmental quanto suprasegmental por meio de marcas formais que, tradicionalmente, são classificadas nos seguintes tipos: “pausas silenciosas hesitativas, pausas preenchidas, repetições hesitativas, prolongamentos hesitativos e falsos inícios” (Koch & Souza e Silva, 1996; Marchuschi, 1999 *apud* Merlo 2006, p. 20). As ocorrências que identificamos nos dados do CFPB e cujo perfil de f_0 analisados podem ser encaixadas na categoria das pausas preenchidas, que são consideradas “fonos de enchimento, que, em português brasileiro, tendem a se manifestar como 'éh', 'ah', 'ahn', e 'mm’” (Merlo 2006, p. 212). Na seção 4.4, em que relatamos os resultados da análise, reporta-se os diferentes tipos de qualidade desses fonos que encontramos nas amostras analisadas, bem como sua frequência de ocorrência.

As características das pausas preenchidas hesitativas encontradas nos dados parece aproximá-las à descrição do conceito de sinal carreador (*carrier signal*) presente na teoria da modulação (*modulation theory of speech*), proposta por Traunmüller (1994). Essa teoria propõe que a informação veiculada pela fala pode ser entendida como resultado da modulação de um sinal carreador. O sinal carreador é caracterizado por Traunmüller como “tendo propriedades características de uma vogal 'neutra', aproximadamente [ə], produzida em um nível baixo de f_0 ”¹⁸, sendo afetado por quatro classes de fatores que atuam simultaneamente para modular suas características: linguísticos, expressivos, orgânicos e condições de transmissão.

Os fatores linguísticos são de tipo convencional/social, como a língua, o dialeto regional/social do falante e informam o conteúdo propriamente linguístico da mensagem, como os sons que conformam as palavras e os padrões prosódicos da mensagem linguística em questão. Os fatores expressivos são de natureza psicossocial e englobam fenômenos paralinguísticos, como emoções e atitudes. Os fatores orgânicos são de natureza extralinguística e transmitem informação sobre características do falante, tais como seu sexo, idade, presença de alguma patologia ou alteração transitória (como uma gripe, por exemplo). Os fatores de transmissão dizem respeito à localização espacial do falante e às condições ambientais para a transmissão do sinal acústico e veiculam informação sobre a presença de ruído ou distorções no canal de transmissão.

¹⁸ No original: “properties characteristic of a 'neutral' vowel, approximately [ə], phonated with a low f_0 ” (Traunmüller 1994, p. 176).

A teoria da modulação propõe que a percepção da fala consiste na operação de reconhecer as modulações introduzidas pelos fatores mencionados anteriormente no sinal carreador. Por meio do processo de desmodulação, o ouvinte reconstruiria, segundo o modelo, os gestos articulatórios responsáveis pela produção dos sons da fala que compõem a mensagem linguística, além de inferir informações a respeito do estado emocional do ouvinte, bem como características físicas do falante e das condições ambientais em que a fala foi produzida.

A proposta deste trabalho é analisar os contornos de f_0 das pausas preenchidas hesitativas e considerá-los como representantes aproximados da f_0 do sinal carreador de cada falante, que seria possível considerar uma propriedade relativamente invariante da produção vocal de um determinado falante.

Reconhece-se que essa aproximação é limitada pelo fato da qualidade acústica dos alongamentos não poder ser descrita exatamente como a de uma vogal neutra, mas sim, na maioria dos casos, como uma das vogais típicas do português brasileiro. Desse modo, o sinal que analisamos ainda estaria afetado em algum grau por uma modulação de natureza linguística. No entanto, o fato de haver relativamente pouca variabilidade nas qualidades vocálicas que são observadas em nossa amostra e o fato do contorno de f_0 nas pausas preenchidas tender a ser relativamente monotônico, jogam a favor da proposta de aproximar as duas coisas, bem como o fato dessas disfluências não representarem um estilo de elocução definido ou emoção marcada.

Considerando tudo isso, torna-se mais plausível imaginar que a f_0 produzida nas pausas preenchidas hesitativas expresse as características mais invariantes do sinal carreador modulado sobretudo pelos fatores orgânicos, de modo que pode refletir mais as características físicas responsáveis pela f_0 típica do falante, como tamanho e densidade das pregas vocais, e menos as modulações adicionadas ao sinal carreador pela imposição de uma cadeia segmental e suprasegmental específica.

É possível ressaltar que outras formas de hesitação também puderam ser identificadas nas amostras, embora não tenham sido objeto de atenção em função dos propósitos do presente trabalho. Como dito anteriormente, por conta de se manifestarem com frequência como segmentos relativamente mais alongados do que uma vogal usada na fala corrente e com perfil quase monotônico de f_0 , são candidatas mais relevantes para os propósitos de tentar revelar características de invariância na produção de f_0 dos falantes do que os outros tipos de hesitações.

3.6. Análise estatística

Apresenta-se nesta seção os procedimentos adotados para a análise estatística dos dados extraídos da amostra de 100 áudios selecionados do *corpus* CFPB. Todas as análises estatísticas e gráficos foram gerados usando o ambiente de computação estatística R e de bibliotecas de funções.

Para verificar se as amostras de f_0 são unimodais, usamos o chamado *dip test* (HARTIGAN; HARTIGAN, 1985), que testa a hipótese nula de unimodalidade contra a hipótese alternativa de não unimodalidade. O teste está disponível na biblioteca *dipTest* para o R.

Para verificar se uma determinada distribuição de dados pode ser adequadamente modelada pela distribuição normal, foi usado o teste estatístico de D'Agostino, que pode ser considerado um dos testes mais robustos entre os disponíveis e leva em conta simultaneamente desvios de assimetria e curtose para determinar a adequação do ajuste da curva normal à amostra em análise (D'AGOSTINO et al., 1990). A hipótese nula do teste é a de que os dados seguem uma distribuição normal e a hipótese alternativa é a de que os dados não seguem uma distribuição normal por desvios de assimetria e/ou curtose. As estatísticas Z_1 e Z_2 , computadas como etapas intermediárias do teste de D'Agostino, foram usadas para determinar a significância estatística dos desvios na assimetria e curtose das amostras, respectivamente. Nos dois casos, a hipótese nula é a de não desvio em relação aos valores esperados para a distribuição normal (0 para assimetria e 3 para curtose) e a hipótese alternativa é a de desvio em relação aos valores esperados.

Testes t de amostras independentes foram usados para a comparação da média dos diversos estimadores estatísticos nos dois estilos de elocução (fala espontânea e leitura de frases). O pressuposto de homogeneidade de variância foi testado usando o teste Fligner-Killeen e nas situações em que ele não foi atendido, a correção dos graus de liberdade de Welch foi aplicada ao teste t . O índice d de Cohen foi usado para estimar o tamanho do efeito nas comparações pareadas.

Testes de análise de variância (ANOVA) de um fator foram usados para a comparação das médias dos diversos estimadores estatísticos nos diferentes níveis da variável região geográfica. O pressuposto de homogeneidade de variância foi testado usando o teste Fligner-Killeen. O índice η^2 foi usado para estimar o tamanho do efeito.

Foi estabelecido 5% como nível para estabelecimento de significância em todos os testes de hipótese.

CAPÍTULO 4 - RESULTADOS

4.1. Caracterização geral das distribuições de f_0 na fala encadeada

A aplicação dos testes de normalidade indica que nenhuma das distribuições de f_0 no estilo fala espontânea é adequadamente modelada pela distribuição normal, quer na escala Hertz, quer na escala de semitons. No estilo leitura de frases, a porcentagem de distribuições que segue a distribuição normal é de 3% na escala Hertz e 5% na escala de semitons. O estilo de elocução leitura tem um pequeno efeito no aumento da proporção de distribuições que seguem a distribuição normal.

Em relação à assimetria, a aplicação do teste estatístico indica que no estilo de fala espontânea o valor obtido desvia significativamente daquele esperado para a distribuição normal em 100% dos casos na escala Hertz e em 98% dos casos na escala de semitons. No caso do estilo de elocução leitura, os números caem para 93% (Hertz) e 84% (semitons). O estilo leitura tende a gerar mais distribuições simétricas, especialmente na escala semitons, embora a maioria expressiva das distribuições continue sendo assimétrica.

Os testes que detectam desvio em relação ao valor típico para a curtose da distribuição normal mostram que no estilo fala espontânea existe desvio significativo em 96% das distribuições medidas na escala Hertz 83% na escala de semitons. No estilo leitura de frases as proporções caem para 78% (Hertz) e 72% (semitons). Tanto o estilo de elocução quanto a conversão para a escala de semitons têm o efeito de aumentar a proporção de distribuições com valores que aderem à distribuição normal.

A aplicação do teste para detecção de não unimodalidade indica que 3% das distribuições no estilo fala espontânea apresentam mais do que uma moda, independentemente da escala de mensuração de f_0 . No estilo leitura de frases, os números são 7% (Hertz) e 5% (semitons). O estilo leitura de frases mostra uma leve tendência a produzir mais distribuições não unimodais do que o estilo fala espontânea. Em análise sobre a presença de bimodalidade em distribuições de f_0 produzidas por falantes de PB, Arantes (2021) mostra que a bimodalidade é mais comum no estilo leitura de palavras isoladas (que o CFPB não contempla) do que no estilo leitura de frases ou fala espontânea. No estudo, o autor argumenta, com o auxílio da análise de contornos de f_0 temporalmente normalizados, que a bimodalidade pode se dever a movimentos ascendentes ou

descendentes rápidos ligados à realização de tons alinhados à sílaba tônica da palavra. Em um levantamento sobre a presença de bimodalidade em amostras de f_0 de sete línguas diferentes, o PB entre elas, em três estilos de elocução (leitura de palavras isoladas, leitura de frases e fala espontânea), Silveira e Arantes (2017) reportam a ocorrência, em especial nos estilos fala espontânea e leitura de frases, de casos de bimodalidade nos quais uma das modas localiza-se entre meia e uma oitava abaixo da mediana da distribuição total da amostra.

Os autores reportam que a inspeção auditiva e visual dos trechos que correspondem aos valores muito baixos que formam essa distribuição em torno da moda mais baixa muitas vezes são ocorrências de vozeamento não modal, em especial voz laringalizada ou crepitante. Não foi feita essa checagem nos casos de bimodalidade detectada nos dados do CFPB, mas esse é um levantamento que poderia ser feito em estudo separado, uma vez que a frequência de ocorrência de episódios de vozeamento não modal pode ter o potencial de ser uma dimensão idiossincrática da produção da fala e, ter algum papel na tarefa de comparação forense de vozes. Conferir Oliveira (2017) a esse respeito.

Os resultados indicam que, tomadas individualmente, as distribuições de f_0 tendem a ser assimétricas e apresentar desvios importantes na curtose, o que torna impróprio seu modelamento pela distribuição normal, que é caracterizada pela simetria e por valores de curtose próximos a 3.

4.2. Estilo de elocução

As tabelas a seguir apresentam os valores médios para cada um dos descritores estatísticos em função dos estilos de elocução leitura e fala semiespontânea, tanto na escala Hertz (tabela V) quanto na escala de semitons (tabela VI).

Tabela V: Valores médios dos descritores estatísticos, medidos na escala Hz em função dos dois estilos de fala

Descritor estatístico	Espontânea	Leitura
Moda	122,1	131,1
Média	136,0	132,4
Mediana	129,9	131,8
Valor de base	106,3	103,9
Desvio-padrão	27,3	20,1
CV (%)	19,7	15,0
Assimetria	1,3	0,5
Curtose	5,8	4,2

Resultados da comparação estatística entre médias (na escala Hz) nos dois estilos de elocução:

Moda: $t(198) = -2,86$; $p = 0,005$; $d = -0,4$; $FK\ x^2(1) = 1,46$; $p = 0,23$.

Média: $t(198) = 1,07$; $p = 0,29$; $d = 0,15$; $FK\ x^2(1) = 0,15$; $p = 0,7$.

Mediana: $t(198) = -0,57$; $p = 0,57$; $d = 0,08$; $FK\ x^2(1) = 0,003$; $p = 0,96$.

Valor de base: $t(198) = 0,98$; $p = 0,33$; $d = 0,14$; $FK\ x^2(1) = 0,72$; $p = 0,4$.

Desvio-padrão: $t(161) = 5,95$; $p < 0,001$; $d = 0,84$; $FK\ x^2(1) = 5,57$; $p = 0,018$.

CV: $t(169) = 7,91$; $p < 0,001$; $d = 1,12$; $FK\ x^2(1) = 7,27$; $p = 0,007$.

Assimetria: $t(198) = 11,9$; $p < 0,001$; $d = 1,69$; $FK\ x^2(1) = 0,0006$; $p = 0,98$.

Curtose: $t(197) = 4,82$; $p < 0,001$; $d = 0,68$; $FK\ x^2(1) = 14,9$; $p < 0,001$.

Tabela VI: Valores médios dos descritores estatísticos, medidos na escala semitons (rel 1 Hz) em função dos dois estilos de fala

Descritor estatístico	Espontânea	Leitura
Moda	83,2	84,5
Média	84,4	84,2
Mediana	84,0	84,3
Valor de base	80,5	80,2
Desvio-padrão	3,2	2,6
CV (%)	3,7	3,1
Assimetria	0,7	0,0
Curtose	3,8	3,3

Resultados da comparação estatística entre médias (na escala de semitons) nos dois estilos de elocução:

Moda: $t(198) = -2,8$; $p = 0,006$; $d = -0,4$; $FK x^2(1) = 0,003$; $p = 0,95$.

Média: $t(198) = 0,69$; $p = 0,49$; $d = 0,098$; $FK x^2(1) = 0,13$; $p = 0,72$.

Mediana: $t(198) = -0,7$; $p = 0,49$; $d = -0,099$; $FK x^2(1) = 0,13$; $p = 0,72$.

Valor de base: $t(198) = 0,89$; $p = 0,38$; $d = 0,13$; $FK x^2(1) = 0,41$; $p = 0,52$.

Desvio-padrão: $t(198) = 6,2$; $p < 0,001$; $d = 0,88$; $FK x^2(1) = 1,38$; $p = 0,24$.

CV: $t(198) = 6,32$; $p < 0,001$; $d = 0,89$; $FK x^2(1) = 1,68$; $p = 0,2$.

Assimetria: $t(198) = 12,1$; $p < 0,001$; $d = 1,71$; $FK x^2(1) = 0,89$; $p = 0,34$.

Curtose: $t(197) = 4,04$; $p < 0,001$; $d = 0,57$; $FK x^2(1) = 10,7$; $p = 0,001$.

O estilo de elocução causa efeito estatisticamente significativo nos estimadores de tendência central apenas na moda ($E < L$)¹⁹ e com tamanho de efeito de magnitude considerada

¹⁹ Leia-se “estilo fala espontânea tem valor menor do que estilo leitura de frases”.

baixa²⁰. Nos demais, a diferença é significativa, sempre na direção $E > L$ e com tamanho de efeito de magnitude moderada a alta. Vê-se que o estilo de elocução fala espontânea tende a não provocar diferenças nos estimadores de tendência central (a menos da moda), mas aumenta a variabilidade geral das distribuições (DP e CV), além da assimetria e da curtose, em comparação com o estilo leitura de frases. Nota-se, finalmente, que a conversão de escala de mensuração da f_0 provoca diminuição nas diferenças numéricas entre os dois estilos, embora as comparações estatísticas mostrem que as diferenças significativas permanecem, tanto a direção dos efeitos quanto o tamanho dos efeitos.

As figuras II, III e IV correspondem aos histogramas gerados para as medidas de tendência central, dispersão, assimetria e curtose dos valores de f_0 em função dos estilos de elocução leitura e fala espontânea. Os resultados são apresentados nas escalas Hertz e semitons.

²⁰ Em relação ao tamanho do efeito, uma prática comum é usar uma convenção proposta por Cohen (1988), o autor responsável pela popularização da medida d , que leva seu nome. Essa convenção baseia-se em um extenso levantamento de valores da medida d feita pelo autor na literatura científica da área da psicologia experimental. A convenção estabelece que o efeito pode ser considerado de magnitude baixa se o valor de d for próximo de 0,2, média se estiver próximo de 0,5 e alta se for próximo de 0,8. Deve-se ter em mente, no entanto, que essa convenção não deve ser interpretada de forma absoluta. É preciso ter em conta o tipo de dado em questão no experimento (se é duração segmental, tempo de reação etc) e resultados de mensurações de tamanho de efeito anteriores reportados na literatura. No caso de tamanhos de efeito que envolvam a frequência fundamental, não é óbvio que essa convenção corresponda a diferenças que sejam, do ponto de vista linguístico ou mesmo meramente perceptual, baixas, médias ou altas. Não consta que essa questão já tenha sido tratada na literatura especializada, mas, por tradição, adota-se a convenção corrente baseada em Cohen.

Figura II - Histogramas das medidas de tendência central em função dos dois estilos de elocução (“E”: espontânea e “L”: leitura de frases), valores de f_0 em Hertz e semitons.

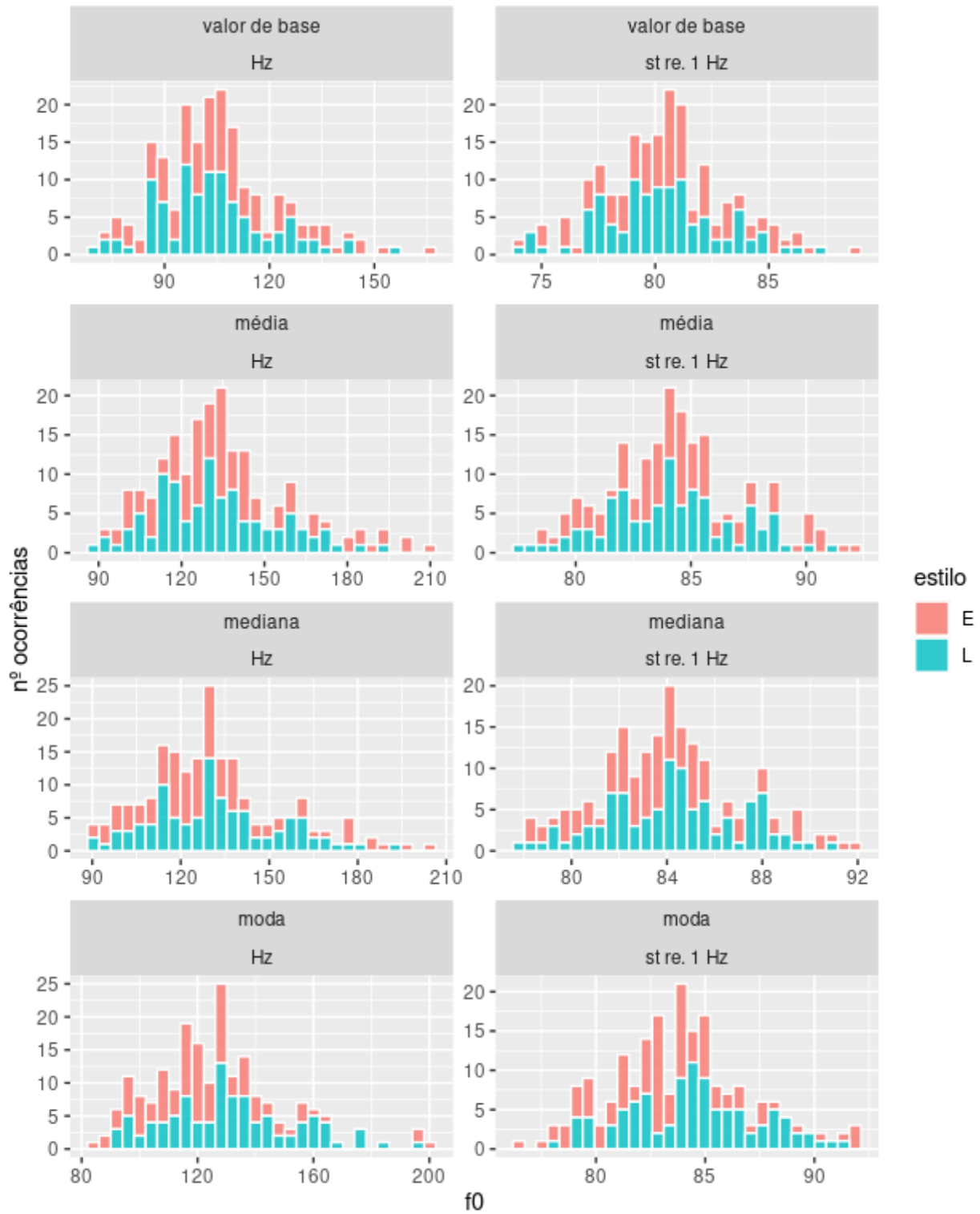


Figura III - Histogramas das medidas de dispersão em função dos dois estilos de elocução (“E”: espontânea e “L”: leitura de frases), valores de f_0 em Hertz e semitons.

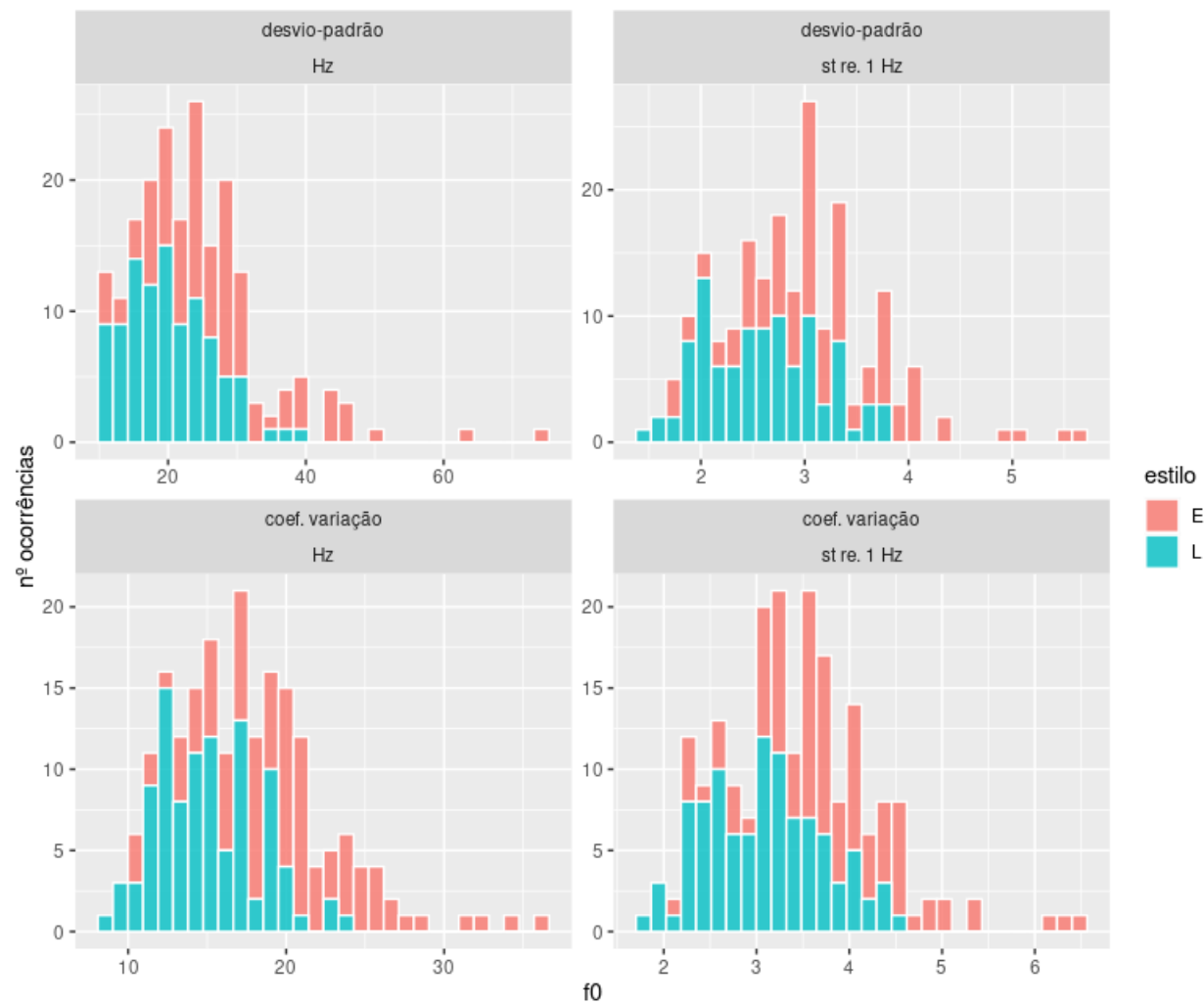
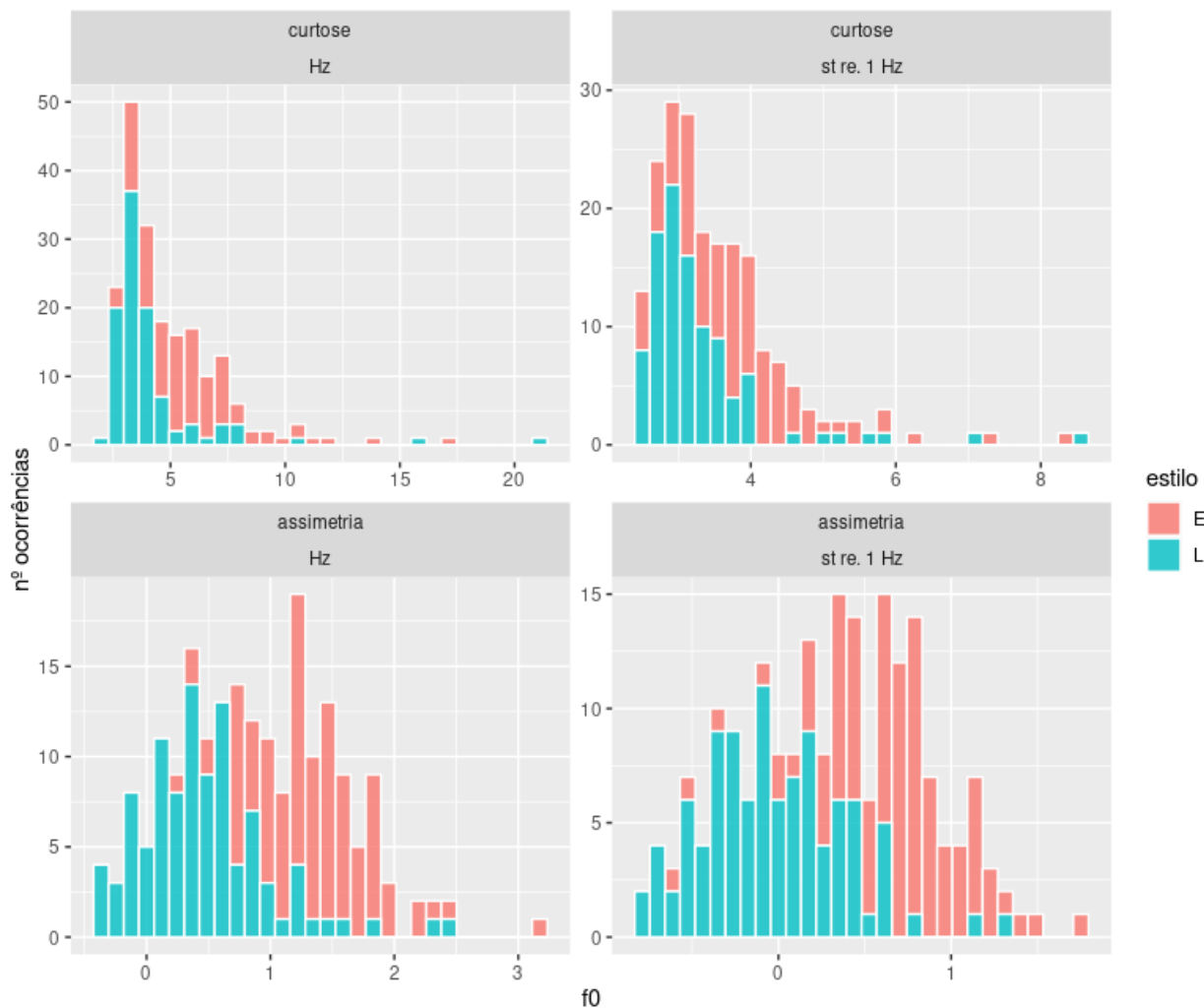


Figura IV - Histogramas de assimetria e curtose em função dos dois estilos de elocução (“E”: espontânea e “L”: leitura de frases), valores de f_0 em Hertz e semitons.



A tabela VII mostra os resultados dos testes de normalidade aplicados às distribuições amostrais dos estimadores estatísticos nas duas escalas de mensuração. Nos casos em que não há diferença significativa entre os dois estilos de elocução para um determinado estimador, as duas amostras foram reunidas para a realização do teste. Os resultados mostram que a maioria das distribuições dos estimadores na escala Hertz não são adequadamente modeladas pela distribuição normal, com exceção da moda, desvio-padrão e CV no estilo leitura de frases. No caso dos dados na escala de semitons, os estimadores de tendência central podem ser modelados

pela distribuição normal, com exceção da moda no estilo fala espontânea. No caso dos estimadores de dispersão, apenas a distribuição de CV no estilo leitura segue a distribuição normal. Para os estimadores de momentos superiores, apenas a distribuição de assimetria no estilo leitura segue a distribuição normal.

De forma geral, a conversão para a escala de semitons tem o efeito de dobrar o número de distribuições que podem ser modeladas pela distribuição normal (de 3 para 6, entre 13 distribuições). Esses resultados são relevantes, conforme destacado na seção 2.6 do capítulo 3, na medida em que é preciso levar em conta a forma da distribuição do estimador de interesse no procedimento de geração do valor da razão de verossimilhança. Na prática usual, assume-se que as distribuições amostrais dos estimadores seguem distribuições normais (ROSE, 2004), como dissemos na seção 2.6 do capítulo 3. Em trabalhos posteriores, seria importante fazer um estudo para mostrar os possíveis efeitos de usar inadequadamente a distribuição normal como modelo teórico para a distribuição de f_0 sobre o cálculo de valores da razão de verossimilhança. O que nossas análises mostram é que grande parte das distribuições amostrais dos estimadores têm mais assimetria do que o previsto pela distribuição normal e isso pode afetar a acuidade da aplicação da razão de verossimilhança. Uma saída para essa situação seria recomendar a troca da distribuição normal por distribuições mais apropriadas. No caso dos dados do CFPB, essa distribuição seria a Burr tipo XII. Uma solução proposta por Morrison (2010), que parece ser adotada no campo, é o uso da técnica de misturas de distribuições normais para modelar distribuições populacionais que não aderem a uma distribuição normal canônica. Poderia ser interessante comparar o desempenho dessas duas soluções (adoção da distribuição Burr e de misturas de normais) em situações realistas.

Tabela VII: Resultados dos testes de normalidade aplicados às distribuições amostrais dos estimadores estatísticos, separados por escala e estilo de fala (E: espontânea, L: leitura).

Resultados significativos ($p < 0,05$) são marcados por “*”

Estimador	Escala	
	Hertz	Semitons
Moda	$\chi^2 = 27,3; p < 0,001$ (E) * $\chi^2 = 3; p = 0,22$ (L)	$\chi^2 = 7,52; p = 0,023$ (E) * $\chi^2 = 1,36; p = 0,51$ (L)
Média	$\chi^2 = 18,1; p < 0,001$ *	$\chi^2 = 2,06; p = 0,36$
Mediana	$\chi^2 = 16,3; p < 0,001$ *	$\chi^2 = 2,41; p = 0,3$
Valor de base	$\chi^2 = 17,8; p < 0,001$ *	$\chi^2 = 1,47; p = 0,48$
Desvio-padrão	$\chi^2 = 40,1; p < 0,001$ (E) * $\chi^2 = 5,57; p = 0,062$ (L)	$\chi^2 = 14,2; p < 0,001$ (E) * $\chi^2 = 6,16; p = 0,046$ (L) *
CV	$\chi^2 = 13,7; p = 0,001$ (E) * $\chi^2 = 2,97; p = 0,23$ (L)	$\chi^2 = 12,8; p = 0,002$ (E) * $\chi^2 = 4,28; p = 0,12$ (L)
Assimetria	$\chi^2 = 9,92; p = 0,007$ (E) * $\chi^2 = 29,2; p < 0,001$ (L) *	$\chi^2 = 6,26; p = 0,044$ (E) * $\chi^2 = 4,7; p = 0,095$ (L)
Curtose	$\chi^2 = 52,8; p < 0,001$ (E) * $\chi^2 = 127; p < 0,001$ (L) *	$\chi^2 = 50,7; p < 0,001$ (E) * $\chi^2 = 97,2; p < 0,001$ (L) *

O resultado do estudo de ajuste de distribuições de probabilidade teóricas às distribuições amostrais dos estimadores estatísticos de f_0 é sumarizado na tabela VIII a seguir. Deve-se notar que as distribuições Gama, Weibull e Burr não puderam ser usadas no ajuste das distribuições de dados de assimetria na escala de semitons, uma vez que essa distribuição empírica contém dados negativos e aquelas distribuições teóricas só admitem valores positivos.

O exame da tabela VIII mostra que das 26 distribuições empíricas dos estimadores estatísticos analisados, a maioria expressiva, 88,5%, é melhor modelada por distribuições assimétricas. Entre essas, a distribuição Burr é de longe a mais frequente (16 entre 23 casos ou quase 70%), seguida da distribuição gama (6 entre 23 casos ou 26%). Das três distribuições empíricas que são melhor modeladas por distribuições simétricas, apenas uma se ajusta melhor à normal: a assimetria, no estilo de elocução leitura, em semitons.

A conversão dos valores de f_0 para a escala de semitons tem pouco efeito em mudar o tipo de distribuição teórica que melhor descreve os dados empíricos. A única mudança observada é que a distribuição da assimetria em semitons no estilo leitura passa ser melhor descrita pela distribuição normal ao invés da distribuição Gumbel, que é a melhor distribuição para esse parâmetro na escala Hertz.

Os resultados mostrados na tabela VIII são convergentes com os que são mostrados na tabela VII, que indicam que a maioria das distribuições empíricas dos estimadores de f_0 falham em passar no teste de normalidade. Pode-se notar, inclusive, que mesmo algumas distribuições que passam no teste de normalidade não são necessariamente melhor modeladas pela distribuição normal. É o caso, por exemplo, da distribuição da mediana de f_0 , que embora tenha um teste de normalidade não significativo, é melhor modelado pela distribuição assimétrica Burr, muito embora a normal apresente um ajuste apenas 1,7 vezes pior do que o conseguido pela normal.

Tabela VIII: Distribuição de probabilidade com melhor ajuste às distribuições de f_0 nos dois estilos de elocução (E: espontânea, L: leitura) e nas duas diferentes escalas. Quando a melhor distribuição não é a normal, indicamos entre parênteses quantas vezes melhor é o grau de ajuste daquela distribuição em comparação com a normal, segundo um teste de *Maximum Likelihood*.

Estimador	Escala	
	Hertz	Semitons
Moda	E: Burr (normal 4 x) L: Gama (normal 1,66 x)	E: Burr (normal 2,24 x) L: Gama (normal 1,01 x)
Média	E/L: Burr (normal 6 x)	E/L: Burr (normal 1,85 x)
Mediana	E/L: Burr (normal 5,16 x)	E/L: Burr (normal 1,7 x)
Valor de base	E/L: Burr (normal 5 x)	E/L: Burr (normal 1,83 x)
Desvio-padrão	E: Burr (normal 8,58 x) L: Gama (normal 1,7 x)	E: Burr (normal 2,68 x) L: Gama (normal 1,25 x)
CV	E: Burr (normal 4,74 x) L: Gama (normal 1,87 x)	E: Burr (normal 3,67 x) L: Gama (normal 1,32 x)
Assimetria	E: Logística (normal 1,38 x) L: Gumbel (normal 5 x)	E: Logística (normal 2,34 x) L: Normal
Curtose	E: Burr (normal 6 x) L: Burr (normal 35, 8 x)	E: Burr (normal 8,52 x) L: Burr (normal 28,3 x)

4.3. Variação regional

Nesta seção é descrito o efeito da origem regional dos falantes sobre os valores dos descritores estatísticos. As figuras V, VI e VII mostram boxplots dos valores dos diferentes descritores estatísticos em função da origem regional dos falantes.

Figura V: Boxplots das medidas de tendência central em função das regiões brasileiras, valores de f_0 em Hertz e semitons.

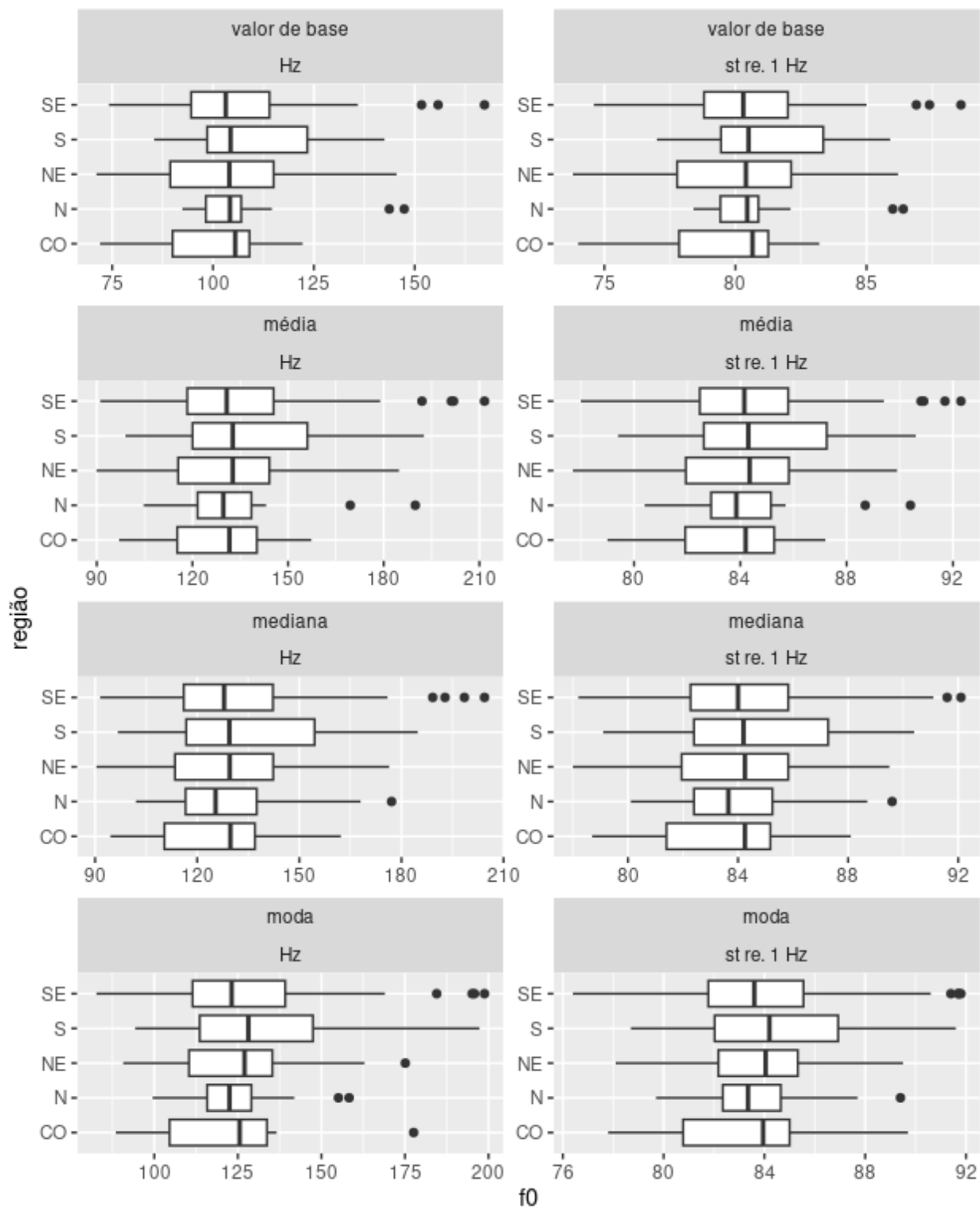


Figura VI: Boxplots das medidas de dispersão em função das regiões brasileiras, valores de f_0 em Hertz e semitons.

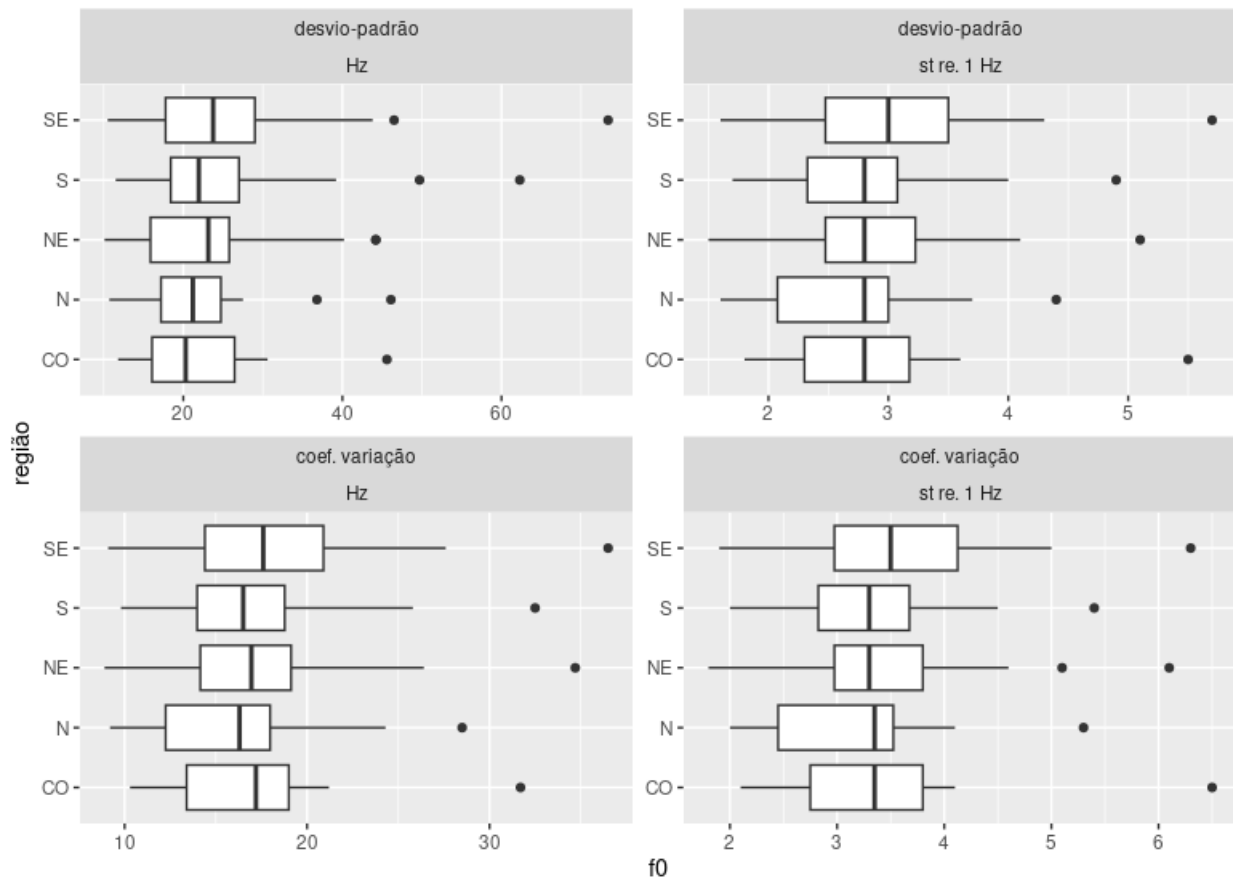
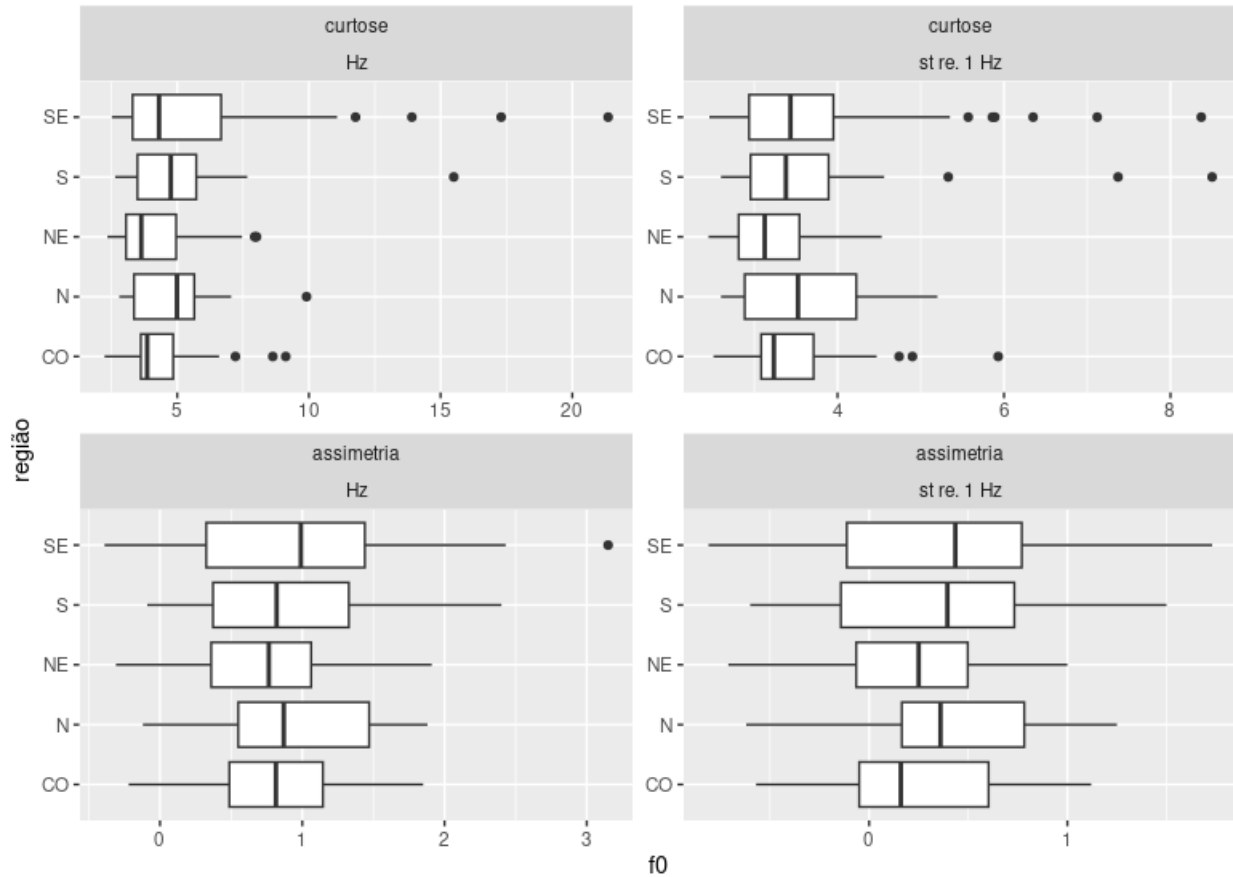


Figura VII: Boxplots das medidas de assimetria e curtose em função das regiões brasileiras, valores de f0 em Hertz e semitons.



A observação dos *boxplots* sugere que a origem regional dos falantes não produz efeitos importantes nos diferentes parâmetros. São apresentados a seguir os resultados da comparação estatística entre as médias (primeiro na escala Hz e depois na escala de semitons) das cinco regiões. Para a análise, as amostras dos dois estilos de elocução foram agregadas.

Moda: $F(4, 195) = 0,64$; $p = 0,63$; $\eta^2 = 0,013$; $FK \chi^2(4) = 2,88$; $p = 0,58$.

Média: $F(4, 195) = 0,64$; $p = 0,63$; $\eta^2 = 0,013$; $FK \chi^2(4) = 2,76$; $p = 0,6$.

Mediana: $F(4, 195) = 0,61$; $p = 0,65$; $\eta^2 = 0,012$; $FK \chi^2(4) = 2,5$; $p = 0,63$.

Valor de base: $F(4, 195) = 0,93$; $p = 0,45$; $\eta^2 = 0,019$; $FK \chi^2(4) = 3,23$; $p = 0,52$.

Desvio-padrão: $F(4, 195) = 0,67$; $p = 0,61$; $\eta^2 = 0,014$; $FK \chi^2(4) = 0,54$; $p = 0,97$.

CV: $t(169) = F(4, 195) = 0,87$; $p = 0,48$; $\eta^2 = 0,018$; $FK \chi^2(4) = 2,04$; $p = 0,73$.

Assimetria: $F(4, 195) = 1,09$; $p = 0,36$; $\eta^2 = 0,022$; $FK \chi^2(4) = 6,86$; $p = 0,14$.

Curtose: $F(4, 195) = 1,98$; $p = 0,099$; $\eta^2 = 0,039$; $FK \chi^2(4) = 10,4$; $p = 0,034$.

Os resultados dos testes confirmam que a origem regional dos falantes não tem efeito estatisticamente significativo sobre os valores dos estimadores, a não ser pela significância marginal no caso da curtose. Pode-se questionar se a diferença entre o número de amostras por região pode ser um fator que influencie os resultados, em especial a falta de efeitos significativos. Um dos possíveis efeitos da diferença no número de amostras em cada condição, nesse caso, as regiões geográficas, seria uma diferença na variância das condições, violando, por exemplo, a homogeneidade de variância, necessária para a aplicação de sua análise. Os resultados dos testes dessa condição, no entanto, mostram que apenas no caso da curtose essa premissa é violada.

4.4. f_0 em pausas preenchidas hesitativas

Foram encontradas 259 ocorrências de pausas preenchidas disfluentes nos 100 trechos analisados. Dos 100 falantes da amostra, 87 apresentaram pelo menos uma ocorrência de alongamento disfluyente. O número de ocorrências por falante varia entre 1 e 8, sendo que 70% dos falantes apresenta entre 1 e 3 ocorrências. Os sons mais frequentes nos alongamento são [ε] e [i], com 50% e 29% das ocorrências, respectivamente.

Reteve-se para análise de f_0 apenas as ocorrências em que o contorno tem uma baixa variabilidade, de modo que se aproximasse da condição de invariabilidade e, assim, refletisse, nos termos da teoria da modulação de Traunmüller, as características do sinal carreador mais do que dos fatores moduladores, em especial as funções linguísticas que impõem variação significativa da f_0 . Para isso, foi estabelecido como critério o limiar de 1 semitom (aproximadamente 6%) para a desvio mediano absoluto do contorno de cada ocorrência de alongamento.

Após a aplicação do limiar, restaram 186 ocorrências ou aproximadamente 72% do total, produzidas por 74 falantes diferentes, sendo que a duração mediana dessas ocorrências é de 434 milissegundos (desvio absoluto mediano de 212 ms), que é um valor muito mais alto do que a duração típica das vogais do PB em trechos fluentes.

Em seguida, foi calculado para cada um dos 74 falantes os valores da média e mediana, em Hertz e semitons, dos contornos de f_0 de suas pausas preenchidas. Esses valores foram comparados, então, com os valores da média, mediana, moda e valor de base dos contornos completos de f_0 de cada falante. A tabela IX mostra a diferença mediana entre a mediana dos contornos das pausas preenchidas hesitativas e as quatro medidas estatísticas dos contornos completos, considerando todos os falantes. A primeira linha mostra os valores na escala Hz e a segunda na escala de semitons.

Tabela IX: Diferença mediana entre a mediana dos contornos das pausas preenchidas e as quatro medidas de valor típico dos contornos da amostra completa de f_0 . Valores nas escalas Hertz e semitons.

	Moda	Média	Mediana	Valor de base
Hertz	6,00	18,30	13,05	7,80
Semitom	1,10	2,30	1,90	1,30

Uma análise de variância (ANOVA) foi aplicada ao valor das diferenças medianas de f_0 na escala de semitons como variável dependente e os quatro estimadores estatísticos como variável independente e o resultado mostra que os estimadores exercem efeito significativo sobre o valor das diferenças [$F(3, 292) = 5,9 p < 0,001$]. Diferenças em relação à moda ($p < 0,001$) e valor de base ($p = 0,011$) são significativamente menores do que em relação à média. Diferença em relação à mediana não difere das diferenças em relação a todos os demais estimadores.

Os resultados são interessantes e potencialmente úteis do ponto de vista prático na medida em que a diferença entre o contorno mediano das hesitações analisadas e o contorno completo produzido pelo falante tende a ser um valor bastante baixo, indicando que a f_0 de

pausas preenchidas podem ser vista como uma aproximação razoável da f_0 típica do falante, assim como estimada pela moda e pelo valor de base, em um número bastante alto de falantes. Como a ocorrência de pausas preenchidas disfluentes parece ser uma ocorrência bastante comum em estilos de elocução espontânea, o fato da f_0 dessas ocorrências se aproximar do nível geral da f_0 estimada a partir de trechos mais longos pode ser útil do ponto de vista prático, já que no contexto de tarefas de comparação forense de vozes não é incomum que a duração das amostras questionadas seja limitada.

4.5. Discussão

De forma geral, o quadro apresentado na seção 4.1 mostra convergência com resultados relatados em Arantes (2021), que mostram a prevalência da assimetria e desvios de curtose em amostras de f_0 , além da presença de distribuições não unimodais. Apesar da diferença no número de contornos estudados - 10 por estilo de fala em Arantes (2021) e 100 por estilo de fala aqui - a proporção de contornos afetados pela não normalidade, presença de assimetria, excesso de curtose e não unimodalidade é bastante semelhante nos dois estudos. Outro aspecto da caracterização das distribuições de f_0 apontado em Arantes (2021) que foi replicado no presente trabalho é o efeito da conversão dos valores de f_0 da escala física (Hertz) para a escala não linear de semitons na diminuição dos valores de assimetria e curtose, bem como no aumento da proporção de distribuições que não apresentam desvios estatisticamente significativos nesses estimadores em relação ao esperado para a distribuição normal.

Os resultados descritos na seção 4.2 mostram ainda mais uma convergência em relação ao que é relatado em Arantes (2021) em relação ao efeito dos estilos de elocução sobre as distribuições de f_0 : efeitos de baixa magnitude nos estimadores de tendência central e efeitos de magnitude maior e diferença estatisticamente significativa nos estimadores e variabilidade e nos momentos superiores (assimetria e curtose). A direção dos efeitos também é a mesma encontrada em Arantes (2021), isto é, o estilo leitura de frases tende a gerar distribuições com valores menores de desvio-padrão, menor grau de assimetria e valores de curtose próximos a 3.

Do ponto de vista da determinação da razão de verossimilhança, um resultado importante a ser ressaltado é o fato de que a distribuição amostral de parte dos estimadores estatísticos não segue a distribuição normal, mesmo com a conversão dos valores para a escala de semitons. Essa

questão pode ter relevância prática, uma vez que parece ser o caso que a aplicação da LR em contexto de tarefas forenses de comparação de vozes, como sugere a apresentação didática de Morrison (2010), tende a assumir que tanto a distribuição populacional quanto a do suspeito seguem distribuições normais. Essa discrepância pode ter impacto nos valores de LR que seria importante apontar e possivelmente estimar.

Os resultados a respeito do efeito da variação regional relatados na seção 4.3 são interessantes na medida em que sugerem a inexistência de uma influência importante da região geográfica sobre o valor dos estimadores estatísticos coletados no presente trabalho. É preciso levar em conta, no entanto, que adotou-se aqui a divisão regional sugerida pelo IBGE, cuja elaboração levou em conta critérios históricos, geográficos e econômicos e não necessariamente corresponde a fronteiras linguísticas. De todo modo, o fato de haver homogeneidade de variância na maioria das análises que envolveram a variável regional pode ser tomado como indicação de que, pelo menos para os estimadores estudados aqui, as regiões definidas pelo IBGE são homogêneas internamente e quando comparadas entre si. Um outro fato que é preciso considerar é o número relativamente baixo de amostras nas regiões Norte (8) e Centro-Oeste (9). Seria interessante se os responsáveis pelo CFPB conseguissem incluir mais falantes dessas regiões ao corpus, de modo que fosse possível verificar se os resultados encontrados agora se mantêm.

Seja como for, o fato de não haver evidência para um efeito importante da proveniência regional dos falantes facilita a prática na tarefa de comparação de voz na medida em que, assumindo que a falta de efeito seja verdadeira, a definição da população relevante não precisaria levar em conta a proveniência geográfica tanto do(s) suspeito(s) quanto do falante na amostra questionada, uma vez que essa variável não impacta a variação dos estimadores de f_0 aqui estudados.

Por fim, conforme comentado na seção 2.3, uma das características da amostra de falantes que compõem o CFPB é a predominância de pessoas com nível de escolaridade alta. É possível que uma variável socioeconômica como essa tenha um efeito na manifestação de f_0 (ECKERT, 1997) e essa hipótese merece ser investigada, pois se for confirmada, esse é um fator que pode ser relevante para a definição da população relevante a ser considerada no cálculo de uma razão de verossimilhança. Não obteve-se conhecimento, no entanto, de estudos que tomem a população brasileira como base e que investiguem se o grau de escolaridade ou outra variável que lhe seja equivalente tem efeito sobre a distribuição de f_0 . Nesse caso, o único a ser feito é

notar qual é a composição da amostra usada no estudo em termos de grau de escolaridade para que, no futuro, quando houver essa investigação, tais dados possam ser usados em uma comparação.

CAPÍTULO 5 - CONSIDERAÇÕES FINAIS

Como proposto, o presente trabalho compilou os dados a respeito da distribuição populacional das características da frequência fundamental em recortes demográficos de falantes masculinos do português brasileiro, contribuindo para o conhecimento a respeito desse importante traço acústico.

Para tanto, utilizou o Corpus Forense do Português Brasileiro (CFPB) como base de dados de fala que permitiu gerar as estatísticas populacionais propostas nos objetivos do trabalho, apresentando uma visão geral das distribuições de f_0 para os falantes masculinos do português brasileiro, tornando essas estatísticas mais acessíveis para profissionais da área, bem como demais pesquisadores, além de comparar os resultados obtidos com a literatura científica prévia.

Foram feitos testes de normalidade em que se observou que o estilo de elocução tem um pequeno efeito no aumento da proporção de distribuições que seguem a distribuição normal. Enquanto o estilo de fala espontânea não é modelado pela distribuição normal, 3% das distribuições de leitura de frases seguem tal distribuição, em Hertz e 5% em semitom.

Com relação à assimetria, observou-se que a maioria expressiva das distribuições é assimétrica e que, entre os estilos de elocução presentes no CFPB, o estilo leitura de frases tende a gerar mais distribuições simétricas. Quanto ao desvio em relação ao valor típico para a curtose, ambos os estilos mostraram aumento na proporção de distribuições com valores que não desviam em relação ao esperado para distribuição normal.

De acordo com o teste de não unimodalidade, a leitura de frases foi o estilo que apresentou leve tendência a produzir mais distribuições não unimodais do que o estilo de fala espontânea.

De forma geral, observou-se bastante semelhança em relação aos resultados de Arantes (2021) com respeito à prevalência de assimetria, desvios de curtose e presença de distribuições não unimodais nas distribuições de f_0 . Além disso, também foram replicados efeitos similares na conversão de valores da escala Hertz para a escala semitom a fim de se atingir a diminuição dos valores de assimetria e curtose, bem como o aumento na proporção dos estimadores que apresentam desvios significativos em relação à distribuição normal.

Observou-se, ainda, efeitos de baixa magnitude nos estimadores de tendência central e de magnitude maior nos estimadores e variabilidade nos momentos superiores. O estilo leitura de frases tende a gerar distribuições de f_0 com maiores valores de desvio padrão, salvo pela assimetria e curtose com valores próximos a 3, assemelhando-se ao estudo proposto por Arantes (2021).

Quando se trata da determinação da razão de verossimilhança, os resultados apontaram que parte dos estimadores estatísticos analisados não seguem a distribuição normal, mesmo quando convertidos para a escala de semitons. Esse fato pode ter relevância prática na medida em que, segundo a apresentação feita por Morrison (2010) do cálculo da LR, é prática comum assumir que tanto os dados da população relevante quanto das amostras do suspeito podem ser modelados por curvas de distribuição normal.

Em relação à origem regional dos falantes, os resultados apontam para o fato de que ela não tem efeito estatisticamente significativo sobre os valores dos estimadores de f_0 , salvo sua significância marginal no caso da curtose. Esse resultado, embora confiável do ponto de vista estatístico, pode ser fruto do fato de que a divisão regional do país proposta pelo IBGE não seja relevante como critério de diferenciação linguística. Também é preciso realizar mais estudos sociofonéticos para determinar quais fatores sociodemográficos afetam mais a expressão de f_0 e que, portanto, deveriam ser usados para definir a população relevante no caso de tarefas de comparação forense de vozes, em especial quando se opta pela metodologia da relevância lógica.

REFERÊNCIAS

AGUILERA, V. de A.; MILANI, G. A. L.; MOTA, J. A. (Org.). Projeto Atlas Lingüístico do Brasil – Documentos I. Salvador: ILUFBA–EDUFBA, 2004.

ARANTES, P. better_f0: A Praat script for better f0 extraction (v1.0). Zenodo. <https://doi.org/10.5281/zenodo.3470108>. 2019. Acesso em 20 de dezembro de 2021.

_____.; Effects of speaking style on the shape of fundamental frequency distributions . Revista da ABRALIN, v. 20, n. 1, p. 1-39, 29 jun. 2021.

_____.; ERIKSSON, A.; GUTZEIT, S. Effect of language, speaking style and speaker on long-term F0 estimation. In: Interspeech 2017. Stockholm: ISCA, 2017. p. 3897–3901.

_____.; ERIKSSON, A.; LIMA, V. G. Minimum sample length for the estimation of long-term speaking rate. In: Speech Prosody 2018. [S.l.: s.n.], p. 661–665, 2018.

_____.; LIMA, V. G. Towards a methodology to estimate minimum sample length for speaking rate. Revista do GEL, v. 14, n. 2, p. 183–197, 2017.

_____.; LINHARES, M. E. N. Efeito da língua, estilo de elocução e sexo do falante sobre medidas globais da frequência fundamental. Letras de Hoje, v. 52, n. 1, p. 26–39, 2017. ISSN 1984-7726.

_____.; parantes/f0-outliers: Version 1.0 (v1.0). Zenodo. <https://doi.org/10.5281/zenodo.5083453>. 2021. Acesso em 20 de dezembro de 2021.

BELL, S.; The facts on file dictionary of forensic science. New York, NY: Facts on File, Inc., 2004.

BRENNER, J. C.; Forensic Science: An Illustrated \dictionary. Boca Raton: CRC Press; 2004.

BOERSMA, Paul. Praat, a system for doing phonetics by computer. *Glott International*, v. 5, n. 9/10, p. 341-345, 2001.

BRAUN, Angelika. Fundamental frequency - how speaker-specific is it? In: BRAUN, ANGELIKA; KÖSTER, JENS-PETER (Org.). *Studies in Forensic Phonetics*. Trier: Wissenschaftlicher Verlag Trier, 1995. p. 9-23.

BUCHSTALLER, I.; KHATTAB, G.; Population samples. In: PODESVA, R. J.; SHARMA, D. (Ed.). *Research methods in linguistics*. Cambridge, UK: Cambridge University Press, 2013. p. 74–95.

CAO, H.; LEI, Y.; Fundamental frequency statistics for young male speakers of mandarin. *Journal of Forensic Science and Medicine*, v. 3, n. 4, p. 217-222, 2017.

CARDOSO, S. A. *Dialetologia*. In: C. FERRARIZI JÚNIOR; MOLLICA, M.C (orgs.). *Sociolinguística, Sociolinguísticas: Uma introdução*. São Paulo: Editora Contexto, 2016.

CASTILHO, A. T. Informações sobre o Projeto de Estudo da Norma Urbana Lingüística Culta (Projeto NURC). *Cadernos de Estudos Lingüísticos*, v. 6, p. 187–190, 1984.

CHAMPOD, C. MEUWLY, D. The inference of identity in forensic speaker recognition. *Speech Communication* 31, 193-203. 2000.

COLEMAN, R. F. WALLS, H. J. The evaluation of scientific evidence. *Criminal Law Review*, 276-287. 1974.

COLLISCHONN, G.; MONARETTO, V. de O. Banco de dados VARSUL: a relevância de suas características e a abrangência de seus resultados. *ALFA: Revista de Linguística*, São Paulo, v. 56, n. 3, 2012.

CONSTANTINI, A.C. **Caracterização prosódica de sujeitos de diferentes variedades de fala do português brasileiro em diferentes relações sinal-ruído**. Tese (doutorado). Universidade Estadual de Campinas, Campinas. 2014. Disponível em: <<https://hdl.handle.net/20.500.12733/1623059>>. Acesso em: 9 ago. 2022.

_____.; BARBOSA, P. A. Prosodic characteristics of different varieties of Brazilian Portuguese. *Revista Brasileira de Criminalística*, v. 4, n. 3, p. 44–53, 2015.

D'AGOSTINO, R. B.; BELANGER, A.; D'AGOSTINO, Jr., R. B. "A suggestion for using powerful and informative tests of normality" (PDF). *The American Statistician*. 44 (4): 316–321. 1990.

DE PAULA MACHADO, Aline. **Análise fonético-acústica para fins forenses em cinco localidades brasileiras a partir da base Brasildata**. 2018. Doutora em Linguística – Universidade Estadual de Campinas, Campinas, 2018. Disponível em: <http://acervus.unicamp.br/index.asp?codigo_sophia=1062209>. Acesso em: 9 ago. 2022.

ECKERT, P. (1997) Age as a sociolinguistic variable. In F. Coulmas (ed.), *Handbook of Sociolinguistics* (p. 151-67). Oxford: Blackwell.

ERIKSSON, A.; Aural/acoustic vs. automatic methods in forensic phonetic case work. In: NEUSTEIN, A.; PATIL, H. A. (Org.). *Forensic Speaker Recognition: Law Enforcement and Counter-terrorism*. [S.l.]: Springer, 2011. p. 41–70.

EVETT, I. W. et al.; The impact of the principles of evidence interpretation on the structure and content of statements. *Science and Justice* 40 (4), 233-239. 2000.

GOLD, E.; FRENCH, P. International practices in forensic speaker comparison. *The International Journal of Speech, Language and the Law*, v. 18, n. 2, p. 293–307, 2011.

_____.; International practices in forensic speaker comparisons: second survey. *International Journal of Speech Language and the Law*, v. 26, n. 1, p. 1–20, jun. 2019. <https://doi.org/10.1558/ijssl.38028>

GONÇALVES, Sebastião Carlos Leite; TENANI, Luciani Ester. Projeto ALiRP: constituição de um banco de dados para o estudo do português falado na região de São José do Rio Preto. *Mosaico (São José do Rio Preto)*, São José do Rio Preto, v. 3, n.2, p. 13-37, 2004

HARTIGAN, J. A.; HARTIGAN, P. M. The Dip Test of Unimodality. *The Annals of Statistics*, v. 13, n. 1, p. 70–84, 1985.

HUDSON, T. et al. F0 statistics for 100 young male speakers of Standard Southern British English. In: *International Congress of Phonetic Sciences*, 16th, 6-10 aug. 2007. *Anais... Saarbrücken*, 2007.

HUGHES, V. S.; The definition of the relevant population and the collection of data for likelihood ratio-based forensic voice comparison. Unpublished PhD thesis, University of York. 2014.

HUSSAIN, S.; IRTZA, S.; HUDSON, T.; JONG, G. F0 statistics for speakers of Urdu, 2012. 48 slides. Disponível em: http://www.pholab.uzh.ch/static/IAFPA/IAFPA2014_Conference_Material.pdf. Acesso em: 10/12/2020.

JESSEN, Michael; KÖSTER, Olaf; GFROERER, Stefan. Influence of vocal effort on average and variability of fundamental frequency. *Speech, Language and the Law*, Equinox Publishing, v. 12, n. 2, p. 174-213, 2005. Doi: 10.1558/sll.2005.12.2.174

JESSEN, Michael. Speaker Classification in Forensic Phonetics and Acoustics. In: MÜLLER, CHRISTIAN (Org.). *Speaker Classification I*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 180–204.

KINOSHITA, Y.; ISHIHARA, S.; ROSE, P. Exploring the discriminatory potential of F0 distribution parameters in traditional forensic speaker recognition. *The International Journal of Speech, Language and the Law*, v. 16, n. 1, p. 91–111, 2009.

KOCH, I. G. V. SOUZA E SILVA, M. C. P. Atividades de composição do texto falado: a elocução formal. In: Castilho, Ataliba Teixeira de & Basilio, Margarida (orgs). *Gramática do Português Falado. Volume I V: Estudos Descritivos*. Campinas: Editora da UNICAMP e São Paulo: FAPESP. p. 379-410. 1996.

LADEFOGED, P. *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago Press, 1971.

LAVER, John. *The phonetic description of voice quality*. New York: Cambridge University Press, 1980.

LENNES, Mietta et al. Comparing Pitch Distributions Using Praat And R. *The Phonetician*, v. 111–112, p. 35–53, 2015.

MARCUSCHI, L. A. A hesitação. In: Neves, Maria Helena de Moura. *Gramática do Português falado. Vol. VII : novos estudos*. São Paulo e Campinas: Humanitas e Editora da Unicamp, 159-194. 1999.

MENDES, R.B. (2013) Projeto SP2010: Amostra da fala paulistana. Disponível em <<http://projetosp2010.fflch.usp.br>>. Acesso em 20/09/2022.

MERLO, S. Hesitações na fala semi-espontânea: análise por séries temporais. *Dissertação de Mestrado*. Campinas, SP. 2006.

MORRISON, G. S. et al.; “Database selection for forensic voice comparison,” in Odyssey 2012: The Language and Speaker Recognition Workshop, June 25–28, Singapore, Proceedings, 74–77. 2012.

_____ ; Forensic voice comparison. In Expert Evidence (Ch 99). (Ed.) I. Freckerlton and H. Selby. Sydney: Thomson Reuters. 2010.

_____ ; Forensic voice comparison and the paradigm shift. Science and Justice, v. 49, n. 4, p. 298–308, 2009.

NOLAN, F. The phonetic bases of speaker recognition. Cambridge: Cambridge University Press, 1983.

OLIVEIRA, J. C. C.; Laringalização no português brasileiro: uma análise em torno do fenômeno laríngeo e implicações para a comparação de locutor. Tese (Mestrado) - Faculdade de Letras, Universidade Federal de Alagoas, Maceió, 2017.

RASO, Tommaso; MELLO, Heliana (Org.). C-oral-Brasil I: corpus de referência do português brasileiro falado informal. Belo Horizonte: Editora UFMG, 2012.

ROSE, P. Forensic speaker identification. London ; New York: Taylor & Francis, 2002.

_____. How effective are long term mean and standard deviation as normalization parameters for tonal fundamental frequency? Speech Communication, v. 10, n. 3, p. 229-247, 1991.

_____. Technical Forensic Speaker Identification from a Bayesian Linguist’s Perspective. In Keynote paper, Forensic Speaker Recognition Workshop, Speaker Odyssey ’ 04. Toledo, Spain, pp. 3–10, 2004.

SANKOFF, D. A quantitative paradigm for the study of communicative competence. In: SANKOFF, D. (Ed.). *The Social Life of Language*. Philadelphia: University of Philadelphia Press, 1980. p. 47–79.

SANTOS, T. Levantamento de informações penitenciárias: INFOPEN - Atualização junho 2016. Brasília, 2017.

SCHIEFER, Ana Maria; ARCURI, Cláudia Fassin. Avaliação da Fluência da Fala. In: MARCHESAN, Irene Queiroz; SILVA, Hilton Justino da; TOMÉ, Marileda Cattelan. *Tratado de Especialidades em Fonoaudiologia*. São Paulo: Guanabara Koogan, 2014.

SILVA, G. A. Proposta de construção de um banco de dados de amostras de fala para uso forense em um arcabouço bayesiano. *Rev. Brasil. Crimin.* 2016; v. 5, n. 1, p. 35-45.

SILVEIRA, I. J. ARANTES, P. Efeitos do vozeamento não modal na descrição estatística de amostras de frequência fundamental. *Anais do VI Colóquio Brasileiro de Prosódia da Fala*, v.4, p. 119 – 122. 2017.

SKARNITZL, R.; VANKOVÁ, J. Fundamental frequency statistics for male speakers of Common Czech. *AUC Philologica*, v. 2017, n. 3, p. 7-17, set. 2017.

TITZE, Ingo. *Principles of voice production*. Englewood Cliffs: Prentice Hall, 1994.

TRAUNMÜLLER, H.; ERIKSSON, A.; The size of F0 excursions in speech production and perception. *Working Papers*. Lund, Sweden: Dept of Linguistics and Phonetics. 43, 1994.

WAND, M. P.; JONES, M. C. *Kernel Smoothing*. Chapman and Hall, London. 1995.

APÊNDICE: LISTA DAS 22 FRASES LIDAS PELOS ENTREVISTADOS

- Olha lá o avião azul.
- Minha mãe namorou um anjo.
- Sônia sabe sambar sozinha.
- Érica tomou suco de pêra e amora.
- Agora, é hora de acabar.
- Você quer me dizer a data?
- Eu precisei de microfone na conferência.
- A paixão dele é a natureza.
- Podia dizer as horas, por favor?
- A fila aumentou ao longo do dia.
- Nunca se pode ficar em cima do muro.
- A proposta foi inspecionada pela gerência.
- Ela saía discretamente.
- Minhas correspondências não estão em casa.
- As queimadas devem diminuir este ano.
- De dia, apague a luz.
- um homem não caminha sem um fim.
- O atabaque do Tito é coberto com pele de gato.
- Procurei Maria na copa.
- Esses são nossos times.
- Daqui a pouco, a gente irá pousar.
- O analfabetismo é a vergonha do país.