UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

APLICAÇÃO DO VALOR DE BASE DA FREQUÊNCIA FUNDAMENTAL VIA ESTATÍSTICA MVKD EM COMPARAÇÃO FORENSE DE LOCUTOR

RONALDO RODRIGUES DA SILVA

ORIENTADOR: JOÃO PAULO CARVALHO LUSTOSA DA COSTA

DISSERTAÇÃO DE MESTRADO EM
ENGENHARIA ELÉTRICA - ÁREA DE CONCENTRAÇÃO
INFORMÁTICA FORENSE E SEGURANÇA DA INFORMAÇÃO

PUBLICAÇÃO: PPGEE.DM - 630/16

BRASÍLIA/DF: DEZEMBRO/2016.

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

APLICAÇÃO DO VALOR DE BASE DA FREQUÊNCIA FUNDAMENTAL VIA ESTATÍSTICA MVKD EM COMPARAÇÃO FORENSE DE LOUCUTOR

RONALDO RODRIGUES DA SILVA

DISSERTAÇÃO DE MESTRADO PROFISSIONAL SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE.

APROVADA POR:

JOÃO PAULO CARVALHO LUSTOSA DA COSTA, Dr., ENE/UNB, TUIL, FRAUNHOFER IIS
(ORIENTADOR)

RICARDO ZELENOVSKY, Dr., ENE/UNB
(EXAMINADOR INTERNO)

EBRAHIM SAMER EL'YOUSSEF, Dr., UFSC
(EXAMINADOR EXTERNO)

Brasília, 13 de Dezembro de 2016.

FICHA CATALOGRÁFICA

SILVA, RONALDO RODRIGUES DA
Aplicação do Valor de Base da Frequência Fundamental via Estatística MVKD em Comparação Forense de Locutor.
[Distrito Federal] 2016.
xiv, 61 p., 297 mm (ENE/FT/UnB, Mestre, Engenharia Elétrica, 2016).
Dissertação de Mestrado - Universidade de Brasília.
Faculdade de Tecnologia. Departamento de Engenharia Elétrica.

1. Frequência Fundamental
2. MVKD
3. Valor de base de F0
4. Comparação Forense de Locutor
I. ENE/FT/UnB
II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

Silva, R. R. (2016). Aplicação do Valor de Base da Frequência Fundamental via Estatística MVKD em Comparação Forense de Locutor. Dissertação de Mestrado em Engenharia Elétrica, Publicação PPGENE.DM - 630/2016, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 61p.

CESSÃO DE DIREITOS

NOME DO AUTOR: Ronaldo Rodrigues da Silva.

TÍTULO DA DISSERTAÇÃO DE MESTRADO: Aplicação do Valor de Base da Frequência Fundamental via Estatística MVKD em Comparação Forense de Locutor.

GRAU / ANO: Mestre / 2016

Ronaldo Rodrigues da Silva
SPO Quadra 7 Lote 23, Setor Policial Sul
70610-200 Brasília - DF - Brasil.

DEDICATÓRIA

À minha família.

AGRADECIMENTOS

A conclusão deste Mestrado foi possível com a colaboração irrestrita de vários colegas que me acompanharam nesta caminhada. Entre os quais destaco e agradeço:

Ao Professor João Paulo Carvalho Lustosa da Costa, meu orientador do ENE/UnB, pelo apoio, direcionamento e pelos conhecimentos transmitidos.

Ao Ricardo Kehrle Miranda que auxiliou nas revisões de artigos e da dissertação.

Ao Professor Pablo Arantes, que incentivou a pesquisa na área e disponibilizou ferramentas fundamentais na consecução das pesquisas.

Aos colegas de trabalho que assumiram atividades durante a minha ausência nos períodos de aula.

A toda a minha família, pela compreensão e apoio.

Ronaldo Rodrigues da Silva

RESUMO

APLICAÇÃO DO VALOR DE BASE DA FREQUÊNCIA FUNDAMENTAL VIA ESTATÍSTICA MVKD EM COMPARAÇÃO FORENSE DE LOCUTOR

Autor: Ronaldo Rodrigues da Silva

Orientador: João Paulo Carvalho Lustosa da Costa

Programa de Pós-graduação em Engenharia Elétrica

Brasília, Dezembro de 2016

Comparação forense de locutor (CFL) é utilizada como uma abordagem complementar na confirmação da autoria de um crime. A metodologia mais difundida mundialmente neste tipo de exame se baseia em análises perceptuais e acústicas.

Uma das medidas acústicas mais utilizadas em CFL é a frequência fundamental (F0). O parâmetro acústico F0 é robusto em áudios de baixa qualidade e é independente do conteúdo das falas, o que o torna um parâmetro interessante de ser utilizado nas análises forenses. Além disso, o algoritmo de extração de F0 apresenta baixa complexidade computacional.

Neste trabalho, propõe-se analisar o poder discriminante da medida de longo termo da frequência fundamental nomeada valor de base de F0, que em trabalhos recentes tem se mostrado menos sujeita a variações associadas ao conteúdo, ao estilo da fala, ao canal utilizado na gravação, além de exigir uma menor quantidade de material para obter uma medida estável em comparação a outras medidas de longo termo, como a média aritmética e o desvio padrão.

Foi avaliado o ganho de poder discriminante ao combinar a medida do valor de base de F0 a outras medidas de longo termo de F0 usualmente utilizadas na área forense por meio de uma abordagem que aplica a estatística de densidade do núcleo de multivariáveis, do inglês *Multivariate Kernel-Density* (MVKD). Os testes foram realizados utilizando um corpus composto de gravações de áudios de falantes masculinos do português brasileiro contendo 60 segundos de produções vozeadas e obteve-se uma Taxa de Erro Igual, do inglês *Equal Error Rate* (EER) de 13 %, superando pesquisas recentes.

ABSTRACT

APPLYING BASE VALUE OF FUNDAMENTAL FREQUENCY VIA MVKD IN FORENSIC SPEAKER COMPARISON

Author: Ronaldo Rodrigues da Silva

Supervisor: João Paulo Carvalho Lustosa da Costa

Programa de Pós-graduação em Engenharia Elétrica

Brasília, December of 2016

Forensic Speaker Comparisons (FSC) are applied as a complementary approach to confirm the authorship of a crime. The methodology most used in FSC is based on perceptual and acoustic analysis.

One of the most frequent measures in FSC is the fundamental frequency F0. The acoustic parameter F0 is robust in low audio quality regardless of the speech content, which is very important to the forensic area. Moreover, its algorithm has a low computational complexity.

In this work, we propose to analyze the discriminatory power of the long-term fundamental frequency parameter named baseline of the F0. This parameter is more stable considering the speech content and style, the recording channel and needs less audio quantity to extract a reliable measure compared to other F0 parameters, as arithmetic mean and the standard deviation which are the most used parameters in the forensic area.

The discriminant gain improvement obtained combining the baseline of the F0 and other long-term fundamental frequency measures was addressed using the statistics of the Multivariate Kernel-Density (MVKD). The experiments were done using a brasilian portuguese male recording corpus containing 60 seconds of voiced speech each sample. We show that our proposed approach achieves an Equal Error Rate (EER) of 13 % outperforming recent researches.

SUMÁRIO

LISTA DE TABELAS

LISTA DE FIGURAS

LISTA DE SÍMBOLOS, NOMENCLATURA E ABREVIAÇÕES

CFL: Comparação Forense de Locutor.

F0: Frequência Fundamental.

MVKD: Multivariate Kernel-Density, em português Estatística de Densidade do Núcleo de Multivariáveis.

EER: Equal Error Rate, em português Taxa de Erro Igual.

FSC: Forensic Speaker Comparison, em português Comparação Forense de Locutor.

LR: Likelihood Ratio, em português Razão de verossimilhança.

LTF0: Medida de Longo Termo da Frequência Fundamental.

F_b: Valor de base da Frequência Fundamental.

FM: Modulação em Frequência.

ENFSI: European Network of Forensic Science Institutes.

DET: Detection Error Tradeoff.

C_{llr}: Log-likelihood-ratio cost.

FAR: False Acceptance Rate, em português Taxa de Falsos Positivos.

FRR: False Rejection Rate, em português Taxa de Falsos Negativos.

Baseline: Valor de base da Frequência Fundamental.

ss: Same Speaker, em português Mesmo Falante.

ds: Different Speaker, em português Diferentes Falantes.

CFPB: Corpus Forense do Português Brasileiro.

UBM: Universal Background Model, em português Modelo Universal.

GMM: Gaussian Mixture Model, em português Modelo de Misturas Gaussianas.

1 INTRODUÇÃO

Boa parte dos exames realizados na área forense visam a determinação da fonte de um vestígio relacionado a um crime. Neste tipo de exame, as características extraídas do vestígio são comparadas com o padrão coletado do suspeito ou do objeto que supostamente o produziu. São exemplos típicos de determinação de fonte, os exames forenses de DNA, de comparação microbalística destinado a determinar se um projetil partiu de uma arma específica, de cotejo grafoscópico que visa identificar o autor de determinado manuscrito, de comparação de marcas deixadas por uma ferramenta ou solado, de comparação de faces, de exame datiloscópico utilizado para identificar um indivíduo a partir da impressão digital obtida no local de crime e, entre outros, o exame de comparação forense de locutor (VALENTE, 2012).

O advento dos gravadores analógicos e posteriormente digitais de baixo custo e, mais recentemente, a disseminação das mídias digitais com grande capacidade de armazenamento tornou corriqueira a disponibilidade de gravações de áudio relacionados ao cometimento de crimes. Além disso, o uso de interceptações telefônicas, de comunicação de dados e de gravações ambientais são, atualmente, uma importante ferramenta de investigação, sendo utilizadas amplamente pelas instituições de segurança pública no Brasil.

Quando há a necessidade de determinar a autoria de determinadas falas presentes em um vestígio relacionado a um crime, no caso, uma gravação de áudio, é realizado o exame de Comparação Forense de Locutor (CFL). Por meio deste exame, são cotejadas as falas questionadas, cuja autoria se deseja determinar, com amostras de fala obtidas de um ou mais suspeitos de tê-las produzido.

Atualmente, os vestígios neste tipo de exame forense são provenientes, em sua grande maioria, de gravações de áudio originárias de interceptações de telefonia fixa ou móvel, de gravações de áudio ambiental realizadas por meio de escutas instaladas pelas equipes de investigação policial, devidamente autorizadas pela justiça e, em menor frequência, de áudios gravados por um dos interlocutores da comunicação telefônica ou que estava presente no ambiente em que a conversa se desenrolou e, ultimamente, gravações provenientes de aplicativos como o *WhatsApp*.

O exame de CFL enfrenta várias limitações práticas inerentes ao ambiente forense, tais como, baixa qualidade acústica dos áudios pela utilização de codificadores com compressão com perdas, como os codificadores psicoacústicos, pequena duração dos áudios questionados, vestígios, que contém as falas cuja autoria se deseja determinar, limitações espectrais relacionadas ao canal utilizado na gravação questionada, baixa relação sinal/ruído, presença de reverberação e superposição de sons interferentes, como, sobreposição de falas de outros indivíduos presentes no mesmo ambiente, ruído de tráfego ou máquinas, sons de rádio ou programas de televisão ao fundo, entre outros. Soma-se a essas limitações, a falta de controle do conteúdo das falas cuja autoria é questionada, limitando sobremaneira a realização de cotejos entre segmentos compatíveis de fala presentes no áudio questionado e no padrão de voz do suspeito.

Todos os consultados que utilizam acústica nos exames de CFL, segundo pesquisa realizada pela Universidade de York, (GOLD; FRENCH, 2011), responderam realizar rotineiramente medidas relacionadas à frequência fundamental (F0), que é o número de ciclos completos de abertura e fechamento das pregas vocais, destes, 94 % usam medidas de média aritmética, 72 % utilizam desvio padrão, 41 % realizam medidas de mediana, 34 % analisam a moda, 25 % utilizam o valor de base de F0 e 6 % analisam o *range* dos valores de F0. Portanto, F0 é um dos mais difundidos parâmetros acústicos utilizados em exames de CFL.

F0 apresenta uma grande variação intra-falante, sendo afetado, entre outros, pelo estilo da fala, pelo esforço vocal e pelo estado emocional. Tais características diminuem o poder discriminativo de F0. De acordo com (KINOSHITA, 2005), a média aritmética de F0 tem uma grande variância, o que implica em indecisão devido às razões de verossimilhança (LR - *Likelihood Ratio*) próximas de um.

Ainda como resultado da pesquisa realizada pela Universidade de York, destaca-se que apenas 25 % dos consultados citaram utilizar a medida de longo termo da frequência fundamental (LTF0) nomeada valor de base de F0, embora pesquisas recentes indiquem ser ela uma medida estatística mais estável comparada a outros parâmetros LTF0 como média aritmética e desvio padrão, que são as mais citadas.

Entre estas pesquisas recentes, (LINDH; ERIKSSON, 2007) conclui ser o valor de base de F0 menos afetado pelo estilo da fala, pelo conteúdo, pelo esforço vocal e pelo canal utilizado na gravação comparado a outros parâmetros estatísticos, como a média aritmética e a mediana. Na mesma linha, os resultados obtidos por (ARANTES;

ERIKSSON, 2014), para o português brasileiro, indicam que a quantidade necessária de fala vozeada para que o valor de base de F0 estabilize, considerando uma queda acentuada na sua variância, fica em torno de 5 segundos, que é aproximadamente a metade da quantidade necessária nas medidas de média aritmética e mediana, também avaliadas na mesma pesquisa.

A estabilidade dos valores da medida do valor de base de F0 comparada às medidas de média e mediana, que são medidas estatísticas mais utilizadas, para amostras de áudio de menor extensão é relevante no ambiente forense, considerando o fato de serem comuns amostras de voz de reduzida duração. Dadas estas características interessantes para o exame de CFL, o parâmetro valor de base de F0 foi o escolhido para aprofundar os estudos neste trabalho.

1.1 Exame de Comparação Forense de Locutor

A fala depende de processos cognitivos, da anatomia dos órgãos fonatórios, das habilidades motoras, entre outras. Duas pessoas da mesma língua não se expressam da mesma forma e as variações linguísticas podem ser descritas por meio de análises perceptual e acústica.

Em pesquisa realizada pela Universidade de York, (GOLD; FRENCH, 2011), constatou-se que a metodologia mais adotada mundialmente nos exames de CFL é conhecida como metodologia tradicional ou método combinado, abrangendo análises perceptuais de níveis segmentais e supra-segmentais da fala em conjunto com análises acústicas. Esta metodologia é a adotada no Brasil, nos Institutos de Criminalística estaduais e na Polícia Federal, que também é responsável pela realização de cursos de formação de peritos em exames de CFL.

A análise perceptual elenca eventos individualizantes do falante, entre outros, a realização articulatória peculiar de determinado fonema, a avaliação da qualidade da voz, a prosódia[1], o *pitch*[2], a taxa de elocução, traços linguísticos característicos de determinada região geográfica (dialeto), características comuns ao estrato social que o indivíduo pertence (socioleto) e peculiaridades de cunho individual (idioletos).

A análise acústica tem a finalidade de medir e, em alguns casos, permitir visualizar

[1] Entoação + Ritmo

[2] Percepção auditiva da frequência fundamental

parâmetros fonético-acústicos como a frequência fundamental, os formantes dos segmentos vocálicos, a composição espectral de determinado som produzido, a medida da taxa de articulação, medida da duração do *Voice Onset Time*[3] (VOT), entre outros. Vários parâmetros técnicos identificados perceptualmente, podem ser confirmados pela análise acústica.

Entre as análises acústicas, F0 tem um papel de destaque no exame de CFL e conforme (KINOSHITA; ISHIHARA; ROSE, 2009), F0 é um dos parâmetros mais robustos às limitações impostas pelo ambiente forense e em (KINOSHITA, 2005), se mostrou uma abordagem de baixa complexidade e robusta em situações de áudios de baixa qualidade. Adicionalmente, medidas de longo termo de F0 não requerem comparações envolvendo mesmas palavras e fonemas.

Na Figura 1.1 é apresentado um compêndio das análises comumente realizadas em CFL. Conforme pode ser observado, existem análises puramente perceptuais, como a sociolinguística, análises acústicas, como a análise de F0 e análises que podem ser realizadas por meio da acústica e perceptualmente, como a taxa de articulação.

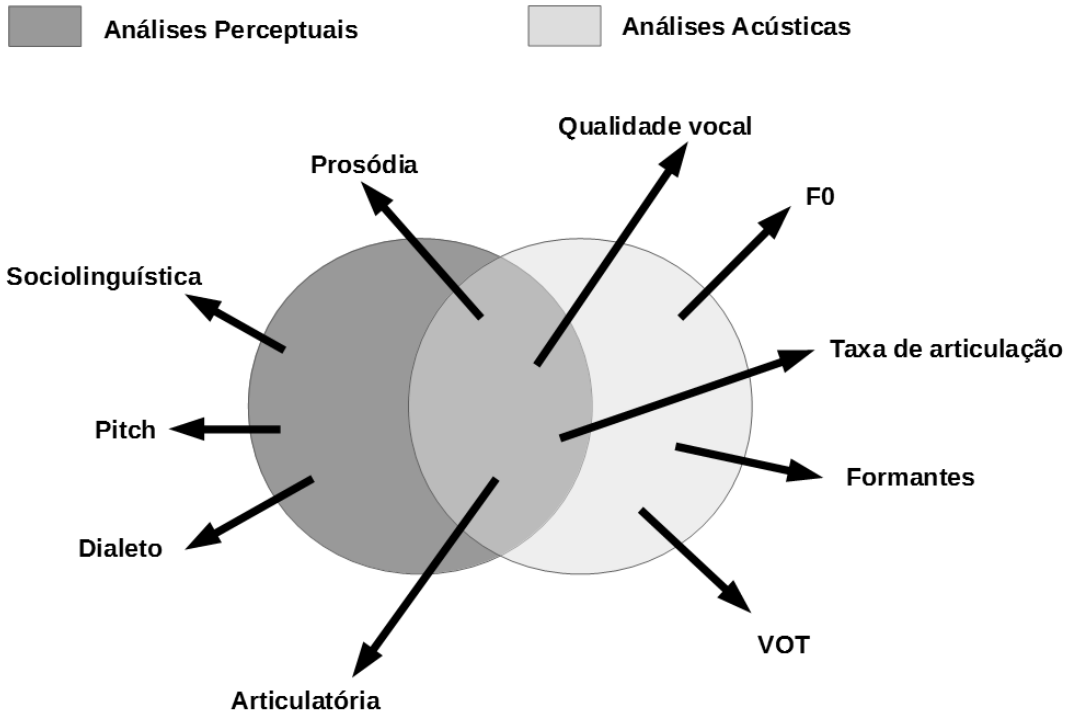


Figura 1.1: Análises realizadas no exame de CFL. Em cinza escuro estão destacados os exames perceptuais, em cinza claro, os exames de natureza acústica. Na intersecção estão as análises realizadas tanto perceptualmente quanto por análise acústica.

[3]Duração de tempo entre a liberação de uma consoante plosiva e o início de vibração das cordas vocais.

Uma vez que o exame de CFL se baseia, em parte, em análises perceptivas, o resultado depende da experiência do perito. Por esse motivo, os resultados comumente são apresentados pela escolha de um dos níveis de uma escala qualitativa, como a proposta pelo *European Network of Forensic Science Institutes* (ENFSI) (ENFSI, 2015), similar à proposta por (CHAMPOD; EVETT, 1999), presente na Tabela 2.1.

Os exames de comparação forense de locutor seguem as etapas do diagrama da Figura 1.2.

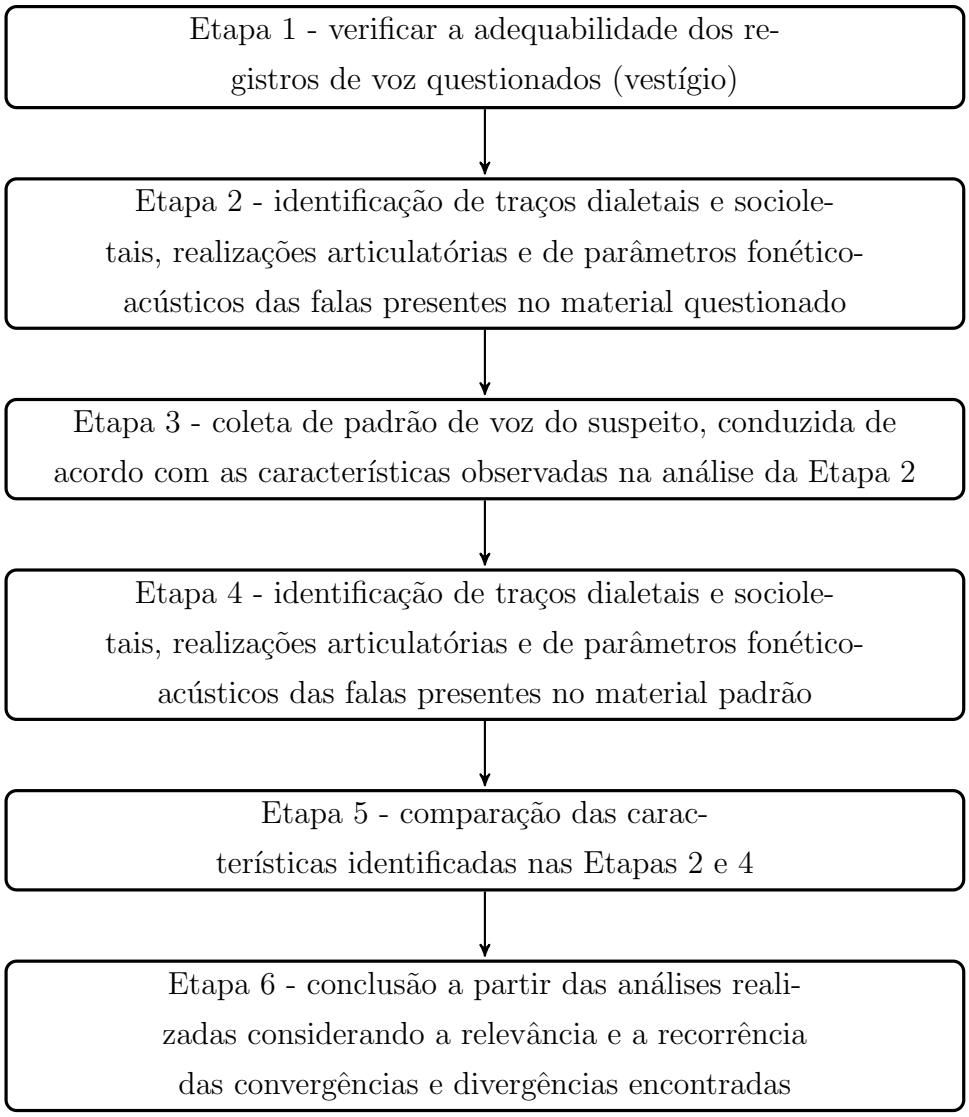


Figura 1.2: Diagrama em blocos com as etapas do exame CFL.

Na Etapa 1, o áudio questionado é analisado para verificar se os registros de voz nele presentes são adequados ao exame. Nesta etapa é verificada a quantidade de falas presente no áudio, não sendo possível, entretanto, estabelecer uma quantidade mínima

de material necessário ao exame, pois é dependente também da especificidade da voz em análise e da diversidade linguística das falas. Além disso, é analisada a presença de ruídos ou falas sobrepostas que impeçam a realização de medidas acústicas.

Na Etapa 2, caso o áudio questionado tenha sido aprovado na Etapa 1, é realizada a análise para identificação de traços dialetais e socioletais, realizações articulatórias e de parâmetros fonético-acústicos das falas presentes, especificando, a partir destas análises, um roteiro para a realização de coleta de padrão de voz do suspeito. Este roteiro será utilizado na coleta para possibilitar, nas etapas seguintes, comparações entre elementos compatíveis.

Na Etapa 3, a coleta de padrão de voz do suspeito de ter produzido as falas questionadas é realizada seguindo o roteiro produzido na etapa anterior. A coleta envolve gravação de falas semi-espontâneas (entrevista) e leitura de frases e tem uma duração variável, dependendo da dificuldade em obter as produções necessárias para a realização das comparações previstas no roteiro.

A Etapa 4 engloba as mesmas análises realizadas na Etapa 2, mas para o padrão de voz do suspeito.

Na Etapa 5, são realizadas as comparações entre os resultados das análises conduzidas nas Etapas 2 e 4. Ao final desta etapa serão relacionadas as convergências e divergências encontradas entre as vozes e falas questionada e padrão.

Na Etapa 6, são determinadas a relevância e a recorrência das convergências e divergências encontradas na Etapa 5. Ao final, o exame indicará, como conclusão, um dos níveis da escala qualitativa adotada.

Anualmente são realizados aproximadamente 60 exames deste tipo na Polícia Federal. Exames realizados por meio desta metodologia consomem, em média, 30 dias de trabalho exclusivo por perito por voz analisada, incluindo o tempo de trabalho necessário para processar cada uma das etapas descritas anteriormente, entre elas, o tempo de deslocamento dos peritos para a realização da coleta de padrão sonoro que, não raramente, ocorrem fora da localidade de trabalho normal. O fato do exame de CFL ser demorado faz com que, muitas vezes, seja solicitado pela defesa com a finalidade de postergar processos judiciais.

1.2 Estado da arte

Por décadas pesquisadores analisam o poder discriminatório de F0 que, inicialmente, foi considerado um parâmetro muito promissor no uso em exames de CFL (KINOSHITA, 2005).

(KINOSHITA; ISHIHARA; ROSE, 2009) atribuem este otimismo aos resultados promissores obtidos nas pesquisas iniciais em reconhecimento de locutor realizados na década de 1970. Contudo, pesquisas posteriores demonstraram que F0 tem uma grande variação intrafalante, diminuindo o seu poder discriminativo.

(BRAUN, 1995), relaciona vários fatores que afetam o valor de F0, como o estilo da fala, o estado emocional, o nível de ruído no ambiente onde as falas foram produzidas e se a pessoa fala ao telefone ou não. Além desses fatores, afetam o valor de F0, a condição de saúde, o esforço vocal, a embriagues, entre outros.

Em 1994, (TRAUNMÜLLER, 1994) propõe a teoria da modulação para os sinais de voz em que uma portadora é modulada por componentes linguísticos e paralinguísticos[4]. A frequência da portadora é nomeada valor de base de F0, F_{b}, correspondendo à frequência pessoal de vibração das cordas vocais. Assim, F_{b} é a frequência das cordas vocais em posição relaxada.

Nos trabalhos (TRAUNMÜLLER; ERIKSSON, 1994a) e (TRAUNMÜLLER; ERIKSSON, 1995) os autores realizam experimentos e constatam inicialmente que o valor da frequência F_{b} corresponde a aproximadamente 1,5 desvios padrão abaixo da média aritmética da F0, índice revisado posteriormente para 1,43 desvios padrão.

Em (LINDH; ERIKSSON, 2007), os autores comparam a estabilidade dos valores de F_{b} em relação às medidas de média μ e de mediana Q_2 de F0, comumente utilizadas por peritos da área na realização de exames de CFL. Como resultado, os autores concluem que o parâmetro F_{b}, em algumas condições, é menos afetado pelo estilo da fala, pelo seu conteúdo, pelo esforço vocal e pelo canal da gravação.

Conforme já citado, em 2011, uma pesquisa realizada pela Universidade de York, (GOLD; FRENCH, 2011), a fim de fazer um levantamento a respeito das metodo-

[4]Elementos associados aos componentes linguísticos utilizados na comunicação, como o timbre, a entonação, a mudança de voz, entre outros.

logias utilizadas por peritos da área forense de comparação de locutor, constatou que todos os respondentes que realizam rotineiramente medidas acústicas, fazem medidas relacionadas à frequência fundamental (F0) e destes, 94 % usam medidas de média aritmética, 72 % utilizam desvio padrão, 41 % realizam medidas de mediana, 34 % analisam a moda, 25 % utilizam o valor de base de F0 e 6 % analisam o *range* dos valores de F0.

Na pesquisa (ARANTES; ERIKSSON, 2014), os autores avaliam a quantidade de produções vozeadas requerida para que as medidas de longo termo de F0 estabilizem pela diminuição significativa da variância, sendo considerada a partir deste ponto uma medida estável. Por exemplo, como mostrado pelos autores, para o corpus de Português Brasileiro, F_{b} estabiliza em aproximadamente 5 segundos, enquanto a média aritmética e a mediana estabilizam em aproximadamente 11 segundos.

Em (AITKEN; LUCY, 2004), *Multivariate Kernel-Density* (MVKD) foi proposto para calcular LRs na presença de variáveis estatisticamente dependentes. Neste estudo, os autores aplicaram a técnica para analisar 3 medidas de composição de fragmentos de vidro encontrados na roupa de um suspeito e em janelas quebradas no local de crime a fim de determinar o grau de similaridade entre elas.

MVKD apresenta boa performance em casos onde poucas medidas por variável estão disponíveis. Por este motivo, o uso de MVKD para analisar variáveis relacionadas à linguística e fonética tem se tornado um procedimento padrão na área de comparação forense de locutores, conforme destaca (MORRISON, 2011).

Em exames de CFL, MVKD determina a similaridade entre amostras de fala provenientes do áudio questionado e do áudio proveniente do suspeito. O resultado da aplicação da função MVKD é uma razão de verossimilhança, LR[5], correspondendo à razão entre a probabilidade da hipótese de que as falas questionadas foram produzidas pelo suspeito e a hipótese de terem sido produzidas por outro falante da população de referência adotada.

Vários estudos recentes aplicam MVKD na análise de parâmetros acústicos, entre eles destacam-se o estudo do poder discriminante de formantes analisado por (ROSE, 2007) usando onze vogais do Inglês Australiano, enquanto (ROSE; WINTER, 2010) compara a performance de MVKD e o modelo de misturas gaussianas - modelo universal,

[5]LR e inferência bayesiana serão melhor exploradas na Seção 2.3

do inglês *Gaussian Mixture Model - Universal Background Model* (GMM-UBM) combinando os primeiros três formantes de cinco monotongos do Inglês Australiano em falantes femininas, (MORRISON, 2009) avalia trajetória de formantes de cinco ditongos em Inglês Australiano, e (HUGHES, 2014) analisa formantes e ditongos em Inglês Britânico e Neozelandês.

(GOLD, 2014) avalia o poder discriminativo da taxa de articulação na produção da fala e combina as LTF0 média aritmética e desvio padrão. Em (KINOSHITA, 2005) os autores realizam testes envolvendo LTF0s média aritmética e desvio padrão utilizando bases de dados reais e artificialmente criadas a fim de determinar o poder discriminativo combinado e verificar o *range* de LRs obtidas na aplicação da fórmula MVKD, concluindo que os resultados obtidos indicam um pequeno poder discriminativo de parâmetros LTF0.

Em (KINOSHITA; ISHIHARA; ROSE, 2009) é analisado o poder discriminativo dos parâmetros LTF0 média aritmética, desvio padrão, assimetria, curtose, moda e densidade modal individualmente e combinados.

Considerando os bons resultados obtidos pelos trabalhos citados, MVKD foi adotado nesta pesquisa.

1.3 Justificativa e motivação

A principal motivação para realizar a pesquisa no Mestrado Profissional foi aprofundar o conhecimento em uma área que trouxesse resultado prático para os exames de Comparação Forense de Locutor (CFL) realizados na Polícia Federal.

As peculiaridades do exame CFL, principalmente o grande consumo de horas para a sua consecução, justificam a necessidade de desenvolver métodos que agilizem a sua realização.

Considerando as características robustas do parâmetro F0 com relação ao ruído, a independência do conteúdo das falas presentes no áudio, a grande utilização deste parâmetro em exames de CFL por parte dos especialistas e os poucos trabalhos investigativos voltados para avaliar o poder discriminativo do parâmetro de valor de base de F0 envolvendo o português brasileiro e que faltam estudos e populações de referência que permitam avaliar o poder discriminativo das medidas utilizadas.

E, considerando que existe uma nova tendência mundial em apresentar os resultados dos exames forenses de determinação de fonte em termos de razões de verossimilhança (ENFSI, 2015), e que o uso desta metodologia permite que os resultados sejam reprodutíveis e não somente baseados na experiência do perito.

Dado o exposto, decidiu-se investigar, nesta pesquisa, o poder discriminativo da medida estatística do valor de base de F0, a quantidade de áudio vozeado em português brasileiro necessário para uma medida estável e a melhor combinação de medidas de longo termo da frequência fundamental para se obter o maior poder discriminativo.

1.4 Metodologia

A fim de atingir os objetivos traçados na seção anterior, será realizada uma revisão bibliográfica para identificar o estado da arte atual das pesquisas relacionadas ao parâmetro acústico de valor de base de F0 aplicadas aos exames de CFL e, por meio de experimentos, comparar os resultados de pesquisas recentes à abordagem aqui utilizada.

A função *Multivariate Kernel Density* (AITKEN; LUCY, 2004) será utilizada para calcular as LRs resultantes de comparações entre amostras de falas de mesmos e de diferentes falantes e contabilizar os resultados em termos de taxa de erro igual, do inglês *Equal Error Rate* (EER), para cada um dos parâmetros, suas combinações e diferentes tamanhos de amostras de áudio, bem como avaliar a magnitude dos valores das LRs obtidas.

Ao final, será possível responder à pergunta se o valor de base de F0 isoladamente e em combinação com outras medidas de longo termo de F0 apresenta melhor poder discriminativo que as medidas estatísticas de F0 tradicionalmente utilizadas no ambiente forense.

1.5 Publicações

Com o objetivo de divulgar os resultados obtidos nesta pesquisa, foi publicado artigo na RBC - Revista Brasileira de Criminalística (SILVA, R. R.; DA COSTA, J. P. C. L.; MIRANDA, R. K.; DEL GALDO, G., 2016a) com o título “Aplicação do valor de base da frequência fundamental via estatística MVKD em comparação forense de locutor”. A Revista Brasileira de Criminalística é um veículo de comunicação importante no meio forense e dedicado a distribuir informações relevantes aos profissionais da área.

O artigo “Applying base value of fundamental frequency via the multivariate kernel-density in forensic speaker comparison” foi apresentado no evento ICSPCS’2016 - *10th International Conference on Signal Processing and Communication Systems* (SILVA, R. R.; DA COSTA, J. P. C. L.; MIRANDA, R. K.; DEL GALDO, G., 2016b) com o objetivo de dar visibilidade aos resultados obtidos nesta pesquisa.

1.6 Organização da dissertação

Esta dissertação está dividida em cinco capítulos incluindo esta introdução. No Capítulo 2, são abordados conceitos sobre a teoria da produção da fala, teoria da modulação aplicada ao sinal de voz, inferência bayesiana e a abordagem utilizada na avaliação do poder discriminativo de F0. São apresentados, ainda, os operadores matemáticos utilizados no decorrer da pesquisa. No Capítulo 3, a abordagem proposta é descrita. No Capítulo 4, a validação experimental da abordagem proposta é apresentada. O Capítulo 5 relaciona as conclusões e propostas para trabalhos futuros.

2 BASE TEÓRICA

Este capítulo aborda a teoria relacionada à produção da fala, descreve a teoria da modulação aplicada aos sinais de voz, faz uma introdução à inferência bayesiana, apresenta a abordagem utilizada na avaliação do poder discriminativo de F0 e relaciona os operadores matemáticos utilizados no decorrer da pesquisa.

2.1 Teoria da produção da fala

O som produzido ao falar é resultado de um complexo processo de produção iniciando na conceituação, passando pelos comandos neuromotores e finalmente aos articuladores da fala que modulam o fluxo de ar, produzindo o som (BARBOSA; MADUREIRA, 2015).

A produção dos sons da fala se deve aos subsistemas respiratório, laríngeo e supralaríngeo, descritos a seguir. O subsistema respiratório é responsável por gerar o fluxo de ar passante pelo aparelho fonador, seja ele ingressivo ou egressivo. Este fluxo de ar é a fonte de energia que permite a produção dos sons e é gerado pela expansão e constrição da caixa torácica, músculos, pulmões e diafragma. No caso da língua portuguesa, os sons são produzidos pela expiração, ou seja, pelo ar egressivo, mantendo uma pressão subglotal que, associada a outros fatores, influencia o valor da frequência fundamental e a intensidade (LAVER, 1994).

O subsistema laríngeo é composto por um conjunto de músculos, cartilagens e ligamentos e exerce funções relacionadas à respiração e à fonação. As pregas vocais, também chamadas de cordas vocais, que são dois pedaços de tecidos em forma de lábios inseridos nessa estrutura, são controladas pelos músculos tireoaritenoideos e pelos ligamentos vocais. A Figura 2.1[6] contém um modelo composto dos subsistemas laríngeo e supralaríngeo, estando, na região indicada pela seta, a localização das pregas vocais.

[6]Reprodução a partir de (BARBOSA; MADUREIRA, 2015) p. 40, com autorização dos autores.

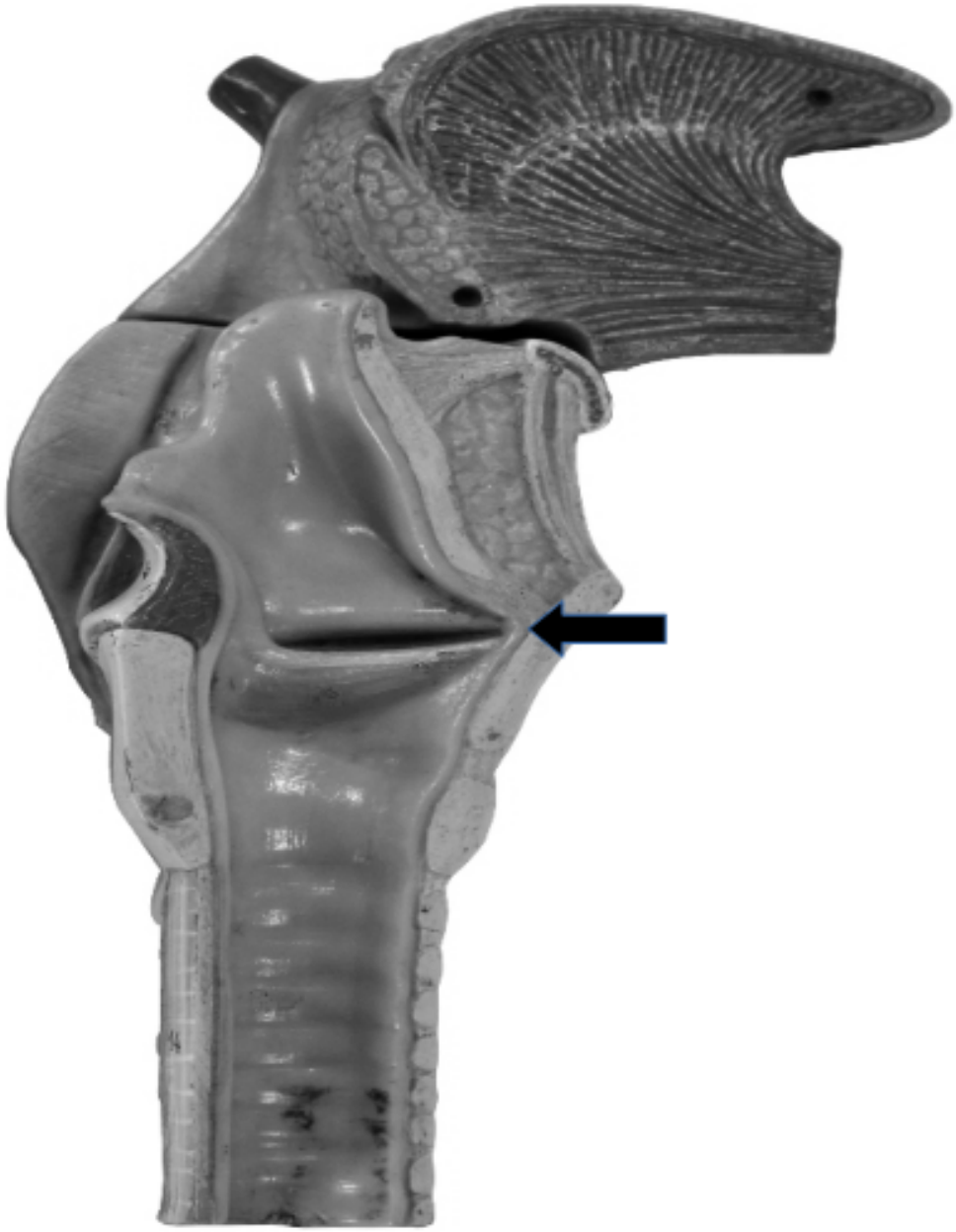


Figura 2.1: As pregas vocais são posicionadas na região indicada pela seta no modelo dos subsistemas laríngeo e supralaríngeo.

A vibração das cordas vocais é resultado de um processo envolvendo a interação de forças aerodinâmicas e de tensão elástica controlada pelos músculos cricoaritenoideos. Mantendo uma diferença de pressão adequada entre as regiões sub e supraglotal de forma a vencer a resistência das pregas vocais, estas se abrem e, com a passagem do fluxo de ar, há uma queda da pressão, fechando novamente as pregas e o processo se repete pelo fechamento e abertura alternados.

A frequência de vibração das pregas vocais, F0, pode variar de 50 Hz a 700 Hz dependendo de vários fatores, entre eles da sua massa, da entoação, do sexo, da idade do falante, do estado emocional no momento da produção da fala, de condições de saúde, entre outros.

O subsistema supralaríngeo é formado pelas estruturas situadas logo acima das pregas vocais até a abertura da boca e das narinas. Este subsistema modula o fluxo de ar proveniente da laringe por meio dos articuladores, produzindo, pela interação com os outros subsistemas, os sons da fala. Este subsistema, também chamado de trato vocal, exerce a função de um filtro, atenuando determinadas faixas de frequência enquanto reforça outras por meio de ressonância nas suas cavidades.

2.1.1 Teoria fonte-filtro

Em 1960, Gunnar Fant (FANT, 1960) propôs a teoria de separação da produção da fala em duas partes, a fonte e o filtro, essa teoria passou a ser conhecida como teoria linear fonte-filtro de produção da fala.

A fonte de energia utilizada na produção dos sons da fala pode ter várias origens: a vibração periódica das pregas vocais utilizada para gerar os sons vozeados como por exemplo a produção do fone [a], a passagem do ar em seção reduzida do trato oral como ocorre na produção do fone [s], ou com a combinação das duas, como na produção de consoantes vozeadas, a exemplo do fone [z].

O filtro corresponde às modificações impostas aos sons gerados pela fonte na passagem pelo trato vocal que é composto pelo trato oral e pelo trato nasal. O trato vocal comporta-se como uma composição de cavidades ressonantes, atenuando ou amplificando faixas específicas de frequência que se alteram conforme a posição dos articuladores, grau de abertura ou de fechamento da boca e da posição do véu palatino.

Conforme essa teoria, o espectro do som irradiado após a passagem pelo trato vocal é resultado da multiplicação em frequência do conteúdo espectral da fonte pela resposta em frequência do trato vocal. Dessa forma, os vários sons da fala são resultado da energia produzida pela fonte após modificada pela passagem pelo trato vocal e as características do trato vocal podem ser inferidas a partir da análise do sinal de saída do sistema. Os dois sistemas, fonte e filtro, podem ser analisados independentemente.

Na Figura 2.2 é reproduzido o espectro correspondente à produção da vogal [ε] sobreposto pelo envelope da resposta em frequência do trato vocal, que amplifica determinadas faixas de frequência e atenua outras, evidenciando os quatro primeiros formantes, correspondendo aos quatro picos no envelope de resposta em frequência.

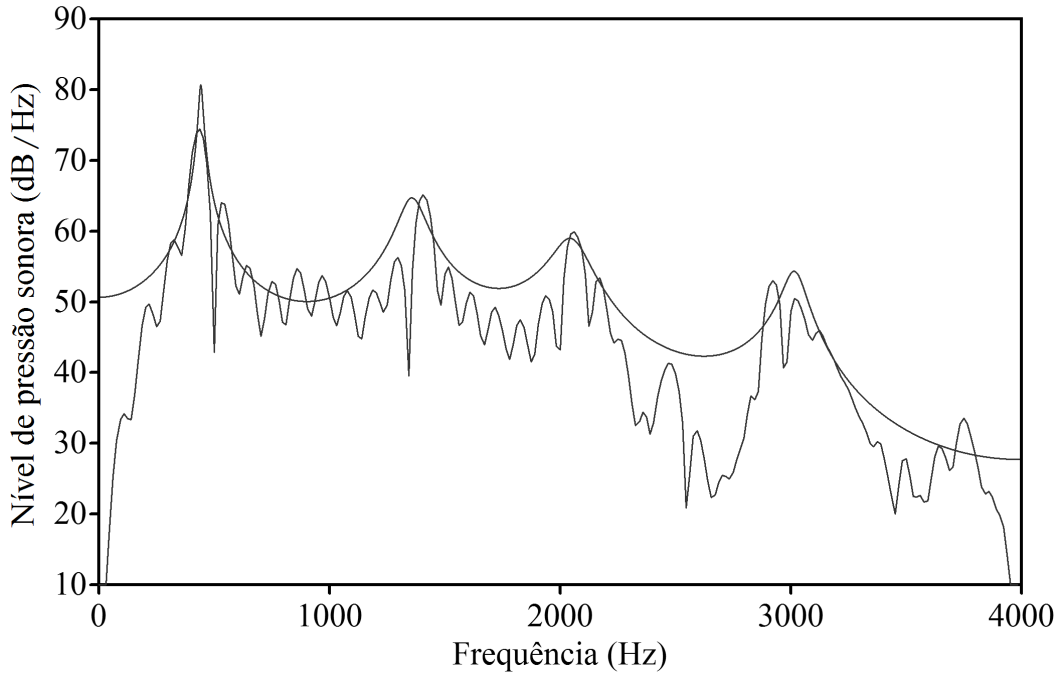


Figura 2.2: Espectro correspondente à produção da vogal [ε] sobreposto pela resposta em frequência do trato vocal, envoltória mais externa.

O foco deste trabalho de mestrado está inteiramente baseado na fonte, no caso, a frequência de vibração das cordas vocais (F0) presente nas produções vozeadas.

2.2 Teoria da modulação

Em 1994, Traunmüller (TRAUNMÜLLER, 1994), propôs a teoria da modulação para os sinais de voz. Por meio da comunicação oral, além do conteúdo linguístico, estão presentes informações paralinguísticas referentes ao estado emocional, ao esforço vocal, à atitude, entre outras.

As características acústicas da produção da fala de um indivíduo são afetadas, entre outras, pela massa das pregas vocais e pelas dimensões do trato vocal e não podem ser removidas do sinal de voz.

Na teoria da modulação proposta pelo autor, ao realizar os gestos da fala[7], o indivíduo perturba o seu trato vocal a fim de produzir a informação linguística desejada, sendo que o trato vocal nos instantes em que não está sendo perturbado reflete a informação pessoal do falante.

Nesse sentido, a teoria da modulação considera o sinal de voz como resultado da modulação de uma portadora por meio dos gestos da fala, transportando os componentes linguísticos e paralinguísticos da comunicação, enquanto, do outro lado, o ouvinte recupera o conteúdo fonético por meio da demodulação deste sinal.

Conforme (LINDH; ERIKSSON, 2007), de acordo com a teoria da modulação, deve existir um valor de base da frequência fundamental, F_{b}, considerado como a frequência de uma portadora que representa a articulação individual do falante. O valor de F_{b} é um melhor preditor do valor da F0 típica do indivíduo, correspondendo à frequência pessoal das cordas vocais, ou o valor neutro de F0, característico do falante. F_{b} é a frequência de vibração das cordas vocais em uma posição relaxada, ou seja, a frequência em que as cordas vocais sempre e naturalmente retornam após uma excursão prosódica[8].

Traçando um paralelo com a modulação em frequência (FM), no que se refere à variação do valor da F0 ao longo da produção da fala, F_{b} seria a frequência da portadora e o sinal modulante corresponderia a uma composição dos componentes linguísticos e paralinguísticos.

Assim, o desvio em frequência $\Delta\mathrm{f(t)}$ é proporcional a uma constante k aplicada ao valor instantâneo $l(t)$ do sinal modulante:

$$\Delta\mathrm{f}(t) = \mathrm{k} \cdot l(t), \tag{2.1}$$

Portanto, a F0 resultante, dado o valor de pico do sinal modulante $\mathrm{V_p}$, será:

$$\mathrm{G}(t) = \mathrm{V_p} \cdot \mathrm{sen}[2\pi(F_{\mathrm{b}} + \Delta\mathrm{f}(t))t], \tag{2.2}$$

onde $\mathrm{G}(t)$ corresponde ao sinal contendo a variação dos valores de F0 ao longo da produção da fala.

Por meio de experimentos, (TRAUNMÜLLER; ERIKSSON, 1994a) e (TRAUNMÜLLER; ERIKSSON, 1995) concluíram que o valor de base da frequência fundamental F_{b} está

[7] Alteração das dimensões físicas do trato vocal e da configuração dos articuladores para a produção de determinado fone (som).

[8] Prosódia = ritmo + entonação.

localizado próximo do limite inferior do *range* de produções de F0 do falante, determinado por

$$F_{\mathrm{b}} = \mu - 1.43\sigma, \tag{2.3}$$

onde μ e σ são os parâmetros LTF0 média aritmética e o desvio padrão de F0, respectivamente.

Dado que F_{b} está próximo do limite inferior de produção de F0, a variação dos valores de F0, nesse sentido, é pequena, limitada pela ocorrência do fenômeno de laringalização[9]. Em contrapartida, no sentido superior, a F0 pode atingir valores bastante elevados, resultando em curvas de distribuição cumulativa de F0 com assimetria positiva.

Posteriormente, em (LINDH; ERIKSSON, 2007), os autores propõem uma abordagem alternativa para calcular F_{b} minimizando os valores extremos, em inglês *outliers*, que afetam significativamente o valor de σ na fórmula original de cálculo (2.3). Supondo uma distribuição normal para F0, (2.3) implica que F_{b} pode ser obtido pelo percentil equivalente a 7,64 % da distribuição de F0.

As metodologias original e alternativa para o cálculo de F_{b} resultam, aproximadamente, no mesmo valor em gravações de áudio de boa qualidade, enquanto, para áudios de baixa qualidade, a metodologia alternativa é mais robusta, e, consequentemente, é a melhor escolha em análises forenses e foi adotada aqui.

Considerando a Teoria da Modulação aqui exposta, é esperado que F_{b} seja mais estável e mais representativo do indivíduo, sendo, portanto, mais discriminante que outros parâmetros LTF0.

2.3 Conceitos básicos de inferência bayesiana

O uso de LR tem se tornado comum entre os peritos forenses em exames de determinação de fonte, sendo utilizado frequentemente em exames de DNA forense, hoje, uma referência neste tipo de abordagem. Devido à indisponibilidade de populações de referência adequadas e de definição de quais características são mais discriminantes em cada tipo de exame, a aplicabilidade desta abordagem ainda é bastante restrita nos demais exames de determinação de fonte, como os exames de comparação facial, de confronto microbalístico, de grafoscopia, de reconhecimento de padrões de solados, entre outros.

[9]Lenta vibração das pregas vocais com as cartilagens aritenoides permanecendo abertas.

O uso de LR permite avaliar o peso, conhecido como valor de uma evidência, por meio da razão entre a probabilidade da evidência dada a hipótese da acusação e a probabilidade da evidência dada a hipótese da defesa.

No contexto de exames de CFL, LR permite avaliar o peso de uma evidência pela razão entre as probabilidades de observá-la sob a hipótese de que as falas questionadas foram produzidas pelo suspeito (H_{a}) e a hipótese de terem sido produzidas por outro falante da população de referência adotada (H_{d}).

Dado que P é a função de probabilidade, I é a informação de contexto, do inglês *background information* e E denota a evidência, a LR é dada por

$$\mathrm{LR} = \frac{P(E|H_{\mathrm{a}}, I)}{P(E|H_{\mathrm{d}}, I)}. \tag{2.4}$$

O Teorema de Bayes pode ser utilizado a fim de combinar os pesos das várias evidências obtidas na análise dos vestígios coletados em um determinado caso. A aplicação do Teorema de Bayes nas ciências forenses interpreta o juízo de condenação ou absolvição considerando dois fatores, o primeiro de que o grau de convicção sobre a culpabilidade/inocência em um caso é alterado conforme são agregadas novas evidências ou ocorrem alterações nos seus resultados e segundo, de que convicções individuais com relação a um mesmo evento variam devido às diferenças de pesos de cada uma das peças incluídas no caso (AITKEN; TARONI, 2004).

A aplicação teórica do Teorema de Bayes é chamada de inferência bayesiana e, na área criminal, é dada como segue:

$$\frac{P(H_{\mathrm{a}}|I)}{P(H_{\mathrm{d}}|I)} \cdot \frac{P(E|H_{\mathrm{a}}, I)}{P(E|H_{\mathrm{d}}, I)} = \frac{P(H_{\mathrm{a}}|E, I)}{P(H_{\mathrm{d}}|E, I)}, \tag{2.5}$$

onde P é a função de probabilidade, H_{a} é a hipótese da acusação, H_{d} é a hipótese da defesa, I é a informação de contexto e E é a evidência.

O primeiro termo da equação (2.5), $\frac{P(H_{\mathrm{a}}|I)}{P(H_d|I)}$, corresponde à probabilidade *a priori*, antes da inclusão do peso da evidência E, o segundo termo, $\frac{P(E|H_{\mathrm{a}},I)}{P(E|H_{\mathrm{d}},I)}$, é o peso da evidência E

e o último termo, $\frac{P(H_a|E,I)}{P(H_d|E,I)}$, corresponde à razão de probabilidade *a priori* combinada com o peso da evidência E, e é chamado de probabilidade *a posteriori*.

O uso do Teorema de Bayes possibilita a combinação de resultados provenientes de diferentes evidências relacionadas a uma mesma investigação que, caso sejam estatisticamente independentes, podem ser simplesmente multiplicadas, obtendo-se uma única LR que agrega a contribuição de todas as evidências.

Como exemplo, um caso que tenha uma LR de 2,5 *a priori* e se deseja acrescentar os pesos referentes aos exames de CFL, de comparação de padrão de solado e de um exame forense de DNA cujas LRs sejam iguais a 2, 0,01 e 200, respectivamente, resultará em uma probabilidade *a posteriori* igual a LR = 2,5 x 2 x 0,05 x 200 = 50, ou seja, é 50 vezes mais provável a hipótese da acusação em contraposição à hipótese da defesa, considerando o peso de todas as evidências e que elas são estatisticamente independentes entre si.

Entretanto, quando analisados parâmetros que apresentam algum grau de dependência entre si, a LR conjunta não pode ser obtida pela simples multiplicação das LRs individuais. Na presente pesquisa, diferentes parâmetros LTF0, estatisticamente dependentes, como por exemplo a média aritmética, a mediana e o valor de base de F0 são combinadas a fim de avaliar o poder discriminativo resultante. Por serem dependentes, a simples multiplicação entre os valores de LR de cada um dos parâmetros LTF0 não pode ser feita. A alternativa foi calcular a LR utilizando a função MVKD proposta por Aitken e Lucy (AITKEN; LUCY, 2004), que, ao calcular o valor da LR, leva em consideração o fato das variáveis envolvidas serem correlacionadas entre si.

A apresentação do resultado das análises por meio de um valor de LR pode não ser corretamente interpretado pelas partes envolvidas no julgamento de um caso. A fim de facilitar a interpretação dos resultados, Champod e Evett (CHAMPOD; EVETT, 1999), propõem uma escala verbal para descrever o peso de uma evidência de acordo com o valor da LR, transcrita na Tabela 2.1.

A vantagem do uso de uma escala verbal está baseada no fato de facilitar a interpretação dos resultados dos exames às partes e aos integrantes do poder judiciário. No exemplo apresentado anteriormente, a evidência de DNA cuja LR foi avaliada em 200, equivalente a um $\log_{10}(\text{LR}) = 2,3$, representa um suporte moderadamente forte à hipótese de mesma fonte, enquanto a LR = 0,05, equivalente a um $\log_{10}(\text{LR}) = -1,3$, da análise

Tabela 2.1: Escala verbal proposta por Champod e Evett

LR	$\log_{10}(\text{LR})$	Expressão verbal
$\text{LR} \geq 10^4$	$\log_{10}(\text{LR}) \geq 4$	Suporte muito forte à hipótese de mesma fonte
$10^3 \leq \text{LR} < 10^4$	$3 \leq \log_{10}(\text{LR}) < 4$	Suporte forte à hipótese de mesma fonte
$10^2 \leq \text{LR} < 10^3$	$2 \leq \log_{10}(\text{LR}) < 3$	Suporte moderadamente forte à hipótese de mesma fonte
$10 \leq \text{LR} < 10^2$	$1 \leq \log_{10}(\text{LR}) < 2$	Suporte moderado à hipótese de mesma fonte
$1 < \text{LR} < 10$	$0 \leq \log_{10}(\text{LR}) < 1$	Suporte limitado à hipótese de mesma fonte
$\text{LR} = 1$	$\delta = 0$	Sem suporte às hipóteses
$10^{-1} < \text{LR} < 1$	$-1 < \log_{10}(\text{LR}) < 0$	Suporte limitado à hipótese de diferentes fontes
$10^{-2} < \text{LR} \leq 10^{-1}$	$-2 < \log(\text{LR}) \leq -1$	Suporte moderado à hipótese de diferentes fontes
$10^{-3} < \text{LR} \leq 10^{-2}$	$-3 < \log_{10}(\text{LR}) \leq -2$	Suporte moderadamente forte à hipótese de diferentes fontes
$10^{-4} < \text{LR} \leq 10^{-3}$	$-4 < \log_{10}(\text{LR}) \leq -3$	Suporte forte à hipótese de diferentes fontes
$\text{LR} \leq 10^{-4}$	$\log_{10}(\text{LR}) \leq -4$	Suporte muito forte à hipótese de diferentes fontes

do solado corresponde a um suporte moderado à hipótese de diferentes fontes. Combinando as várias evidências do exemplo e a probabilidade *a priori*, tem-se $\text{LR} = 50$, $\log_{10}(\text{LR}) = 1,7$, correspondendo a um suporte moderado à hipótese da acusação.

A *European Network of Forensic Science Institutes* (ENFSI) que congrega 66 laboratórios forenses de 36 países, além de acordos com outras organizações de relevância na área, tem por finalidade compartilhar conhecimento, trocar experiências e realizar acordos mútuos no campo das ciências forenses. A ENFSI publicou em novembro de 2015 o guia *ENFSI guideline for evaluative reporting in forensic science* (ENFSI, 2015) que recomenda que todos os exames forenses conduzidos pelos laboratórios associados

apresentem como resultado LRs e a respectiva correspondência na escala verbal.

Essa tendência mundial na direção de adotar LR nos exames forenses, reforça a necessidade de realizar pesquisas, como a do presente trabalho, desenvolver novas metodologias e criar populações de referência adequadas aos exames.

2.4 Avaliação do poder discriminativo por meio do uso de EER, curvas DET e C_{llr}

A acuracidade de um sistema biométrico é avaliada pela medida da quantidade de erros de identificação que ocorrem ao realizar uma grande quantidade de comparações envolvendo amostras provenientes do mesmo e de diferentes indivíduos. Em sistemas de identificação biométrica do tipo 1 x n, os dados biométricos de um indivíduo são comparados àqueles presentes em um banco de dados a fim de identificá-lo. Nestes sistemas, é estabelecido um limiar de decisão que, se atingido, corresponderá à identificação positiva, ou, alternativamente, a comparação que atingir o maior *score* entre todos os integrantes do banco de dados corresponderá à identificação.

São parâmetros importantes de sistemas biométricos, a taxa de falsos positivos, em inglês *False Acceptance Rate* (FAR) e a taxa de falsos negativos, em inglês *False Rejection Rate* (FRR). A FAR é definida como a taxa entre a quantidade de comparações que são erroneamente estimadas como provenientes do mesmo indivíduo e o total de confrontos envolvendo diferentes indivíduos. Enquanto a FRR é a taxa entre a quantidade de confrontos erroneamente indicados pelo sistema como provenientes de diferentes indivíduos e o total de confrontos envolvendo mesmo indivíduo.

Os valores das taxas FAR e FRR de um sistema biométrico são determinados por um limiar de decisão δ escolhido, que, nesta pesquisa, corresponde ao valor da LR utilizada para decidir se em uma determinada comparação o vestígio e o padrão são provenientes do mesmo ou de diferentes falantes. Ao utilizar um limiar de decisão elevado, a taxa de FAR será reduzida em detrimento de um aumento na taxa de falsos negativos FRR. Em contrapartida, usar limiar de decisão pequeno resultará em altas taxas de falsos positivos FAR e pequenas taxas de falsos negativos FRR.

A determinação do ponto ideal de funcionamento de um sistema, pela escolha do δ, depende da finalidade do mesmo. Um sistema de segurança em que um falso positivo represente um alto custo, como em sistemas bancários, o limiar de decisão deverá

ser ajustado a fim de reduzir tanto quanto possível FAR, ainda que ocorram várias negativas de acesso a clientes verdadeiros, ou seja, δ deve ter um valor alto. Já sistemas de acesso onde se deseja minimizar o número de falsos negativos, como por exemplo o acesso a uma biblioteca, o limiar pode ser reduzido.

Em aplicações forenses, deve-se minimizar a taxa de falsos positivos ainda que ocorra um maior número de falsos negativos, pois há um alto custo envolvido na condenação de um inocente comparado ao custo de inocentar indevidamente um criminoso.

Variando o limiar δ obtém-se um ponto de ajuste em que FAR = FRR, chamada de Taxa de Erro Igual, do inglês *Equal Error Rate* (EER). A EER é usada para avaliar o desempenho de sistemas biométricos. Quanto menor o valor da EER, melhor o poder discriminante do sistema.

O valor da EER pode ser obtido pela intersecção de duas curvas, a primeira correspondendo à plotagem dos valores de FAR em função do limiar δ e a segunda curva, de valores de FRR, também em função do δ.

A determinação do valor da EER é utilizada neste trabalho para avaliar o poder discriminativo dos parâmetros acústicos LTF0. Para tal, vários confrontos envolvendo amostras de voz de mesmo falante e de diferentes falantes são realizados. Dados os resultados obtidos dos confrontos e um limiar de decisão δ, as taxas FAR e FRR são computadas e, variando o valor do δ, obtém-se a EER, ou seja, o ponto de ajuste em que FAR = FRR. Quanto menor o valor da EER, maior o poder discriminante do parâmetro sendo avaliado.

Na área forense, onde normalmente não existe um conjunto fechado de candidatos, ou seja, não existe um banco de dados contendo todas as possíveis fontes da amostra suspeita, é interessante que as curvas FAR e FRR estejam o máximo afastadas na horizontal uma da outra, dando maior grau de suporte aos resultados.

Exemplificativamente, na Figura 2.3, a EER é o ponto de interceptação entre as duas curvas, correspondendo a uma EER de aproximadamente 5 %. As curvas FAR e FRR na Figura 2.3 estão afastadas na horizontal, possuindo suporte, conforme Tabela 2.1, pelo menos moderadamente forte ($\log_{10}(\delta) \geq 2$) em mais de 80 % das comparações efetivamente envolvendo mesma fonte, calculado por 1 - FRR do limiar escolhido, e igual suporte ($\log_{10}(\delta) \leq -2$) para mais de 80 % das comparações envolvendo diferentes

fontes, sendo δ o valor do LR escolhido como limiar de decisão.

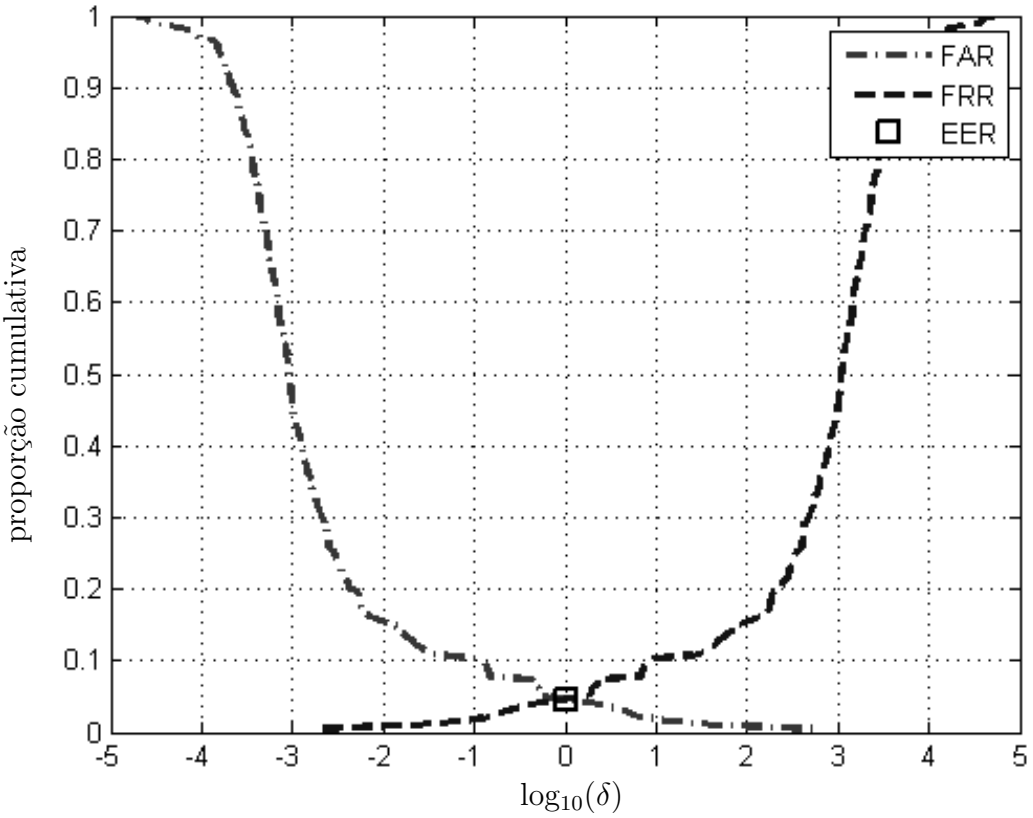


Figura 2.3: FAR e FRR versus $\log_{10}(\delta)$ apresentando uma EER de 5 % e grau de suporte pelo menos moderadamente forte em mais de 80 % das comparações, sendo δ a LR utilizada como limiar de decisão.

Na Figura 2.4 é apresentado um contra-exemplo onde, além da EER ser elevada, da ordem de 42 %, a maioria das LRs resultam em suporte próximo da indecisão, conforme Tabela 2.1.

As comparações devem apresentar grandes valores positivos de $\log_{10}$(LR) quando envolve mesma fonte e grandes valores negativos de $\log_{10}$(LR) em comparações envolvendo fontes diferentes, dando maior suporte às evidências. Afim de determinar qual combinação de parâmetros LTF0 possuem o melhor desempenho, foi utilizada a abordagem proposta por (BRÜMMER; PREEZ, 2006), adotada, entre outros, por (MORRISON, 2009), calculando o *log-likelihood-ratio cost* (C_{llr}) de cada uma das combinações de LTF0.

Finalmente, outra forma de comparar o poder discriminativo de diferentes sistemas biométricos é comumente realizada pela avaliação de curvas DET, do inglês *Detection Error Tradeoff*, correspondentes aos valores de FAR e FRR em função do valor do

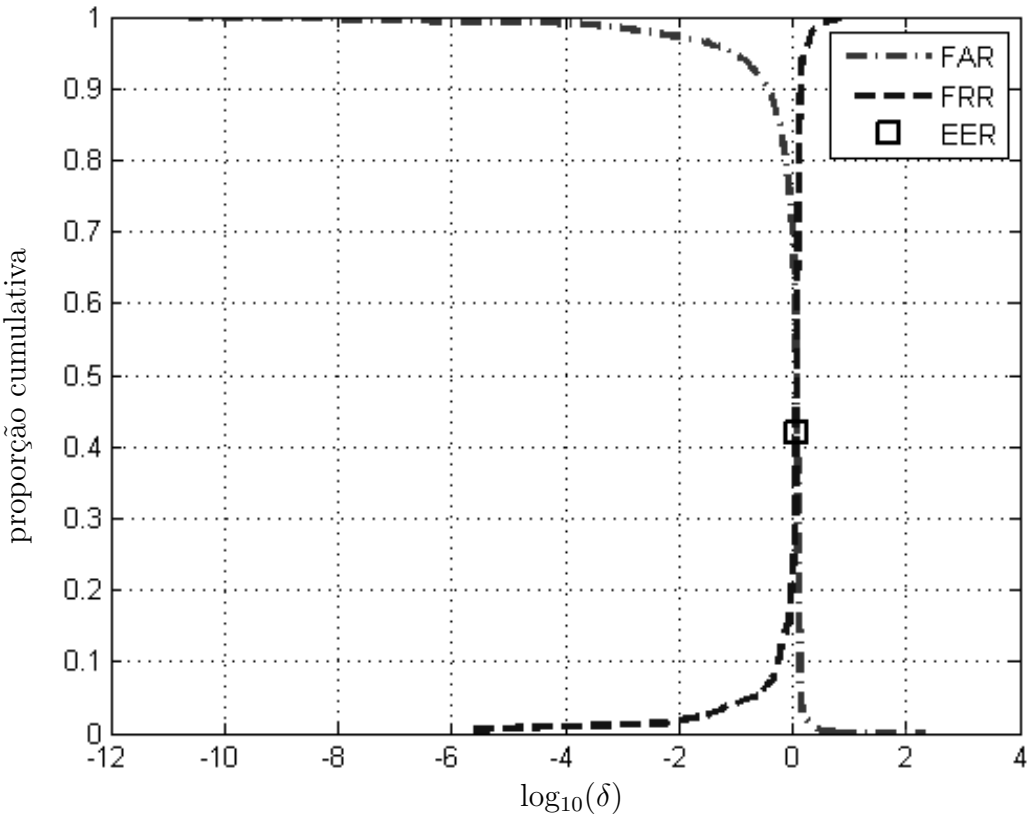


Figura 2.4: FAR e FRR versus $\log_{10}(\delta)$ apresentando uma EER de aproximadamente 42 % e a maioria das comparações próximas da indecisão. Dados extraídos da Seção 4.2, onde a curtose ($\hat{w}$) é obtida em seções de 15 segundos conforme penúltima linha da Tabela 4.1.

limiar δ, neste trabalho uma LR, de cada um dos sistemas, permitindo uma visualização conjunta do poder discriminativo de cada sistema em função do valor do limiar.

Na Figura 2.5 é apresentada a comparação do poder discriminante dos parâmetros $\hat{F}_{\mathrm{b}}$ e $\hat{\mu}$ por meio de curvas DET. Neste gráfico, quanto menor a área sob a curva, melhor o poder discriminante. Além disso, este tipo de gráfico permite comparar a capacidade discriminativa de cada sistema de acordo com o limiar δ escolhido. No exemplo da referida figura, $\hat{F}_{\mathrm{b}}$ é claramente melhor discriminativo que $\hat{\mu}$.

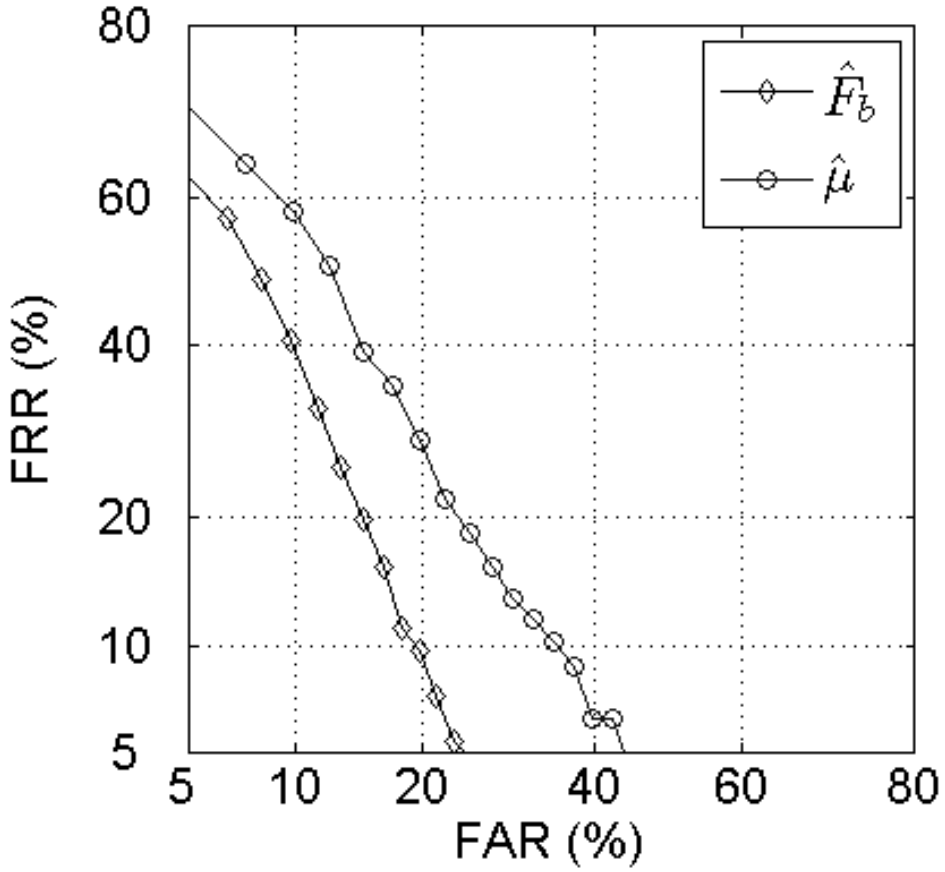


Figura 2.5: Comparativo do poder discriminante dos parâmetros LTF0 $\hat{F}_\mathrm{b}$ e $\hat{\mu}$ por meio de curvas DET. Dados extraídos da Seção 4.2, com dados obtidos utilizando seções de 15 segundos. Quanto menor a área sob a curva, melhor o poder discriminante.

2.5 Operadores matemáticos

2.5.1 Expressões matemáticas dos parâmetros LTF0 utilizados na pesquisa

Os parâmetros LTF0 escolhidos para serem investigados neste trabalho são a média aritmética μ, mediana Q_2, desvio padrão σ, curtose w, assimetria, do inglês *skewness*, η, valor de base de F0 F_b, moda ψ e densidade modal γ.

Dada uma gravação de áudio contendo somente produções vozeadas, após removidos os trechos de silêncio e de produções surdas, e dado o valor da menor F0 que o falante consegue produzir e manter $\mathrm{F0_{min}}$, o número N correspondente à quantidade de medidas de F0 presentes no áudio de t segundos é $N = \lfloor \frac{\mathrm{F0_{min}} \cdot t}{0{,}75} \rfloor$.

Considerando que f_n é a $n-$ésima medida de F0 para $n = 1, ..., N$ e $\mathbf{f} = [\mathrm{f}_1, \mathrm{f}_2, ..., \mathrm{f}_N]$ é o vetor contendo N medidas de F0, então os parâmetros LFT0 são obtidos por:

A média aritmética $\hat{\mu}$ para N medidas de F0 é computada por

$$\hat{\mu} = \frac{\sum_{n=1}^{N} \mathrm{f}_n}{N}. \tag{2.6}$$

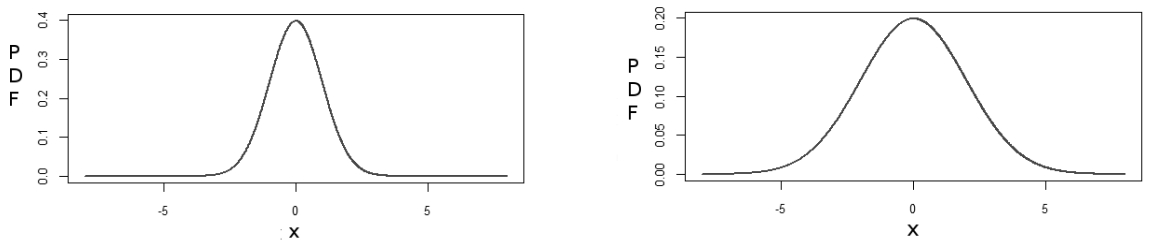
Dado o vetor ordenado de medidas de F0 $\mathbf{f}$ de tamanho N, a mediana ($\hat{Q}_2$) é o segundo quartil de $\mathbf{f}$, ou seja, é o valor da amostra que separa o vetor em duas metades contendo o mesmo número de elementos no caso do vetor possuir número ímpar de elementos e é a média dos dois elementos centrais caso possua número par de elementos,

$$\hat{Q}_2 = \begin{cases} \mathrm{f}_{\frac{N+1}{2}} & \text{se } N \text{ for ímpar} \\ \frac{\mathrm{f}_{\frac{N}{2}} + \mathrm{f}_{\frac{N+2}{2}}}{2} & \text{se } N \text{ for par} \end{cases} \tag{2.7}$$

O desvio padrão ($\hat{\sigma}$) representa o grau de dispersão em relação à média dos dados sob análise e é computado por

$$\hat{\sigma} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (\mathrm{f}_n - \hat{\mu})^2}. \tag{2.8}$$

A Figura 2.6 contém a simulação de distribuições normais de probabilidade da variável x contendo desvios padrão $\hat{\sigma} = 1$ (a) e $\hat{\sigma} = 2$ (b), sendo possível observar a maior dispersão dos dados no gráfico de maior desvio padrão.



(a) Exemplo de distribuição normal com $\hat{\sigma} = 1$ (b) Exemplo de distribuição normal com $\hat{\sigma} = 2$

Figura 2.6: Distribuições de probabilidade normal com $\hat{\sigma} = 1$ (a) e $\hat{\sigma} = 2$ (b)

Dado o vetor ordenado de medidas de F0 $\mathbf{f}$ de comprimento N, o valor de base de F0, $\hat{F}_\mathrm{b}$, corresponde ao valor do elemento em que 7,64 % das amostras do vetor possuem

valores inferiores a ele e é obtido pelo quantil 0,0764 de **f**, ou seja, o valor de F_{b} corresponde a aproximadamente ao sétimo percentil da distribuição dos valores de F0,

$$\hat{F}_{\mathrm{b}} = \mathrm{f}_{\lfloor 0{,}0764N \rceil}. \tag{2.9}$$

Na Figura 2.7 é mostrada a distribuição de densidade de probabilidade de valores de F0 da amostra BA23036 do Corpus Forense do Português Brasileiro utilizado na pesquisa e a indicação do valor de $\hat{F}_{\mathrm{b}}$, no caso em torno de 132 Hz, correspondendo a aproximadamente ao sétimo percentil da distribuição.

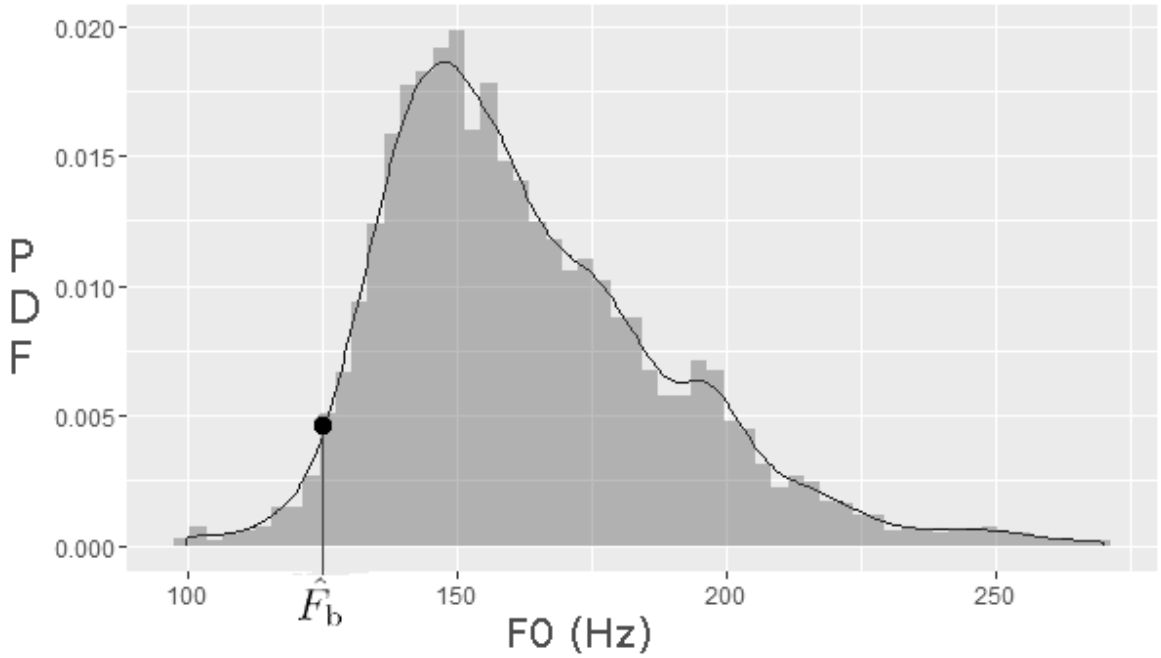


Figura 2.7: Distribuição de densidade de probabilidade dos valores de F0 com a indicação do valor do parâmetro $\hat{F}_{\mathrm{b}}$

Curtose ($\hat{w}$) indica o quão plana é a curva de distribuição de probabilidade e é obtida por

$$\hat{w} = \frac{m_4}{\hat{\sigma}^4} - 3, \tag{2.10}$$

onde $\hat{\sigma}$ é o desvio padrão e m_4 é o quarto momento obtido por

$$m_4 = \frac{\sum_{n=1}^{N} n(\mathrm{f}_n - \hat{\mu})^4}{N}. \tag{2.11}$$

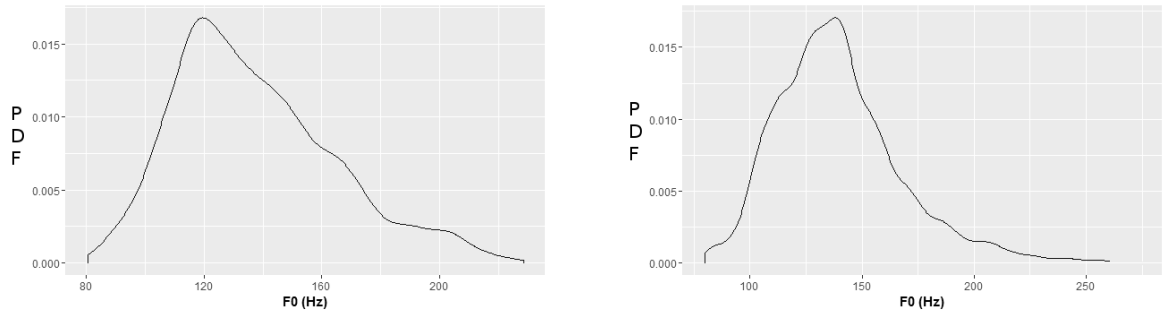
Assimetria ($\hat{\eta}$) indica a assimetria da distribuição de probabilidade, definida por

$$\hat{\eta} = \frac{\mathrm{m}_3}{\hat{\sigma}^3}, \tag{2.12}$$

onde $\hat{\sigma}$ é o desvio padrão e m_3 é o terceiro momento obtido por:

$$m_3 = \frac{\sum_{n=1}^{N} n(\mathrm{f}_n - \hat{\mu})^3}{N}. \tag{2.13}$$

Exemplificativamente, na Figura 2.8 são plotadas duas distribuições do Corpus Forense do Português Brasileiro contendo diferentes valores de curtose e assimetria. Pode ser notado na Figura 2.8a que a distribuição correspondente ao padrão AM1728 possui uma curtose $\hat{w}$ maior que a da distribuição do padrão PR4644 da Figura 2.8b, resultando em um gráfico com um pico menos agudo ou seja, é uma distribuição mais "plana". Em contrapartida, a distribuição do padrão AM1728 da Figura 2.8a possui uma assimetria maior que a do padrão PR4644 da Figura 2.8b.



(a) Padrão AM1728 com $\hat{\eta} = 0,90$ e $\hat{w} = 4,37$ (b) Padrão PR4644 com $\hat{\eta} = 0,68$ e $\hat{w} = 3,10$

Figura 2.8: Distribuição de densidade de probabilidade dos padrões AM1728 (a) e PR4644 (b) do Corpus Forense do Português Brasileiro

Para extrair a densidade modal e a moda das medidas de F0 do vetor $\mathbf{f}$, a densidade de probabilidade das medidas de F0 para cada gravação é estimada usando *binned kernel density*, de acordo com

$$\hat{f}(x, h, \mathbf{f}) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x - \mathrm{f}_i}{h}\right). \tag{2.14}$$

onde x é o valor de F0 no eixo da frequência fundamental, h é a largura de banda e K é o *kernel* utilizado, no caso desta pesquisa, *kernel* gaussiano. Um exemplo de distribuição de densidade de probabilidade das medidas de F0 utilizando *binned kernel density* é apresentada na Figura 2.9.

O valor apropriado da largura de banda h é selecionado utilizando o *plug-in* dpik da biblioteca Kernsmooth (WAND, 2015) do software R®[10].

[10]software R versão 3.1.3 (*Smooth Sidewalk*) baixada em 08/05/2015 de https://cran.r-project.org/src/base/R-3/.

Densidade modal ($\hat{\gamma}$) é a densidade de probabilidade da moda, obtido por

$$\hat{\gamma} \;=\; \max_{\mathrm{x}}(\hat{f}). \tag{2.15}$$

A moda ($\hat{\psi}$) é o valor x da densidade modal obtida por (2.15) que corresponde ao valor mais frequente de F0 do vetor **f**.

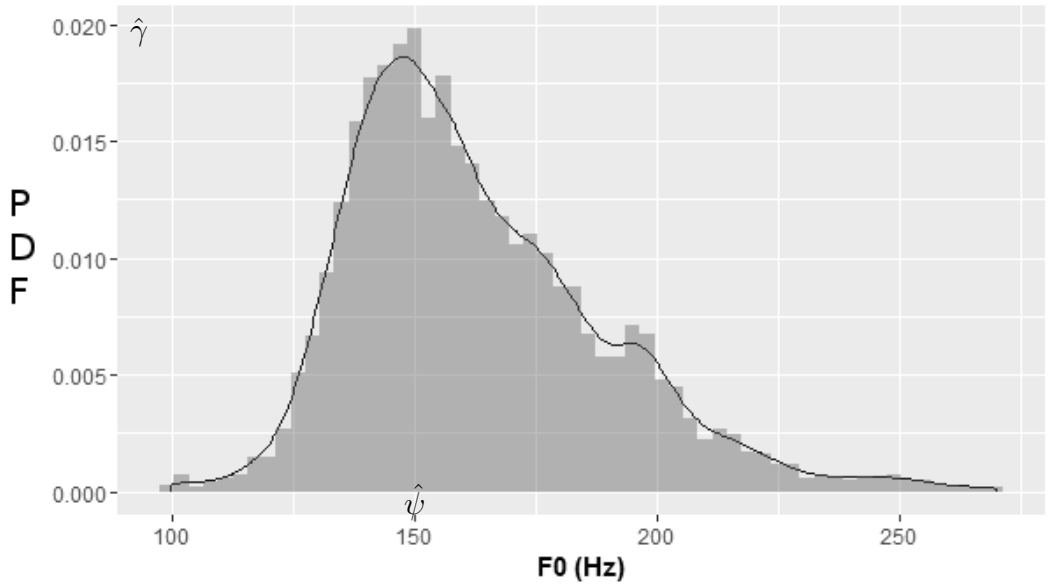


Figura 2.9: Exemplo de distribuição de densidade de probabilidade utilizando *binned kernel density* com a indicação da moda $\hat{\psi}$ e da densidade modal $\hat{\gamma}$

Pode ser observado na Figura 2.9, que contém uma distribuição de probabilidade das medidas de F0, os valores da moda $\hat{\psi} = 150$ Hz e da densidade modal $\hat{\gamma} = 0,02$.

2.5.2 MVKD - *Multivariate Kernel-Density*

A fim de calcular a LR na presença de dependência estatística entre as variáveis utilizadas, é aplicada a função MVKD proposta por (AITKEN; LUCY, 2004), com algumas simplificações, dado que nesta pesquisa todas as variáveis utilizadas, no caso, os parâmetros LTF0, possuem a mesma quantidade de medidas, garantindo comparações homogêneas.

Para tal, define-se o conjunto $\mathcal{K}$ composto por $k_1,\ k_2,\ \ldots,\ k_K$ parâmetros LFT0 a serem combinados, r_1 e r_2 correspondem à evidência $\{r_1 \;=\; 1,\ \ldots,\ R\}$ e ao padrão do suspeito $\{r_2 \;=\; 1,\ \ldots,\ R\}$, a população de referência $\mathcal{R}$ inclui todos os padrões de suspeitos,

exceto os padrões dos falantes sendo comparados, i.e. $\mathcal{R} = \{1, \ldots, R-2\} \mid \{r_1, r_2\} \not\subset \mathcal{R}$ com R igual ao número de padrões de suspeito do corpus e z_s é o número de medidas por tamanho de seção t_s do áudio de extensão T e é computado por $\lfloor \frac{T}{t_s} \rfloor$, para $s = 1, \ldots, S$, indicando o tamanho da seção.

As matrizes do vestígio, do padrão do suspeito e da população de referência são obtidas, respectivamente, como

$$\mathbf{D}_{\mathcal{K},r_1,s} = \begin{pmatrix} \mathrm{d}_{k_1,r_1,s,1} & \mathrm{d}_{k_2,r_1,s,1} & \cdots & \mathrm{d}_{k_K,r_1,s,1} \\ \mathrm{d}_{k_1,r_1,s,2} & \mathrm{d}_{k_2,r_1,s,2} & \cdots & \mathrm{d}_{k_K,r_1,s,2} \\ \vdots & \vdots & \cdots & \vdots \\ \mathrm{d}_{k_1,r_1,s,z_s} & \mathrm{d}_{k_2,r_1,s,z_s} & \cdots & \mathrm{d}_{k_K,r_1,s,z_s} \end{pmatrix}, \tag{2.16}$$

com $\mathbf{D}_{\mathcal{K},r_1,s} \in \mathcal{R}^{z_s \times K}$, cujos elementos são $\mathrm{d}_{k_d,r_1,s,z_i}$, $z_i = 1, \ldots, z_s$ e o índice k_d indica o d-ésimo parâmetro LTF0 do conjunto $\mathcal{K}$.

$$\mathbf{E}_{\mathcal{K},r_2,s} = \begin{pmatrix} \mathrm{e}_{k_1,r_2,s,1} & \mathrm{e}_{k_2,r_2,s,1} & \cdots & \mathrm{e}_{k_K,r_2,s,1} \\ \mathrm{e}_{k_1,r_2,s,2} & \mathrm{e}_{k_2,r_2,s,2} & \cdots & \mathrm{e}_{k_K,r_2,s,2} \\ \vdots & \vdots & \cdots & \vdots \\ \mathrm{e}_{k_1,r_2,s,z_s} & \mathrm{e}_{k_2,r_2,s,z_s} & \cdots & \mathrm{e}_{k_K,r_2,s,z_s} \end{pmatrix}, \tag{2.17}$$

sendo $\mathbf{E}_{\mathcal{K},r_2,s} \in \mathcal{R}^{z_s \times K}$, cujos elementos são $\mathrm{e}_{k_d,r_2,s,z_i}$.

$$\mathbf{P}_{\mathcal{K},r,s} = \begin{pmatrix} \mathrm{p}_{k_1,r,s,1} & \mathrm{p}_{k_2,r,s,1} & \cdots & \mathrm{p}_{k_K,r,s,1} \\ \mathrm{p}_{k_1,r,s,2} & \mathrm{p}_{k_2,r,s,2} & \cdots & \mathrm{p}_{k_K,r,s,2} \\ \vdots & \vdots & \cdots & \vdots \\ \mathrm{p}_{k_1,r,s,z_s} & \mathrm{p}_{k_2,r,s,z_s} & \cdots & \mathrm{p}_{k_K,r,s,z_s} \end{pmatrix}, \tag{2.18}$$

sendo $\mathbf{P}_{\mathcal{K},r,s} \in \mathcal{R}^{z_s \times K}$, a matriz contendo as medidas do falante r da população de referência cujos elementos são p_{k_d,r,s,z_i}, para $r = 1, \ldots, R-2$, correspondendo ao r-ésimo falante da população de referência.

A partir da matriz 2.18, obtém-se a matriz $\mathbf{P}_{\mathcal{K},\mathcal{R},s} \in \mathcal{R}^{z_s \times K \cdot (R-2)}$, formada pela concatenação das matrizes $\mathbf{P}_{\mathcal{K},r,s}$ de todos os falantes presentes na população de referência, conforme

$$\mathbf{P}_{\mathcal{K},\mathcal{R},s} = \begin{pmatrix} \mathbf{P}_{\mathcal{K},1,s} & \mathbf{P}_{\mathcal{K},2,s} & \mathbf{P}_{\mathcal{K},3,s} & \cdots & \mathbf{P}_{\mathcal{K},R-2,s} \end{pmatrix}, \tag{2.19}$$

onde $R-2$ corresponde ao número de elementos na população de referência.

Cada linha da matriz $\mathbf{D}_{\mathcal{K},r_1,s}$ pode ser escrita como um vetor $\mathbf{d}_{\mathcal{K},r_1,s,j} \in \mathcal{R}^K$ contendo uma das amostras de medidas dos LTF0$_\mathcal{K}$ parâmetros do vestígio r_1

$$\mathbf{d}_{\mathcal{K},r_1,s,j} = [\mathrm{d}_{k_1,r_1,s,j}, \mathrm{d}_{k_2,r_1,s,j}, \ldots, \mathrm{d}_{k_K,r_1,s,j}], \tag{2.20}$$

para $j = 1, \ldots, z_s$.

O vetor $\bar{\mathbf{d}}_{\mathcal{K},r_1,s} \in \mathcal{R}^K$, contendo a média de cada um dos K LTF0 do vestígio r_1 é obtido pela média dos elementos de índice j da matriz 2.16, como segue

$$\bar{\mathbf{d}}_{\mathcal{K},r_1,s} = \left[\frac{1}{z_s} \sum_{j=1}^{z_s} \mathbf{d}_{k_1,r_1,s,j}, \frac{1}{z_s} \sum_{j=1}^{z_s} \mathbf{d}_{k_2,r_1,s,j}, \ldots, \frac{1}{z_s} \sum_{j=1}^{z_s} \mathbf{d}_{k_K,r_1,s,j} \right]. \tag{2.21}$$

Cada linha da matriz $\mathbf{E}_{\mathcal{K},r_2,s}$ pode ser escrita como um vetor $\mathbf{e}_{\mathcal{K},r_2,s,j} \in \mathcal{R}^K$ contendo uma das amostras de medidas dos LTF0$_\mathcal{K}$ parâmetros do padrão de suspeito r_2

$$\mathbf{e}_{\mathcal{K},r_2,s,j} = [\mathrm{e}_{k_1,r_2,s,j}, \mathrm{e}_{k_2,r_2,s,j}, \ldots, \mathrm{e}_{k_K,r_2,s,j}], \tag{2.22}$$

para $j = 1, \ldots, z_s$.

O vetor $\bar{\mathbf{e}}_{\mathcal{K},r_2,s} \in \mathcal{R}^K$, contendo a média de cada um dos K LTF0 do padrão de suspeito r_2 é obtido pela média dos elementos de índice j da matriz 2.17, como segue

$$\bar{\mathbf{e}}_{\mathcal{K},r_2,s} = \left[\frac{1}{z_s} \sum_{j=1}^{z_s} \mathbf{e}_{k_1,r_2,s,j}, \frac{1}{z_s} \sum_{j=1}^{z_s} \mathbf{e}_{k_2,r_2,s,j}, \ldots, \frac{1}{z_s} \sum_{j=1}^{z_s} \mathbf{e}_{k_K,r_2,s,j} \right]. \tag{2.23}$$

Cada linha da matriz $\mathbf{P}_{\mathcal{K},r,s}$ pode ser escrita como um vetor $\mathbf{p}_{\mathcal{K},r,s,j} \in \mathcal{R}^K$ contendo as medidas dos LTF0$_\mathcal{K}$ parâmetros do falante r da população de referência

$$\mathbf{p}_{\mathcal{K},r,s,j} = [\mathrm{p}_{k_1,r,s,j}, \mathrm{p}_{k_2,r,s,j}, \ldots, \mathrm{p}_{k_K,r,s,j}], \tag{2.24}$$

para $j = 1, ..., z_s$.

O vetor $\bar{\mathbf{p}}_{\mathcal{K},r,s} \in \mathcal{R}^K$, contendo a média de cada um dos K LTF0 do falante r da população de referência é obtido pela média dos elementos de índice j da matriz 2.18, como segue

$$\bar{\mathbf{p}}_{\mathcal{K},r,s} = \left[\frac{1}{z_s} \sum_{j=1}^{z_s} \mathbf{p}_{k_1,r,s,j}, \frac{1}{z_s} \sum_{j=1}^{z_s} \mathbf{p}_{k_2,r,s,j}, \ldots, \frac{1}{z_s} \sum_{j=1}^{z_s} \mathbf{p}_{k_K,r,s,j} \right], \tag{2.25}$$

para $r = 1, \ \ldots, \ R-2$, correspondendo ao r-ésimo falante da população de referência.

A média geral de todos os falantes da população de referência $\bar{\mathbf{p}}_{\mathcal{K},\mathcal{R},s} \in \mathcal{R}^K$ é obtida como segue

$$\bar{\mathbf{p}}_{\mathcal{K},\mathcal{R},s} = \frac{1}{R-2}\sum_{r=1}^{R-2}\bar{\mathbf{p}}_{\mathcal{K},r,s}. \tag{2.26}$$

A matriz de variância/covariância intrafalante da população de referência $\hat{\mathbf{U}}_s \in \mathcal{R}^{K\times K}$, é computada por

$$\hat{\mathbf{U}}_s = \frac{\sum_{r=1}^{R-2}\sum_{j=1}^{z_s}[\mathbf{p}_{\mathcal{K},r,s,j} - \bar{\mathbf{p}}_{\mathcal{K},r,s}]^{\mathrm{T}}[\mathbf{p}_{\mathcal{K},r,s,j} - \bar{\mathbf{p}}_{\mathcal{K},r,s}]}{(R-2)(z_s-1)}, \tag{2.27}$$

enquanto $\hat{\mathbf{C}}_s \in \mathcal{R}^{K\times K}$ é a matriz de variância/covariância interfalantes da população de referência

$$\hat{\mathbf{C}}_s = \frac{\sum_{r=1}^{R-2}[\bar{\mathbf{p}}_{\mathcal{K},r,s} - \bar{\mathbf{p}}_{\mathcal{K},\mathcal{R},s}]^{\mathrm{T}}[\bar{\mathbf{p}}_{\mathcal{K},r,s} - \bar{\mathbf{p}}_{\mathcal{K},\mathcal{R},s}]}{R-3} - \frac{\sum_{r=1}^{R-2}\sum_{j=1}^{z_s}[\mathbf{p}_{\mathcal{K},r,s,j} - \bar{\mathbf{p}}_{\mathcal{K},r,s}]^{\mathrm{T}}[\mathbf{p}_{\mathcal{K},r,s,j} - \bar{\mathbf{p}}_{\mathcal{K},r,s}]}{z_s(R-2)(z_s-1)}. \tag{2.28}$$

Normalizando $\hat{\mathbf{U}}_s$, definimos

$$\hat{\mathbf{G}}_s = \frac{\hat{\mathbf{U}}_s}{z_s}, \tag{2.29}$$

e $\bar{\mathbf{y}}_s$ é definido como

$$\bar{\mathbf{y}}_s = \frac{\bar{\mathbf{d}}_{\mathcal{K},r_1,s} + \bar{\mathbf{e}}_{\mathcal{K},r_2,s}}{2}. \tag{2.30}$$

Usando as definições (2.20) a (2.30) computa-se a LR de cada comparação entre o vestígio, r_1, e o padrão do suspeito, r_2, armazenada no elemento $m_{\mathcal{K},r_1,r_2,s}$ da matriz, aplicando a função MVKD aos parâmetros LTF0 selecionados para uma determinada seção s,

$$m_{\mathcal{K},r_1,r_2,s} = \frac{V_1(\mathbf{D}_{\mathcal{K},r_1,s}, \mathbf{E}_{\mathcal{K},r_2,s}, \mathbf{P}_{\mathcal{K},\mathcal{R},s})}{V_2(\mathbf{D}_{\mathcal{K},r_1,s}, \mathbf{E}_{\mathcal{K},r_2,s}, \mathbf{P}_{\mathcal{K},\mathcal{R},s})}, \tag{2.31}$$

onde

$$V_1 = (2\pi)^{-K}|\hat{\mathbf{G}}_s|^{-1}|\hat{\mathbf{C}}_s|^{-1/2}[(R-2)h^K]^{-1}\cdot|2\hat{\mathbf{G}}_s^{-1}+(h^2\hat{\mathbf{C}}_s)^{-1}|^{-1/2}\cdot \exp\left\{-\frac{1}{2}(\bar{\mathbf{d}}_{\mathcal{K},r_1,s}-\bar{\mathbf{e}}_{\mathcal{K},r_2,s})\frac{1}{2}\hat{\mathbf{G}}_s^{-1}(\bar{\mathbf{d}}_{\mathcal{K},r_1,s}-\bar{\mathbf{e}}_{\mathcal{K},r_2,s})^{\mathrm{T}}\right\}\cdot \sum_{r=1}^{R-2}\exp\left[-\frac{1}{2}\left(\bar{\mathbf{y}}_s-\bar{\mathbf{p}}_{\mathcal{K},r,s}\right)\left\{\frac{1}{2}\hat{\mathbf{G}}_s+h^2\hat{\mathbf{C}}_s\right\}^{-1}\left(\bar{\mathbf{y}}_s-\bar{\mathbf{p}}_{\mathcal{K},r,s}\right)^{\mathrm{T}}\right], \quad (2.32)$$

e

$$V_2 = (2\pi)^{-K}|\hat{\mathbf{C}}_s|^{-1}[(R-2)h^K]^{-2}\cdot \left[|\hat{\mathbf{G}}_s|^{-1/2}|\hat{\mathbf{G}}_s^{-1}+(h^2\hat{\mathbf{C}}_s)^{-1}|^{-1/2}\cdot\sum_{r=1}^{R-2}\exp\left\{-\tfrac{1}{2}(\bar{\mathbf{d}}_{\mathcal{K},r_1,s}-\bar{\mathbf{p}}_{\mathcal{K},r,s})(\hat{\mathbf{G}}_s+h^2\hat{\mathbf{C}}_s)^{-1}(\bar{\mathbf{d}}_{\mathcal{K},r_1,s}-\bar{\mathbf{p}}_{\mathcal{K},r,s})^{\mathrm{T}}\right\}\right]\cdot \left[|\hat{\mathbf{G}}_s|^{-1/2}|\hat{\mathbf{G}}_s^{-1}+(h^2\hat{\mathbf{C}}_s)^{-1}|^{-1/2}\cdot\sum_{r=1}^{R-2}\exp\left\{-\tfrac{1}{2}(\bar{\mathbf{e}}_{\mathcal{K},r_2,s}-\bar{\mathbf{p}}_{\mathcal{K},r,s})(\hat{\mathbf{G}}_s+h^2\hat{\mathbf{C}}_s)^{-1}(\bar{\mathbf{e}}_{\mathcal{K},r_2,s}-\bar{\mathbf{p}}_{\mathcal{K},r,s})^{\mathrm{T}}\right\}\right], \quad (2.33)$$

onde $|\cdot|$ é o determinante da matriz e h é o valor ótimo de amortecimento do *kernel*, computado por

$$h = \sqrt[K+4]{\frac{4}{2K+1}}\cdot\frac{1}{\sqrt[K+4]{R-2}}, \quad (2.34)$$

sendo K o número de variáveis por falante que, aqui, corresponde ao número de parâmetros LTF0 considerados. V1 corresponde ao numerador da equação de cálculo da LR (2.4) e V2 corresponde ao denominador da mesma equação.

Ao aplicar a função MVKD à comparação entre um vestígio e um padrão de suspeito utilizando os parâmetros LTF0 mediana e valor de base de F0, teremos, ilustrativamente, na Figura 2.10, a distribuição da densidade de probabilidade, PDF, dos pares de LTF0 da população de referência, a PDF do padrão do suspeito e o par de LTF0 do vestígio.

O cruzamento entre o vestígio e a PDF do padrão do suspeito corresponde ao numerador da equação da LR (2.4) que, no exemplo, corresponde a 3,1. O cruzamento entre o vestígio e a PDF da população de referência corresponde ao denominador da equação da LR (2.4) e é igual a 1,1, resultando em uma LR $= \frac{P(E|H_\mathrm{a})}{P(E|H_\mathrm{d})} = \frac{3{,}1}{1{,}1} = 2,8$.

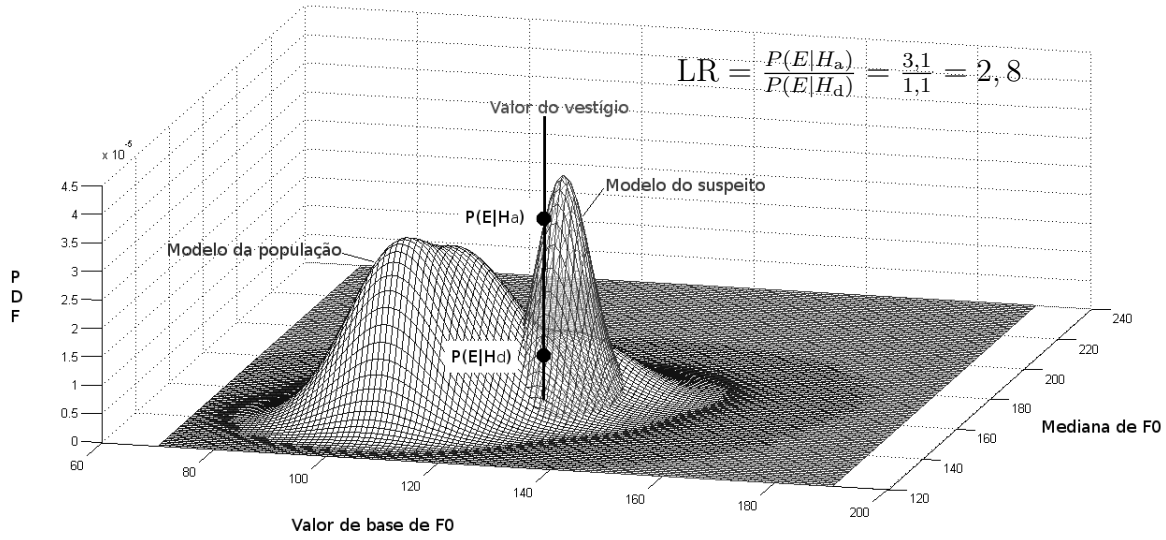


Figura 2.10: Exemplo visual da aplicação da função MVKD.

2.6 Sumário

Neste capítulo foram apresentadas a teoria envolvida na produção da fala, a teoria da modulação aplicada a sinais de voz, a inferência bayesiana aplicada ao ambiente forense, a abordagem utilizada na avaliação do poder discriminativo de sistemas biométricos e a descrição dos operadores matemáticos utilizados neste trabalho.

3 ABORDAGEM PROPOSTA

O objetivo da abordagem aqui proposta é investigar o poder discriminante de F_{b}, reduzir o valor da EER encontrando a melhor combinação de parâmetros LTF0 utilizando a função MVKD, investigar a influência do tamanho das seções das medidas de LTF0 nos resultados e analisar qual combinação apresenta melhor suporte na escala presente na Tabela 2.1. A Figura 3.1 contém os três primeiros passos do diagrama em blocos da abordagem proposta.

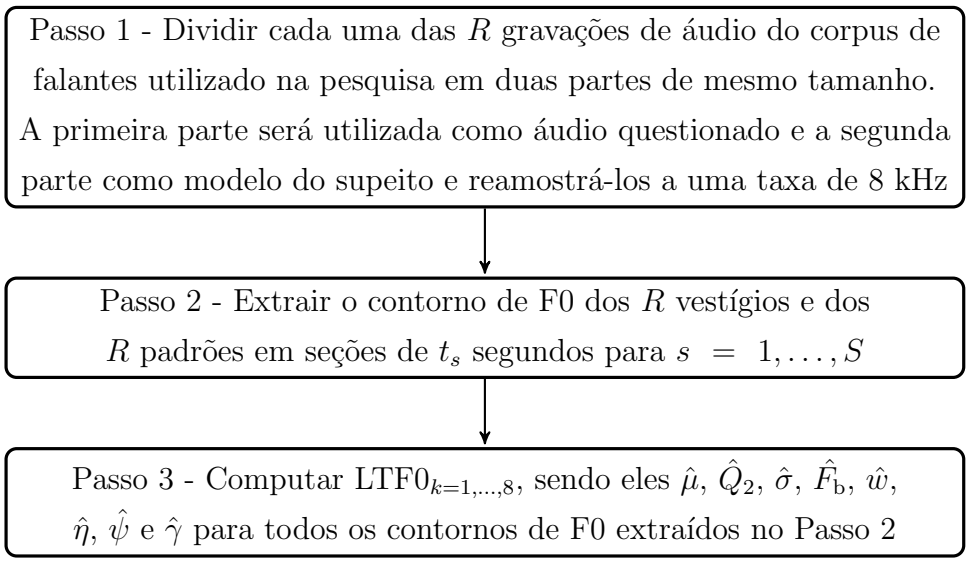


Figura 3.1: Diagrama em blocos da abordagem proposta - Passos 1 a 3.

No Passo 1 da Figura 3.1, cada um dos R áudios usados na pesquisa é dividido em duas partes de mesmo tamanho T, contendo falas líquidas de apenas um falante, para servirem como vestígio e padrão do suspeito, resultando em $2R$ amostras de áudio, sendo R utilizadas como vestígios e R como padrão de suspeito. Separando as amostras, pode-se comparar cada um dos R vestígios com cada um dos R padrões em uma abordagem $n \times n$.

Em várias aplicações forenses, as gravações são obtidas a partir de chamadas telefônicas que têm uma banda de passagem de aproximadamente 4 kHz. Portanto, as $2R$ amostras de áudio são reamostradas a uma taxa de 8 kHz, caso estejam em taxa de amostragem superior, uma vez que os componentes de alta frequência não agregam informação adicional devido ao limite imposto pelo canal de comunicação e também por não pre-

judicarem a extração das medidas de F0.

No Passo 2 da Figura 3.1, os contornos de F0 dos R vestígios e dos R padrões são extraídos em seções de t_s segundos para $s = 1, \ldots, S$ pela aplicação do método de autocorrelação (BOERSMA, 1993).

As medidas de F0 ocorrem em passos de $\frac{0{,}75}{\mathrm{F0_{min}}}$ segundos e uma janela de análise de $\frac{3}{\mathrm{F0_{min}}}$ segundos. $\mathrm{F0_{min}}$ é a menor F0 que um determinado falante consegue sustentar, e o seu valor típico se situa em torno de 75 Hz.

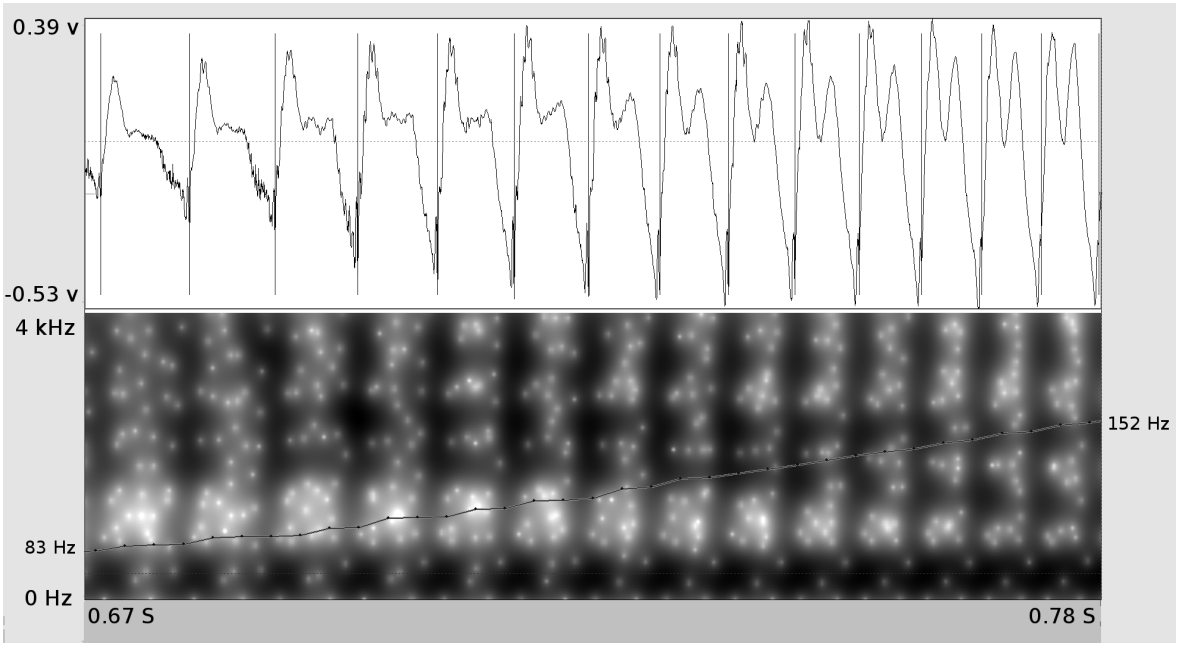


Figura 3.2: Detecção de F0 pelo método de autocorrelação do ditongo *iu* extraído da produção *viu?*. As barras verticais no oscilograma delimitam os períodos da F0. A linha inclinada sobrepondo o espectrograma indica o valor da F0 correspondente ao tamanho do período detectado.

A Figura 3.2 mostra um exemplo de detecção de F0 utilizando o software Praat® (BOERSMA; WEENINK, 2013) no ditongo *iu* extraído da produção *viu?*. Na parte superior da figura é apresentado um oscilograma onde barras verticais delimitam os períodos de F0. Note que o período vai diminuindo no decorrer da produção, aumentando o valor de F0 a fim de obter a entoação característica em contexto da produção de uma pergunta. A parte inferior da figura reproduz o espectrograma da mesma produção onde está plotado o contorno de detecção de F0, correspondendo à linha inclinada, cujos valores variam de aproximadamente 83 a 152 Hz conforme escala no extremo direito do espectrograma. Ainda na mesma figura, é possível observar, no espectrograma, a evolução dos quatro primeiros formantes que são destacados pelo tom

mais escuro presente em cada barra vertical.

Os contornos extraídos são inspecionados visualmente a fim de identificar erros de detecção de F0, entre eles, o salto de uma oitava em relação ao valor correto, trechos vozeados não detectados ou trechos desvozeados erroneamente marcados com valores válidos de F0.

No Passo 3 da Figura 3.1, cada um dos contornos extraídos e manualmente corrigidos são processados para estimar os oito parâmetros LTF0[11], $\text{LTF0}_{k=1,\ldots,8}$, nomeados conforme Tabela 3.1

Tabela 3.1: Nomenclatura utilizada nos parâmetros LTF0

Parâmetro LTF0	**Código**	**Símbolo**
Média aritmética	LTF0_1	$\hat{\mu}$
Mediana	LTF0_2	$\hat{Q}_2$
Desvio padrão	LTF0_3	$\hat{\sigma}$
Valor base de F0	LTF0_4	$\hat{F}_\text{b}$
Curtose	LTF0_5	$\hat{w}$
Assimetria	LTF0_6	$\hat{\eta}$
Moda	LTF0_7	$\hat{\psi}$
Densidade modal	LTF0_8	$\hat{\gamma}$

As medidas de cada um dos parâmetros LTF0 são armazenadas em vetores, um para cada tamanho de seção t_s utilizado, gerando, assim, S vetores de medidas para cada um dos oito LTF0 e para cada um dos R vestígios e R padrões.

Os Passos 1 a 3 da abordagem estão ilustrados na Figura 3.3.

[11]Para a estimação dos parâmetros LTF0, foi utilizado o software R® versão 3.1.3 (*Smooth Sidewalk*) baixada em 08/05/2015 de https://cran.r-project.org/src/base/R-3/.

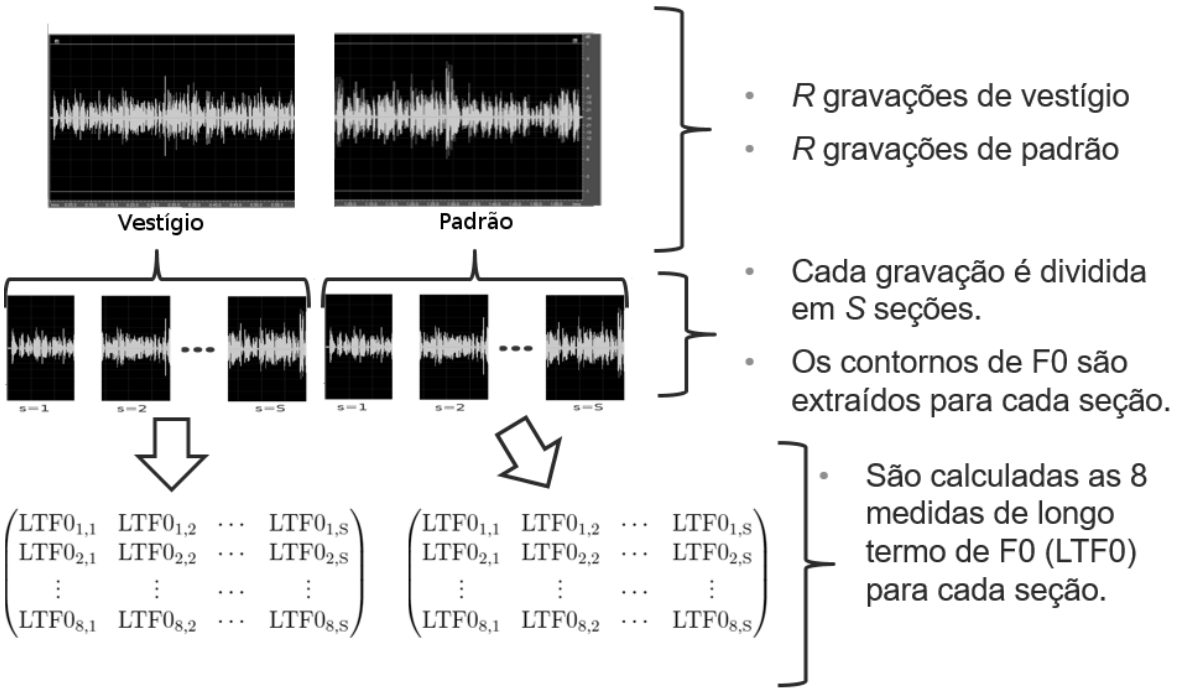


Figura 3.3: Diagrama correspondente aos Passos 1 a 3 da abordagem.

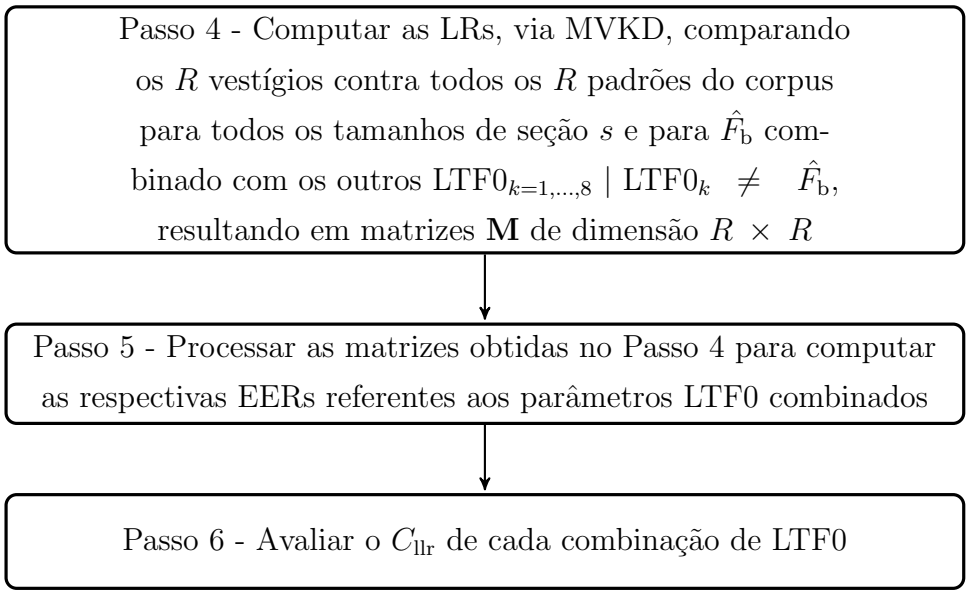


Figura 3.4: Diagrama em blocos da abordagem proposta - Passos 4 a 6.

O Passo 4 da Figura 3.4 investiga o ganho em termos discriminativos ao combinar $\hat{F}_{\mathrm{b}}$ a outras medidas LTF0, usando MVKD.

Os vetores obtidos no Passo 3 são utilizados para gerar as matrizes (2.16), (2.17) e (2.18), que são as entradas da função MVKD (2.31) utilizada para os cálculos da LR.

São realizadas comparações envolvendo cada um dos R vestígios contra cada um dos R padrões em cada uma das S seções escolhidas, totalizando $R \times R$ comparações, sendo R envolvendo mesmo falante e $R(R-1)$ envolvendo comparações entre falantes distintos.

Diferentemente de abordagens comumente aplicadas, onde as amostras disponíveis são divididas em dois grupos, um utilizado para a realização das comparações e outro para população de referência, a fim de maximizar o uso do corpus da pesquisa, é utilizada a abordagem *leave-one-out*, ou seja, a população de referência utilizada para o cálculo da LR de cada comparação é o conjunto de todos os R padrões, exceto os padrões dos dois falantes sendo testados, ou seja, $R-2$ padrões a cada comparação.

Os valores das $R \times R$ comparações por seção são armazenadas em matrizes onde a diagonal principal contém as medidas envolvendo comparações de amostras de voz de mesmo falante enquanto as demais posições da matriz contém LRs envolvendo comparações entre diferentes falantes.

A Figura 3.5 apresenta o diagrama do passo 4 da abordagem.

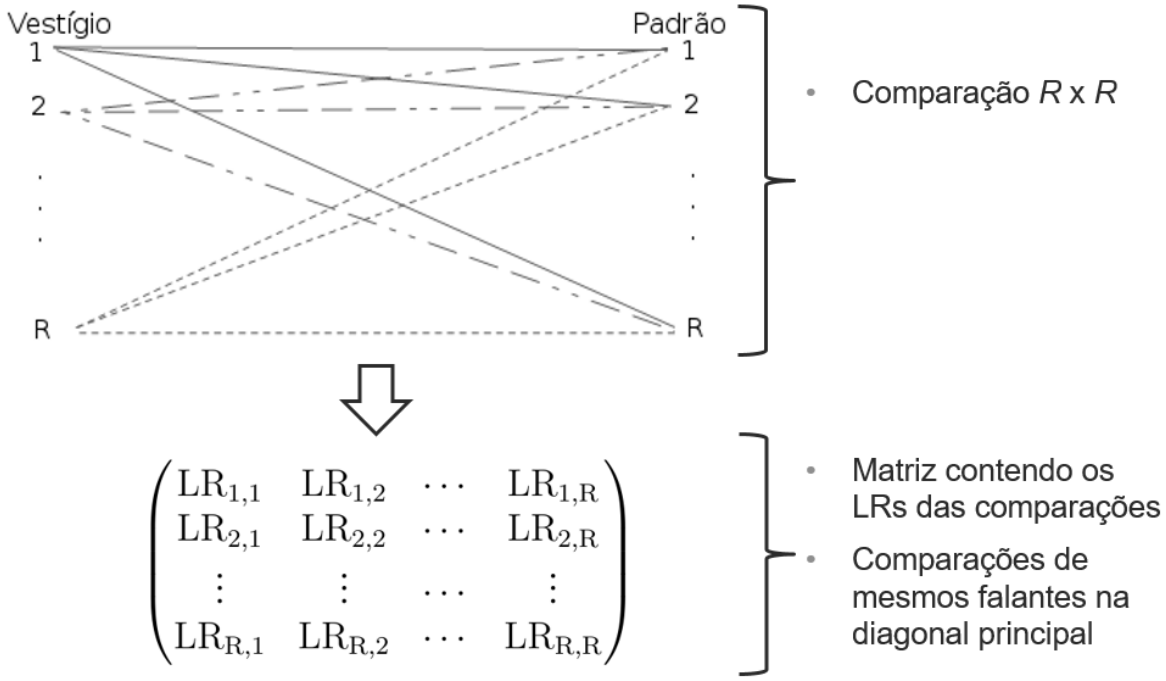


Figura 3.5: Diagrama correspondente ao Passo 4 da abordagem.

No Passo 5 da Figura 3.4, FAR e FRR para o conjunto de $\mathcal{K}$ parâmetros LTF0 são

computadas, para uma determinada seção s, como segue,

$$\mathrm{FAR}(\mathcal{K},\delta,\mathrm{s}) = \frac{1}{2\cdot R(R-1)}\sum_{r_1=1}^{R}\sum_{\substack{r_2=1\\ r_2\neq r_1}}^{R}\lceil \mathrm{sign}(m_{\mathcal{K},r_1,r_2,s}-\delta)+0,5\rceil, \tag{3.1}$$

$$\mathrm{FRR}(\mathcal{K},\delta,\mathrm{s}) = \frac{1}{2\cdot R}\sum_{r_1=r_2=1}^{R}\lceil \mathrm{sign}(\delta-m_{\mathcal{K},r_1,r_2,s})+0,5\rceil, \tag{3.2}$$

para o vestígio r_1, para o padrão de suspeito r_2, o limiar de decisão δ e $m_{\mathcal{K},r_1,r_2,s}$ que é a LR calculada por meio da equação 2.31, sendo $\mathrm{sign}(x)$ a função que retorna o valor $+1$ para $x > 0$ e -1 para $x \leq 0$.

Como exemplo, na Figura 3.6, a EER é o ponto de interceptação entre as duas curvas: FAR obtida por (3.1) e FRR obtida por (3.2), cujos valores variam com $\log_{10}(\delta)$. Note que, na Figura 3.6, o eixo y contém a proporção de valores correspondentes a FAR e FRR, a depender do limiar δ escolhido. Adicionalmente, note que, neste gráfico, o valor da EER = 13 % quando $\log(\delta) = 0{,}23$, i.e. $\delta = 1{,}17$.

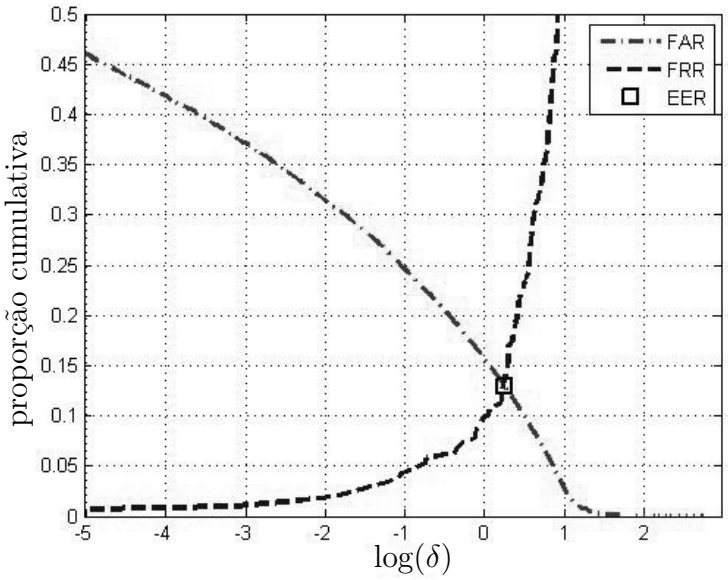


Figura 3.6: FAR e FRR versus $\log(\delta)$ usando dados da Seção 4.4, onde a mediana ($\hat{Q}_2$) e o valor de base ($\hat{F}_\mathrm{b}$) são obtidos em seções de 15 segundos. A EER, quando FAR = FRR, é igual a 13 % neste exemplo.

Ainda da análise das curvas FAR e FRR da mesma figura, pode-se constatar que a maioria das comparações envolvendo mesmo falante, curva FRR, resultaram em suporte limitado à hipótese de mesmo falante, $\log_{10}(\delta)$ entre 0 e 1, de acordo com a Tabela 2.1,

enquanto a curva FAR tem a maioria das comparações com suporte muito forte à hipótese de diferentes falantes.

Variando o limiar de decisão δ, pode-se computar a EER, conforme a expressão

$$\mathrm{EER}(\mathcal{K}, s) \;=\; \min_{\delta} |\mathrm{FAR}(\mathcal{K}, \delta, s) - \mathrm{FRR}(\mathcal{K}, \delta, s)|. \tag{3.3}$$

A EER de menor valor corresponde à combinação de parâmetros LTF0 de melhor poder discriminante.

Adicionalmente o poder discriminativo é analisado pela plotagem conjunta das curvas DET dos parâmetros analisados.

No Passo 6 da Figura 3.4, o suporte que os valores de LR obtidos em cada comparação entre as amostras de fala para cada combinação de parâmetros LTF0 é analisada por meio do cálculo do respectivo C_{llr}.

O cálculo do C_{llr} de cada conjunto de parâmetros LTF0 $\mathcal{K}$ é computado por

$$C_{\mathrm{llr}}(\mathcal{K}, s) \;=\; \frac{1}{2}\left(\frac{1}{R}\sum_{i=1}^{R}\log_2\left[1+\frac{1}{\mathrm{LR_{ss}}}\right] + \frac{1}{(R(R-1))}\sum_{j=1}^{R(R-1)}\log_2\left[1+\mathrm{LR_{ds}}\right]\right) \tag{3.4}$$

Onde R é o número de comparações envolvendo mesmos falantes, $R(R-1)$ é o total de comparações envolvendo diferentes falantes, $\mathrm{LR_{ss}}$ e $\mathrm{LR_{ds}}$ são os valores das LRs obtidas nas comparações envolvendo mesmos falantes e diferentes falantes, respectivamente.

Conforme (3.4), os erros de comparação são penalizados, não de maneira binária como na análise de FAR e FRR, mas atribuindo uma penalidade proporcional ao valor da LR obtida. O primeiro termo de (3.4) penaliza os Falsos Negativos, na medida do valor da $\mathrm{LR_{ss}}$ que, quanto mais próxima de zero, maior o somatório obtido, já o segundo termo penaliza os Falsos Positivos proporcionalmente ao valor da $\mathrm{LR_{ds}}$ obtida.

Ao avaliar dois sistemas de comparação de locutor utilizando a mesma base de dados, aquele que apresentar o menor C_{llr} corresponde ao sistema de melhor desempenho.

Analisando o valor das EERs, as curvas DET e os C_{llr} obtidos, é determinada a combinação de LTF0 com melhor poder discriminativo.

3.1 Sumário

Neste capítulo foi apresentada a abordagem proposta para avaliar o poder discriminativo obtido ao combinar parâmetros LTF0. Como resultado esperado ao aplicar a abordagem, será identificada a melhor combinação de LTF0 em termos de menor taxa EER e menor C_{llr}.

4 VALIDAÇÃO EXPERIMENTAL

A validação experimental da abordagem proposta é realizada no decorrer deste capítulo. A abordagem é aplicada a um corpus de amostras de falas em português brasileiro e os resultados são comparados com aqueles obtidos em pesquisas recentes.

Este capítulo é dividido em quatro seções. Na Seção 4.1, são apresentadas informações sobre os áudios utilizados na validação da abordagem proposta, na Seção 4.2, o poder discriminativo dos parâmetros LTF0 analisados individualmente é apresentado, na Seção 4.3, os resultados obtidos são comparados a artigos recentes e na Seção 4.4, é investigada a influência nos valores das EERs ao combinar $\hat{F}_{\mathrm{b}}$ a outros parâmetros LTF0 usando a função MVKD.

4.1 Dados e variáveis

Os experimentos e validações ao longo deste trabalho foram realizados usando gravações de falantes obtidas do Corpus Forense do Português Brasileiro (CFPB)[12]. Este corpus consiste de 206 gravações de falantes masculinos e 50 femininos, incluindo falas semi-espontâneas obtidas por meio de entrevistas e leitura de frases. Cada gravação tem aproximadamente cinco minutos líquidos de falas semi-espontâneas e um minuto de leitura de frases que visam contemplar todos os sons do português brasileiro.

O CFPB foi criado pela Polícia Federal do Brasil incluindo amostras de falantes de todas as regiões do país, a maioria delas, proveniente de trabalhadores da Polícia Federal e é utilizado para conduzir pesquisas na área forense. O presente trabalho usa o subgrupo de 206 gravações semi-espontâneas masculinas[13] do CFPB, provenientes de todas as regiões do Brasil.

A escolha de trabalhar com o corpus CFPB, nesta pesquisa, em detrimento de utilizar bases de vozes publicadas foi a falta de disponibilidade de corpus com grande quanti-

[12]O corpus CFPB foi criado pela Polícia Federal para ser utilizado em pesquisas na área forense internamente ou cedido a pesquisadores por meio de convênio com universidades.

[13]Este trabalho utilizou somente gravações masculinas uma vez que mais de 90 % das CFLs no âmbito da Polícia Federal envolvem somente falantes masculinos.

dade de amostras de falas em português brasileiro obtidas de forma homogênea. Como o objetivo da pesquisa é comparar o poder discriminante dos parâmetros LTF0, é interessante que as outras variáveis que possam influir nos valores de F0 sejam mantidas constantes, entre elas, o canal, a duração das amostras e o estilo de fala. Após determinado o parâmetro com melhor desempenho nas condições ideais, pesquisas futuras poderão avaliar o efeito das demais variáveis nos resultados obtidos.

O CFPB é homogêneo quanto a essas variáveis, uma vez que todas as gravações possuem a mesma duração e estilo de fala e são realizadas utilizando a mesma marca e modelo de microfone (Shure® - SM58), a mesma marca de placa de captura (Edirol® UA25 e UA25EX) e o mesmo software de captura (Adobe Audition® 3.0).

De acordo com Passo 1 da Figura 3.1, cada uma das 206 gravações selecionadas do CFPB foram reamostradas de 22,05 kHz, que é taxa original de amostragem do corpus, para 8 kHz e divididas em duas partes de 2 minutos de falas líquidas após removidos todos os trechos de silêncio, a primeira parte utilizada como vestígio e a segunda parte como padrão.

Conforme o Passo 2 da Figura 3.1, o software Praat®[14] (BOERSMA; WEENINK, 2013) foi usado para extrair os contornos de F0 dos 206 vestígios e dos 206 padrões de suspeito pelo método de autocorrelação proposto por (BOERSMA, 1993) em passos de $\frac{0{,}75}{\mathrm{F0_{min}}}$ segundos, que é igual a 0,01 segundos para uma $\mathrm{F0_{min}}$ típica de 75 Hz, resultando em 100 medidas de F0 por segundo de produções vozeadas.

Em seguida, uma inspeção visual, por meio das ferramentas disponibilizadas no software Praat®, foi realizada em todos os contornos de F0 para identificar erros remanescentes e corrigi-los manualmente.

A inspeção visual dos contornos é realizada no presente trabalho para que as medidas de LTF0 sejam dependentes diretamente das suas definições e não sejam afetadas por eventuais erros de extração de F0.

Na Figura 4.1 é apresentado um exemplo de inspeção visual utilizando o software Praat® onde dois valores de F0 em torno de 600 Hz são erroneamente detectados e

[14]Utilizou-se o *script* do Praat (better_f0-2012-03.16.praat) versão 1.3 provido por Pablo Arantes e disponibilizado em https://code.google.com/archive/p/praat-tools/downloads. Este *script* minimiza os erros de extração de F0 escolhendo os melhores valores mínimos e máximos de F0.

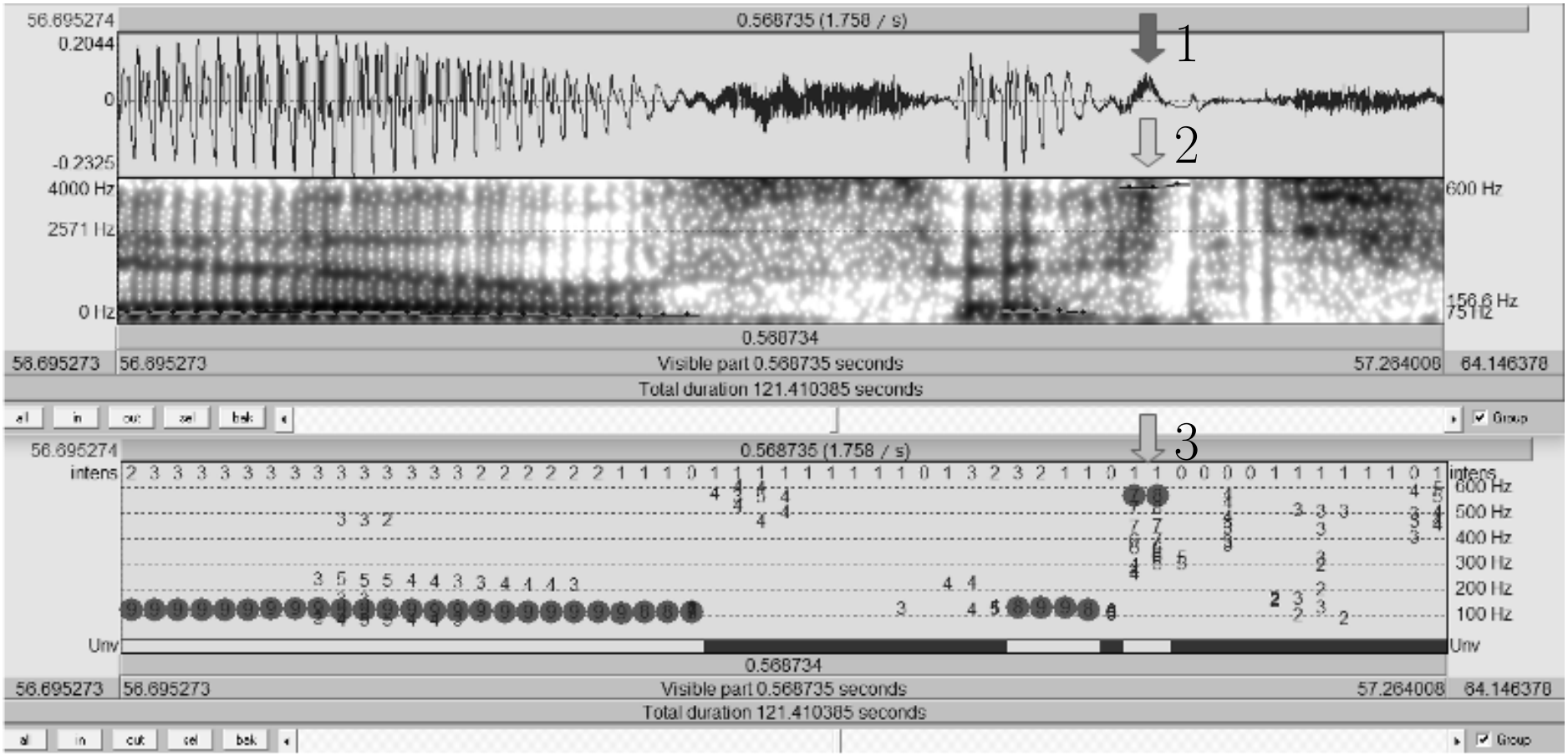


Figura 4.1: Correção de F0 utilizando o software Praat®. A seta 1 indica a produção de uma fricativa não vozeada na janela do oscilograma, a seta 2, no espectrograma, indica a detecção equivocada de F0 em torno de 600 Hz e a seta 3 localizada na janela de edição de F0 indica as medidas que deverão ser eliminadas.

devem ser eliminados.

Ao final da correção manual de cada um dos contornos de F0, em média apenas 60 % dos dois minutos de cada áudio resulta em produções vozeadas com valores válidos de F0, correspondendo a aproximadamente 72 segundos de medidas líquidas de F0. A fim de igualar a quantidade de material em todas as amostras de áudio e garantir comparações homogêneas, foram selecionados os primeiros 60 segundos equivalentes de medidas de F0 por vestígio e por padrão.

Pelo Passo 3 da Figura 3.1, cada um dos contornos de F0 extraídos e corrigidos nos passos anteriores foram divididos em seções $t_s = 5$, 10, 15, 20 e 30 segundos para $s = 1, \ldots, 5$. Cada uma das seções foram processadas utilizando o software R® para extrair os oito parâmetros LTF0 listados na Tabela 3.1: $\hat{\mu}$, $\hat{Q}_2$, $\hat{\sigma}$, $\hat{F}_{\mathrm{b}}$, $\hat{w}$ e $\hat{\eta}$ usando as bibliotecas "E1071"(MEYER et al., 2015) e "Stats"(R Core Team, 2014) e $\hat{\psi}$ e $\hat{\gamma}$ foram extraídas usando a biblioteca "KernSmooth" (WAND, 2015) do mesmo modo descrito por (KINOSHITA; ISHIHARA; ROSE, 2009).

Cada um dos parâmetros LTF0 resultou em cinco vetores contendo z_s medidas de LTF0 cada, onde $z_s = \frac{60}{t_s}$ e t_s é o tamanho da seção. Exemplificativamente, ao utilizar seção t_s com 15 segundos, obtém-se 4 medidas de cada LTF0, a primeira referente aos 15 segundos iniciais, a segunda referente aos 15 segundos subsequentes e assim por diante, conforme pode ser observado na Figura 4.2, onde cada retângulo delimita o intervalo de áudio que será utilizado para cada uma das medidas de 1 a z_{s}.

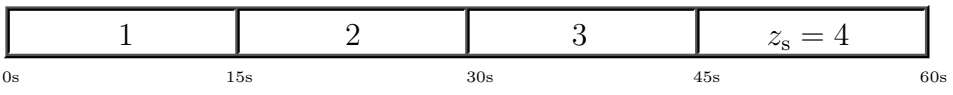


Figura 4.2: Esquema de divisão de cada amostra de áudio em intervalos de medidas. Para $t_{\mathrm{s}} = 15$ s, $T = 60$ s, são obtidas quatro medidas, considerando os intervalos delimitados na figura.

4.2 Poder discriminativo dos parâmetros LTF0 analisados individualmente

Inicialmente foi avaliado o poder discriminativo de cada um dos oito parâmetros LTF0 analisados neste trabalho. Para tal, conforme Passo 4 da Figura 3.4, cada um dos cinco vetores por LTF0, correspondendo a $s = 1, \ldots, 5$, e por falante foram processados para calcular LRs usando MVKD, conforme (2.31), proposto por (AITKEN; LUCY, 2004) e implementado em Matlab® por (MORRISON, 2007).

De acordo com a abordagem proposta, em cada comparação, as amostras dos padrões de falantes sob teste foram removidas e todas as outras foram utilizadas como população de referência, metodologia *leave one out*. Foram, portanto, realizadas 42436 comparações (206 x 206), 206 delas do mesmo falante e 42230 comparações envolvendo diferentes falantes.

Para cada comparação entre os falantes, envolvendo vestígio r_1 e o padrão r_2 obtém-se um valor de LR que é armazenado em uma matriz $\mathbf{M}$. As comparações envolvendo mesmos falantes, i.e., quando $r_1 = r_2$, ocupam a diagonal principal de $\mathbf{M}$.

A seguir, para cada matriz obtida, variando o limiar de decisão δ em (3.1) e (3.2), é determinado o valor da EER, quando FAR = FRR.

Na Tabela 4.1 são apresentadas as EERs obtidas em todos os testes envolvendo parâmetros LTF0 isoladamente, enquanto a Figura 4.3 possibilita uma comparação visual dos mesmos dados. A EER obtida pelo parâmetro $\hat{F}_{\mathrm{b}}$ ($\mathrm{LTF0}_4$) superou o resultado de todos os outros $\mathrm{LTF0}_{\mathrm{k}}$, para $k = 1, \ldots, 8$ e $k \neq 4$. O valor de base de F0 $\hat{F}_{\mathrm{b}}$ ($\mathrm{LTF0}_4$) obteve uma EER média de 16,1 %, o que é aproximadamente 6 % menor que o segundo melhor parâmetro LTF0 analisado, média aritmética $\hat{\mu}$ ($\mathrm{LTF0}_1$).

Tabela 4.1: EERs dos parâmetros LTF0

	EER (%)					
LTF0 (Símbolo)	$s = 1$	$s = 2$	$s = 3$	$s = 4$	$s = 5$	Média do LTF0
LTF0_4 ($\hat{F}_\text{b}$)	16,5	15,9	16,1	16,1	15,7	**16,1**
LTF0_1 ($\hat{\mu}$)	22,3	22,2	22,0	21,9	21,4	22,0
LTF0_7 ($\hat{\psi}$)	22,2	20,9	22,1	21,7	23,3	22,0
LTF0_2 ($\hat{Q}_2$)	22,3	21,8	22,7	21,7	21,5	22,0
LTF0_3 ($\hat{\sigma}$)	32,0	32,5	32,5	33,0	33,0	32,6
LTF0_8 ($\hat{\gamma}$)	33,4	33,0	32,5	33,1	34,0	33,2
LTF0_6 ($\hat{\eta}$)	45,2	42,1	40,3	43,7	41,9	42,6
LTF0_5 ($\hat{w}$)	42,0	43,7	41,7	42,2	43,1	42,5
Média geral	29,5	29,0	**28,7**	29,2	29,2	29,1

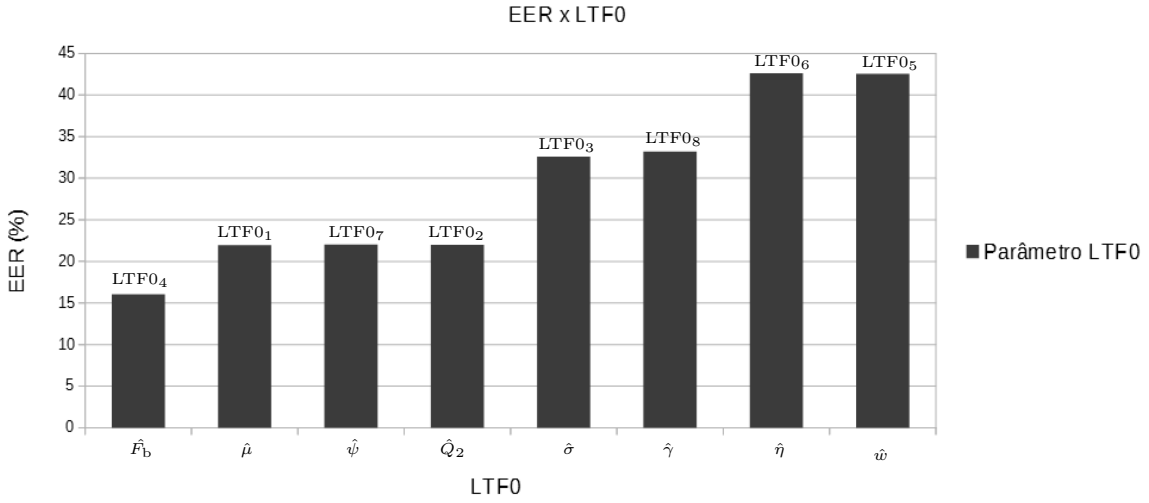


Figura 4.3: EERs obtidas pelos parâmetros LTF0 isoladamente, a partir da Tabela 4.1.

De acordo com a Tabela 4.1, não há discrepância significativa entre os valores de EER envolvendo o mesmo parâmetro LTF0 mas com diferentes durações das seções.

Para comparar o desempenho discriminativo dos parâmetros LTF0 de maneira global, foram traçadas as curvas DET dos 8 parâmetros analisados em trechos de 15 s, $s = 3$, por ter atingido a menor média entre os vários tamanhos de trechos, resultando na Figura 4.4. Comparando as curvas DET obtidas, identifica-se claramente o melhor desempenho do valor de base da F0 $\hat{F}_\text{b}$, que possui a menor área sob seu

traçado. Destacam-se também as curvas correspondentes à assimetria $\hat{\eta}$ e à curtose $\hat{w}$, respectivamente, com os piores desempenhos.

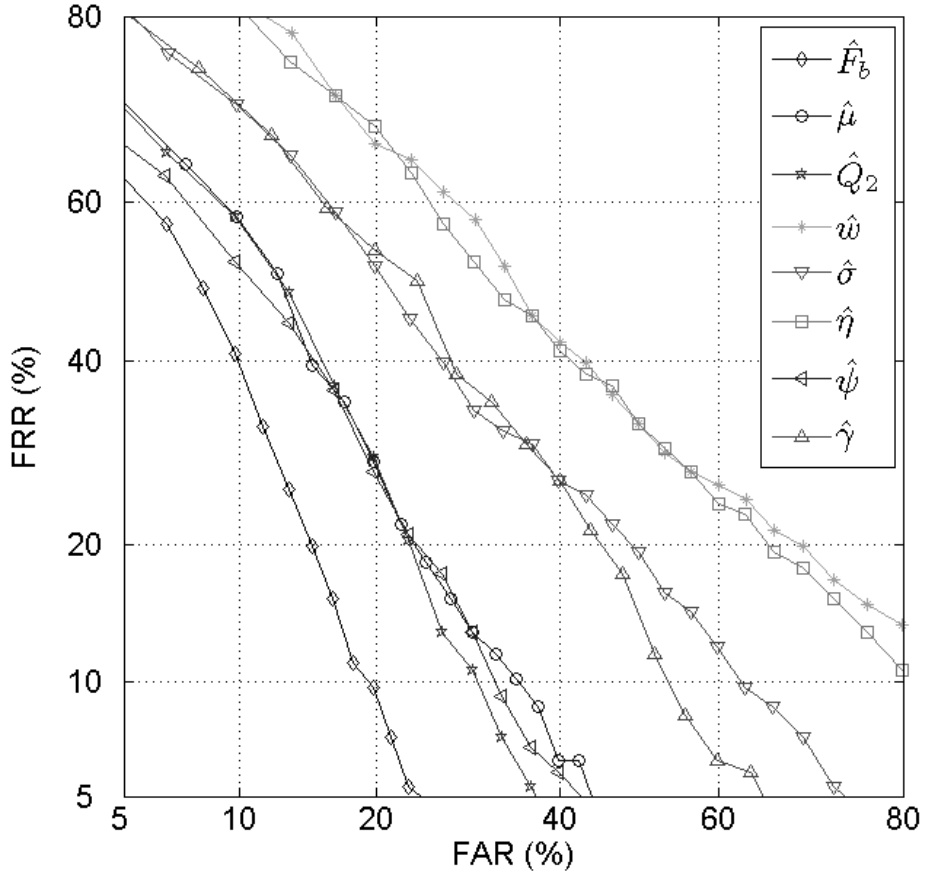


Figura 4.4: Curvas DET dos parâmetros LTF0 analisados isoladamente. O parâmetro $\hat{F}_\mathrm{b}$ obteve o melhor desempenho discriminativo

Nas figuras 4.5 e 4.6 são apresentadas as curvas FAR e FRR em função do limiar de decisão $\log(\delta)$ dos parâmetros LTF0 $\hat{F}_\mathrm{b}$ e $\hat{\mu}$, respectivamente. Comparando os gráficos, percebe-se a menor EER do parâmetro $\hat{F}_\mathrm{b}$ e um ligeiro deslocamento da curva FRR para a direita em relação ao parâmetro $\hat{\mu}$, evidenciando a maior quantidade de LR > 10, ou seja, $\log_{10}(\delta)>1$, indicando possuir maior suporte em comparações de mesmos falantes.

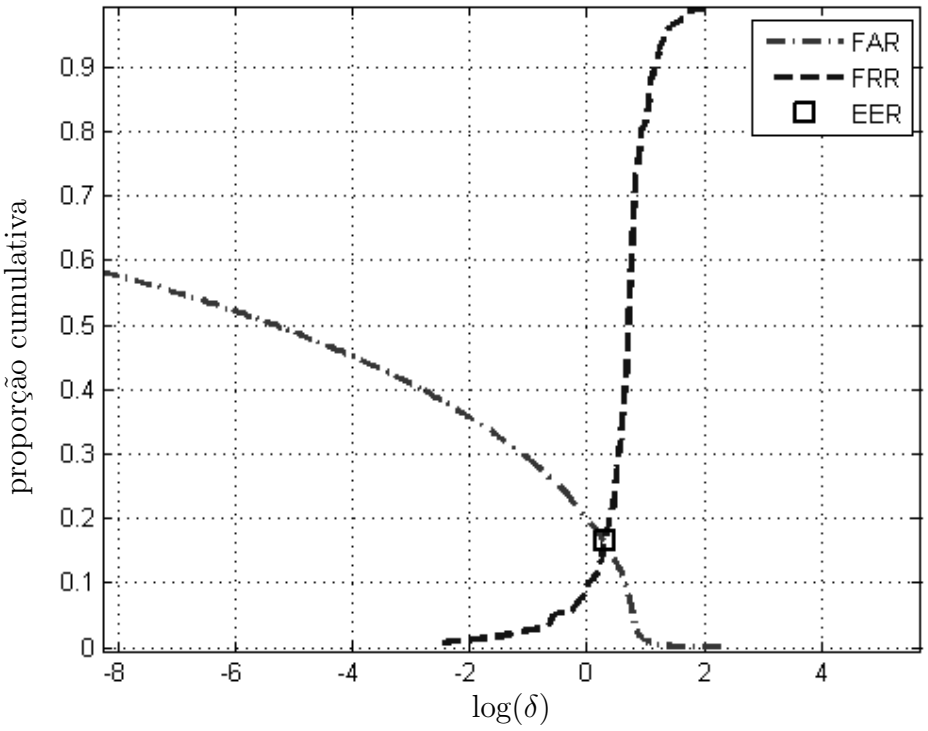


Figura 4.5: FAR e FRR versus $\log(\delta)$ do parâmetro $\hat{F}_{\mathrm{b}}$ em seções de 15 segundos conforme Tabela 4.1. A EER é igual a 16,1 %.

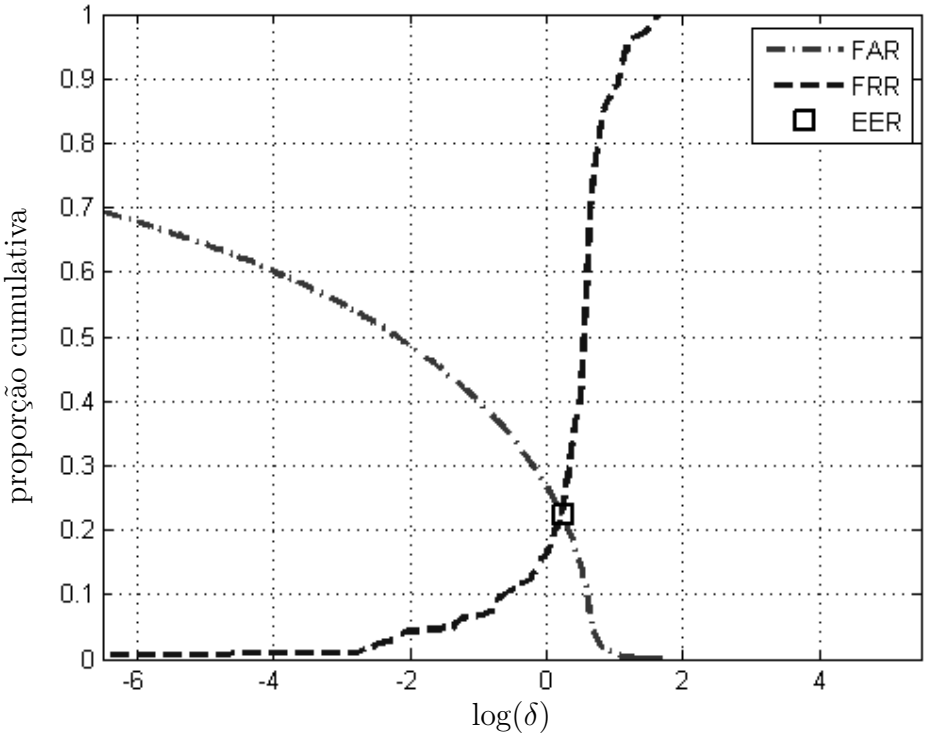


Figura 4.6: FAR e FRR versus $\log(\delta)$ do parâmetro $\hat{\mu}$ em seções de 15 segundos conforme Tabela 4.1. A EER é igual a 22,0 %.

4.3 Aplicando abordagens de pesquisas recentes ao corpus CFPB

As abordagens adotadas por (GOLD, 2014) e (KINOSHITA; ISHIHARA; ROSE, 2009) foram utilizadas como referência para avaliar o desempenho obtido pela nossa abordagem usando o corpus CFPB.

Em (GOLD, 2014), a autora estuda o poder discriminativo da combinação dos parâmetros LTF0 $\hat{\mu}$ e $\hat{\sigma}$ usando MVKD para calcular LRs envolvendo 100 gravações de falantes ingleses com uma média de 6 minutos de duração cada, sendo 50 deles usados como população de referência. Usando a mesma metodologia, combinando os parâmetros $\hat{\mu}$ e $\hat{\sigma}$ nas 206 gravações semi-espontâneas de falantes masculinos do CFPB, obteve-se uma EER de 17,2 %, conforme Tabela 4.2. Percebe-se que o poder discriminativo do parâmetro $\hat{F}_{\mathrm{b}}$ apresenta uma EER média de 16,1 %, sendo, portanto, mais discriminante que o uso dos parâmetros $\hat{\mu}$ e do $\hat{\sigma}$ combinados.

Tabela 4.2: EER obtida aplicando a abordagem proposta por (GOLD, 2014) ao CFPB.

	EER (%)					
	$s = 1$	$s = 2$	$s = 3$	$s = 4$	$s = 5$	Média
$\hat{\mu} + \hat{\sigma}$ = (GOLD, 2014)	17,4	17,3	17,5	17,0	17,0	17,2
$\hat{F}_{\mathrm{b}}$	16,5	15,9	16,1	16,1	15,7	**16,1**

Conforme já pontuado anteriormente, a pesquisa realizada por Gold (GOLD; FRENCH, 2011) da universidade de York apurou serem a média e o desvio padrão os parâmetros LTF0 mais populares entre os peritos forenses. Os resultados aqui obtidos mostram ser mais adequado trabalhar com $\hat{F}_{\mathrm{b}}$.

Em (KINOSHITA; ISHIHARA; ROSE, 2009), os autores analisam o poder discriminativo de F0 usando 201 gravações de falantes japoneses de 10 a 25 minutos de duração cada. Todas as 201 gravações foram usadas como população de referência sem remover os falantes comparados, ou seja, sem utilizar a abordagem *leave-one-out*.

Inicialmente os autores analisaram o poder discriminativo dos parâmetros LTF0 média aritmética, desvio padrão, assimetria, moda, curtose e densidade modal, individualmente aplicados ao corpus CFPB, e obtiveram os resultados presentes na última linha

da Tabela 4.3. Ressalta-se que os parâmetros LTF0 mediana, $LTF0_2$, e valor de base de F0, $LTF0_4$, não foram avaliados por (KINOSHITA; ISHIHARA; ROSE, 2009).

Tabela 4.3: Comparativo entre as EERs obtidas por (KINOSHITA; ISHIHARA; ROSE, 2009) e obtidas no corpus CFPB.

	EER %					
	$LTF0_1(\hat{\mu})$	$LTF0_3(\hat{\sigma})$	$LTF0_5(\hat{w})$	$LTF0_6(\hat{\eta})$	$LTF0_7(\hat{\psi})$	$LTF0_8(\hat{\gamma})$
Kinoshita	22,3	28,1	29,4	28,0	21,6	26,7
CFPB	22,0	32,6	42,5	42,6	22,0	33,2

De acordo com a Tabela 4.3, todos os valores são similares, exceto pelos parâmetros $LTF0_5$ (assimetria) e $LTF0_6$ (curtose). Uma possível razão está associada ao fato das gravações do CFPB serem menores que aquelas utilizadas no trabalho (KINOSHITA; ISHIHARA; ROSE, 2009) ou ainda pode estar relacionado ao idioma.

Cabe ressaltar que uma comparação direta dos resultados deve ser feita com cautela dado que a população de referência utilizada nos trabalhos citados está em idioma diverso do português brasileiro, o que pode afetar a distribuição de F0 e consequentemente os valores de LTF0.

Na mesma pesquisa, usando MVKD, foram combinados os 6 parâmetros LTF0: $\hat{\mu}$, $\hat{\sigma}$, $\hat{\eta}$, $\hat{w}$, $\hat{\psi}$ e $\hat{\gamma}$. Reproduzindo a mesma combinação de parâmetros usando o CFPB resultou em uma EER de 14,8 % (Tabela 4.4), melhor que o uso de $\hat{F}_{\mathrm{b}}$ isoladamente. A fim de verificar se a inclusão do parâmetro LTF0 $\hat{F}_{\mathrm{b}}$ aos seis parâmetros LTF0 utilizados por (KINOSHITA; ISHIHARA; ROSE, 2009) melhorava o poder discriminativo, este foi, então, incluído e, como esperado, houve uma queda na EER média que ficou em 14,6 %, melhorando o poder discriminativo.

Tabela 4.4: Abordagem proposta por (KINOSHITA; ISHIHARA; ROSE, 2009) aplicada ao CFPB.

	EER (%)					
Combinação de LTF0 $\{\mathcal{K}\}$	$s = 1$	$s = 2$	$s = 3$	$s = 4$	$s = 5$	Média
EER($\mathcal{K} = \{1, 3, 5, 6, 7, 8\}$, s) = (KINOSHITA; ISHIHARA; ROSE, 2009)	15,4	15,0	14,6	14,5	14,4	14,8
EER($\mathcal{K} = \{1, 3, 4, 5, 6, 7, 8\}$, s) = (KINOSHITA; ISHIHARA; ROSE, 2009)+$\hat{F}_{\mathrm{b}}$	14,0	14,2	15,1	14,5	15,1	14,6
$\hat{F}_{\mathrm{b}}$	16,5	15,9	16,1	16,1	15,7	16,1

4.4 Abordagem proposta usando a melhor combinação de LTF0

Considerando o melhor desempenho obtido ao incluir $\hat{F}_{\mathrm{b}}$ à abordagem utilizada por (KINOSHITA; ISHIHARA; ROSE, 2009), passou-se a investigar qual combinação entre os 8 parâmetros LTF0 analisados resultaria em menor EER, conforme Passos 4 e 5 da Figura 3.4.

Inicialmente, verificou-se que a simples combinação de todos os 8 parâmetros LTF0 não resulta na melhor EER, uma vez que esta combinação obteve uma EER média pior que a abordagem de Kinoshita (KINOSHITA; ISHIHARA; ROSE, 2009), conforme a última linha da Tabela 4.5.

Analisando a Tabela 4.1, além de $\hat{F}_{\mathrm{b}}$ $\mathrm{LTF0}_4$, somente $\hat{\mu}$ $\mathrm{LTF0}_1$, $\hat{Q}_2$ $\mathrm{LTF0}_2$ e $\hat{\psi}$ $\mathrm{LTF0}_7$ obtiveram EERs abaixo de 30 % no CFPB, além disso, os mesmos parâmetros se destacaram nas curvas DET da Figura 4.4. Estas quatro LTF0 foram selecionadas para os testes e todas as possíveis combinações envolvendo $\hat{F}_{\mathrm{b}}$ e os outros três parâmetros foram feitas utilizando MVKD em seções t_s = 5, 10, 15, 20 e 30 segundos, para s = 1, ..., 5. Para isso, três matrizes contendo vetores coluna de cada um dos parâmetros LTF0 do vestígio de acordo com (2.16), padrão do suspeito de acordo com (2.17) e população de referência de acordo com (2.18) foram usadas como entrada da função MVKD (2.31).

Tabela 4.5: EER das combinações propostas de parâmetros LFT0 aplicadas ao CFPB

	EER (%)					
Combinação de LTF0 $\{\mathcal{K}\}$	$s = 1$	$s = 2$	$s = 3$	$s = 4$	$s = 5$	Média
EER($\mathcal{K} = \{2, 4\}$, s)	14,9	14,2	**13,0**	14,6	13,5	14,0
EER($\mathcal{K} = \{2, 4, 7\}$, s)	14,9	14,1	13,1	14,5	13,7	14,1
EER($\mathcal{K} = \{1, 2, 4\}$, s)	15,4	14,5	13,5	14,0	13,6	14,2
EER($\mathcal{K} = \{1, 2, 4, 7\}$, s)	14,6	15,0	13,2	14,9	14,5	14,4
EER($\mathcal{K} = \{1, 4\}$, s)	15,5	14,2	15,5	15,5	14,6	15,1
EER($\mathcal{K} = \{1, 4, 7\}$, s)	15,0	15,0	13,5	14,9	13,5	15,1
EER($\mathcal{K} = \{4, 7\}$, s)	15,0	14,6	14,6	16,1	15,5	15,1
EER($\mathcal{K} = \{1, 2, 3, 4, 5, 6, 7, 8\}$, s)	15,1	15,0	15,4	15,0	15,1	15,1

Como mostrado na Tabela 4.5, usando o mesmo corpus CFPB e MVKD, o valor de base de F0 $\hat{F}_{\mathrm{b}}$ (LTF0$_4$) combinado com a mediana $\hat{Q}_2$ (LTF0$_2$) superou o desempenho de todas as outras combinações com uma EER de 13 % usando $s = 3$, ou seja, trechos de 15 s.

Na figura 4.7 são apresentadas as curvas FAR e FRR em função do limiar de decisão $\log(\delta)$ da combinação dos parâmetros LTF0 $\hat{F}_{\mathrm{b}}$ e $\hat{Q}_2$. Percebe-se que a EER que ficou em 13 % correspondendo a uma melhora de 3,1 % em relação à EER do parâmetro $\hat{F}_{\mathrm{b}}$ utilizado isoladamente.

Adicionalmente foram plotadas as curvas DET das combinações envolvendo os quatro melhores parâmetros LTF0 na Figura 4.8 e constatou-se que nenhuma das combinações se destacou em relação às outras, conforme a Tabela 4.5 já indicava. Sendo que as curvas de $\hat{F}_{\mathrm{b}}$ combinado com a mediana $\hat{Q}_2$ e de $\hat{F}_{\mathrm{b}}$ combinado com a mediana $\hat{Q}_2$ e moda $\hat{\psi}$ apresentam um desempenho ligeiramente superior, duas primeiras linhas da Tabela 4.5. Cabe destacar o pior desempenho da combinação $\hat{\mu}$ com $\hat{\sigma}$, justamente envolvendo os parâmetros LTF0 mais citados pelos peritos na pesquisa (GOLD; FRENCH, 2011).

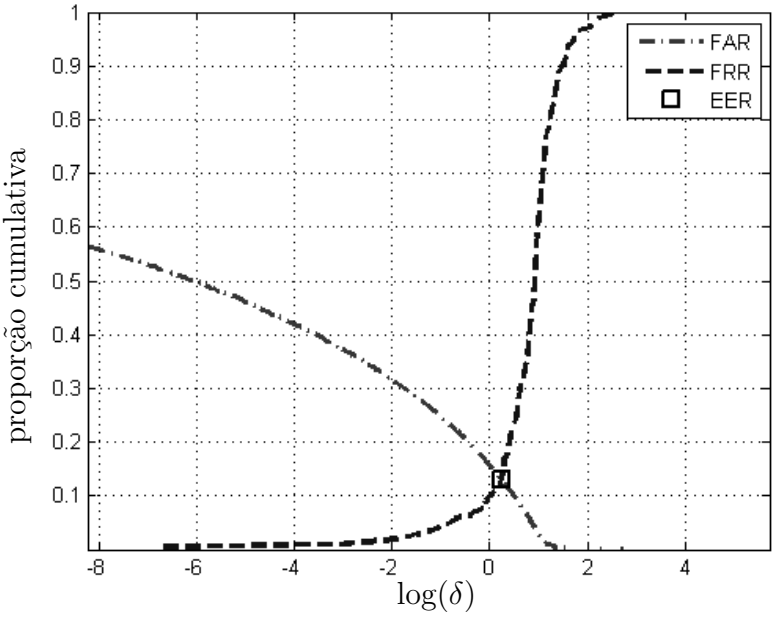


Figura 4.7: FAR e FRR versus $\log(\delta)$ da combinação dos parâmetros $\hat{F}_{\mathrm{b}}$ e $\hat{Q}_2$ em seções de 15 segundos. A EER é igual a 13 %.

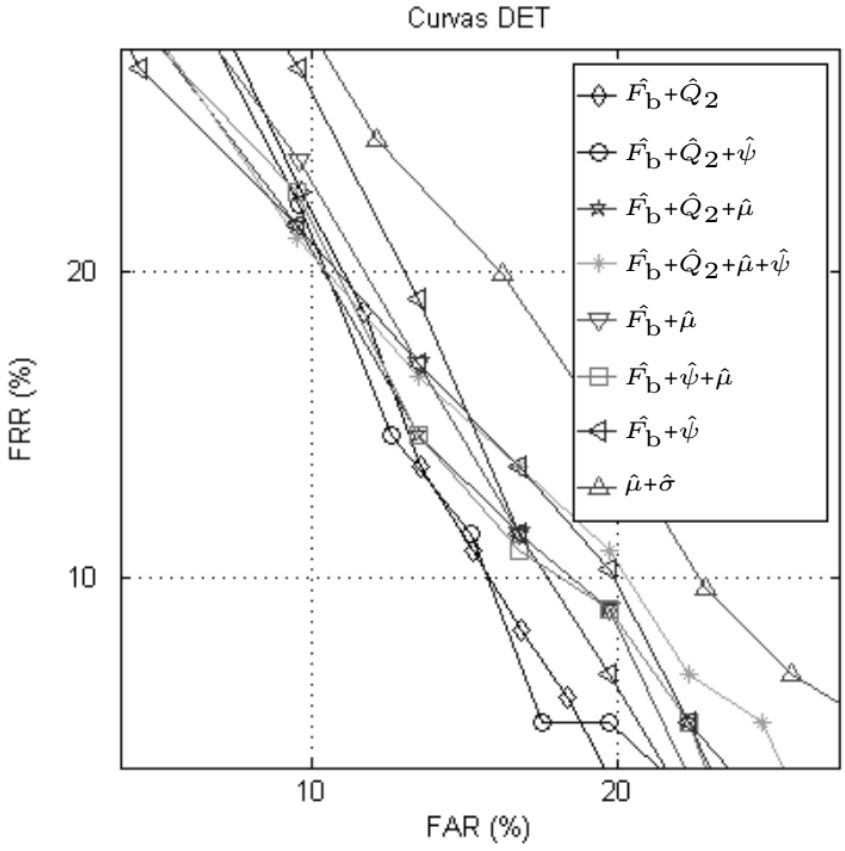


Figura 4.8: Curvas DET dos parâmetros LTF0 combinados.

Conforme Passo 6 da Figura 3.4, foi calculado o valor dos C_{llr} (3.4), de cada uma das comparações envolvendo $\hat{F}_{\text{b}}$, $\hat{Q}_2$, $\hat{\psi}$ e $\hat{\mu}$, relacionadas na Tabela 4.5.

Na Tabela 4.6 são apresentados os C_{llr} obtidos nas combinações de LTF0. Constata-se que o menor $C_{\text{llr}} = 0,618$ resulta da combinação $\hat{F}_{\text{b}}$ com a mediana $\hat{Q}_2$, reforçando os achados anteriormente obtidos na avaliação das EER e das curvas DET.

Tabela 4.6: C_{llr} das combinações propostas de parâmetros LFT0 aplicadas ao CFPB

	C_{llr}					
Combinação de LTF0 $\{\mathcal{K}\}$	$s = 1$	$s = 2$	$s = 3$	$s = 4$	$s = 5$	Média
EER($\mathcal{K} = \{2, 4\}$, s)	0,685	0,621	**0,618**	0,642	0,849	0,683
EER($\mathcal{K} = \{2, 4, 7\}$, s)	0,950	0,952	0,743	0,952	1,113	0,940
EER($\mathcal{K} = \{1, 2, 4\}$, s)	0,950	0,952	0,920	0,953	1,113	0,978
EER($\mathcal{K} = \{1, 2, 4, 7\}$, s)	0,916	0,900	1,151	0,810	0,984	0,952
EER($\mathcal{K} = \{1, 4\}$, s)	0,849	0,833	0,819	0,846	0,897	0,849
EER($\mathcal{K} = \{1, 4, 7\}$, s)	0,810	0,766	0,863	0,736	0,804	0,796
EER($\mathcal{K} = \{4, 7\}$, s)	0,663	0,790	1,167	0,871	1,189	0,936
EER($\mathcal{K} = \{1, 3\}$, s) = (GOLD, 2014)	0,773	0,722	0,875	0,822	0,995	0,798

Portanto, os estudos aqui realizados indicam ser mais vantajoso a utilização da combinação do valor de base de F0, $\hat{F}_{\text{b}}$, com a mediana, $\hat{Q}_2$, combinados via MVKD em análises de F0.

4.5 Sumário

Este capítulo apresentou os resultados obtidos ao realizar combinações entre os parâmetros LTF0 investigados neste trabalho. Constatou-se que a combinação de $\hat{F}_{\text{b}}$ com $\hat{Q}_2$ resulta na menor EER e no menor valor de C_{llr}, sendo, entre todas as combinações de parâmetros LTF0 avaliadas, a de melhor desempenho discriminativo.

5 CONCLUSÕES

O poder discriminante dos parâmetros de longo termo da frequência fundamental (LTF0) foi avaliado usando um subgrupo do Corpus Forense do Português Brasileiro (CFPB) contendo 206 gravações de falas semi-espontâneas masculinas divididas em duas partes contendo um minuto de produções vozeadas referentes ao vestígio e ao padrão do suspeito.

Multivariate Kernel-Density (MVKD) foi utilizado no cálculo das razões de verossimilhança (LR) das comparações realizadas entre os vestígios e padrões.

Inicialmente foi analisado o poder discriminante dos parâmetros LTF0 média aritmética, mediana, desvio padrão, assimetria, curtose, moda, densidade modal e valor de base de F0 utilizando seções de 5, 10, 15, 20 e 30 segundos de áudio. O valor de base de F0 obteve melhor poder discriminante com a menor EER média de 16,1 %. O tamanho da seção utilizada não resultou em diferenças significativas nos resultados, exceto para seções de 5 s que apresentaram um desempenho ligeiramente inferior às demais.

Em seguida, foi avaliado o poder discriminativo combinando o valor de base de F0 aos parâmetros média aritmética, mediana e moda. Após extensivos experimentos, a menor EER de 13 % foi obtida combinando o valor de base de F0 com a mediana usando 4 seções de 15 segundos em cada gravação (Tabela 4.5). Novamente, o tamanho da seção influenciou pouco nos resultados e, da mesma forma, as seções de 5 s apresentaram desempenho um pouco abaixo das demais.

Por fim, foi avaliado o desempenho das combinação de LTF0 para determinar aquela que resultava em melhores suportes às comparações pelo cálculo dos respectivos valores de C_{llr}, constatando novamente que a combinação do valor de base de F0 com a mediana resultava no melhor desempenho.

Os resultados desta pesquisa indicam que em exames de comparação forense de locutor no português brasileiro, a medida acústica valor de base de F0 deve ser usada preferencialmente às outras medidas comumente utilizadas pelos peritos forenses da área de comparação de locutor, entre elas, a média e o desvio padrão que foram as mais citadas

na pesquisa (GOLD; FRENCH, 2011).

Constatou-se ainda que o uso do valor de base de F0 combinado com a mediana via MVKD resulta em um melhor poder discriminativo, sendo recomendado o seu uso em detrimento de outras combinações normalmente utilizadas na área forense e aqui avaliadas.

5.1 Recomendações para pesquisas futuras

Esta pesquisa foi realizada em gravações de áudio com qualidade de estúdio realizadas em uma única sessão por falante. Para o objetivo desta pesquisa, que foi o de comparar resultados entre os diversos parâmetros LTF0, trata-se da situação ideal, uma vez que elimina fatores que possam degradar e mascarar os resultados. Como pesquisa futura sugere-se avaliar a degradação nos resultados em decorrência da inserção de ruído, pela utilização de gravações realizadas em mais de uma ocasião entre padrão e vestígio, variação de canal de gravação, estilo da fala, entre outros.

Sugere-se também avaliar o poder discriminativo de diferentes percentis da distribuição de valores de F0, considerando que o valor de base de F0, que corresponde ao percentil 7,64 %, apresentou maior poder discriminativo e não foram avaliados, nesta pesquisa, outros valores de percentis e o correspondente efeito no poder discriminativo.

Outra importante área de pesquisa seria adaptar a função MVKD para o uso de tensores e verificar se há um ganho tensorial, melhorando os resultados das LRs.

E finalmente, realizar um comparativo dos resultados obtidos com a abordagem aqui proposta em casos reais com aqueles obtidos utilizando a abordagem tradicional em exames de CFL.

REFERÊNCIAS BIBLIOGRÁFICAS

AITKEN, C. G. G.; LUCY, D. Evaluation of trace evidence in the form of multivariate data. *Journal of the Royal Statistical Society, 53 (1)*, p. 109–122, 2004.

AITKEN, C. G. G.; TARONI, F. *Statistics and the Evaluation of Evidence for Forensic Scientists*. 2nd. ed. [S.l.]: Chichester: Wiley, 2004.

ARANTES, P.; ERIKSSON, A. Temporal stability of long-term measures of fundamental frequency. In: *International Conference on Speech Prosody, 7th, 2014, Dublin*. [S.l.]: Campbell, Gibbon, and Hirst (eds.), 2014. p. 1149–1152. ISSN 2333-2042.

BARBOSA, P. A.; MADUREIRA, S. *Manual de fonética acústica experimental*. 1st. ed. [S.l.]: Editora Cortez, 2015.

BOERSMA, P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In: *IFA Proceedings, vol. 17*. [S.l.: s.n.], 1993. p. 97–110.

BOERSMA, P.; WEENINK, D. *Praat: doing phonetics by computer [Computer program]. Version 5.3.70*. [S.l.]: Retrieved 5 April 2014 from http://www.praat.org/, 2013.

BRAUN, A. Fundamental frequency? How speaker-specific is it? *BEIPHOL 64: Studies in Forensic Phonetics*, p. 9–23, 1995.

BRÜMMER, N.; PREEZ, J. du. Application independent evaluation of speaker detection. *Computer Speech and Language, 20*, p. 230–275, 2006.

CHAMPOD, C.; EVETT, I. W. Commentary on broeders 1999. *Forensic Linguistics, 7(2)*, p. 238–243, 1999.

ENFSI. *ENFSI guideline for evaluative reporting in forensic science*. [S.l.], 2015. Downloaded: April 2016. Disponível em: <http://www.enfsi.eu/documents/external-publications>.

FANT, G. *Acoustic theory of speech production*. [S.l.], 1960. The Hague: Mouton.

GOLD, E. *Calculating likelihood ratios in forensic speaker comparison cases using phonetic and linguistic features*. Tese (Doutorado) — University of York, 2014.

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

HUGHES, V. *The definition of the relevant population and the collection of data for likelihood ratio-based forensic voice comparison*. Tese (Doutorado) — University of York, 2014.

KINOSHITA, Y. Does lindley's LR estimation formula work for speech data? Investigation using long-term F0. *International Journal of Speech, Language and the Law, 12*, p. 235–254, 2005.

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

LAVER, J. M. D. *Principles of Phonetics*. [S.l.]: Cambridge University Press, 1994.

LINDH, J.; ERIKSSON, A. Robustness of long time measures of fundamental frequency. *INTERSPEECH 2007*, 2007.

MEYER, D. et al. *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*. [S.l.], 2015. R package version 1.6-7. Disponível em: <http://CRAN.R-project.org/package=e1071>.

MORRISON, G. S. *MatLab implementation of Aitken and Lucy's (2004) forensic likelihood ratio software using multivariate-kernel-density estimation*. [S.l.], 2007. Downloaded: November 2015. Disponível em: <http://geoff-morrison.net/#MVKD>.

MORRISON, G. S. Likelihood-ratio voice comparison using parametric representations of the formant trajectories of diphthongs. *Journal of the Acoustical Society of America, 125*, p. 2387–2397, 2009.

MORRISON, G. S. A comparison of procedures for the calculation of forensic likelihood ratios from acoustic-phonetic data: multivariate kernel density (mvkd) versus gaussian mixture model-universal background model (gmm-ubm). *Speech Communication, 53 (2)*, p. 242–256, 2011.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2014. Disponível em: <http://www.R-project.org/>.

ROSE, P. Forensic speaker discrimination with australian english vowel acoustics. In: *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany*. [S.l.: s.n.], 2007. p. 1817–1820.

ROSE, P.; WINTER, E. Traditional forensic voice comparison with female formants: Gaussian mixture model and multivariate likelihood ratio approaches. In: *Proceedings of the 13th Australasian International Conference on Speech Science and Technology. Melbourne, Australia*. [S.l.: s.n.], 2010. p. 42–45.

SILVA, R. R.; DA COSTA, J. P. C. L.; MIRANDA, R. K.; DEL GALDO, G. Aplicação do valor de base da frequência fundamental via estatística mvkd em comparação forense de locutor. *Revista Brasileira de Criminalística, v.5, n. 3*, 2016.

SILVA, R. R.; DA COSTA, J. P. C. L.; MIRANDA, R. K.; DEL GALDO, G. Applying base value of fundamental frequency via the multivariate kernel-density in forensic speaker comparison. *10th International Conference on Signal Processing and Communication Systems*, 2016.

TRAUNMÜLLER, H. Conventional, biological, and environmental factors in speech communication: A modulation theory. *Phonetica, 51*, p. 170–183, 1994.

TRAUNMÜLLER, H.; ERIKSSON, A. The frequency range of the voice fundamental in the speech of male and female adults (unpublished manuscript) downloaded: August 2014 from: http://www2.ling.su.se/staff/hartmut/f0_m&f.pdf. 1994a.

TRAUNMÜLLER, H.; ERIKSSON, A. The perceptual evaluation of F0-excursions in speech as evidenced in liveliness estimations. *J. Acoust. Soc. Am., 97*, p. 1905–1915, 1995.

VALENTE, C. Perspectivas da fonética forense num cenário de quebra do dogma da unicidade. *Anais da Conferência Internacional de Ciências Forenses em Multimídia e Segurança Eletrônica*, p. 7–27, 2012.

WAND, M. *KernSmooth: Functions for kernel smoothing for Wand and Jones 1995*. [S.l.], 2015. R package version 2.23-12. Disponível em: <http://CRAN.R-project.org/package=KernSmooth>.