v. 5, n. 3, p. 30-38, 2016
ISSN 2237-9223

REVISTA BRASILEIRA DE CRIMINALÍSTICA

DOI: http://dx.doi.org/10.15260/rbc.v5i3.134

Aplicação do Valor de Base da Frequência Fundamental Via Estatística MVKD em Comparação Forense de Locutor

R.R. Silva [a,b,*], J.P.C.L.da Costa [b,c,d], R.K. Miranda [b,d], G. Del Galdo [c,d]

[a] *Instituto Nacional de Criminalística, Polícia Federal, Brasília (DF), Brasil*
[b] *Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília (DF), Brasil*
[c] *Fraunhofer Institute for Integrated Circuits IIS, Ilmenau (TH), Alemanha*
[d] *Ilmenau University of Technology, Ilmenau (TH), Alemanha*

**Endereço de e-mail para correspondência: rodrigues.rrs@dpf.gov.br. Tel.: +55-61-2024-9782.*

Recebido em 24/06/2016; Revisado em 03/09/2016; Aceito em 03/09/2016

Resumo

Comparação forense de locutor (CFL) é um exame de determinação de fonte no qual são comparadas amostras de fala de origem conhecida, provenientes de um suspeito, com amostras de fala presentes em um ou mais áudios questionados, cuja autoria se deseja determinar, produzindo uma prova pericial que pode reforçar ou se contrapor à hipótese de que o suspeito é o autor da fala questionada. Em levantamento realizado em 2011 por pesquisadores da Universidade de York entre peritos de treze países, incluindo o Brasil, a metodologia mais usada nos exames de CFL se baseava em análises perceptuais e acústicas. Entre as medidas acústicas, a Frequência Fundamental (F0) era utilizada por quase 70 % dos entrevistados. F0 possui características importantes para a área forense, como a robustez em áudios de baixa qualidade e, em medidas de longo termo, a independência do conteúdo das falas. Conforme estudos recentes, a medida estatística valor de base de F0 é menos afetado pelo estilo e conteúdo da fala, pelo canal utilizado na gravação e exige menor quantidade de áudio para obter uma medida estável, comparada às medidas de F0 comumente utilizadas, entre elas, a média aritmética e o desvio padrão. Neste trabalho é analisado o poder discriminante do valor de base de F0 aplicado ao Corpus Forense do Português Brasileiro (CFPB), obtendo-se uma Taxa de Erro Igual, do inglês *Equal Error Rate* (EER) 5,9 % melhor que o segundo parâmetro mais discriminativo, a média aritmética. A combinação do valor de base de F0 a outras medidas de longo termo de F0 por meio da estatística de densidade do núcleo de multivariáveis, do inglês *Multivariate Kernel-Density* (MVKD), resultou, em todas as simulações, em ganho de poder discriminativo, sendo a combinação entre as medidas do valor de base com a mediana de F0 aquela que obteve o melhor resultado, com uma EER de 13 %, superando valores relatados em pesquisas recentes. Os resultados obtidos demonstram que o valor de base de F0 é o melhor parâmetro a ser utilizado em análises de F0.

Palavras-Chave: F0; Valor de base de F0; MVKD; LTF0; EER.

Abstract

Forensic Speaker Comparison (FSC) is a source identity examination in which speech recordings from a known speaker, the suspect, and from an unknown offender are compared, producing an evaluation of the level of support of the evidence to the same source hypothesis compared to the different source hypothesis. According to a survey of FSC forensic experts from 13 countries conducted by researchers of the University of York, a combination of auditory and acoustic phonetic analysis is the most frequently used method in FSC, and fundamental frequency is a feature measured by almost 70 % of the expert examiners. The statistical parameter base value of F0 has shown, in recent research, to be less affected by the speech style, the content, the recording channel and the speaker effort than other F0 long-term measures, such as arithmetic mean and standard deviation. In this research the discriminating power of the base value of F0, applied to the Brazilian Portuguese Speaker Corpus (CFPB), is investigated. The Equal Error Rate (EER) obtained is 5.9 % better than the second best statistical parameter, the arithmetic mean. The combination of the base value to other statistical measures of F0 using the Multivariate Kernel Density (MVKD), improved the discriminating power in all simulations. The base value combined to the F0 median achieved an Equal Error Rate (EER) of 13 %, outperforming recent researches and indicating that the base value is the more suitable measure for F0 acoustic analysis.

Keywords: F0; Base value of F0; MVKD; LTF0; EER.

1. INTRODUÇÃO

Quando há a necessidade de determinar a autoria de 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 comparadas as falas questionadas, cuja autoria se deseja determinar, com amostras de fala obtidas de um ou mais suspeitos de tê-las produzido.

O exame de CFL enfrenta várias limitações práticas inerentes ao ambiente forense, tais como pequena duração dos áudios questionados, baixa qualidade acústica, limitações espectrais, baixa relação sinal/ruído, presença de reverberação e superposição de sons ambientais. Somam-se a essas limitações, a falta de controle do conteúdo das falas cuja autoria é questionada, limitando sobremaneira a realização de comparações entre segmentos compatíveis de fala presentes no áudio questionado e no padrão de voz do suspeito.

Em pesquisa realizada pela Universidade de York, [1], constatou-se que a metodologia mais adotada entre os peritos da área de CFL consultados é o método combinado, abrangendo análises perceptuais e acústicas. Esta metodologia é a adotada no Brasil, nos Institutos de Criminalística estaduais e na Polícia Federal. Todos os consultados na pesquisa que utilizam acústica nos exames de CFL responderam realizar rotineiramente medidas relacionadas à frequência fundamental (F0), sendo que 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 intervalo de variação dos valores de F0. Destaca-se que apenas 25 % dos consultados citaram utilizar, como medida de longo termo da frequência fundamental (LTF0), o valor de base de F0, embora pesquisas recentes [2,3] indiquem ser ela uma medida estatística mais estável comparada a outros parâmetros LTF0, como a média aritmética e o desvio padrão, que foram as mais citadas.

Portanto, F0 é um dos mais difundidos parâmetros acústicos utilizados em exames de CFL. Conforme Kinoshita *et al.* [4], F0 é um dos parâmetros mais robustos às limitações impostas pelo ambiente forense. Adicionalmente, F0 não requer comparações envolvendo mesmas palavras e fonemas.

F0 é o número de ciclos completos de abertura e fechamento das pregas vocais por segundo, apresentando uma grande variação intrafalante, sendo afetado, entre outros, pelo estilo da fala, pelo esforço vocal e pelo estado emocional. Tais características diminuem o poder discriminativo de F0.

Lindh e Eriksson [2] concluem 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 LTF0. Conforme Traunmüller [5], de acordo com a teoria da modulação, deve existir um valor de base da frequência fundamental, F_b, considerado como uma portadora cuja frequência representa a articulação individual do falante. O valor de F_b é um melhor preditor do valor da F0 intrínseca 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.

Por meio de experimentos, alguns pesquisadores [6,7] concluíram que F_b está localizado próximo do limite inferior do intervalo de produções de F0 do falante, determinado por:

$$F_{\mathrm{b}} = \mu - 1.43\sigma, \quad (1)$$

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

Posteriormente, Lindh e Eriksson [2] propõem uma abordagem alternativa para calcular F_b, minimizando o impacto de valores extremos (*outliers*), que afetam significativamente o valor de σ na fórmula original de cálculo (Eq. 1). Supondo uma distribuição normal para F0, a Eq. 1 implica que F_b pode ser obtido pelo percentil equivalente a 7,64% da distribuição de F0. A metodologia alternativa de cálculo é mais robusta, e, consequentemente, é a melhor escolha em análises forenses e foi adotada aqui.

Os resultados obtidos por Arantes e Eriksson[3], para o português brasileiro, indicam que a quantidade necessária de fala vozeada para que F_b 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. Estes resultados são relevantes no ambiente forense, considerando o fato de serem comuns amostras de voz de reduzida duração.

Multivariate Kernel-Density (MVKD) foi proposto para calcular razões de verossimilhança, do inglês *Likelihood Ratio* (LR) na presença de variáveis correlacionadas [8]. MVKD apresenta bom desempenho em casos onde poucas medidas por variável estão disponíveis.

O uso de MVKD para analisar variáveis relacionadas à linguística e à fonética tem se tornado um procedimento padrão na área de comparação forense de locutores, conforme destaca Morrison [9]. Neste mesmo trabalho, o autor compara o desempenho de MVKD com a abordagem baseada em modelo de misturas gaussianas – modelo universal, do inglês *Gaussian Mixture Model - Universal Background Model* (GMM-

UBM) e obtém melhor acurácia por meio de GMM-UBM na base de dados por ele utilizada. Contudo, MVKD continua sendo amplamente utilizado em pesquisas recentes envolvendo medidas fonético-acústicas tradicionais [10,11] e sua aplicação é direta, sem a necessidade de configuração de vários parâmetros de otimização como ocorre na abordagem GMM-UBM. Considerando o exposto, MVKD foi adotado nesta pesquisa para investigar o poder discriminativo de F_b isoladamente e combinado a outras medidas LTF0.

2. CONCEITOS BÁSICOS DE INFERÊNCIA BAYESIANA

O uso de razões de verossimilhança (LR) tem se tornado comum entre os peritos forenses em exames de determinação de fonte, sendo seu uso recomendado pela *European Network of Forensic Science Institutes* (ENFSI) [12] em todos os exames forenses. 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 probabilidade, I é a informação de contexto, do inglês *background information* e E denota a evidência, a LR é dada por

$$\mathrm{LR} = \frac{P(E|H_a, \mathrm{I})}{P(E|H_d, \mathrm{I})}. \quad (2)$$

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 desse teorema 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 [13]. A aplicação teórica do teorema de Bayes é chamada de inferência bayesiana e, na área criminal, é expressa como segue:

$$\frac{P(H_a|\mathrm{I})}{P(H_d|\mathrm{I})} \cdot \frac{P(E|H_a, \mathrm{I})}{P(E|H_d, \mathrm{I})} = \frac{P(H_a|E, \mathrm{I})}{P(H_d|E, \mathrm{I})}, \quad (3)$$

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 Eq. 3 corresponde à razão de probabilidade a priori, antes da consideração da evidência E, o segundo termo exprime o peso da evidência E e o último termo corresponde à razão de probabilidade a priori combinada com o peso da evidência E, e é chamada de razão 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.

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, dependentes, são combinados e a abordagem adotada foi calcular a LR utilizando a função MVKD proposta por Aitken e Lucy [8].

3. ABORDAGEM PROPOSTA

O objetivo da abordagem aqui proposta é investigar o poder discriminante de F_b, encontrar a melhor combinação de parâmetros LTF0 utilizando a função MVKD e analisar qual combinação apresenta melhor acurácia.

Para avaliar o poder discriminativo dos parâmetros LTF0, na presente pesquisa, vários confrontos envolvendo amostras de voz de mesmo falante e de diferentes falantes são realizados e obtidas suas respectivas LRs. Dados os resultados obtidos dos confrontos e um limiar de decisão δ, as taxas de falsos positivos, em inglês *False Acceptance Rate* (FAR), e de falsos negativos, em inglês *False Rejection Rate* (FRR), são computadas. A FAR é definida como a taxa entre a quantidade de confrontos que são erroneamente estimados como provenientes do falante suspeito e o total de confrontos efetivamente envolvendo diferentes falantes. Enquanto a FRR é a taxa entre a quantidade de confrontos em que o falante suspeito é rejeitado apesar de as falas terem sido produzidas por ele e o total de confrontos envolvendo mesmos falantes. Variando δ obtém-se um ponto de ajuste em que FAR = FRR, chamado de *Equal Error Rate* (EER), comumente utilizado na avaliação do desempenho discriminativo de sistemas automáticos de comparação de locutor.

Além da determinação da EER, a comparação do poder discriminativo de diferentes parâmetros é comumente realizada pela avaliação de curvas DET (*Detection Error Tradeoff*) correspondentes aos valores de FAR e FRR em função do valor do limiar δ, permitindo uma visualização conjunta do poder

discriminativo de cada parâmetro em função do valor do limiar escolhido.

Um bom parâmetro discriminante, na área de comparação forense de locutor, deve apresentar grandes valores positivos de $\log_{10}$(LR) em comparações envolvendo mesmos falantes e grandes valores negativos de $\log_{10}$(LR) em comparações envolvendo falantes diferentes, dando maior suporte às evidências e, possuindo, portanto, maior acurácia ou exatidão.

A fim de determinar qual combinação de LTF0 possui maior acurácia, foi utilizada a abordagem proposta por Brümmere Preez [14], adotada, entre outros, por Morrison [15], calculando o *log-likelihood-ratio cost* (C_{llr}) de cada uma das combinações de LTF0.

Inicialmente, cada um dos R áudios usados na pesquisa são divididos 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 x n.

Os contornos de F0 dos R vestígios e dos R padrões são extraídos, pela aplicação do método de autocorrelação [16], em seções de t_s segundos para s = 1,..., S.

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. 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.

A seguir, cada um dos contornos extraídos é processado para estimar os oito parâmetros LTF0 escolhidos nesta pesquisa, por meio do software R® (https://cran.r-project.org/src/base/R-3/), $\text{LTF0}_{k\,=\,1,\ldots,\ 8}$, nomeados conforme Tab. 1.

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

Por meio da função MVKD [8], usando como entrada os vetores obtidos, computa-se o elemento $m_{\{K,r_1,r_2,s\}}$ correspondendo à LR obtida pela comparação dos falantes r_1 (vestígio) e r_2 (padrão) utilizando os LTF0 selecionados, representados pelo conjunto $K = \{\text{LTF0}_{k_1}, \text{LTF0}_{k_2}, ..., \text{LTF0}_{k_N}\}$, onde $k_1, k_2, \ldots k_N$ são os índices dos LTF0 escolhidos conforme Tab. 1, para uma determinada seções e N variando de 1 a 8.

Tabela 1. Nomenclatura utilizada nos parâmetros LTF0.

LTF0	Código	Símbolo
Média Aritmética	LTF0_1	$\hat{\mu}$
Mediana	LTF0_2	$\hat{Q}_2$
Desvio Padrão	LTF0_3	$\hat{\sigma}$
Valor de base de F0	LTF0_4	$\hat{F}_b$
Curtose	LTF0_5	$\hat{\omega}$
Assimetria (Skewness)	LTF0_6	$\hat{\eta}$
Moda	LTF0_7	$\hat{\psi}$
Densidade modal	LTF0_8	$\hat{\gamma}$

São realizadas R x R comparações, sendo R envolvendo mesmo falante e $R(R\text{-}1)$ envolvendo comparações entre falantes distintos para cada uma das S seções escolhidas.

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 falantes sendo comparados.

Os valores das R x 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 contêm LRs envolvendo comparações entre diferentes falantes.

Então, os valores de FAR e FRR para o conjunto K de parâmetros LTF0 são computados, para uma determinada seção s, como segue:

$$\text{FAR}(K,\delta,s) = \frac{1}{2\cdot R(R-1)}\sum_{r_1=1}^{R}\sum_{\substack{r_1=1\\ r_2\neq r_1}}^{R}\left[\text{sign}\left(m_{K,r_1,r_2,s} - \delta\right) + 0.5\right], \quad (4)$$

$$\text{FRR}(K,\delta,s) = \frac{1}{2\cdot R}\sum_{r_1=r_2=1}^{R}\left[\text{sign}\left(\delta - m_{K,r_1,r_2,s}\right) + 0.5\right], \quad (5)$$

para o vestígio r_1, para o padrão de suspeito r_2, o limiar de decisão δ e o valor da LR $m_{\{K,r1,r2,s\}}$ obtida em cada comparação, sendo sign(x) a função que retorna o valor +1 para $x > 0$ e -1 para $x \leq 0$.

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

$$\text{EER}(K,s) = \min_{\delta}\left|\text{FAR}(K,\delta,s) - \text{FRR}(K,\delta,s)\right|. \quad (6)$$

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

$$C_{llr}(K,s) = \frac{1}{2}\left(\frac{1}{R}\sum_{i=1}^{R}\log_2\left[1+\frac{1}{\mathrm{LR}_{ss}}\right]+\frac{1}{R(R-1))}\sum_{j=1}^{R(R-1)}\log_2[1+\mathrm{LR}_{ds}]\right) \quad (7)$$

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

Conforme Eq. 7, 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 somatório da Eq. 7 penaliza os valores de LR_{ss} que, nas comparações entre mesmos falantes, são baixos: quanto mais próximas de zero essas LR_{ss}, maior o somatório obtido. Já o segundo somatório penaliza de igual forma, com direção inversa, as LR_{ds}, de forma que o C_{llr}, que é sempre um valor positivo, será tanto melhor quanto mais próximo de zero.

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 acurácia.

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

4. VALIDAÇÃO EXPERIMENTAL

Esta seção é dividida em quatro subseções. Na Subseção 4.1, apresentam-se informações sobre os áudios utilizados na validação da abordagem proposta. Na Subseção 4.2, o poder discriminativo dos parâmetros LTF0 analisados individualmente são apresentados. Na Subseção 4.3, comparam-se os resultados obtidos a artigos recentes. Na Subseção 4.4, investiga-se a EER e o C_{llr} obtido ao combinar diferentes LTF0 usando 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). Este corpus consiste de 206 gravações de falantes masculinos e 50 femininos, incluindo falas semi-espontâneas (entrevistas) e leitura de sentenças, compreendendo amostras de falantes de todas as regiões do país. Cada gravação tem aproximadamente cinco minutos líquidos de falas semi-espontâneas e um minuto de leitura de sentenças que visam contemplar todos os sons do português brasileiro.

O presente trabalho usa as 206 gravações semi-espontâneas masculinas do CFPB, uma vez que mais de 90% das CFLs no âmbito da Polícia Federal envolvem somente falantes masculinos.

A escolha de trabalhar com o corpus CFPB, nesta pesquisa, foi o fato de conter uma grande quantidade de amostras de falas em português brasileiro obtidas de forma homogênea, 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), placas de captura da mesma marca (Edirol® UA25EX e UA25) e o mesmo software de captura (Adobe Audition® 3.0). Como o objetivo da pesquisa é comparar o poder discriminante dos parâmetros LTF0, é interessante que as outras variáveis que podem influir nos valores de F0 sejam mantidas constantes, entre elas, o canal, a duração das amostras e o estilo de fala.

Comumente, nas aplicações forenses, as gravações são obtidas a partir de chamadas telefônicas que têm uma banda de passagem de aproximadamente 4 kHz. Assim, cada uma das 206 gravações selecionadas do CFPB foi re-amostrada de 22,050 kHz, que é taxa original de amostragem do corpus, para 8 kHz e dividida em duas partes de 1 minuto de falas líquidas, após removidos todos os trechos de silêncio e produções não vozeadas, a primeira parte utilizada como vestígio e a segunda parte como padrão.

O software Praat® [17] foi usado para extrair os contornos de F0 dos 206 vestígios e dos 206 padrões. Para essa extração, foi utilizado o script better_f0-2012-03.16.praat, v. 1.3, provido por Pablo Arantes e disponibilizado em https://code.google.com/archive/p/praat-tools/downloads. Note que este script minimiza os erros de extração de F0 escolhendo os melhores valores mínimos e máximos de F0.

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 seguir, cada um dos contornos de F0 extraídos e corrigidos foram divididos em seções t_s=5, 10, 15, 20 e 30 segundos para s= 1,...,5.

Cada uma das seções foi processada utilizando o software R® para extrair os oito parâmetros LTF0 constantes da Tab. 1: $\mathrm{LTF0}_{1,\ \ldots,6}$ usando as bibliotecas "E1071" [18] e "Stats" [19]; e $\mathrm{LTF0}_7$ e $\mathrm{LTF0}_8$ foram extraídos usando a biblioteca "Kern Smooth" [20] do mesmo modo descrito por Kinoshita *et al.*[4].

Cada um dos parâmetros LTF0 resultaram em cinco vetores contendo z_s medidas de LTF0 cada, onde $z_s = 60/t_s$ e t_s é o tamanho da seção.

4.2. Poder discriminativo dos parâmetros LTF0 analisados individualmente

Inicialmente foi avaliado o poder discriminativo de cada um dos oito LTF0 analisados neste trabalho. Para tal, cada um dos cinco vetores por LTF0 (correspondendo a *s*=1, ..., 5) e por falante foram processados para calcular LRs usando MVKD, implementado em Matlab® [21].

A Tab. 2 sumariza as EERs obtidas em todos os testes envolvendo parâmetros LTF0 isoladamente.

Tabela 2.EERs dos parâmetros LTF0.

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

A EER obtida pelo valor de base de F0 ($\mathrm{LTF0_4}$) superou o resultado de todos os outros LTF0 e obteve uma EER média de 16,1%, o que é 5,9% melhor que o segundo melhor parâmetro LTF0 analisado, a média aritmética ($\mathrm{LTF0_1}$).

De acordo com a Tab. 2, 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, obtendo a Fig. 1. Comparando as curvas DET obtidas, identifica-se claramente o melhor desempenho de F_b, que possui a menor área sob seu traçado.

4.3. Aplicando abordagens de pesquisas recentes ao corpus CFPB

As abordagens adotadas na literatura [4,11] foram utilizadas como referência para avaliar a performance usando o corpus CFPB.

Gold [11] estuda o poder discriminativo da combinação dos parâmetros média ($\mathrm{LTF0_1}$) e desvio padrão ($\mathrm{LTF0_3}$) usando MVKD para calcular LRs envolvendo 100 gravações de falantes ingleses masculinos com uma média de 6 minutos de duração cada (50 deles usados como população de referência).

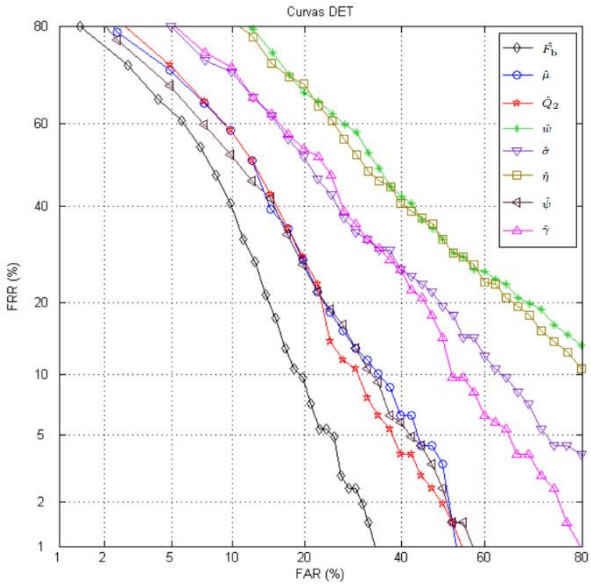

Figura 1. Curvas DET dos parâmetros LTF0 analisados isoladamente. O parâmetro F_b obteve o melhor desempenho discriminativo.

Usando a mesma metodologia, combinando os mesmos parâmetros nas 206 gravações semi-espontâneas de falantes masculinos do CFPB, obteve-se uma EER de 17,2% (penúltima linha da Tab. 3). Ao analisar a Tab. 2, percebe-se que o poder discriminativo do valor de base de F0 ($\mathrm{LTF0_4}$) isoladamente apresenta uma EER média de 16,1%, sendo, portanto, mais discriminante que o uso da média combinada ao desvio padrão.

Tabela 3.EERs dos parâmetros LTF0 combinados.

	EER (%)					
$\mathrm{LTF0_{K=1,\ldots,8}}$	**s = 1**	**s = 2**	**s = 3**	**s = 4**	**s = 5**	**Média**
$\mathrm{LTF0_{2,4}}$	14,9	14,2	13,0	14,6	13,5	14,0
$\mathrm{LTF0_{2,4,7}}$	14,9	14,1	13,1	14,5	13,7	14,1
$\mathrm{LTF0_{1,2,4}}$	15,4	14,5	13,5	14,0	13,6	14,2
$\mathrm{LTF0_{1,2,4,7}}$	14,6	15,0	13,2	14,9	14,5	14,4
$\mathrm{LTF0_{1,4}}$	15,5	14,2	15,5	15,5	14,6	15,1
$\mathrm{LTF0_{1,4,7}}$	15,0	15,0	13,5	14,9	13,5	15,1
$\mathrm{LTF0_{4,7}}$	15,0	14,6	14,6	16,1	15,5	15,1
$\mathrm{LTF0_{1,3,5,6,7,8}}$ ([4])	15,4	15,0	14,6	14,5	14,4	14,8
$\mathrm{LTF0_{1,3,4,5,6,7,8}}$ ([4]+F_b)	14,0	14,2	15,1	14,5	15,1	14,6
$\mathrm{LTF0_{1,3}}$([11])	17,4	17,3	17,5	17,0	17,0	17,2
$\mathrm{LTF0_{1,2,3,4,5,6,7,8}}$	15,1	15,0	15,4	15,0	15,1	15,1

Kinoshita e colaboradores [4] analisam o poder discriminativo de F0 usando 201 gravações de falantes japoneses masculinos de 10 a 25 minutos de duração

cada. Usando MVKD, foram combinados os 6 parâmetros LTF0: média, desvio padrão, moda, densidade modal, assimetria e curtose. Reproduzindo a mesma combinação de parâmetros usando o CFPB resultou em uma EER de 14,8 % (Tab. 3), melhor que o uso do valor de base de F0 ($LTF0_4$) isoladamente (Tab. 2).

A fim de verificar se a inclusão do parâmetro $LTF0_4$ aos seis parâmetros LTF0 utilizados por Kinoshita *et al.* [4] melhorava o poder discriminativo, este foi, então, incluído e, como esperado, houve uma queda na EER média que ficou em 14,6% (Tab. 3), melhorando o poder discriminativo.

4.4. Abordagem proposta usando a melhor combinação de LTF0

Considerando a melhor performance obtida ao incluir F_b à abordagem utilizada por Kinoshita *et al.* [4], passou-se a investigar qual combinação entre os 8 parâmetros LTF0 analisados resultaria em menor EER.

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 *et al.* [4], conforme a última linha da Tab. 3.

Analisando a Tab. 2, além do valor de base de F0 ($LTF0_4$), somente a média aritmética ($LTF0_1$), a mediana ($LTF0_2$) e a moda ($LTF0_7$) obtiveram EERs abaixo de 30% no CFPB. Ademais, os mesmos parâmetros se destacaram nas curvas DET da Fig. 1. Estas quatro LTF0 foram selecionadas para os testes e todas as possíveis combinações envolvendo o valor de base de F0 e os outros três parâmetros foram feitas utilizando MVKD em seções t_s = 5, 10, 15, 20 e 30 segundos.

Como mostrado na Tab. 3, usando o mesmo corpus CFPB e MVKD, o valor de base de F0 ($LTF0_4$) combinado com a mediana ($LTF0_2$) superou a performance de todos os outros parâmetros com uma EER de 13% usando s = 3, ou seja, trechos de 15 s.

Adicionalmente foram plotadas as curvas DET das combinações envolvendo os quatro melhores parâmetros LTF0 na Fig. 2 e constatou-se que nenhuma das combinações se destacou em relação às outras, conforme a Tab. 3 já indicava. Sendo que as curvas do valor de base de F0 combinado com a mediana ($LTF0_{2,4}$) e do valor de base combinado com a mediana e a moda ($LTF0_{2,4,7}$) apresentaram uma performance ligeiramente superior, primeiras linhas da Tab. 3.

Cabe destacar o pior desempenho da combinação da média aritmética com o desvio padrão ($LTF0_{1,3}$), justamente envolvendo os parâmetros LTF0 mais citados pelos peritos que responderam à pesquisa [1], correspondendo à curva mais afastada da origem na Fig. 2.

Então, foi calculado o valor do C_{llr}, conforme Eq. 7, de cada uma das combinações envolvendo os quatro melhores LTF0, média aritmética, mediana, valor de base e moda de F0 a fim de determinar aquela que apresenta melhor desempenho.

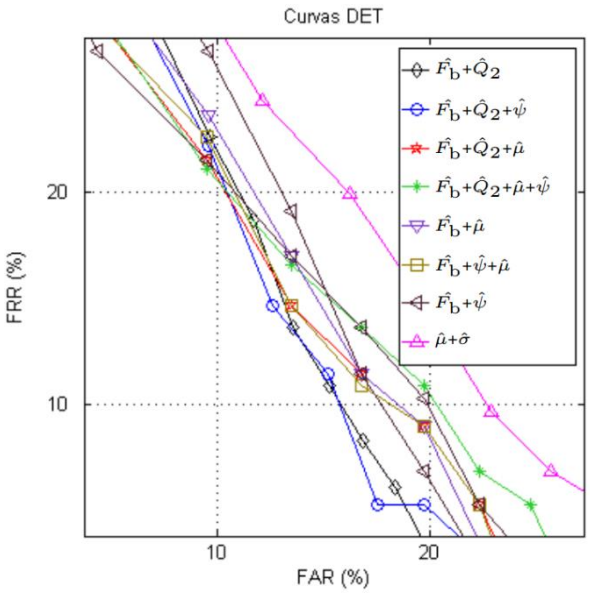

Figura 2. Curvas DET dos parâmetros LTF0 combinados.

A Tab. 4 apresenta os C_{llr} obtidos nas combinações de LTF0. Constata-se que o menor C_{llr} = 0,618 resulta da combinação entre o valor de base e a mediana de F0, reforçando os achados anteriormente obtidos na avaliação das EERs e das curvas DET.

Tabela 4. C_{llr} das combinações propostas de parâmetros LFT0 aplicadas ao CFPB.

	C_{llr}					
$LTF0_{K=1,\ldots,8}$	**s = 1**	**s = 2**	**s = 3**	**s = 4**	**s = 5**	**Média**
$LTF0_{2,4}$	0,685	0,621	0,618	0,642	0,849	0,683
$LTF0_{2,4,7}$	0,950	0,952	0,743	0,952	1,113	0,940
$LTF0_{1,2,4}$	0,950	0,952	0,920	0,953	1,113	0,978
$LTF0_{1,2,4,7}$	0,916	0,900	1,151	0,810	0,984	0,952
$LTF0_{1,4}$	0,849	0,833	0,819	0,846	0,897	0,849
$LTF0_{1,4,7}$	0,810	0,766	0,863	0,736	0,804	0,796
$LTF0_{4,7}$	0,663	0,790	1,167	0,871	1,189	0,936
$LTF0_{1,3}$ ([11])	0,773	0,722	0,875	0,822	0,995	0,798

5. CONCLUSÕES

O poder discriminante dos parâmetros de longo termo da frequência fundamental (LTF0) foram avaliados usando um subgrupo do Corpus Forense do

Português Brasileiro (CFPB) contendo 206 gravações de falas semi-espontâneas masculinas divididas em duas partes contendo, cada uma, um minuto de produções vozeadas (vestígio e 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.

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 (13%) foi obtida combinando o valor de base com a mediana de F0 usando seções de 15 segundos (Tab. 3).

Por fim, foi avaliado o desempenho das combinações de LTF0 a fim de determinar aquela que resultava em maior acurácia por meio do 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, a medida acústica valor de base de F0, F_b, deve ser usada preferencialmente às outras medidas de longo termo de F0 comumente utilizadas, entre elas, a média e o desvio padrão que foram as mais citadas na pesquisa [1].

Constatou-se ainda que o uso do valor de base de F0 combinado com a mediana das medidas de F0 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.

AGRADECIMENTOS

Os autores agradecem as agências brasileiras de pesquisa e inovação FAPDF (Fundação de Apoio à Pesquisa do Distrito Federal), FINEP (Acordo RENASIC / PROTO 01.12.0555.00), CAPES e CNPq sob o projeto FORTE - CAPES Programa Ciências Forenses (Pró-Forense) 25/2014 e o programa ciências sem fronteiras - Tecnologia Aeroespacial apoiada pelo CNPq, CAPES bolsa de Pós-Doutorado no Exterior (PDE) número 207644/2015-2 e CAPES doutorado sanduíche número 88887.115692/2016-00.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] E. Gold, J.P. French. International practices in forensic speaker comparison. *Int. J. Speech Lang. La.***18(2)**, 293-307, 2011.

[2] J. Lindh, A. Eriksson. Robustness of long time measures of fundamental frequency. *Interspeech* **2007**, 2025-2028, 2007.

[3] P. Arantes, A. Eriksson. Temporal stability of long-term measures of fundamental frequency. *Anais da 7ª Conferência International Conference on Speech Prosody* 1149-1152, 2014.

[4] Y. Kinoshita, S.Ishihara, P. Rose. Exploring the discriminatory potential of F0 distribution parameters in traditional forensic speaker recognition. *Int. J. Speech, Lang. La.* **16**, 91-111, 2009.

[5] H. Traunmüller. Conventional, biological, and environmental factors in speech communication: A modulation theory. *Phonetica* **51**, 170-183, 1994.

[6] H. Traunmüller, A. Eriksson. The Frequency Range of the Voice Fundamental in the Speech of Male and Female Adults. *Unpublished manuscript*. Disponível em: http://www2.ling.su.se/staff/hartmut/f0_m%26f.pdf. Acesso em: Agosto de 2014.

[7] H. Traunmüller, A. Eriksson. The perceptual evaluation of F0-excursions in speech as evidenced in live lines estimations. *J. Acoust. Soc. Am.* **97**, 1905-1915, 1995.

[8] C.G.G. Aitken, D. Lucy. Evaluation of trace evidence in the form of multivariate data. *J. Royal Stat. Soc.* **53(1)**, 109-122, 2004.

[9] G.S. Morrison. 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 Commun*. **53(2)**, 242-256, 2011.

[10] V. Hughes. *The definition of the relevant population and the collection of data for likelihood ratio – based forensic voice comparison*. Tese de Doutorado, University of York, 2014.

[11] E. Gold. *Calculating likelihood ratios in forensic speaker comparison cases using phonetic and linguistic features*. Tese de Doutorado, University of York, 2014.

[12] ENFSI. *ENFSI guideline for evaluative reporting in forensic science*. Disponível em http://www.enfsi.eu/documents/external-publications. Acesso em: 05/04/2016.

[13] C.G.G. Aitken, F. Taroni. *Statistics and the Evaluation of Evidence for Forensic Scientists*. Chichester, Wiley, segunda edição: capítulo 1, 2004.

[14] N. Brümmer, J. du Preez. Application independente valuation of speaker detection. *Comp. Speech Lang.* **20**, 230-275, 2006.

[15] G.S Morrison. Likelihood –ratio voice comparison using parametric representations of the formant trajectories of diphthongs. *J. Acoust. Soc. Am*. **125**, 2387-2397, 2009.

[16] P. Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Anais da conferência IFA*. **17**, 97-110, 1993.

[17] P. Boersma, D. Weenink. *Praat: doing phonetics by computer*. Programa de computador. Versão 5.3.70, Retirado em 05/04/2014, de: http://www.praat.org/.

[18] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, F. Leisch. e1071: *Misc Functions of the Department of Statistics, Probability Theory Group* (Formerly: E1071), TU Wien. R package version 1.6-7. Retirado em 12/12/2015 de http://CRAN.R-project.org/package=e1071, 2015.

[19] R. Core Team. *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Retirado em 05/03/2015 de http://www.R-project.org/, 2014.

[20] M. Wand. *Kern Smooth: Functions for Kernel smoothing for Wandand Jones 1995*. R package version 2.23-12. Retirado em 12/12/2015 de http://CRAN.R-project.org/package=KernSmooth, 2015.

[21] G.S. Morrison. *Mat Lab implementation of Aitken and Lucy's (2004) forensic likelihood ratio software using multivariate-kernel-density estimation, 2007*. Retirado em 01/11/2015 de http://geoff-morrison.net/#MVKD, 2007.