Classificação de Desvios Vocais baseadas em Características do Modelo Linear de Produção da Fala

Maria de F. K. B. Couras, Pablo H. U. de Pinho, Silvana L. do N. C. Costa, Suzete E. N. Correia

***Resumo*—Neste trabalho, são empregadas técnicas de processamento digital de sinais, baseadas no modelo linear de produção da fala, para analisar a qualidade vocal. É avaliado o potencial discriminativo dos parâmetros obtidos da análise de predição linear (*Linear Predictive Coding* - LPC) - coeficientes LPC, coeficientes cepstrais e mel-cepstrais na classificação de desvios vocais (rugosidade, soprosidade e tensão). Com o auxílio da curva ROC (*Receiver Operator Characteristic Curve*), é realizada a classificação dos sinais, obtendo-se a melhor acurácia média de 86% na discriminação entre vozes tensas e soprosas, com os parâmetros mel-cepstrais, quantizados em 512 níveis.**

Palavras-Chave— Modelo Linear de Produção da Fala, Desvios vocais, Quantização Vetorial.

Abstract— **In this work, digital signal processing techniques, based on the linear model of speech production, are used to analyze vocal quality. The discriminatory potential of the parameters obtained from Linear Predictive Coding (LPC) - LPC coefficients, cepstral coefficients and mel-cepstral in the classification of vocal deviations (roughness, breathiness and tension) are evaluated. With the aid of the ROC (Receiver Operator Characteristic Curve) curve, the classification of the signals is performed, obtaining 86% of average accuracy in the discrimination between voices tension and breathiness with the parameters mel-cepstrais, quantized in 512 levels.**

Keywords— Linear Model of Speech Production, Voice deviations, Vector Quatization.

I. INTRODUÇÃO

A qualidade vocal é o termo empregado para designar o conjunto de características que identificam uma voz [1]. A avaliação da qualidade vocal possui duas formas de análise que se complementam: a avaliação perceptivo-auditiva e a análise acústica. A avaliação perceptivo-auditiva é realizada por um profissional treinado para ouvir e identificar características presentes no sinal de voz que indiquem se há alterações ou não na qualidade vocal [2]. A análise acústica emprega técnicas de processamento digital de sinais para extrair características representativas do sinal de voz que podem representar a presença de um desvio da voz. Além de ser um método não invasivo de avaliação [3, 4].

Diversas pesquisas têm sido realizadas nos últimos anos com o intuito de investigar medidas que possam avaliar a qualidade vocal de forma objetiva, aplicando a análise acústica. Tradicionalmente, medidas como frequência fundamental (F_0), o *jitter*, o *shimmer*, as medidas de ruído glótico (*Glottal to Noise Excitation ratio* (GNE), *Harmonics-to-Noise Ratio* (HNR), *Normalized Noise Energy* (NNE) são empregadas na análise clínica com relativo sucesso [5-7], detectando e quantificando o grau de desvios vocais. No entanto, medidas que dependam da frequência fundamental, podem não ser muito eficientes quando os desvios vocais apresentam grau elevado ou são decorrentes de patologias laríngeas que inserem alto nível de ruído glótico nos sinais analisados [3], [8].

Algumas medidas que não dependem da frequência fundamental, como os coeficientes obtidos da análise preditiva linear (LPC – *Linear Predictive Coding*), os coeficientes cepstrais e mel-cepstrais têm sido empregadas como parâmetros de classificação, em sistemas de reconhecimento de fala, sistemas de identificação e verificação de locutor, como também na discriminação entre sinais de vozes afetados por patologias laríngeas [3], [7], [9, 10].

Costa et al. utilizaram técnicas de processamento de sinais de voz baseadas no modelo linear de produção da fala aliadas às características da análise não-linear para caracterizar sinais de voz de laringes patológicas e saudáveis. Os autores observaram que os métodos empregados apresentaram resultados promissores, mas ainda é necessário definir, com exatidão, quais as melhores características para cada patologia considerada. Pode ser verificado que as pesquisas têm apresentado a análise acústica como uma forma eficaz, segura e não invasiva sendo importante no auxílio ao diagnóstico médico e acompanhamento de tratamento pré e pós-cirúrgicos de patologias laríngeas [11].

Lopes et al. realizaram um estudo à respeito da utilização das medidas tradicionais (frequência fundamental, formantes, *jitter*, *shimmer* e GNE) na classificação dos desvios vocais. Foram utilizados 302 sinais de voz da base de dados utilizada na presente pesquisa. Todos os sinais são femininos e da vogal /Ɛ/ (“é”) sustentada. A classificação foi realizada através do classificador análise discriminante quadrático utilizando a média e o desvio padrão das características extraídas. Os resultados obtidos demonstram que a medida acústica GNE, mostrou-se a única capaz de discriminar a intensidade do desvio vocal e a qualidade vocal predominante, com acurácia de aproximadamente 71% e houve um ganho no desempenho da classificação com a combinação das medidas acústicas tradicionais e formânticas atingindo acurácia de 84% [12].

A realização de um estudo a respeito da classificação de desvios vocais utilizando parâmetros do modelo linear, baseadas na análise LPC, Cepstral e Mel-cepstral ainda é um campo em aberto. Sua análise é simples computacionalmente e por esses parâmetros serem bastante empregados em outras áreas [3], [9, 10], os tornam potencialmente relevantes para serem investigados na caracterização de desvios vocais. Neste trabalho é realizada uma avaliação da qualidade vocal por meio de medidas a partir da análise linear do modelo de produção vocal, da análise cepstral e mel-cepstral para a detecção de desvios vocais. A avaliação dos sinais é realizada através da curva ROC (*Receiver Operator Characteristic Curve*).

O artigo está organizado da seguinte forma: Na Secção II são descritas as análises LPC, Cepstral e Mel-cepstral. Na Seção III é apresentada a descrição da base de dados e da metodologia empregada. Na Seção IV são apresentados os

Couras, M. F. K. B., Pinho, P. H. U., Costa, S. C. e Correia, S.E.N, Programa de Pós-graduação em Engenharia Elétrica, Instituto Federal da Paraíba, João Pessoa-PB, Brasil, E-mails: kallynna.mary@gmail.com,pablohenriqueifpb@gmail.com, silvana@ifpb.edu.br e suzete@ifpb.edu.br. Este trabalho foi parcialmente financiado pela COPEX e PRPIPG - IFPB.

resultados obtidos e discussão, seguida da Seção V, em que são apresentadas as conclusões.

II. ANÁLISE LINEAR DE PRODUÇÃO VOCAL

O modelo linear de produção da fala é muito útil para entender a relação entre aspectos articulatórios e acústicos da fala. Esse modelo foi construído baseado nas características do modelo vocal humano, em que a fonte de excitação e o aparelho vocal são considerados como dois sistemas separados [13].

A. *Análise por Predição Linear (LPC)*

A análise de voz por predição linear (análise LPC) fornece um conjunto de parâmetros da fala que representam o trato vocal. Este método estima cada amostra de voz baseando-se em uma combinação linear de p amostras anteriores. Quanto maior o valor de p, mais preciso é o modelo. Na Equação (1) é definida a saída do preditor linear com coeficientes de predição, $a(k)$ [3], [14].

$$\tilde{s}(n) = \sum_{k=1}^{p} a(k)s(n-k), \quad (1)$$

em que p é a ordem do preditor e k o atraso das amostras.

Os dois métodos padrões para o cálculo dos coeficientes do preditor são o da autocorrelação e o da covariância. Ambos são baseados na minimização do erro $e(n)$, ou sinal residual, como dado pela Equação (2) [14, 15].

$$e(n) = s(n) - \sum_{k=1}^{p} a(k)s(n-k) \quad (2)$$

Nesta pesquisa, para calcular os coeficientes LPC é empregado o método da autocorrelação. Os coeficientes do filtro são determinados pela minimização da energia do erro de predição, E (Equação (3)), entre o sinal $s(n)$ e o sinal estimado ($\tilde{s}(n)$) (Equação (1)) onde $e(n)$ é o erro do preditor estimado na Equação (2) [14].

$$E = \sum_n e(n)^2 \quad (3)$$

Realizando a minimização da energia do erro através da derivada de E em relação a $\alpha(k)$ obtém-se um sistema de p equações com p incógnitas, que pode ser representado por uma matriz, expressa na Equação (4).

$$\begin{bmatrix} R_s[1] \\ R_s[2] \\ \dots \\ R_s[p] \end{bmatrix} = \begin{bmatrix} R_s[0] & R_s[1] & \dots & R_s[p-1] \\ R_s[1] & R_s[0] & \dots & R_s[p-2] \\ \dots & \dots & \dots & \dots \\ R_s[p-1] & R_s[p-2] & \dots & R_s[0] \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \dots \\ a(k) \end{bmatrix} \quad (4)$$

em que R é a matriz de autocorrelação do sinal degradado e $a(k)$ os coeficientes de predição linear.

B. *Análise Cepstral*

Para realizar o estudo das alterações laríngeas, a análise cepstral de sinais de voz pode ser útil, pois permite trabalhar com o sinal da excitação da glote separadamente das repercussões ressonantes do trato vocal [9], [16].

Os coeficientes cepstrais representam as condições da fonte (laringe) e do filtro (o trato vocal), separadamente. O sinal de voz (Equação (5)) é resultado da convolução da excitação, $ex(n)$ com a resposta do trato vocal, $v(n)$. Então seria útil separar ou deconvoluir as duas componentes [16].

$$s(n) = ex(n) * v(n) \quad (5)$$

A deconvolução cepstral converte um produto de dois espectros na soma de dois sinais, separando-os por um processo de filtragem linear, *liftering* facilitando o estudo individual das modificações ocorridas na fonte e no filtro. Das propriedades matemáticas envolvidas nesta operação destacam-se a FFT (*Fast Fourier Transform*) e as funções logarítmicas, que resultam em uma função chamada cepstral ou cepstro, responsável pela dissociação do sinal de voz. A transformação desejada é logarítmica, dada pela Equação (6):

$$\log\big(Ex(\omega).V(\omega)\big) = \log(Ex(\omega)) + \log(V(\omega)) \quad (6)$$

$Ex(\omega)$ e $V(\omega)$: transformadas de Fourier da forma de onda da excitação e da resposta do trato vocal, respectivamente.

A deconvolução resulta em uma função chamada cepstro, responsável pela separação da fonte e do filtro. Na prática, o *cepstrum* complexo não é necessário, sendo suficiente o *cepstrum* real, definido como a transformada inversa do logaritmo do espectro de magnitude (Equação (7)) [3], [16].

$$c(n) = \frac{1}{2\pi}\int_0^{2\pi} \log\left|X\left(e^{j\omega}\right)\right| e^{j\omega} d\omega \quad (7)$$

Para sinais reais $x(n)$, $c(n)$ é a parte par do cepstro $\hat{x}(n)$.

C. *Análise Mel-cepstral*

Os coeficientes mel-cepstrais (*Mel-frequency Cepstral Coefficients* – MFCC) surgiram devido a estudos que mostraram que a percepção humana das frequências de tons puros ou de sinais de voz não segue uma escala linear. Para cada tom com frequência f, medida em Hz, define-se um tom subjetivo medido em uma escala chamada escala mel. O mel é uma unidade de medida da frequência percebida de um tom [3]. As frequências percebidas Mel podem ser encontras a partir da frequência linear em Hz da seguinte maneira (Equação (8)) [3], [15]:

$$F_{mel} = 2595.\log_{10}\left(1 + \frac{F_{linear}(Hz)}{700}\right) \quad (8)$$

em que F_{linear} é a frequência em Hz e F_{mel} é a frequência percebida (em mel).

Para realizar o cálculo dos coeficientes mel-ceptrais primeiramente é obtido o módulo ao quadrado da transformada de Fourier ($|FFT(x(n))|^2$) do sinal, $x(n)$ para cada segmento do sinal, quando processado a curto intervalo de tempo. Posteriormente. é aplicado um banco de filtros em escala mel com o formato triangular não separados linearmente. A quantidade de filtros, Nf, é determinada por uma relação com a frequência de amostragem, Fa, sendo $Nf = ((Fa)*(3ln(Fa)))$. Em seguida, é feito o cálculo do logaritmo da energia de saída de cada filtro para a obtenção do cepstro e, por fim, é realizada a obtenção dos coeficientes mel-cepstrais $c_{mel}(n)$. De maneira simplificada, os coeficientes podem ser determinados através da Equação (9) [3], [15].

$$c_{mel}(n) = \sum_{k=1}^{N_f} \log\big(S_{FFT}(k)\big).\cos\left[n\left(k-\frac{1}{2}\right)\right].\frac{\pi}{Nf}, \quad (9)$$

N_f: número de filtros digitais; $c_{mel}(n)$: n-coeficientes mel-cepstrais; S_{FFT} (k): sinal de saída do bando de filtros digitais que é obtido através da Equação (10).

$$S_{FFT}(k) = \sum_{j=1}^{NFFT} W(j).X(j), \quad k = 1, \dots, Nf, \quad (10)$$

$W(j)$: janelas de ponderação triangulares associadas a escala mel; $X(j)$: espectro da *FFT* para n pontos [3], [15].

III. MATERIAIS E MÉTODOS

Nesta seção é realizada a descrição da base de dados e da metodologia empregada no trabalho.

A. *Base de Dados*

A base de dados utilizada nesta pesquisa foi desenvolvida e disponibilizada pelo Laboratório Integrado de Estudos da Voz (LIEV) da Universidade Federal da Paraíba, situada no Campus João Pessoa, Paraíba. A base faz parte de um projeto avaliado e aprovado pelo Comitê de Ética em Pesquisa do Centro de Ciências da Saúde/UFPB (parecer número 52492/12) [12]. Foram gravados sinais de vozes dos pacientes referentes à pronúncia da vogal sustentada /ɛ/ ("é"). A coleta dos dados foi realizada em um ambiente tratado acusticamente. A taxa de amostragem dos sinais é de 44.100 amostras/s, quantizados com 16 bits/amostra. Esses sinais, inicialmente, foram classificados por meio da análise perceptivo-auditiva, empregando uma escala analógico-visual (EAV) de acordo com o grau geral de intensidade do desvio vocal (grau 1 para voz saudável, grau 2 para voz com desvio leve, grau 3 para voz com desvio moderado e grau 4 para desvio intenso), não havendo, nessa base de dados, casos de sinais classificados como grau geral 4.

Foram selecionados 120 sinais apenas do sexo feminino por que, segundo alguns estudos, a incidência de distúrbios vocais ocorre principalmente em mulheres [17]. Dos sinais selecionados: 30 são saudáveis e 90 são de vozes desviadas, sendo destes últimos, 30 sinais de vozes com o desvio rugosidade, 30 com soprosidade e 30 com o desvio tensão.

B. *Metodologia*

Na Figura 1 é apresentado o diagrama em blocos da metodologia empregada neste trabalho. Inicialmente, é realizada a aquisição do sinal que, nesta pesquisa foram obtidas da base dados apresentada na seção anterior. A seguir é realizada a etapa de pré-processamento e a extração de características. Foram considerados os coeficientes LPC, os coeficientes cepstrais e os coeficientes mel-cepstrais, extraídos através de rotinas implementadas no ambiente MATLAB©. As características extraídas foram submetidas à quantização vetorial, com *N* níveis (variando entre 16 e 1024) e dimensão *k* = 46 (frequência de amostragem +2) [14]. Posteriormente foram calculadas as medidas de distorção para cada *codebook* (dicionário), onde foram empregados sete dicionários na quantização vetorial. Essas distorções foram armazenadas como padrões de referência com o intuito de realizar a classificação dos sinais saudáveis ou desviados, ou ainda, discriminando o desvio vocal (rugosidade, soprosidade e tensão) utilizando a curva ROC (*Receiver Operator Characteristic Curve*).

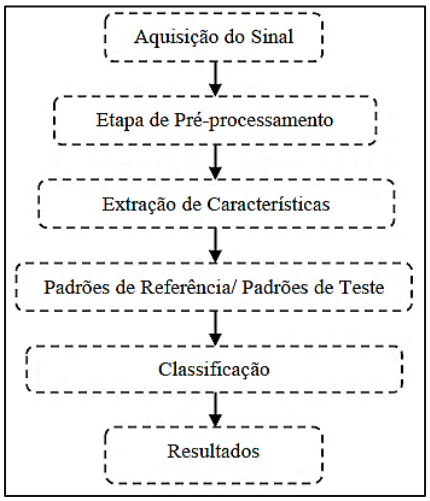

Fig. 1. Diagrama em blocos da metodologia empregada.

Pré-processamento: Nesta etapa o sinal de voz foi segmentado em quadros de 20*ms*, assegurando a sua estacionariedade, no intervalo considerado, com sobreposição de 50%. Foi realizado o janelamento (*Hamming*) e pré-ênfase [3]. Posteriormente, foram extraídos os coeficientes LPC, cepstrais e mel-cepstrais

Extração de Características: As características extraídas foram os coeficientes LPC, cepstrais e mel-cepstrais. Foram extraídos 46 coeficientes de cada segmento do sinal, onde a quantidade de coeficientes extraídos foi de acordo com a frequência de amostragem dos sinais utilizados (frequência de amostragem +2) [14]. Após a extração de características foi realizada a quantização vetorial. A quantização vetorial é realizada para a criação do dicionário que contém um conjunto de dados de tamanho finito, composto por vetores de dimensão k [19]. A quantização vetorial reduz a dimensão dos dados a serem utilizados na etapa de classificação dos sinais em voz saudável ou voz desviada. Foi gerado um dicionário para cada classe, para determinar a similaridade entre as elocuções dos sinais a serem analisados [3].

Nesta pesquisa, para o projeto dos dicionários do quantizador vetorial, é utilizado o algoritmo LBG (*Linde-Buzo-Gray*) [20]. O dicionário inicial é constituído a partir de um conjunto de amostras iniciais da sequência de treino, tomadas de forma aleatória a partir da matriz dos coeficientes LPC, cepstrais ou mel-cepstrais concatenados. A ordem do preditor (p = 46) corresponde à dimensão do quantizador. Foram criados sete dicionários, cada um com um determinado número de níveis (16, 32, 64, 128, 256, 512 e 1024 níveis), visando investigar qual número de níveis fornece o melhor resultado na classificação. A medida de distorção utilizada é a Distância do Erro Médio Quadrático Mínimo [21]. Esta distorção é calculada pela Equação (11)

$$d(v_N, c_M) = \frac{1}{K}\sum_{i=1}^{K}|v_i - c_i|^2 \qquad (11)$$

sendo $d(v_N, c_M)$ a distorção do erro médio quadrático, K a dimensão dos espaços Euclidianos de entrada e de reprodução, v_i corresponde aos vetores de entrada e c_i corresponde aos *codevectors*.

C. *Classificação*

Para realizar a avaliação da classificação e demonstrar os resultados obtidos com as distorções dos coeficientes LPC, cepstrais e mel-cepstrais, foram utilizados gráficos das curvas ROC. A Curva ROC é um método muito utilizado para avaliação de desempenho em testes de diagnósticos médicos [3]. Para o cálculo da curva ROC foi empregada uma regra de decisão em que foi avaliado se 2/3 dos sinais pertenciam ou não à classe. A regra de decisão é baseada na busca por um ponto de corte, de forma que valores menores ou iguais a este ponto de corte são classificados com pertencentes a classe e, analogamente, valores com resposta ao teste maiores que o ponto de corte são classificados como não pertencentes a classe (ou vice-versa). Dessa forma, para diferentes pontos de corte, dentro dos possíveis valores que o teste produz, é possível estimar características como sensibilidade e especificidade.

A acurácia (ACUR) mede a capacidade do classificador de identificar corretamente a presença (Verdadeiro Positivo –VP) ou a ausência do distúrbio vocal (Verdadeiro Negativo – VN), dada por: ACUR = (VP+VN)/(VP+VP+FN+FP), em que FN representa a taxa de Falsos Negativos (Não detecta o desvio quando ele está presente) e FP é a taxa de Falsos Positivos

(Detecção do desvio quando o mesmo não existe de fato). A sensibilidade representa a proporção de pessoas com o distúrbio vocal de interesse que têm o resultado do teste positivo, dada por SENS = VP/ (VP + FN) e a especificidade (ESP = VN/(VN+FP) representa a proporção de pessoas com a ausência do distúrbio vocal, cujo teste dá negativo [22]. Com este método de classificação, as curvas de diferentes testes diagnósticos podem ser comparadas. Quanto mais próxima estiver a curva do canto superior esquerdo, melhor o resultado de classificação[3].

A partir dos dicionários criados foram calculadas as medidas de distorção para cada nível e cada classe. A partir das distorções obtidas foi realizada a classificação utilizando a curva ROC. Foram considerados sete casos de classificação considerando as distorções calculadas para cada dicionário: Desviada Vs. Saudável (DES Vs. SDL), Rugosa Vs. Saudável (RUG Vs. SDL), Tensa Vs. Saudável (TEN Vs. SDL), Soprosa Vs. Saudável (SOP Vs. SDL), Rugosa Vs. Tensa (RUG Vs. TEN), Rugosa Vs. Tensa (RUG Vs. TEN) e Tensa Vs. Soprosa (TEN Vs. SOP).

IV. RESULTADOS

Na Tabela I são apresentados os resultados obtidos com a curva ROC, para as distorções com os coeficientes cepstrais e mel-cepstrais, para os sete casos de classificação considerados, apresentando qual foi o número de níveis de quantização que apresentou a melhor acurácia na classificação. As taxas de classificação, utilizando os coeficientes LPC não obtiveram bons resultados de acordo com os critérios de Hosmer e Lemeshow [23] (taxas inferiores a 70%), por este motivo, não foram expostas.

TABELA I. RESULTADO PARA A CLASSIFICAÇÃO COM AS MEDIDAS DE DISTORÇÃO DOS COEFICIENTES - CURVA ROC.

CLASSE	ACUR (%)	SEN (%)	ESP (%)	Medida	Nível
DES Vs. SDL	77,84	70,69	85,00	*MEL*	16
RUG Vs. SDL	75,00	65,00	85,00	*MEL*	16
TEN Vs. SDL	**84,73**	89,47	80,00	*CEPS*	1024
SOP Vs. SDL	**82,10**	84,21	80,00	*CEPS*	256
RUG Vs. TEN	79,34	85,00	73,68	*MEL*	128
RUG Vs. SOP	74,60	65,00	84,21	*CEPS*	32
TEN Vs. SOP	**86,84**	84,21	89,47	*MEL*	1024

Observa-se, na Tabela I, que os melhores resultados de classificação foram obtidos com as distorções dos coeficientes cepstrais (*CEPS*) e mel-cepstrais (*MEL*). Os melhores resultados (acurácia >80%) na classificação foram a distinção entre: vozes tensas e saudáveis; soprosas e saudáveis; tensas e soprosas. Nos dois primeiros casos foram os coeficientes cepstrais que proporcionaram os melhores resultados e, no terceiro caso, com os coeficientes mel-cepstrais.

É possível identificar que os números de níveis que forneceram maior acurácia foram 256 e 1024. Vale ressaltar que houve uma grande variação no número de níveis que resultaram maior acurácia nos sete casos de classificação. Observa-se que o uso de um maior número de níveis, embora possa elevar as taxas de classificação, em alguns casos, aumenta o custo computacional.

Nas Figuras 2, 3 e 4, são apresentadas as curvas ROC obtidas para os resultados expostos na Tabela I. A Figura 2 corresponde à curva ROC para a classificação entre vozes desviadas e saudáveis. A Figura 3 corresponde às curvas ROC da classificação entre vozes rugosas e saudáveis, tensas e saudáveis e entre soprosas e saudáveis e a Figura 4 corresponde às curvas ROC para a classificação entre vozes rugosas e soprosas, rugosas e tensas e entre tensas e soprosas.

Os resultados das curvas demonstrado nas Figuras 2, 3 e 4 confirmam a superioridade dos coeficientes cepstrais e mel-cepstrais em relação aos coeficientes LPC na classificação dos sinais, onde a classificação utilizando os coeficientes mel-cepstrais obteve a maior acurácia entre os três métodos. Quanto ao número de níveis, houve uma grande variação na classificação, não sendo possível ainda definir um valor único que sirva para os sete casos de classificação.

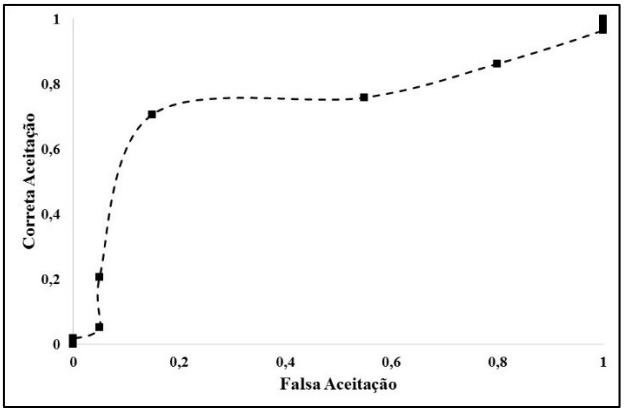

Fig. 2. Curva ROC para a classificação entre DES Vs. SDL, 16 Níveis Coeficientes Mel-Cepstrais.

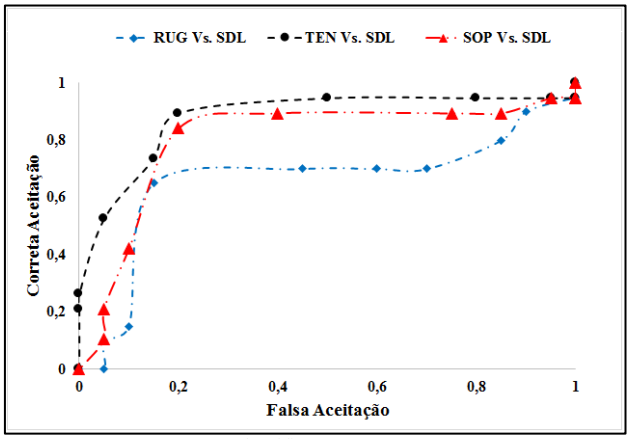

Fig. 3. Curva ROC para as classificações entre RUG Vs. SDL (16 Níveis. Coeficientes Mel-Cepstrais); TEN Vs. SDL (1024 níveis, Coeficientes Cepstrais; e entre SOP Vs. SDL (256 Níveis, Coeficientes Cepstrais).

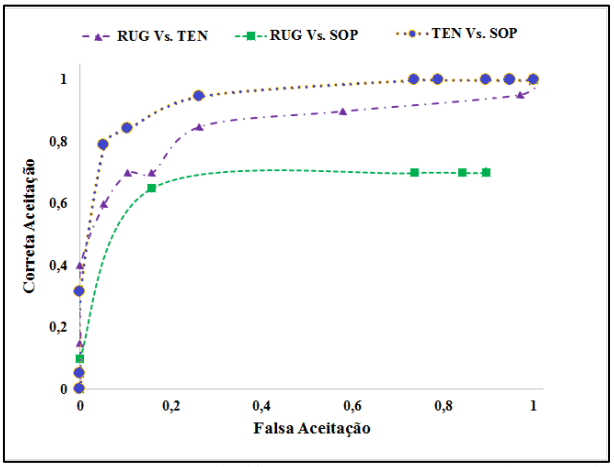

Fig. 4. Curva ROC para a classificação entre RUG Vs. TEN (128 Níveis dos Coeficientes Mel-Cepstrais; entre RUG Vs. SOP (32 Níveis, Coeficientes Cepstrais); e entre TEN Vs. SOP (1024 Níveis, Coeficientes Mel-Cepstrais).

Na Tabela II é apresentado a comparação dos resultados obtidos com os coeficientes LPC, cepstrais e mel-cepstrais com os resultados obtidos por Lopes et al [12].

TABELA II. COMPARAÇÃO DOS RESULTADOS DE LOPES ET AL. E OS RESULTADOS OBTIDOS COM A CURVA ROC.

CLASSE	ACUR (%) Lopes et al.	ACUR (%) com os coeficientes
RUG Vs. SDL	78,57	75,00
SOP Vs. SDL	84,05	82,10
RUG Vs. TEN	73,75	79,34
TEN Vs. SOP	75,71	86,84

Comparando os resultados obtidos com o trabalho de Lopes et al. observa-se que a classificação obtida apresentou acurácia superior em alguns casos (classificação entre vozes rugosas e tensas e entre vozes tensas e soprosas), já em outros casos de classificação a acurácia foi semelhante, demonstrando que os coeficientes LPC, cepstrais e mel-cepstrais possuem potencial para classificação dos desvios vocais.

V. CONCLUSÕES

Este trabalho apresentou diversas contribuições como a comparação das medidas tradicionais com as medidas obtidas da Análise LPC, Análise cepstral e mel-cepstral na classificação dos desvios vocais. Foi observado quais demonstram maior potencial na discriminação dos desvios vocais, como também a investigação do uso da quantização vetorial na classificação dos desvios vocais., observando quais implicações ela poderia trazer na discriminação dos sinais e sua influência sobre os coeficientes LPC, cepstrais e mel-cepstrais.

Discriminar entre os desvios vocais não é uma tarefa trivial, devido ao fato de que, na maioria dos casos, uma voz não apresenta um desvio vocal isolado. Geralmente há mais de um desvio presente, sendo um atuante de forma mais predominante do que outro.

Observa-se que a classificação utilizando os coeficientes cepstrais e mel-cepstrais foi eficiente, onde os coeficientes LPC não estiveram presentes em nenhum dos casos de classificação. Comparada ao trabalho de Lopes et al., observa-se que, em alguns casos, os resultados são semelhantes, sendo que em Lopes et al., algumas das características utilizadas dependem da obtenção da frequência fundamental (*jitter* e *shimmer*). Nesta pesquisa, tanto a análise cepstral como a análise mel-cepstral são eficientes na classificação dos desvios vocais, apresentando a vantagem de não depender da frequência fundamental. São de fácil extração, podendo ser empregadas no desenvolvimento de uma ferramenta adicional para o diagnóstico de desvios vocais.

AGRADECIMENTOS

Os autores agradecem a PRPIPG e a COPEX do Instituto Federal da Paraíba, Campus João Pessoa, pelo financiamento parcial da pesquisa e ao PPgEE pelo apoio científico.

REFERÊNCIAS

[1] M. Behlau, *Voz, O Livro do Especialista*, vol. 1. Reimpressão, Rio de Janeiro: Revinter, 2008.

[2] R. H Colton, J. K. Casper, R. Leonard, *Understanding voice problems: A physiological perspective for diagnosis and treatment.* Wolters Kluwer Health, 2006.

[3] S. L. do N. C. Costa, *Análise Acústica, Baseada no Modelo Linear de Produção da Fala, para Discriminação de Vozes Patológicas.* 161 f. Tese de Doutorado, Universidade Federal de Campina Grande, PB, 2008.

[4] W. C. de A. Costa, S. L. do N. C. Costa, F. M. de Assis, B. G. Aguiar Neto, "Classificação de Sinais de Vozes Saudáveis e Patológicas por meio da Combinação entre Medidas da Análise Dinâmica Não Linear e Codificação Preditiva Linear", *Revista Brasileira de Engenharia Biomédica*, [Online]. Vol. 29, nº 1, pp. 3-14, Mar, 2013.

[5] E. R. Carrasco, G. Oliveira, M. Behlau, "Análise Perceptivo-Auditiva e Acústica da Voz de Indivíduos Gagos*" Revista CEFAC*, São Paulo, 2010.

[6] M. P. A. Bandeira, O. P. Neto, *Análise da Frequência Fundamental Comparativa das Vozes Disfônicas e Normais do Professor.* Disponível em: http://unicastelo.br/epginic2016/edicoes_anteriores/files/2014/EPG/Engenharias/291%20-%20EPG225.pdf

[7] P. B. Baravieira, *Aplicação de uma rede neural artificial para avaliação da rugosidade e soprosidade vocal*. 101 f. Tese de doutorado, Universidade de São Paulo, SP, 2016.

[8] L. J. I. P. Godino, V. Gomez, V. M. Blanco, "Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters. Biomedical Engineering", *IEEE Transactions*, vol. 53, no. 10, pp. 1943–1953, 2006.

[9] J. M. Fechine, *Reconhecimento Automático de Identidade Vocal Utilizando Modelagem Híbrida: Paramétrica e Estatística.* 237 f. Tese de Doutorado, Universidade Federal da Paraíba, 2000.

[10] R. Tavares, N. A. Brunet, S. Correia, S. C. Costa, B. G. Aguiar Neto, J. M. Fechine, "Combinação de Classificadores Baseados em Análise LPC e Cepstral para a Detecção de Paralisia nas Dobras Vocais", *Anais do XXII Congresso Brasileiro de Engenharia Biomédica,* 2010.

[11] , S. C. Costa, W. C. de A. Costa, S. E. N. Correia, J. M. F. R. de Araújo, V. J. D. Vieira, "Análise de Sinais de Voz para Caracterização de Patologias na Laringe", *Revista de Tecnologia da Informação e Comunicação.* Vol. 4, nº. 2, Out. 2014 [Online].

[12] L. W. Lopes, D. da S. Evangelista, F. P. França, L. B. Simões, J. D. da Silva, V. J. D. Vieira, "Acurácia das medidas acústicas tradicionais e formânticas na discriminação de vozes saudáveis e desviadas", *Anais do XXIV Congresso Brasileiro de Fonoaudiologia.* Out. 2016.

[13] A. Alcaim, C. A. dos S. Oliveira, *Fundamentos do Processamento de Sinais de Voz e Imagem.* Ed. Interciência, 1ªEd. Rio de Janeiro, RJ, 2011.

[14] L. R. Rabiner, and R. W. Schafer, *Digital Processing of Speech Signals*. Prentice Hall, Upper Saddle River, New Jersey, 1978.

[15] D. O'Shaughnessy, *Speech Communications: Human and Machine*, 2nd Edition, NY, IEEE Press, 2000.

[16] I. C. Zwetsch, R. D. R Fagundes, T. E, Russomano, D. Scolari, "Processamento digital de sinais no diagnóstico diferencial de doenças laríngeas benignas", *Scientia Medica*, Porto Alegre: PUCRS, v. 16, n. 3, jul./set. 2006.

[17] L. A. A. Mota, C. M. B. Santos, J. M. de Vasconcelos, B. C. Mota, H. de S. C. Mota, "Aplicação da técnica de emissão em tempo máximo de fonação em paciente com disfonia espasmódica adutora: relato de caso", *Revista Sociedade Brasileira de Fonoaudiologia*, 2012, vol. 17, nº.3, pp. 351-356.

[18] S. Haykin, *Redes Neurais, Princípios e Práticas,* Porto Alegre, RS. BOOKMAN, 2001.

[19] F. O. Simões, M. U. Neto, J. B. Machado, E. J. Nagle, F. O. Runstein, L. de C. T. Gomes, "Compressão de fala utilizando quantização vetorial e redes neurais não supervisionadas" *Cadernos CPqD Tecnologia*, Campinas, v. 5, n. 1, p. 33-48, jan./jun. 2009.

[20] Y. Linde, A. Buzo, R.M. Gray, "An Algorithm for Vector Quantizer Design", *IEEE Transactions on Communications*, Vol. COM - 28, No. 1, pp. 84-95, Jan, 1980.

[21] J. Makhoul, S. Roucos, H. Gish, "Vector Quantization in Speech Coding", *Proceedings of the IEEE,* Vol. 73, nº.11, pp.1551-1588, Nov, 1985.

[22] E. Z. Martinez; F. Louzada-Neto; B. B. Pereira, "A Curva ROC para Testes Diagnósticos", *Caderno de Saúde Coletiva*, Rio de Janeiro, vol. 11, nº 1, pp. 7-31, 2007.

[23] D. W. Jr. Hosmer, S. Lemeshow, *Applied logistic regression,* 2.ed. New York: John Wiley & Sons; 2000.