

Detecção de Distúrbios Vocais através de Parâmetros Cepstrais Multibanda de Vogal Sustentada

Marco A. R. Alves, Gabriel A. R. Silva, Bruno C. Bispo, María E. Dajer e Pedro M. Rodrigues

Resumo— Este trabalho tem como objetivo detectar distúrbios vocais relacionados com nódulo vocal, edema de Reinke e patologias neurológicas através de parâmetros cepstrais multibanda da vogal sustentada /a/. A detecção é realizada entre pares de grupos de estudo e a análise multibanda é efetuada usando a transformada Wavelet. Para cada par de grupos, uma seleção de parâmetros é realizada. Estatísticas dos parâmetros selecionados são utilizadas como entrada para quatro classificadores com validação cruzada *leave-one-out*. Acurácias de classificação iguais ou superiores a 96,70% são obtidas para 6 pares de grupos de estudo enquanto apenas 74,20% é alcançado para o par Edema/Nódulo.

Palavras-Chave— Distúrbios vocais, nódulo vocal, edema de Reinke, patologias neurológicas, parâmetros cepstrais, wavelet.

Abstract— This work aims to detect vocal disorders related to vocal nodule, Reinke's edema and neurological pathologies through multiband cepstral measures of the sustained vowel /a/. Detection is performed between pairs of study groups and multiband analysis is accomplished using the Wavelet transform. For each pair of groups, a parameters selection is performed. Statistics of the selected parameters are used as input for four classifiers with *leave-one-out* cross validation. Classification accuracies equal or greater than 96.70% are obtained for 6 pairs of study groups while only 74.20% is achieved for the Edema/Nodule pair.

Keywords— Voice disorders, vocal fold, Reinke's Edema, neurological pathologies, cepstral measures, wavelet.

I. INTRODUÇÃO

A voz é um elemento da fala essencial para a comunicação humana e sua qualidade impacta diretamente a vida profissional e social, podendo suscitar o desenvolvimento de problemas de ordem psicológica e emocional [1]. Acredita-se que cerca de 10% da população apresenta distúrbios vocais, sendo que este valor aumenta para os 50% quando contabilizados apenas os profissionais com funções comunicacionais [2].

Com o intuito de auxiliar o diagnóstico de disfunções vocais, pesquisadores começaram a investigar parâmetros de sinais de voz que pudessem representar índices para diferentes aspectos da voz, incluindo patologias relacionadas com a fala [3]. Inicialmente, parâmetros para a análise acústica da voz encontravam-se no domínio do tempo, principalmente jitter e shimmer [3]. Porém, visto que dependem do conhecimento do início e fim dos ciclos vocais, há uma perda de acurácia para sinais altamente perturbados [3], [4]. Além disso, para calcular esses parâmetros, é necessário que a vogal seja analisada como

um pico alto e estável, pois qualquer alteração no sinal pode ser considerada como uma perturbação vocal [5].

Para superar tais limitações, medidas baseadas no espectro e cepstro foram consideradas tanto para vogal sustentada quanto para fala contínua [6]. De acordo com [3], [7], parâmetros cepstrais são os índices mais promissores para detectar a severidade de disфонia na análise de sinais vocais. A pesquisa bibliográfica realizada mostrou que o parâmetro cepstral mais aplicado em sinais de fala para detecção de distúrbios vocais são os coeficientes mel-cepstrais (MFCC, do inglês *Mel-Frequency Cepstral Coefficients*).

Os MFCCs foram os únicos parâmetros usados em [8] para detectar distúrbios vocais causados pela doença de Parkinson, em [9] para identificar pacientes com patologias vocais, incluindo cistos, pólipos, nódulos e paralisia, e em [10] para discriminar indivíduos saudáveis e com paralisia nos nervos laríngeos bem como indivíduos saudáveis e com edema de Reinke.

A combinação dos MFCCs e parâmetros não-cepstrais também são encontrados na literatura para esse propósito. Em [11], eles foram utilizados na diferenciação entre três grupos de estudos: pacientes diagnosticados com doença de Parkinson, pacientes com patologias da laringe como nódulos, pólipos ou edemas; e pacientes com fenda lábio-palatina. No caso de [12], foram aplicados para diferenciar três classes de sinais: vozes saudáveis, vozes com patologias neuromusculares como paralisia das cordas vocais; e patologias fisiológicas como nódulos e edemas. Nestes trabalhos, os MFCCs foram obtidos a partir da banda completa dos sinais de fala.

Combinações de MFCCs calculados em sub-bandas em conjunto com parâmetros não-cepstrais também são encontradas na literatura. Em [13], os MFCCs foram calculados por meio de uma transformada Wavelet para detectar pacientes com doenças neurológicas. E em [14], eles foram calculados por meio da Wavelet Packet para detectar patologias na laringe, incluindo edemas, nódulos e pólipos nas pregas vocais.

Outros parâmetros cepstrais encontrados na literatura para detectar patologias vocais são a proeminência do pico cepstral (CPP, do inglês *Cepstral Peak Prominence*) e a sua versão suavizada CPPS (do inglês *CPP Smoothed*). Em [3], o CPPS foi utilizado para discriminar entre indivíduos saudáveis e com patologias vocais, incluindo 10 patologias. Em [15], o CPP foi utilizado para detectar indivíduos com nódulos vocais e saudáveis. Com o mesmo propósito, CPP e CPPS foram utilizados separadamente em [16]. Nesses casos, os parâmetros foram obtidos a partir da banda completa dos sinais de fala.

Todos os trabalhos mencionados possuem em comum o fato de utilizarem apenas um tipo de parâmetro cepstral. Esse trabalho objetiva combinar parâmetros cepstrais de vogal sustentada para discriminar entre pares de grupos de estudo formados por pacientes com nódulo vocal, pacientes com edema de Reinke

Marco A. R. Alves, Gabriel A. R. Silva e Pedro M. Rodrigues, Universidade Católica Portuguesa, CBQF - Centro de Biotecnologia e Química Fina - Laboratório Associado, Escola Superior de Biotecnologia, Porto, Portugal, e-mail: prodrigues@porto.ucp.pt; Bruno C. Bispo, Departamento de Engenharia Elétrica e Eletrônica, Universidade Federal de Santa Catarina, Florianópolis-SC, Brasil, e-mail: bruno.bispo@ufsc.br; María E. Dajer, Departamento de Engenharia Elétrica, Universidade Tecnológica Federal do Paraná, Cornélio Procopio-PR, Brasil, e-mail: medajer@utfpr.edu.br.

TABELA I
INFORMAÇÃO DA BASE DE DADOS.

Grupo	Número de indivíduos
Controle	20
Edema de Reinke	16
Nódulo vocal	15
Neurológica	14

e pacientes com disfonia de origem neurológica. Com intuito de explorar uma possível maior capacidade discriminante de bandas frequenciais, os parâmetros são obtidos a partir de uma análise multibanda usando a transformada Wavelet.

O artigo está organizado da seguinte maneira: a Seção II descreve a base de sinais de voz; a Seção III explica a análise cepstral multibanda realizada; a Seção IV apresenta os parâmetros cepstrais avaliados; a Seção V apresenta e discute os resultados obtidos; e, por fim, a Seção VI conclui o trabalho.

II. BASE DE DADOS DE VOZ

A base de dados de voz foi fornecida pelo Grupo de Pesquisa em Engenharia Médica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (GP/EM/CNPq). Ela consiste em 65 gravações da vogal sustentada /a/ proferidas por homens e mulheres adultos (1 gravação por indivíduo).

As gravações foram realizadas no Ambulatório de Voz do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (HC-FMUSP) com autorização do comitê de ética em pesquisa de seres humanos da Universidade Federal de São Carlos, protocolo 256/2010. Elas foram digitalizadas com uma frequência de amostragem de 22050 Hz. A vogal /a/ foi escolhida devido à sua ampla utilização na análise acústica da voz [3]. Como recomendado em [17], a duração de cada fonema foi sempre superior a 2 s.

Neste trabalho, foram considerados quatro grupos de estudo: controle (indivíduos saudáveis), pacientes diagnosticados com edema de Reinke, com nódulos vocais e com disfonia de origem neurológica. Os sinais do último grupo contêm patologias como doença de Huntington, doença de Parkinson, acidente vascular cerebral (AVC), mononeurite múltipla, mitocondropatia, distrofia de Duchenne, distonia cervical, distrofia miotônica e esclerose lateral amiotrófica (ELA). O número de indivíduos em cada grupo de estudo está descrito na Tabela I.

Os sinais de voz tiveram amplitude normalizada segundo

$$x(n) = \frac{x(n)}{\sum_{n=0}^{N-1} x^2(n)}, \quad (1)$$

onde N é o tamanho do sinal, e componente contínua removida. Em seguida, os sinais foram segmentados em trechos não-sobrepostos com duração de 100 ms usando a janela de Hamming. Essa duração foi escolhida para garantir que os sinais do último nível de decomposição tenham amostras suficientes para computação do cepstro.

III. ANÁLISE CEPSTRAL MULTIBANDA

A. Decomposição Multibanda com Transformada Wavelet

A transformada wavelet discreta (DWT, do inglês *Discrete Wavelet Transform*) de um sinal de tempo discreto e energia finita é a sua decomposição em um conjunto de funções base

obtidas a partir de um número finito de sequências protótipas e suas versões deslocadas no tempo [18].

Essa expansão estruturada é implementada por meio de um banco de filtros em bandas de oitavas e criticamente decimado [18], [19]. Considerando apenas as frequências positivas, a m -ésima sub-banda está limitada a [19]

$$W_m = \begin{cases} [0, \pi/2^S], & m = 0, \\ [\pi/2^{S-m+1}, \pi/2^{S-m}], & m = 1, 2, \dots, S, \end{cases} \quad (2)$$

onde S é o número de sub-bandas ou níveis de decomposição e π é a frequência angular normalizada.

A DWT utiliza uma função escala de análise $\tilde{\phi}_1(n)$ e uma função wavelet de análise $\tilde{\psi}_1(n)$ que são respostas ao impulso dos filtros de análise passa-baixa e passa-alta de meia banda, respectivamente. Definindo as seguintes fórmulas de recursão

$$\tilde{\phi}_{i+1}(n) = \tilde{\phi}_i(n/2) * \tilde{\phi}_1(n) \quad (3)$$

$$\tilde{\psi}_{i+1}(n) = \tilde{\phi}_i(n) * \tilde{\psi}_1(n/2^i), \quad (4)$$

o filtro de análise equivalente da m -ésima sub-banda é dado por

$$h_m(n) = \begin{cases} \tilde{\phi}_S(n), & m = 0 \\ \tilde{\psi}_{S+1-m}(n), & m = 1, 2, \dots, S. \end{cases} \quad (5)$$

E o sinal da m -ésima sub-banda do sinal $x(n)$ é dado por

$$x_m(n) = \begin{cases} \sum_{k=-\infty}^{\infty} x(k)h_m(2^S n - k), & m = 0, \\ \sum_{k=-\infty}^{\infty} x(k)h_m(2^{S-m+1} n - k), & m = 1, 2, \dots, S. \end{cases} \quad (6)$$

Neste trabalho, a decomposição em sub-banda foi aplicada até o terceiro nível, $S = 3$, aos segmentos dos sinais de voz. Os sinais das sub-bandas foram reamostrados para a frequência de amostragem original através do método de interpolação wavelet [20].

B. Análise cepstral

A análise cepstral é uma técnica de processamento não-linear de sinais baseada numa transformação homomórfica que mapeia convolução em adição, resultando no cepstro. O cepstro foi inicialmente proposto como uma alternativa à função de autocorrelação para detectar ecos em sinais sísmicos. Diferentes definições matemáticas são encontradas na literatura, levando a algumas versões do cepstro [21].

O cepstro real de um sinal $x(n)$ é definido como [21]

$$c(n) = \mathcal{F}^{-1} \{ \log [| \mathcal{F} \{ x(n) \} |] \}, \quad (7)$$

onde $\mathcal{F}\{\cdot\}$ and $\mathcal{F}^{-1}\{\cdot\}$ denotam a Transformada Rápida de Fourier (FFT, do inglês *Fast Fourier Transform*) de N -pontos e sua inversa. Embora qualquer base possa ser usada nesse cálculo, neste trabalho optou-se pelo logaritmo natural como em [21]. No cepstro, o índice de amostra n é usualmente chamado de quefrência [21].

Neste trabalho, os cepstros foram calculados para cada sub-banda $m = \{0, 1, 2, 3\}$ dos segmentos dos sinais de voz.

IV. PARÂMETROS CEPSTRAIS

A. Coeficientes Mel-Cepstrais

Os coeficientes mel-cepstrais se diferenciam do cepstro real pela conversão do espectro de magnitude, $|\mathcal{F}\{x(n)\}|$, para uma escala de frequências que se aproxima ao sistema auditivo humano, denominada Mel, antes da aplicação do logaritmo.

Essa conversão é realizada por meio de um banco de filtros composto por filtros passa-banda triangulares sobrepostos em 50% e com largura de banda variando com a frequência, começando em cerca de 100 Hz para frequências abaixo de 1 kHz e depois aumentando logaritmicamente acima de 1 kHz. A soma do conteúdo de cada espectro de magnitude filtrado é associada ao bin da FFT correspondente à frequência central da banda. Os restantes bins da FFT são considerados nulos.

Denotando a operação descrita acima por $\Lambda\{\cdot\}$, os coeficientes mel-cepstrais são definidos como

$$c_{mel}(n) = \mathcal{F}^{-1}\{\log\{\Lambda[|\mathcal{F}\{x(n)\}|]\}\}. \quad (8)$$

Devido às propriedades de $\log\{\Lambda[|\mathcal{F}\{x(n)\}|]\}$, a FFT inversa é normalmente substituída pela transformada discreta do cosseno (DCT, do inglês *Discrete Cosine Transform*). Detalhes sobre a computação dos MFCCs podem ser vistos em [21].

Neste trabalho, para cada sub-banda $m = \{0, 1, 2, 3\}$ dos segmentos dos sinais de voz, 13 MFCCs médios foram obtidos e utilizados como parâmetros. Isso foi realizado ao calcular 13 MFCCs para cada trecho com duração de 20 ms e sobreposição de 30 ms, obtidos com janela de Hamming, dos sinais de sub-bandas utilizando um banco com 40 filtros. A média dos MFCCs ao longo dos trechos resultou nos 13 MFCCs médios.

B. Distâncias Cepstrais

Medidas de distância entre cepstros são encontradas na literatura e utilizadas como parâmetros discriminantes em diferentes aplicações, principalmente em reconhecimento de voz. Algumas dessas medidas são definidas abaixo considerando os cepstros $c_i(n)$ e $c_j(n)$ com comprimento N .

- Distância cepstral RMS normalizada [22]:

$$D1_{i,j} = 4.3429 \times \sqrt{[c_i(1) - c_j(1)]^2 + 2 \sum_{n=2}^N [c_i(n) - c_j(n)]^2}. \quad (9)$$

- Distância cepstral RMS normalizada sem o primeiro coeficiente [23]:

$$D2_{i,j} = 4.3429 \times \sqrt{2 \sum_{n=2}^N [c_i(n) - c_j(n)]^2}. \quad (10)$$

- Distância cepstral Euclidiana [24]

$$D3_{i,j} = \sqrt{\sum_{n=1}^N [c_i(n) - c_j(n)]^2}. \quad (11)$$

- Distância cepstral ponderada pela quefrência [25]:

$$D4_{i,j} = \sqrt{\sum_{n=1}^N n \times [c_i(n) - c_j(n)]^2}, \quad (12)$$

a qual também é comumente denominada medida de distância RPS (do inglês *root power sums*).

- Distância cepstral ponderada pela raiz da quefrência [24]:

$$D5_{i,j} = \sqrt{\sum_{n=1}^N \sqrt{n} \times [c_i(n) - c_j(n)]^2}. \quad (13)$$

- Distância cepstral ponderada pelo quadrado da quefrência [24]:

$$D6_{i,j} = \sqrt{\sum_{n=1}^N n^2 \times [c_i(n) - c_j(n)]^2}. \quad (14)$$

Neste trabalho, para cada segmento dos sinais de voz, as 6 distâncias foram calculadas entre os cepstros reais de todos os pares $\{i, j\}$ das sub-bandas, onde $i, j = \{0, 1, 2, 3\}$ e $i \neq j$.

C. Picos Cepstrais

Para sinais quasi-periódicos, como sons de vogais, o cepstro apresenta picos nas quefrências correspondentes a múltiplos inteiros do período fundamental do sinal mais proeminentes do que um sinal menos periódico [21]. Isso deve-se a uma estrutura harmônica melhor definida [21].

O primeiro pico é explorado nos parâmetros CPP e CPPS. Neste trabalho, entretanto, optou-se por utilizar como parâmetros a diferença em amplitude (DAP) e quefrência (DQP) entre os dois primeiros picos cepstrais, as energias desses picos (EP1 e EP2) e a energia do cepstro entre esses picos (EEP). A diferença em quefrência (DQP) correspondente ao período glotal. Esses parâmetros foram calculados para cada sub-banda $m = \{0, 1, 2, 3\}$ dos segmentos dos sinais de voz.

V. METODOLOGIA E RESULTADOS

Séries temporais, ao longo dos segmentos de voz, dos valores de cada parâmetro foram obtidas para cada indivíduo e tratadas como distribuições. Para cada distribuição, as seguintes estatísticas foram calculadas: média, mediana, desvio padrão, variância, valor máximo e valor mínimo.

A. Seleção de Parâmetros

Como os valores de cada parâmetro foram calculados para todas as sub-bandas dos segmentos e dependiam da wavelet $\psi_1(n)$, uma busca por um subconjunto com maior capacidade discriminante de cada parâmetro foi realizada para reduzir o número de características a serem utilizadas na classificação.

Para cada um dos 13 MFCCs, buscou-se pela combinação da wavelet e sub-banda que resultasse na estatística com maior capacidade discriminante. O mesmo foi efetuado para DAP, DQP, EP1, EP2 e EEP. No caso das distâncias cepstrais, para cada par de sub-banda $\{i, j\}$, buscou-se pela combinação de wavelet e distância cepstral $Dk_{i,j}$, $k = 1, 2, \dots, 6$, que resultasse na estatística com maior capacidade discriminante. As famílias de wavelets avaliadas foram Daubechies (db), Coiflets (coif), Symlets (sym), Fejer-Korovkin (fk), biortogonais (bior) e biortogonais reversas (rbio).

Essa busca foi realizada para cada par de grupos de estudo, após normalizar os parâmetros por grupo de estudo da mesma forma que em (1), através do classificador binário X-ROC [26]. Esse classificador utiliza a distância quadrática entre médias de

distribuições para encontrar um bom limiar para discriminar duas classes e a curva ROC (do inglês *Receiver Operating Characteristic*) para analisar o resultado da classificação [26]. As maiores acurácias do classificador para cada parâmetro e par de grupos são apresentadas na Tabela II, juntamente com a configuração que gerou esses resultados de classificação. As estatísticas responsáveis pelos resultados de cada parâmetro não são exibidas por motivos de organização do texto.

Uma análise da Tabela II permite algumas considerações. Em relação aos MFCCs, DAP, DQP, EP1, EP2 e EEP, a sub-banda 0 não obteve o melhor resultado para nenhum par de grupos de estudo. No caso das distâncias cepstrais, os resultados de todas as distâncias para os pares de sub-banda que incluem a sub-banda 0 ($\{0, 1\}$, $\{0, 2\}$ e $\{0, 3\}$) foram sempre inferiores às outras combinações e, por isso, não são exibidos. Isso indica que o cepstro da sub-banda 0 não possui uma boa capacidade para discriminar os pares de grupos analisados. E, para os pares de sub-bandas $\{1, 2\}$, $\{1, 3\}$ e $\{2, 3\}$, as distâncias D_2 , D_3 e D_4 não obtiveram o melhor resultado para nenhum par de grupos, indicando também uma baixa capacidade discriminante desses parâmetros.

Classificações acima de 80% foram obtidas para todos os pares de grupos de estudo, exceto Edema/Nódulo. Isso indica que os parâmetros cepstrais multibanda analisados nesse trabalho podem não serem muito adequados para discriminar entre essas patologias. Por fim, destaca-se a grande capacidade discriminante das distâncias cepstrais e da diferença em amplitude entre os dois primeiros picos cepstrais para os pares de estudo Edema/Controle e Nódulo/Neurológica, onde acurácias de classificação superiores a 99% são obtidas.

B. Classificação Utilizando Parâmetros Selecionados

Para cada par de grupos de estudo, definidas a combinação de wavelet e distância cepstral para os pares de sub-bandas $\{1, 2\}$, $\{1, 3\}$ e $\{2, 3\}$, e de wavelet e sub-banda para os outros parâmetros (13 MFCCs, DAP, DQP, EP1, EP2 e EEP) que resultaram na estatística de cada parâmetro com maior capacidade discriminante, combinações que estão descritas na Tabela II, as outras estatísticas da mesma série temporal de

cada parâmetro também são selecionadas para classificação. A combinação das 6 estatísticas dos 21 parâmetros resulta em 126 características por indivíduo para cada par de grupos.

E, para cada par de grupos de estudo, o correspondente conjunto de características foi utilizado como entrada de classificadores binários. Quatro classificadores foram avaliados, nomeadamente, árvores de decisão (DT, do inglês *Decision Trees*), máquinas de vetor suporte (SVM, do inglês *Support Vector Machines*), K vizinhos mais próximos (KNN, do inglês *K-nearest neighbors*) cosseno e euclidiano. Em todos os casos, para verificar a capacidade de generalização dos classificadores, uma validação cruzada *leave-one-out* foi utilizada.

Os resultados de classificação são exibidos na Tabela III. Observa-se que uma acurácia de 100% foi obtida para todos os pares de grupo de estudo que envolvem o grupo de Controle, nomeadamente, Patologias/Controle, Edema/Controle, Neurológica/Controle, Nódulo/Controle. Para os pares de grupo que envolvem duas patologias, 100% de acurácia foi obtido apenas para o caso Nódulo/Neurológica. Esses resultados foram obtidos utilizando apenas classificadores SVM ou DT.

Para o caso Edema/Neurológica, uma elevada acurácia de classificação, especificamente 96,70%, foi obtida utilizando o classificador KNN cosseno. Por outro lado, no caso Edema/Nódulo, a maior acurácia obtida foi 74,20% com o classificador KNN euclidiano. Curiosamente, os classificadores SVM e DT, que foram os únicos a obterem 100% de acurácia nos outros casos, não alcançaram os melhores resultados para Edema/Neurológica e Edema/Nódulo. O baixo desempenho no par Edema/Nódulo pode ser justificado pela baixa capacidade discriminante de todos os parâmetros cepstrais avaliados, como discutido na Seção V-A.

Em comparação com os métodos do estado da arte baseados em parâmetros cepstrais, os resultados entre 91,20% e 100% obtidos na classificação Neurológica/Controle são superiores aos 82,5% e 94,3% obtidos em [8] e [13], respectivamente. No caso de Nódulos/Controle, os resultados entre 97,10% a 100% obtidos nesse trabalho superam a acurácia de 85,71% obtida em [16]. E no caso de Edema/Controle, o melhor resultado obtido neste trabalho, 100%, iguala a acurácia alcançada em

TABELA II
MAIOR ACURÁCIA DE CADA PARÂMETRO PARA CADA PAR DE GRUPOS DE ESTUDO EM FUNÇÃO DO TIPO DE WAVELET (WAVE.) E SUB-BANDA(M)/DISTÂNCIA(K).

Parâmetro	Patologias/Controle			Edema/Neurológica			Edema/Nódulo			Edema/Controle			Neurológica/Controle			Nódulo/Controle			Nódulo/Neurológica		
	%	wave.	m/k	%	wave.	m/k	%	wave.	m/k	%	wave.	m/k	%	wave.	m/k	%	wave.	m/k	%	wave.	m/k
DAP	82,09	db1	3	78,80	db1	2	62,84	fk8	3	99,57	db1	3	74,13	db1	3	81,98	db1	1	100	db1	1
DQP	63,87	bior2.4	3	69,47	db1	3	73,54	bior1.1	3	75,11	db2	3	63,20	bior6.8	n1	75,88	rbio2.2	3	78,93	db1	1
EEP	72,62	db5	3	83,41	rbio1.1	2	64,40	sum4	2	89,02	rbio3.3	3	66,26	bior1.1	1	83,60	db3	3	88,86	bior3.3	2
EP1	71,71	rbio3.7	3	82,09	fk4	2	65,37	sym16	1	88,80	rbio3.3	3	64,16	rbio6.8	3	83,33	rbio1.5	2	88,38	bior3.3	2
EP2	69,06	rbio3.3	3	78,37	bior3.5	2	64,20	sym12	1	84,46	bior2.2	3	63,81	rbio3.3	3	81,03	bior2.4	3	86,44	sym2	1
MFCC 1	58,56	bior1.1	2	67,18	sym11	2	60,98	sym11	3	76,30	bior1.1	2	80,72	rbio3.5	3	65,60	bior1.1	2	81,71	coif3	2
MFCC 2	68,13	bior1.3	3	67,35	rbio3.1	3	63,20	bior1.1	1	82,46	rbio1.1	3	83,55	bior3.1	2	66,09	rbio1.1	3	86,74	rbio3.1	1
MFCC 3	72,14	rbio1.3	3	68,78	fk4	3	64,56	bior3.1	2	81,30	bior1.1	3	84,94	fk6	2	61,85	rbio3.1	3	85,22	db11	1
MFCC 4	76,49	rbio1.1	1	76,57	rbio1.1	3	68,32	sym13	3	86,37	db1	1	81,97	fk6	2	81,63	fk4	1	88,65	fk4	1
MFCC 5	75,55	db1	2	69,21	rbio3.3	1	66,87	bior1.1	1	82,70	fk4	2	81,76	bior1.1	3	73,96	bior1.1	2	89,44	db1	2
MFCC 6	65,78	rbio3.5	2	71,32	sym11	2	63,15	sym8	3	79,37	bior3.5	1	82,68	rbio2.6	2	70,44	bior3.5	1	83,82	sym1	2
MFCC 7	79,56	rbio3.1	1	80,49	rbio1.1	1	61,92	rbio1.3	2	86,34	rbio1.1	1	81,68	fk4	2	79,10	rbio1.1	1	87,55	rbio1.1	3
MFCC 8	67,58	rbio3.3	1	71,67	rbio3.1	1	63,67	sym16	3	77,82	sym16	1	81,58	fk22	1	72,05	rbio3.5	1	82,95	rbio3.1	3
MFCC 9	67,34	rbio3.1	2	72,35	sym13	2	64,56	fk14	3	77,21	rbio3.5	2	84,04	rbio3.3	2	70,97	rbio3.5	1	82,67	db13	2
MFCC 10	73,76	db1	1	71,97	rbio1.1	1	64,84	sym13	3	83,64	rbio1.1	1	80,87	rbio3.5	2	78,54	rbio1.1	1	86,69	bior1.1	3
MFCC 11	73,26	rbio3.1	1	70,94	rbio3.1	1	66,15	rbio1.1	1	82,89	db1	1	81,71	db1	3	63,86	rbio1.1	1	82,92	rbio2.2	2
MFCC 12	63,44	rbio3.5	3	6989	rbio3.1	3	63,20	sym10	3	77,46	db9	1	84,83	rbio3.1	2	71,69	db8	1	82,84	sym11	2
MFCC 13	64,02	fk22	1	70,08	sym14	2	69,21	fk4	2	77,01	rbio2.2	2	81,18	rbio2.2	2	66,72	rbio2.2	2	82,64	sym14	2
$Dk_{2,3}$	81,66	db1	1	73,08	db1	1	60,89	bior2.6	1	100	db1	1	73,43	db1	1	75,61	db1	1	100	db5	6
$Dk_{1,3}$	81,60	db1	1	72,96	db1	1	61,28	bior6.8	6	99,89	db1	1	73,43	db1	1	74,66	db1	1	100	db9	5
$Dk_{1,2}$	81,60	db1	1	73,56	db8	1	60,89	db15	6	99,89	db1	1	73,43	db1	5	75,88	db1	1	100	db10	1

TABELA III
ACURÁCIA MÉDIA DOS CLASSIFICADORES PARA CADA
PAR DE GRUPOS DE ESTUDO.

Par de Grupos	KNN cos.	KNN eucl.	SVM	DT
Patologias/Controle	96,92%	96,92%	100,00%	100,00%
Edema/Neurológica	96,70%	83,30%	90,00%	76,70%
Edema/Nódulo	58,10%	74,20%	61,30%	41,90%
Edema/Controle	97,20%	97,20%	97,20%	100,00%
Neurológica/Controle	91,20%	91,20%	100,00%	100,00%
Nódulo/Controle	97,10%	97,10%	100,00%	97,10%
Nódulo/Neurológica	89,70%	89,70%	100,00%	93,10%

[10]. Outras comparações não podem ser realizadas devido a diferenças entre os grupos de estudo discriminados.

De uma forma global, os resultados obtidos demonstraram que a combinação de parâmetros cepstrais multibanda de sinais de vogal sustentada é adequada para detectar distúrbios vocais relacionados com nódulo vocal, edema de Reinke e patologias neurológicas. No entanto, apesar dos resultados promissores, o número de gravações por grupo de estudo disponível na base de dados utilizada não ultrapassa 20, sendo um pouco limitado. Em trabalhos futuros, os resultados devem ser atualizados com uma maior população para garantir a generalização.

VI. CONCLUSÕES

Este trabalho investigou a detecção de distúrbios vocais relacionados com nódulo vocal, edema de Reinke e patologias neurológicas através de parâmetros cepstrais multibanda da vogal sustentada /a/. A detecção foi realizada entre pares de grupos de estudo e a análise multibanda foi efetuada usando a transformada Wavelet. Para cada par de grupos de estudo, uma seleção dos parâmetros foi realizada.

Estatísticas dos parâmetros selecionados foram utilizadas como entrada para quatro classificadores, nomeadamente DT, SVM, KNN cosseno e euclidiano, com validação cruzada *leave-one-out*. Acurácias de classificação de 100% foram obtidas para os pares Patologias/Controle, Nódulo/Controle, Edema/Controle, Neurológica/Controle e Nódulo/Neurológica. Para o par Edema/Neurológica, a maior acurácia obtida foi 96,70%. E uma acurácia de classificação de apenas 74,20% foi alcançada para o par Edema/Nódulo.

AGRADECIMENTOS

Os autores gostariam de agradecer o Grupo de Pesquisa em Engenharia Médica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (GPEM/CNPq) por fornecer a base de dados de voz. Esse trabalho foi apoiado pelos Fundos Nacionais da Fundação para a Ciência e a Tecnologia (FCT), Portugal, através do projeto UIDB/50016/2020.

REFERÊNCIAS

- [1] S. Krischke, S. Weigelt, U. Hoppe, V. Köllner, M. Klotz, U. Eysholdt, and F. Rosanowski, "Quality of life in dysphonic patients," *Journal of Voice*, vol. 19, no. 1, pp. 132–137, March 2005.
- [2] R. H. G. Martins, H. A. do Amaral, E. L. M. Tavares, M. G. Martins, T. M. Gonçalves, and N. H. Dias, "Voice disorders: Etiology and diagnosis," *Journal of Voice*, vol. 30, no. 6, pp. 761.e1–761.e9, November 2016.
- [3] A. Castellana, A. Carullo, S. Corbellini, and A. Astolfi, "Discriminating pathological voice from healthy voice using cepstral peak prominence smoothed distribution in sustained vowel," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 3, pp. 646–654, March 2018.
- [4] S. Bielamowicz, J. Kreiman, B. R. Gerratt, M. S. Dauer, and G. S. Berke, "Comparison of voice analysis systems for perturbation measurement," *Journal of Speech Language and Hearing Research*, vol. 39, no. 1, pp. 126–124, February 1996.
- [5] Y. Zhang and J. J. Jiang, "Acoustic analyses of sustained and running voices from patients with laryngeal pathologies," *Journal of Voice*, vol. 22, no. 1, pp. 1–9, January 2008.
- [6] S. Y. Lowell, R. H. Colton, R. T. Kelley, and Y. C. Hahn, "Spectral- and cepstral-based measures during continuous speech: capacity to distinguish dysphonia and consistency within a speaker," *Journal of Voice*, vol. 25, no. 5, pp. 223–232, September 2011.
- [7] S. Y. Lowell, R. H. Colton, R. T. Kelley, and S. A. Mizia, "Predictive value and discriminant capacity of cepstral- and spectral-based measures during continuous speech," *Journal of Voice*, vol. 27, no. 4, pp. 393–400, July 2013.
- [8] A. Benba, A. Jilbab, and A. Hammouch, "Analysis of multiple types of voice recordings in cepstral domain using MFCC for discriminating between patients with parkinson's disease and healthy people," *International Journal of Speech Technology*, vol. 19, no. 3, pp. 449–456, March 2016.
- [9] Z. Ali, M. Alsulaiman, G. Muhammad, I. Elamvazuthi, and T. A. Mesallam, "Vocal fold disorder detection based on continuous speech by using MFCC and GMM," in *2013 7th IEEE GCC Conference and Exhibition (GCC)*. IEEE, November 2013.
- [10] G. L. Ribeiro, R. Gomes, S. C. Costa, and W. A. Costa, "Análise mel-cepstral na discriminação de patologias laringeas," in *XXIV Congresso Brasileiro de Engenharia Biomédica*, October 2014.
- [11] J. R. Orozco-Arroyave, E. A. Belalcazar-Bolanos, J. D. Arias-Londono, J. F. Vargas-Bonilla, S. Skodda, J. Ruz, K. Daqrouq, F. Honig, and E. Noth, "Characterization methods for the detection of multiple voice disorders: Neurological, functional, and laryngeal diseases," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 6, pp. 1820–1828, November 2015.
- [12] H. Cordeiro, J. Fonseca, I. Guimaraes, and C. Meneses, "Voice pathologies identification speech signals, features and classifiers evaluation," in *2015 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*. IEEE, September 2015.
- [13] U. Rani K and M. S. Holli, "Wavelet transform features to hybrid classifier for detection of neurological-disordered voices," *Journal of Clinical Engineering*, vol. 42, no. 2, pp. 89–98, 2017.
- [14] R. Behroozmand and F. Almasganj, "Comparison of neural networks and support vector machines applied to optimized features extracted from patients~ speech signal for classification of vocal fold inflammation," in *Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology, 2005*. IEEE, 2005.
- [15] B. R. Kumar, J. S. Bhat, and N. Prasad, "Cepstral analysis of voice in persons with vocal nodules," *Journal of Voice*, vol. 24, no. 6, pp. 651–653, November 2010.
- [16] L. A. Tokuhara, B. C. Bispo, M. E. Dajer, E. A. R. Theodoro, and P. M. Rodrigues, "Avaliação da capacidade discriminante de distribuições da proeminência do pico cepstral para identificar nódulos vocais," in *Proceedings of XXXVII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, Petrópolis, Brazil, September 2019.
- [17] R. F. Coleman, "Sources of variation in phonetograms," *Journal of Voice*, vol. 7, no. 1, pp. 1–14, March 1993.
- [18] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Englewood Cliffs, New Jersey: Prentice Hall, 1995.
- [19] H. S. Malvar, *Signal Processing with Lapped Transforms*. Norwood, Massachusetts: Artech House, 1992.
- [20] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Processing Magazine*, vol. 8, no. 4, pp. 14–38, October 1991.
- [21] J. R. Deller Jr., J. H. L. Hansen, and J. G. Proakis, *Discrete-Time Processing of Speech Signals*. Piscataway, New Jersey: IEEE Press, 2000.
- [22] A. H. Gray Jr. and J. D. Markel, "Distance measures for speech processing," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, no. 5, pp. 380–391, October 1976.
- [23] P. M. Rodrigues, D. Freitas, J. P. Teixeira, B. Bispo, D. Alves, and C. Garrett, "Electroencephalogram hybrid method for alzheimer early detection," *Procedia Computer Science*, vol. 138, pp. 209–214, 2018.
- [24] Y. Tohkura, "A weighted cepstral distance measure for speech recognition," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 10, pp. 1414–1422, October 1987.
- [25] K. K. Paliwal, "On the performance of the quefrency-weighted cepstral coefficients in vowel recognition," *Speech Communication*, vol. 1, no. 2, pp. 151–154, August 1982.
- [26] C. T. Nakas and C. T. Yiannoutsos, "Ordered multiple-class ROC analysis with continuous measurements," *Statistics in Medicine*, vol. 23, no. 22, pp. 3437–3449, November 2004.