

Classificação de desvios vocais utilizando o método de decomposição em modos empíricos

Samuel Ribeiro de Abreu, Suzete Élide N. Correia, Silvana Luciene do N. C. Costa

Resumo—Desvios vocais são frequentemente monitorados por especialistas em voz para avaliação da qualidade vocal e tratamento fonoterápico. Um método para análise de sinais não estacionários é o método de decomposição em modos empíricos. Este estudo visa a classificação de sinais de voz como saudável ou com desvios rugosidade ou sopro. Foram empregadas características temporais e espectrais das funções de modo intrínseco extraídas dos sinais. Para a classificação, foi utilizada uma rede neural *multilayer perceptron* com algoritmo de aprendizado supervisionado do gradiente conjugado escalonado. Uma acurácia de 95,00% foi obtida na detecção entre sinais de vozes saudáveis e soprosos.

Palavras-Chave—Desvios Vocais, Decomposição em Modos Empíricos, Rede Neural.

Abstract—Vocal deviations are often monitored by voice specialists for voice quality assessment and speech therapy treatment. A method for analysis of non-stationary signals is the empirical mode decomposition method. This study aims to classify vocal signals as healthy or containing roughness and breathiness. Temporal and spectral characteristics of the intrinsic functions extracted from the signals were employed. For the classification, a multilayer perceptron neural network was used with supervised learning of algorithm scaled conjugate gradient. A 95 % accuracy was obtained by detection between healthy and breathiness signals.

Keywords—Vocal Deviations, Empirical Mode Decomposition, Neural Network.

I. INTRODUÇÃO

A avaliação de qualidade da voz é um elemento importante para o diagnóstico de desvio vocal. Até hoje, a avaliação perceptual auditiva ainda é o principal método de medição da qualidade da voz no ambiente clínico, a qual é realizada por especialistas treinados e seu julgamento depende da experiência do profissional [1].

Nos últimos anos vem aumentando o interesse dos pesquisadores e profissionais que tratam a voz (fonoaudiólogo e otorrinolaringologista) pelo método de análise acústica para avaliação da qualidade vocal, criação de ferramentas para o auxílio de diagnósticos de desordens vocais e classificadores de patologias laríngeas. Este tipo de análise pode auxiliar o profissional nas tomadas de decisões para a determinação de alterações da voz, avaliações de cirurgias, tratamentos medicamentoso e de reabilitação. A escolha acertada do método para identificar o grau dos desvios, ou a presença de patologias laríngeas, ainda é objeto de estudo [2].

Samuel Ribeiro de Abreu, Suzete Élide N. Correia, Silvana Luciene do N. C. Costa, Programa de Pós-graduação em Engenharia Elétrica, Instituto Federal da Paraíba (IFPB), João Pessoa-PB, Brasil, E-mail: abreu.samuel@academico.ifpb.edu.br, suzete@ifpb.edu.br, silvana@ifpb.edu.br.

Técnicas de análise de sinais em tempo-frequência constituem um importante avanço na análise de sinais. Tais técnicas são apropriadas aos sinais biomédicos, pois são sinais que se caracterizam por conteúdos em frequências, são não lineares e não estacionários [3].

Um método para análise de sinais não estacionários é o método de decomposição em modos empíricos [4], o qual tem sido extensamente aplicado em muitas áreas da ciência e engenharia, incluindo engenharia biomédica [5]. A decomposição em modos empíricos decompõe os sinais em funções do tempo, do qual a informação espectral também pode ser obtida através do espectro de *Hilbert* [6].

Em Kallem et al. [6] foi realizado um estudo para classificar sinais entre vozes saudáveis e patológicas utilizando a decomposição em modos empíricos. Nesta pesquisa, foram analisadas características temporais e espectrais dos sinais, através do espectro de *Hilbert*, obtendo uma classificação de 95,70% de acurácia.

Já em Yu et al. [7], foi utilizada a decomposição em modos empíricos para a análise e classificação de sinais de vozes (vogal sustentada /a/) diagnosticadas de acordo com a intensidade do desvio vocal, na escala GRBAS. Este estudo teve nos seus resultados a melhor classificação de 81,43% de acurácia.

No presente estudo, a decomposição em modos empíricos é empregada para detectar os desvios vocais rugosidade (irregularidade na vibração das pregas vocais) e sopro (escape de ar audível na voz) em sinais de voz. Características temporais e espectrais foram extraídas das funções de modo intrínseco. Para classificação foi empregada a rede neural *multilayer perceptron* com o algoritmo de aprendizado supervisionado do gradiente conjugado escalonado [8], a fim de detectar padrões de desvios vocais.

O artigo está organizado da seguinte forma: são apresentados, inicialmente, na Seção II, a transformada *Hilbert-Huang*, na Seção III materiais e métodos e na Seção IV os resultados, seguidos das conclusões, na Seção V.

II. TRANSFORMADA DE *Hilbert-Huang*

A transformada de *Hilbert-Huang* (HHT, do inglês *Hilbert-Huang Transform*) é um método de análise em tempo-frequência e é composto por duas etapas: a primeira decompõe o sinal em funções de modo intrínseco (IMF, do inglês *Intrinsic Mode Functions*) utilizando o método de decomposição em modos empíricos (EMD, do inglês *Empirical Mode Decomposition*) que permite obter um conjunto de funções base de forma adaptativa; a segunda é a análise do espectro

de *Hilbert* (HSA, do inglês *Hilbert Spectral Analysis*) que possibilita a obtenção de frequência instantânea do sinal [4].

Método de Decomposição em Modos Empíricos - É uma técnica adaptativa que permite a decomposição de sinais não lineares e não estacionários em funções de modo intrínseco. Uma função de modo intrínseco tem que satisfazer duas condições: i) o número de máximos ou mínimos locais é igual ao número de vezes que a função cruza o zero ou diferente de uma unidade; ii) em qualquer instante de tempo, a média local da envoltória formada pela interpolação do extremos (máximos e mínimos) é igual a zero [4].

A essência da EMD é localizar localmente, de forma empírica, todos os modos de oscilações através das características temporais do sinal. O processo sistemático para extrair as IMFs é descrito a seguir [4]:

- 1) Identificar todos os mínimos e máximos de um determinado sinal de tempo discreto $x[n]$;
- 2) Obter as envoltórias superior (E_{sup}) e inferior (E_{inf}), interpolando-se os pontos de máximos e mínimos locais, respectivamente. Para isto adota-se a interpolação polinomial de terceiro grau utilizando o método *splines*;
- 3) Calcular a média $E_{med} = \frac{E_{sup} + E_{inf}}{2}$ das envoltórias superior e inferior e o sinal candidato a IMF $h[n]$ é obtido através de: $h[n] = x[n] - E_{med}$;
- 4) As três etapas anteriores são executadas em um laço, chamado de *sifting*, até que $h[n]$ seja reduzido a uma IMF $c1[n]$, que deve está em conformidade com as duas condições para existência da IMF, que foram descritas anteriormente. A primeira IMF contém as maiores frequências de oscilação dos dados originais de $x[n]$;
- 5) A primeira IMF $c1[n]$ é subtraída de $x[n]$ para obter o resíduo $r1[n]$;
- 6) O resíduo $r1[n]$ será o novo sinal para o ponto de partida, as etapas mencionadas anteriormente são repetidas para encontrar todas as IMFs $c_i[n]$ e para que o resíduo r_K ou se torne uma constante, ou uma função monótona, ou uma função com um único máximo e mínimo do qual nenhuma outra IMF pode ser extraída.

Portanto, no final da decomposição, $x[n]$ é representado como a soma de K IMFs e um resíduo r_K , conforme apresentado na Equação 1 [6].

$$x[n] = \sum_{i=1}^K c_i[n] + r_K[n] \quad (1)$$

Amplitude e Frequência Instantânea - A HHT é aplicada nas IMFs obtidas de modo a obter a amplitude e a frequência instantânea do sinal [4]. Primeiro, o sinal analítico z_i correspondente a cada IMF é obtido, como demonstrado pela Equação 2 [6].

$$z_i[n] = c_i[n] + jH(c_i[n]) \quad (2)$$

Em que $H(c_i[n])$ é a transformada de *Hilbert* de $c_i[n]$. Reescrevendo a Equação 2 para forma polar [6]:

$$z_i[n] = a_i[n]e^{j\theta_i[n]} \quad (3)$$

A amplitude instantânea é representada por a_i e a fase instantânea é representada por θ_i . A amplitude instantânea e

fase instantânea são calculadas conforme as Equações 4 e 5, respectivamente.

$$a_i[n] = \sqrt{c_i^2[n] + H^2(c_i[n])} \quad (4)$$

$$\theta_i[n] = \text{tg}^{-1} \frac{H(c_i[n])}{c_i[n]} \quad (5)$$

A frequência instantânea, ω_i , pode ser obtida conforme Equação 6 [6].

$$\omega_i[n] = \frac{d\theta_i[n]}{dn} \quad (6)$$

Uma vez que a HHT tenha sido aplicada a todos as IMFs, e calculadas as amplitudes e frequências instantâneas, o sinal original $x[n]$ pode ser expresso conforme a Equação 7 [6].

$$x[n] = \sum_{i=1}^K a_i[n]e^{j\omega_i[n]d[n]} \quad (7)$$

A parte real da Equação 7 permite uma distribuição tempo-frequência-amplitude aliada ao espectro de *Hilbert*, $H[\omega, n]$, que é expresso por [6]:

$$H[\omega, n] = \text{Re} \sum_{i=1}^K a_i[n]e^{j\omega_i[n]d[n]} \quad (8)$$

III. MATERIAIS E MÉTODOS

A. Base de Dados

O banco de dados dos sinais utilizados neste estudo é proveniente do Laboratório Integrado de Estudos da Voz da Universidade Federal da Paraíba, com o parecer de número 52492/12, aprovado pelo Comitê de Ética da UFPB [9].

A coleta da voz foi realizada em uma cabine acústica, com ruído ambiental inferior a 50 dB NPS. A taxa de amostragem dos sinais é de 44.100 amostras/s. As vozes foram editadas no *software Sound Forge* versão 10.0, sendo eliminados os dois segundos iniciais e finais da emissão da vogal, devido a maior irregularidade nesses trechos. A normalização foi realizada no controle “*normalize*” do *Sound Forge*, a fim de obter uma padronização na saída de áudio entre -6 e 6 dB [9].

Da base de dados foram selecionados sinais de vozes masculinas e femininas totalizando 87 sinais, sendo 30 sinais de vozes saudáveis e 57 sinais de vozes com desvio, sendo estes, 29 sinais de vozes com o desvio rugosidade e 28 sinais com desvio soprosidade. Para a análise acústica, foram selecionados trechos de 2 segundos dos sinais.

B. Metodologia

Para cada sinal de voz da base de dados foram obtidas as IMFs através do método EMD, como descrito na Seção II. Realizados os cálculos para obter as IMFs, foram calculadas as amplitudes e frequências instantâneas de acordo com as Equações 2-8. Na etapa seguinte foi realizada a extração de características. Como cada sinal gerou um número de IMFs diferentes, decidiu-se então trabalhar com as dez primeiras IMFs de cada sinal, tomando como base o estudo de Kaleem et al. [6]. A Figura 1 ilustra um conjunto de 10 IMFs obtidas para um sinal de voz saudável, rugosa e soprosa,

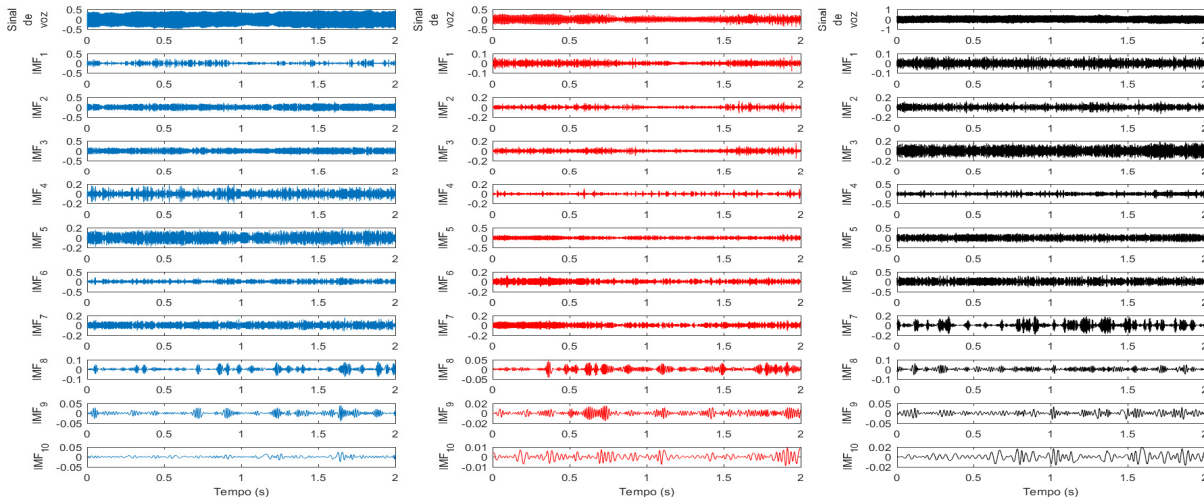


Fig. 1. Três sinais de vozes decompostos em 10 IMFs. Em azul está o sinal com voz saudável; em vermelho está o sinal de voz com rugosidade; e em preto o sinal de voz com sopro.

respectivamente. Após a extração de características, os sinais foram classificados e, por fim, os resultados foram obtidos.

Para a decomposição dos sinais em IMFs foram usadas funções da biblioteca disponibilizadas por Flandrin et al. [10]. Para calcular as amplitudes e frequências instantâneas foram usadas funções disponibilizadas por Auger et al. [11].

Extração de características - Foram extraídas características temporais e espectrais de cada IMF. A primeira característica extraída foi a energia do sinal, que representa a intensidade sonora do sinal [2]. A Equação 9 mostra como é calculada a energia (E_i) do sinal para cada IMF.

$$E_i = \sum_{n=1}^N c_i^2[n], i = 1, 2, 3, \dots, 10 \quad (9)$$

Para as Equações de 9-13, N representa o número total de amostras da IMF (cada IMF possui 88.200 amostras) e o índice i representa a IMF associada. A energia foi escolhida como característica, pois a amplitude do sinal em vozes saudáveis tem menos variações do que em sinais com desvios [6].

Devido a voz com desvio sopro ter ruídos em frequências mais altas [12], espera-se que estes ruídos se espalhem pelo espectro HSA. A Figura 2 (a) mostra o espectro HSA da IMF₂ para os sinais de vozes saudáveis, com desvio rugosidade e com desvio sopro. Pode ser observado que, para esta IMF, o sinal com desvio sopro tem mais ruídos em altas frequências do que os sinais de vozes saudáveis e com o desvio rugosidade.

Já os ruídos das vozes com desvio rugosidade ocorrem em frequências mais baixas [12]. A Figura 2 (b) mostra o espectro HSA da IMF₆ para os sinais de vozes saudável, com desvio rugosidade e com desvio sopro. Pode ser observado que, para esta IMF, o sinal com desvio rugosidade tem mais ruídos em baixas frequências do que os sinais de vozes saudáveis e com desvio sopro.

Com base nas informações do comportamento dos sinais de vozes na escala energia-tempo-frequência, foram extraídas quatro características do espectro HSA de cada IMF, duas em

relação a contribuição do tempo e duas em relação a contribuição da frequência. Com relação às contribuições referentes ao tempo, tem-se a dispersão da densidade de energia temporal instantânea ($DETI$), que é demonstrada pela Equação 10 [13].

$$DETI_i = \frac{1}{N} \sum_{n=1}^N |a_i[n] - \bar{a}_i|^2, i = 1, 2, 3, \dots, 10 \quad (10)$$

Na Equação 10, \bar{a} representa a média aritmética temporal. Nas Equações 10, 11 e 13, a representa amplitude instantânea.

A outra característica analisada é a amplitude instantânea ponderada pelo tempo (APT), apresentada na Equação 11 [6].

$$APT_i = \frac{1}{N} \sum_{n=1}^N (n \times a_i[n]), i = 1, 2, 3, \dots, 10 \quad (11)$$

Com relação às contribuições referentes à frequência, tem-se a dispersão da frequência instantânea (DFI), descrita pela Equação 12 [13].

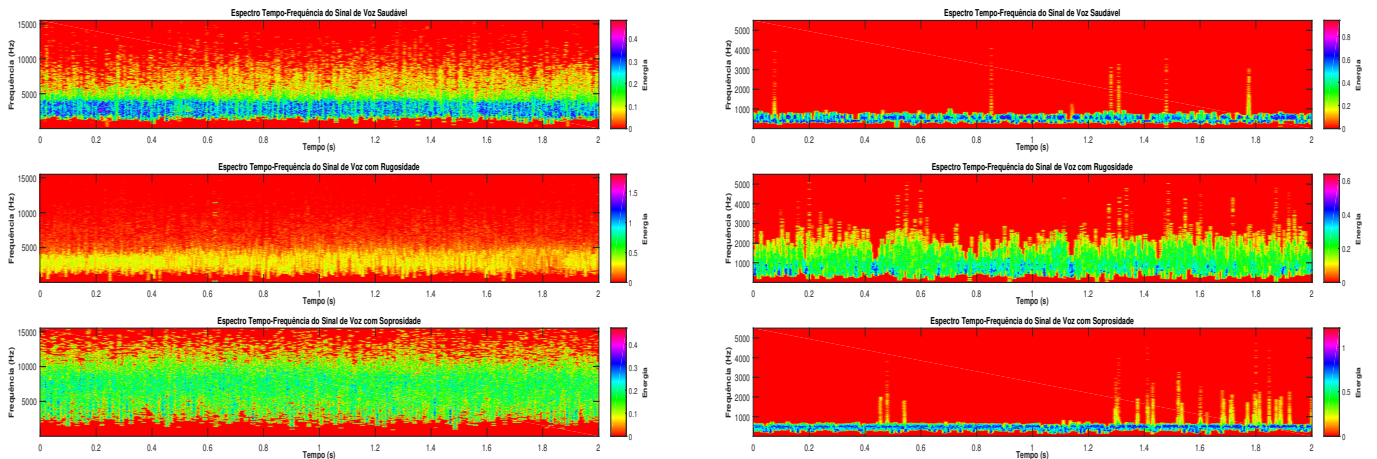
$$DFI_i = \frac{1}{N} \sum_{n=1}^N |\omega_i[n] - \bar{\omega}_i|^2, i = 1, 2, 3, \dots, 10 \quad (12)$$

Na Equação 12, $\bar{\omega}$ representa a média aritmética da frequência instantânea. Nas Equações 12 e 13, ω representa frequência instantânea.

A outra característica analisada é a amplitude instantânea ponderada pela frequência instantânea (APF). A Equação 13 demonstra esta relação [6].

$$APF_i = \frac{1}{N} \sum_{n=1}^N (a_i[n] \times \omega_i[n]), i = 1, 2, 3, \dots, 10 \quad (13)$$

Classificação - Após extrair as cinco características para cada IMF de todos os sinais, foi realizada uma análise estatística do potencial discriminativo de cada característica com o intuito de identificar as medidas mais relevantes para cada caso de classificação, proporcionando uma redução na dimensionalidade dos dados. Nesta análise, foram realizados



(a) Espectro de *Hilbert* em relação a IMF_2 dos sinais de vozes saudável, rugoso e soproso.

(b) Espectro de *Hilbert* em relação a IMF_6 dos sinais de vozes saudável, rugoso e soproso.

Fig. 2. Exemplo da análise do espectro de *Hilbert* do sinais de vozes saudáveis, com desvio rugosidade e desvio soprosoidade.

os testes de *Kolmogorov-Smirnov* para verificar a normalidade da característica, *t-student* para testes paramétricos e *Mann-Whitney* para testes não paramétricos. O nível de significância considerado foi de 5% [14].

As Tabelas I e II mostram, através do valor-p, os resultados do potencial discriminativo de cada IMF para o estudo de caso: i) saudável \times soproso e ii) saudável \times rugoso. Somente as características que tiveram a hipótese nula rejeitada (em negrito nas Tabelas I e II) foram usadas no classificador.

TABELA I
POTENCIAL DISCRIMINATIVO SAUDÁVEL \times SOPROSO.

Característica	IMF_1	IMF_2	IMF_3	IMF_4	IMF_5
E_i	0,056	0,002	0,003	0,440	0,351
$DETI_i$	0,068	0,001	0,002	0,850	0,307
APT_i	0,446	0,004	0,004	0,374	0,512
DFI_i	0,088	0,201	0,015	0,593	0,073
APF_i	0,060	0,097	0,028	0,056	0,210
Característica	IMF_6	IMF_7	IMF_8	IMF_9	IMF_{10}
E_i	0,752	0,283	0,430	0,783	0,167
$DETI_i$	0,720	0,926	0,134	0,575	0,217
APT_i	0,779	0,201	0,700	0,567	0,299
DFI_i	0,297	0,408	0,114	0,506	0,014
APF_i	0,241	0,236	0,053	0,076	0,140

TABELA II
POTENCIAL DISCRIMINATIVO SAUDÁVEL \times RUGOSO.

Característica	IMF_1	IMF_2	IMF_3	IMF_4	IMF_5
E_i	0,544	0,096	0,105	0,709	0,010
$DETI_i$	0,488	0,088	0,375	0,236	0,020
APT_i	0,589	0,278	0,132	0,651	0,002
DFI_i	0,995	0,183	0,075	0,011	0,448
APF_i	0,008	0,933	0,170	0,133	0,001
Característica	IMF_6	IMF_7	IMF_8	IMF_9	IMF_{10}
E_i	0,053	0,181	0,732	0,632	0,937
$DETI_i$	0,014	0,166	0,875	0,813	0,447
APT_i	0,027	0,457	0,863	0,664	0,989
DFI_i	0,665	0,426	0,212	0,146	0,003
APF_i	0,022	0,093	0,260	0,400	0,626

O classificador utilizado foi o *multilayer perceptron* com o algoritmo de aprendizado supervisionado do gradiente conjugado escalonado. Os sinais de vozes utilizados foram separados, aleatoriamente, em conjuntos distintos, dos quais $\frac{2}{3}$ foram utilizados para treinamento e $\frac{1}{3}$ para teste, com validação cruzada com dez testes distintos. Foram utilizados 5 neurônios na camada oculta e 2 na camada de saída. Na camada de entrada, a quantidade de neurônios variou de acordo com o número de características empregadas, podendo ser de 1 a 10 características.

O desempenho do classificador foi medido usando três medidas de desempenho: acurácia (ACU), sensibilidade (SENS) e a especificidade (ESP). A acurácia mede a capacidade do teste de identificar corretamente quando há e quando não há presença do desvio vocal. A sensibilidade é a capacidade do teste em identificar corretamente o desvio vocal. A especificidade é a capacidade do teste em excluir corretamente aqueles que não possuem o desvio vocal [2].

IV. RESULTADOS

Nas Tabelas III e IV estão apresentados os melhores resultados da classificação entre vozes saudáveis e soproso, obtidos com medidas individuais e combinadas. Essas combinações foram feitas uma a uma com a intenção de obter o melhor resultado para classificação de padrões desses sinais.

TABELA III
MELHOR DESEMPENHO DE MEDIDAS INDIVIDUAIS NA CLASSIFICAÇÃO ENTRE SINAIS DE VOZES SAUDÁVEIS E SOPROSO.

Característica	ACU (%)	SENS (%)	ESP (%)
E_2	76,00 \pm 4,35	89,16 \pm 5,83	72,00 \pm 5,21
E_3	75,33 \pm 4,84	92,50 \pm 5,33	71,00 \pm 5,95
$DETI_2$	74,66 \pm 5,64	87,50 \pm 6,71	73,00 \pm 6,46
$DETI_3$	74,33 \pm 4,66	87,66 \pm 6,40	72,66 \pm 5,51
APT_2	73,00 \pm 5,60	76,16 \pm 6,87	79,33 \pm 4,11
APT_3	79,00 \pm 5,64	83,33 \pm 5,95	80,00 \pm 6,93
DFI_3	70,66 \pm 6,14	75,83 \pm 7,07	70,00 \pm 6,93
DFI_{10}	75,66 \pm 4,60	86,66 \pm 5,44	71,16 \pm 5,22
APF_3	76,00 \pm 4,65	86,67 \pm 5,74	76,00 \pm 5,69

TABELA IV

MELHORES DESEMPENHOS DE MEDIDAS COMBINADAS NA CLASSIFICAÇÃO ENTRE SINAIS DE VOZES SAUDÁVEIS E SOPROSOS.

Características	ACU (%)	SENS (%)	ESP (%)
$E_2 + APT_3$	88,00±2,63	85,00±4,08	97,50±2,50
$E_2 + APT_2 + APT_3$	95,00±3,55	97,50±2,50	96,00±4,00
$E_2 + APT_2 + APT_3 + APF_3$	95,00±2,54	97,50±2,50	95,00±3,33
$E_2 + E_3 + APT_2 + APT_3 + APF_3$	93,33±2,72	95,00±3,33	95,00±3,33

As medidas que mais se destacaram para classificação saudável \times soproso, foram: energia da IMF₂ (E_2), energia da IMF₃ (E_3), amplitude instantânea ponderada pelo tempo da IMF₂ (APT_2), amplitude instantânea ponderada pelo tempo da IMF₃ (APT_3) e amplitude instantânea ponderada pela frequência instantânea da IMF₃ (APF_3). Os melhores resultados foram das combinações 3 a 3 e 4 a 4, obtendo 95% de acurácia. Observa-se que as características que se destacaram estão relacionadas as IMF₂ e IMF₃, relacionadas as altas frequências.

Nas Tabelas V e VI estão apresentados os melhores resultados da classificação entre vozes saudáveis e rugosas, obtidos com medidas individuais e combinadas.

TABELA V

MELHOR DESEMPENHO DE MEDIDAS INDIVIDUAIS NA CLASSIFICAÇÃO DOS SINAIS DE VOZES SAUDÁVEIS E RUGOSOS.

Característica	ACU (%)	SENS (%)	ESP (%)
E_5	76,00±4,65	91,66±5,59	70,16±4,63
$DETI_5$	72,66±2,97	89,16±4,48	73,16±6,17
$DETI_6$	71,33±4,97	81,00±6,86	75,50±6,15
APT_5	78,00±4,97	81,83±6,36	80,00±5,85
APT_6	71,66±5,58	78,50±7,92	75,83±7,05
DFI_4	71,33±4,92	75,00±6,45	75,16±2,20
DFI_{10}	70,66±7,45	71,00±7,33	83,33±8,90
APF_1	77,66±5,21	87,66±6,40	76,66±5,91
APF_5	78,00±4,97	85,16±6,35	79,33±6,05
APF_6	73,00±5,02	73,66±6,34	81,66±6,67

TABELA VI

MELHORES DESEMPENHOS DE MEDIDAS COMBINADAS NA CLASSIFICAÇÃO ENTRE SINAIS DE VOZES SAUDÁVEIS E RUGOSOS.

Características	ACU (%)	SENS (%)	ESP (%)
$APT_5 + APF_1$	85,00±5,24	89,16±5,83	84,16±5,25
$E_5 + APT_5 + APF_1$	84,33±6,06	86,67±5,85	86,00±5,18
$E_5 + DETI_6 + APT_5 + APF_1$	81,33±4,61	86,00±5,13	85,00±5,52
$E_5 + DETI_6 + APT_5 + APF_1 + APF_5$	85,33±3,83	95,00±3,33	84,50±5,44

As medidas que mais se destacaram para classificação saudável \times rugoso, foram: energia da IMF₅ (E_5), dispersão da densidade de energia temporal instantânea da IMF₆ ($DETI_6$), amplitude instantânea ponderada pelo tempo da IMF₅ (APT_5), amplitude instantânea ponderada pela frequência instantânea da IMF₁ (APF_1) e amplitude instantânea ponderada pela frequência instantânea da IMF₅ (APF_5). Os melhores resultados ficaram em média com 85,00% de acurácia. Observa-se que as características que se destacaram estão relacionadas as IMF₅ e IMF₆, relacionadas as baixas frequências.

A contribuição deste estudo está na quantidade de características utilizadas no vetor de classificação para a obtenção

dos resultados, pois para atingir a acurácia de 95,00% na classificação entre vozes saudáveis e soprosas foi utilizado um vetor com 3 características e para atingir 85,00% de acurácia na classificação entre vozes saudáveis e rugosas foi utilizado um vetor com 2 características. Quando comparados com Kaleem et al. [6] que para chegar em seus resultados utilizaram um vetor com 47 características e Yu et al. [7] que utilizaram um vetor de 65 características.

V. CONCLUSÕES

O método EMD mostrou-se eficiente, obtendo como melhor resultado de classificação uma acurácia 95,00% na detecção entre sinais saudáveis e soprosos. Com estes resultados, pode-se concluir que o método EMD pode ser considerado na análise acústica como uma ferramenta auxiliar na fonoaudiologia.

AGRADECIMENTOS

Ao IFPB, ao Departamento de Fonoaudiologia da UFPB pela disponibilidade do banco de dados e a CAPES.

REFERÊNCIAS

- [1] J. Kreiman, B. R. Gerratt and K. Precod, "Listener experience and perception of voice quality", *Journal of Speech, Language, and Hearing Research*, vol. 33, no. 1, pp. 103–115, 1990.
- [2] S. L. N. C. Costa, "Análise Acústica, Baseada no Modelo Linear de Produção da Fala, para discriminação de vozes patológicas", Ph.D. Dissertation, Departamento de Engenharia Elétrica, Universidade de Federal de Campina Grande, Campina Grande, 2008.
- [3] R. M. F. Pinto, "Novas Abordagens ao Estudo de Sinais Biomédicos: Análise em Tempo-Frequência e Transformada de Hilbert-Huang", Ph.D. Dissertation, Faculdade de Ciências, Universidade de Lisboa, Lisboa, 2011.
- [4] N.E. Huang et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis", *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, vol. 454, no. 1971, pp. 903–995, March 1998.
- [5] B. Mijovic et al., "Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis", *IEEE transactions on biomedical engineering*, vol. 57, no. 9, pp. 2188–2196, 2010.
- [6] M. Kaleem et al., "Pathological speech signal analysis and classification using empirical mode decomposition", *Medical biological engineering computing*, vol. 51, no. 7, pp. 811–821, 2013.
- [7] P. Yu et al., "Multidimensional acoustic analysis for voice quality assessment based on the GRBAS scale", In. *Chinese Spoken Language Processing (ISCSLP), 2014 9th International Symposium on*, IEEE, pp. 321–325, 2014.
- [8] G. K. L. P. Queiroz, "Análise Dinâmica não linear e Análise de Quantificação de Recorrência Aplicadas na Classificação de Desvios Vocais", M.S. thesis, Faculdade de Engenharia, Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, João Pessoa, 2017.
- [9] L. W. Lopes et al., "Effectiveness of recurrence quantification measures in discriminating patients with and without voice disorders", In. *10th International Conference on Voice Physiology and Biomechanics*, pp. 14–17, 2016.
- [10] P. Flandrin et al., "Empirical Mode Decomposition", 2007, Disponível em: <http://perso.ens-lyon.fr/patrick.flandrin/emd.html>. Acesso em: 10 de fevereiro de 2018.
- [11] F. Auger et al., "Time-frequency toolbox", 1995, Disponível em: <http://tftb.nongnu.org>. Acesso em: 03 de março de 2018.
- [12] L. W. Lopes et al., "A medida HNR: sua relevância na análise acústica da voz e sua estimação precisa", In. *Anais da 1 Jornadas sobre Tecnologia e Saúde*, 2008.
- [13] M. F. Kaleem et al., "Telephone-quality pathological speech classification using empirical mode decomposition", In. *Engineering in Medicine and Biology Society (EMBC), 2011 Annual International Conference of the IEEE*, IEEE, pp. 7095–7098, 2011.
- [14] V. J. D. Vieira, "Avaliação de Distúrbios da Voz por meio de Análise de Quantificação de Recorrência", M.S. thesis, Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, João Pessoa-PB, 2014.