

Classificação de sinais de voz através da aplicação da transformada *Wavelet Packet* e redes neurais artificiais

Augusto F. T. da Silva ^[1], Samuel R. de Abreu ^[2], Silvana Cunha Costa ^[3], Suzete E. N. Correia ^[4]

[1] augustofdc92@gmail.com. [2] abreu.s.ribeiro@gmail.com. [3] silvana@ifpb.edu.br. [4] suzete@ifpb.edu.br. Instituto Federal da Paraíba, Campus João Pessoa.

RESUMO

Patologias como edemas, nódulos e paralisia são bastante recorrentes e influenciam diretamente as disfunções vocais. A análise acústica tem sido empregada para avaliar as desordens causadas nos sinais de voz, detectando a presença de patologias na laringe, por meio de técnicas de processamento digital de sinais. Este trabalho tem como objetivo distinguir sinais de voz, entre saudáveis e afetados por patologias laríngeas, empregando a transformada *Wavelet Packet* na etapa de extração de características. Medidas de energia e entropia, em seis níveis de resolução, obtidas através da *wavelet* de *Daubechies* de ordem 4, são usadas na discriminação dos sinais de voz. A classificação é feita por meio de Redes Neurais Artificiais. Foram obtidas taxas de acurácia acima de 90%, com a medida de entropia, na discriminação entre vozes saudáveis e afetadas por patologias nas pregas vocais (nódulos, edema de Reinke e paralisia).

Palavras-chave: Análise acústica. Transformada *Wavelet Packet*. Redes Neurais Artificiais. Patologias Laríngeas.

ABSTRACT

Pathologies such as edema, nodules and paralysis are quite recurrent and directly influence vocal dysfunctions. The acoustic analysis has been used to evaluate the disorders caused in the voice signals, detecting the presence of pathologies in the larynx, through digital signal processing techniques. This work aims to distinguish healthy voice signals from the ones affected by laryngeal pathologies, using the Wavelet Packet transform in the feature extraction step. Energy and entropy measures, in six resolution levels, obtained through the Daubechies wavelet of order 4 are used in the discrimination of the voice signals. The classification is done through Artificial Neural Networks. Accuracies above 90% were obtained, with the entropy measure, in the discrimination between healthy voices and affected ones by pathologies in the vocal folds (nodules, Reinke's edema and paralysis).

Keywords: *Acoustic Analysis. Wavelet Packet Transform. Artificial Neural Networks. Laryngeal Pathologies.*

1 Introdução

Muitos sinais na natureza, particularmente os biológicos, tais como sinais de voz, imagens e eletrocardiogramas, possuem características não estacionárias, ou seja, suas propriedades variam com o tempo, como fluxos, tendências, mudanças abruptas, início e final de um evento etc. (AKAY, 1998; CASTELANO, 2006).

Técnicas de processamento de sinais representam, descrevem e processam sinais através da combinação linear de funções elementares. A análise de Fourier utiliza uma série de senos e cossenos de diferentes frequências. Essa é uma técnica clássica que apresenta duas alternativas para análise de sinais: no domínio do tempo e no domínio da frequência, com o intuito de, através delas, se conhecerem as especificidades que caracterizam esses sinais (PINTO, 2009).

A representação temporal indica de que forma a amplitude do sinal varia com o tempo e, em certos casos, fornece algumas indicações acerca da variação do conteúdo em frequência do sinal. A representação no domínio da frequência permite averiguar a existência de diferentes componentes de frequência no sinal bem como obter informação sobre a magnitude e fase desses componentes (PINTO, 2009). Ao se aplicar a transformada de Fourier a um sinal, a informação do tempo é perdida, tendo-se apenas a informação da frequência (PARRAGA, 2002).

A transformada de Fourier não é indicada para representação de sinais não estacionários, tendo em vista que os componentes de frequência permanecem inalterados com o tempo. Para sinais estacionários, que não mudam com o tempo, esse problema é indiferente. Como características importantes dos sinais não estacionários encontram-se nas variações temporais, a análise de Fourier torna-se inadequada para esta detecção (CASTELANO, 2006).

A transformada *Wavelet* tem se mostrado uma ferramenta eficiente na análise de um grande número de sinais, devido a propriedades como: (i) permissão de análise simultânea de características no domínio do tempo e da frequência, (ii) a possibilidade de usar análise multirresolucional – que permite a análise de sinais em resoluções distintas, de modo que em cada escala aspectos diferentes possam ser observados; (iii) o fato de as *Wavelets* não serem únicas, ou seja, existem na literatura vários tipos dessas funções, que podem ser selecionadas de acordo com a aplicação; (iv) promoção de cálculo eficiente, através da filtra-

gem do sinal por filtros digitais passa-baixas e passa-altas (MALLAT, 1999; TANG, 2000).

A transformada *Wavelet Packet* proposta por Coifman e Wickerhauser (1992) é uma generalização da transformada *Wavelet* discreta, na qual a resolução tempo-frequência pode ser escolhida de acordo com o sinal. Na transformada *Wavelet* rápida, o sinal, no processo de filtragem, é dividido em coeficientes de aproximação, resultante da filtragem passa-baixas, e em coeficientes de detalhes, da filtragem passa-altas. Na decomposição, apenas os coeficientes de aproximação são novamente divididos em coeficientes de aproximação e detalhes. Na análise *Wavelet packet*, tanto os coeficientes de aproximação quanto os de detalhes são divididos, permitindo a análise de sinais em mais faixas de frequência (ALMEIDA, 2010).

Características baseadas na energia e na entropia de Shannon, obtidas a partir da análise com *Wavelets*, têm sido recentemente empregadas na avaliação de patologias laringeas. As *Wavelets* têm sido aplicadas de diversas maneiras, a exemplo da transformada *Wavelet* discreta (SANTOS *et al.*, 2014; CORREIA, COSTA, COSTA, 2011) e as *Wavelets packets* (AKBARI, ARJMANDI, 2014; ARJMANDI, POOYAN, 2012; ALMEIDA, 2010; BEHROOZMAND, ALMASGANJ, 2007).

O objetivo deste trabalho consiste em empregar a transformada *Wavelet Packet* na classificação de sinais de voz entre saudáveis e patológicos. Para tanto, foram extraídas medidas de energia e entropia dos coeficientes de detalhes e aproximação, em seis níveis de resolução, através da *Wavelet* de Daubechies de ordem 4. Redes neurais artificiais foram usadas para a discriminação dos sinais de voz.

O artigo encontra-se organizado como segue. Na Seção 2, há uma descrição acerca de sinais saudáveis e com patologias na laringe. Na Seção 3, é feita uma revisão bibliográfica e descrição das medidas obtidas por meio da transformada *Wavelet Packet*. Os materiais e métodos são apresentados na Seção 4, seguidos dos resultados, na Seção 5. Por fim, são apresentadas as conclusões na Seção 6.

2 Patologias Laríngeas

A voz tem uma grande importância para os seres vivos, pois, além de ser responsável por propiciar a comunicação verbal, ela também é capaz de demonstrar nossos sentimentos, como irritação, alegria, tristeza, confiança etc., além de características físicas,

como gênero e idade (MAGALHÃES *et al.*, 2008). Ela é resultado da interação harmônica de diversos sistemas fisiológicos que, a partir da passagem do fluxo de ar vindo dos pulmões, gera uma vibração nas pregas vocais, produzindo sinais acústicos de característica única para cada indivíduo (TOSI, 1979).

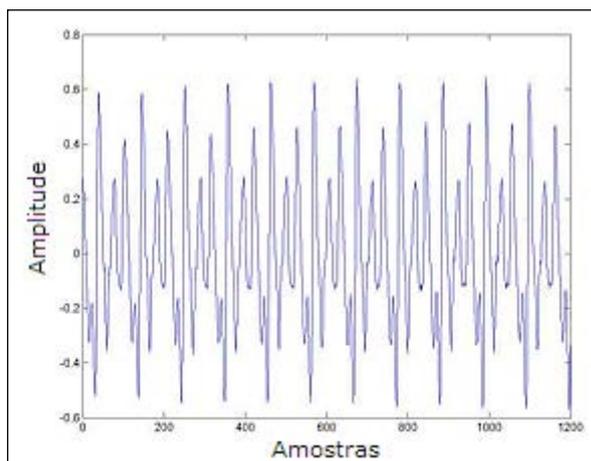
Tem sido observado que diversas patologias que ocorrem na laringe alteram a dinâmica do sistema de produção da fala, gerando mudanças nos padrões vibratórios nas pregas vocais, modificando a qualidade da voz. As Figuras de 1-4 ilustram, respectivamente, um sinal de voz saudável e afetado por edema de Reinke, nódulo e paralisia, nas pregas vocais.

Na Figura 1, sinal representativo de um trecho de 1200 amostras da vogal /a/ sustentada, observa-se

periodicidade e nível de amplitude maiores e mais regulares que no sinal das Figuras 2 e 3, cujos sinais de voz foram afetados pelas lesões de massa nas pregas vocais (edema de Reinke e nódulos), causando perda, tanto na periodicidade quanto na amplitude do sinal, provocada pelo esforço ao falar e pela vibração irregular das pregas vocais. A paralisia, patologia de origem neurológica, também provoca desordens no sinal de voz.

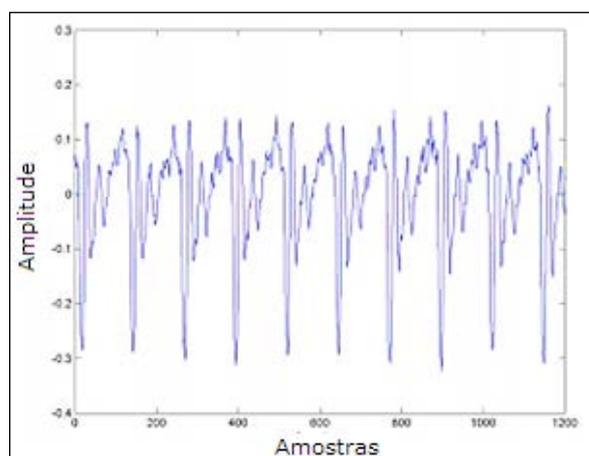
É importante verificar os efeitos da patologia laringea no conteúdo de frequências do sinal. Para tanto, nesta pesquisa, se fará uso da transformada *Wavelet Packet*, descrita a seguir.

Figura 1 – Sinal de voz saudável



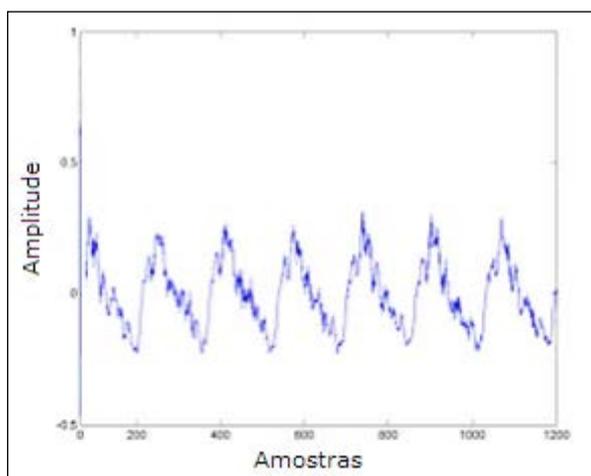
Fonte: Dados da pesquisa

Figura 3 – Sinal de voz afetado com nódulo nas pregas vocais



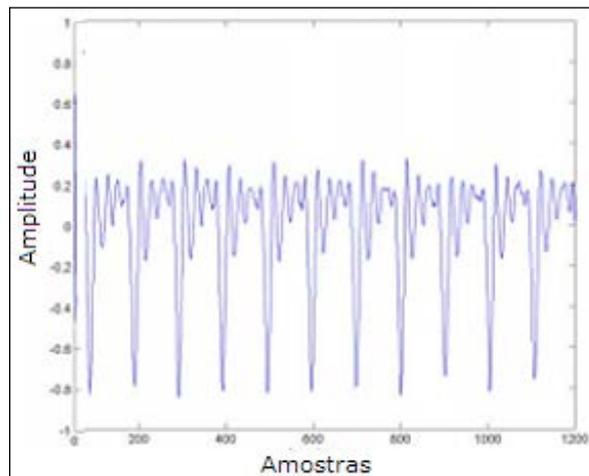
Fonte: Dados da pesquisa

Figura 2 – Sinal de voz afetado por edema de Reinke nas pregas vocais



Fonte: Dados da pesquisa

Figura 4 – Sinal de voz afetado por paralisia nas pregas vocais



Fonte: Dados da pesquisa

3 Transformada *Wavelet Packet*

A transformada *Wavelet Packet* tem sido utilizada nos últimos anos em processamento de sinais, permitindo melhorar a classificação desses sinais (PARRAGA, 2002). Na detecção de patologias laringeas, alguns trabalhos existentes na literatura, empregam *Wavelets Packets* como extratores de características e dos mais diversos tipos de classificadores, a exemplo de redes neurais artificiais e Máquinas de Vetor de Suporte.

A maioria dos autores utiliza a função custo entropia como a que apresenta melhores resultados. Parraga (2002) usa essa função custo aliada à família de *Wavelets Symlet* de ordem 5. Por sua vez, Shuck, Guimarges e Wisbe (2003), além de utilizarem essa mesma família *Wavelet* e função custo, também empregam redes neurais artificiais para resultados mais precisos.

Estudos mais recentes, como o de Arjmandi e Pooyan (2012), utilizam a energia e entropia, porém a *Wavelet* usada foi a *Daubechies* de ordem 4. A Análise Discriminante Linear (LDA) e a Análise de Componentes Principais (PCA) são usadas para reduzir a dimensão das características extraídas. Para a classificação dos sinais de voz, foi utilizado o algoritmo Máquina de Vetor de Suporte.

Em seu estudo, Behroozmand e Almasganj (2007) utilizam cinco níveis de decomposição para extrair medidas de energia e entropia a partir dos coeficientes da *Wavelet Packet*, além de utilizarem Máquinas de Vetor de Suporte para a classificação dos sinais, otimizando seus resultados. Almeida (2010) também utiliza o classificador Máquina de Vetor de Suporte (SVM) a partir da extração das características dos sinais de voz, através dos coeficientes da *Wavelet Packet*.

Akbari e Arjmandi (2014) analisam as medidas de energia e entropia de Shannon, extraídas dos sinais de voz, a partir dos coeficientes da *Wavelet Packet*, empregando a Análise Discriminante Linear Multiclasse e Redes Neurais Multicamadas

A transformada *Wavelet Packet* utiliza uma família de funções *Wavelet* e as suas funções de escalonamento associadas, para decompor o sinal original em sub-bandas subsequentes. O processo de decomposição é recursivamente aplicado nas sub-bandas de baixa e alta frequências para gerar o próximo nível da hierarquia (CHOUEITER; GLASS, 2004). Se uma base *Wavelet* ortonormal é escolhida, os coeficientes

calculados serão independentes e podem resultar em grupos distintos de vetores característicos (BURRUS; GOPINATH; GUO, 1998).

Na análise *Wavelet Packet*, tanto os coeficientes de aproximação quanto os de detalhes são subdivididos, de modo que, em cada resolução j , 2^j sub-bandas são obtidas, cada uma delas correspondendo a uma faixa de frequência (MALLAT, 1999).

A energia dos coeficientes *Wavelet Packet*, para cada sub-banda i da resolução j , é obtida por:

$$E_i = \frac{1}{n^2} \sum_{k=1}^n |C_{j,k}^i|^2 \tag{1}$$

, sendo n a quantidade de coeficientes da sub-banda, k o fator de translação e $C_{j,k}^i$ os coeficientes *Wavelet Packet*.

A entropia de Shannon pode ser calculada usando-se os coeficientes da *Wavelet Packet*, segundo a Equação 2. Esta medida irá avaliar o grau de desordem do sinal.

$$H_i = - \sum_{k=1}^n |C_{j,k}^i|^2 \log |C_{j,k}^i|^2 \tag{2}$$

4 Materiais e métodos

A metodologia do trabalho está de acordo com o diagrama em blocos ilustrado na Figura 5.

Figura 5 – Diagrama em blocos com a metodologia desenvolvida



Fonte: Elaborada pelos autores

4.1 Base de dados

Foram analisados 167 sinais, entre saudáveis e patológicos, da base de dados da *Disordered Voice Database, Model 4337*, da *Kay Elemetrics*, gravada pelo *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab*. Desse total, 53 eram sinais de vozes saudáveis (SDL), 44 com edema de Reinke (EDM), 18 com nódulos (NDL) e 52 com paralisia (PRL) nas pregas vocais. Todos os sinais foram amostrados a uma taxa de 25000 amostras/s.

4.2 Extração de características

Para cada sinal de voz proveniente da base de dados, foram extraídas medidas de energia e de entropia, obtidas a partir dos coeficientes de aproximação e detalhes da transformada *Wavelet Packet*. A família *Wavelet* adotada foi a *Daubechies* de ordem 4 (db4). Esses sinais foram classificados em saudáveis ou com patologias nas pregas vocais, com todas as patologias numa única classe. Os sinais foram, ainda, discriminados quanto ao tipo de patologia existente (edema de Reinke, nódulo e paralisia).

Foram escolhidos seis níveis de resolução, para a decomposição da transformada *Wavelet*, de modo a realizar a análise em faixas de frequência diferentes. O diagrama da decomposição está ilustrado até o terceiro nível, na Figura 6, em que $A(j,k)$ e $D(j,k)$ repre-

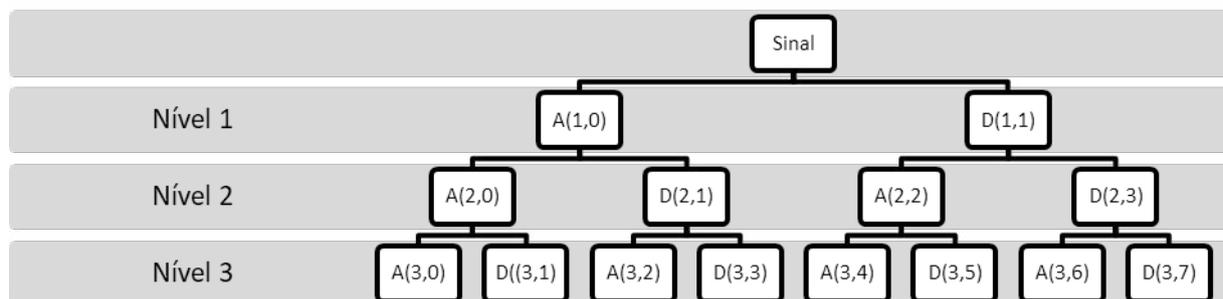
sentam os coeficientes de aproximação no j -ésimo nível de decomposição, respectivamente. O valor de k indica a ordem como estão dispostos os coeficientes no vetor de características em cada nível j .

4.3 Classificação

As medidas de energia e entropia extraídas são empregadas, individualmente, para a classificação de sinais de voz saudável e afetada por patologias nas pregas vocais. Para cada nível de resolução, uma rede neural com validação cruzada foi treinada. Os sinais de voz foram divididos, aleatoriamente, em 10 subconjuntos mutuamente exclusivos e de mesmo tamanho, um para teste e os demais para treino. O procedimento foi realizado dez vezes, alternando-se o subconjunto de teste e, ao final do processo, uma acurácia média foi calculada para avaliar o desempenho.

A quantidade de neurônios empregada na camada de entrada correspondeu à quantidade de medidas de energia ou entropia obtidas para o nível de resolução analisado. Dois neurônios foram empregados na camada de saída, um para cada classe, e a quantidade de neurônios na camada escondida foi determinada pela média aritmética entre os neurônios de entrada e saída.

Figura 6 – Decomposição da Transformada *Wavelet Packet* até o terceiro nível de decomposição.



Fonte: Dados da pesquisa

5 Resultados

Quatro casos de classificação foram considerados para a discriminação dos sinais de voz: Saudáveis x Patológicos (SDL x PTL), Saudáveis e afetados por Nódulos (SDL x NDL), Saudáveis e afetados por Edema de Reinke (SDL x EDM) e Sinais Saudáveis e afetados por Paralisia (SDL x PRL). A classe de sinais patológicos agrupa todos os sinais de voz afetados por patologias (Edema de Reinke, nódulos e paralisia).

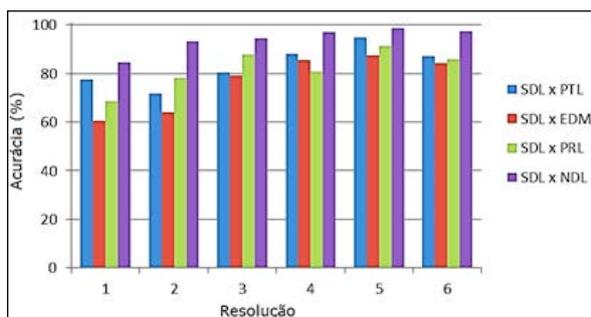
A Acurácia (Ac) foi utilizada como a medida de classificação dos sinais analisados, que é a relação de precisão entre o número de casos classificados de forma correta e o total de resultados da classificação, de forma percentual (COSTA, 2012).

Nas Figuras 7 e 8, são apresentados, em cada coluna, os valores de acurácia média, obtidas a partir da energia e da entropia extraídas dos coeficientes *Wavelet*, respectivamente. A cor de cada coluna re-

apresenta um caso de classificação realizado para seis níveis de decomposição.

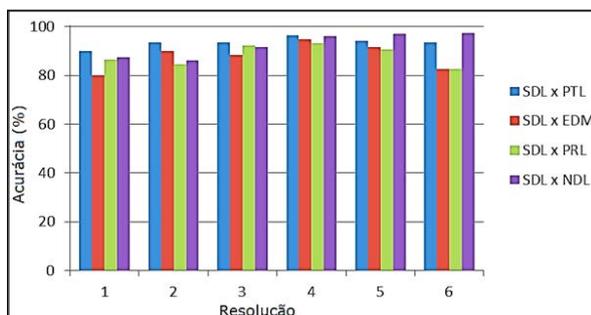
Uma análise estatística através de *boxplots* foi realizada para se identificarem as faixas de frequência que permitem a discriminação entre sinais de voz saudáveis e afetados por patologias laríngeas. Para a medida de entropia, as Figuras 9 a 11 representam as faixas de frequência de cada nó da árvore, no quinto nível de decomposição. Nas Figuras 9 e 10, percebe-se uma grande distinção entre sinais saudáveis comparados com qualquer patologia analisada. Já na Figura 11, observa-se uma diferença relevante entre as próprias patologias, com maior evidência entre o sinal saudável, paralisia e Edema de Reinke.

Figura 7 – Valores Médios da Acurácia (%) determinados pela Energia dos sinais de voz em seis níveis de resolução



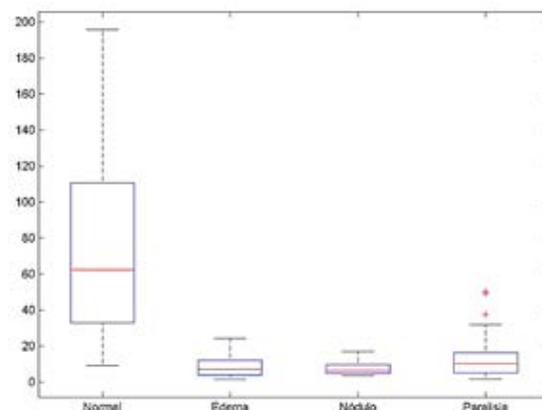
Fonte: Dados da pesquisa

Figura 8 – Valores Médios da Acurácia (%) determinados pela Entropia dos sinais de voz em seis níveis de resolução



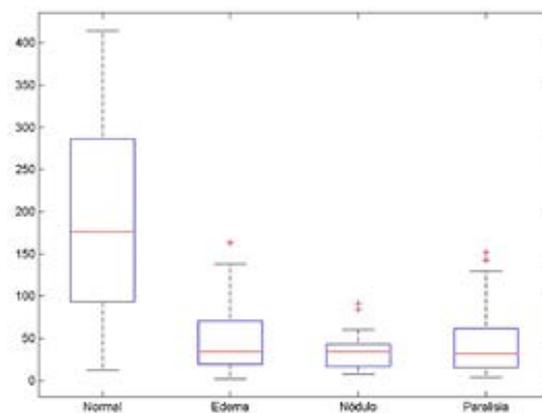
Fonte: Dados da pesquisa

Figura 9 – Entropia no nó A(5,6), que corresponde à faixa de frequência de 2343,75 – 2734,375 Hz



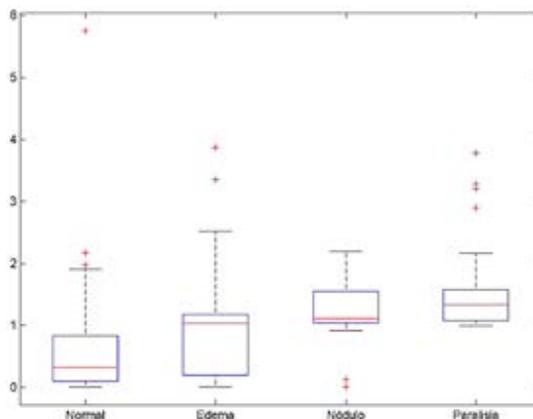
Fonte: Dados da pesquisa.

Figura 10 – Entropia no nó (5,7), que corresponde à faixa de frequência de 2734,375 – 3125 Hz



Fonte: Dados da pesquisa.

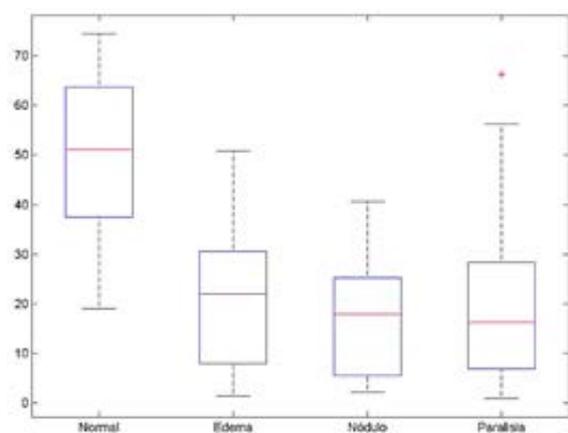
Figura 11 – Entropia no nó A (5,16), que corresponde à faixa de frequência de 6250 – 6640,625 Hz



Fonte: Dados da pesquisa.

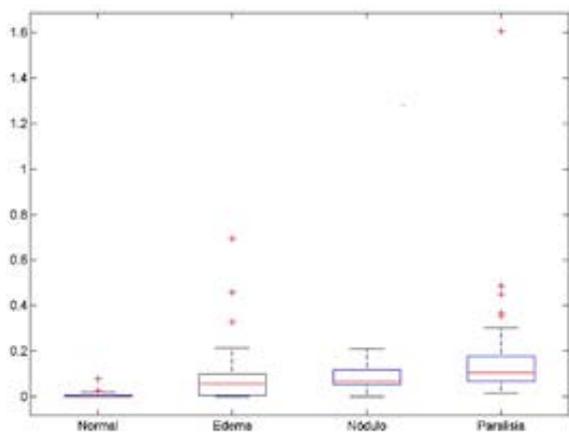
Para a medida de energia, as Figuras 12 a 14 representam as faixas de frequência dos coeficientes da *Wavelet Packet*, selecionados no quarto nível de decomposição. Observa-se, nos gráficos, que a energia nos sinais saudáveis se acumula nas baixas frequências, sendo praticamente inexistente nas altas frequências. Nos sinais patológicos, existe, em algumas faixas de alta frequência, um valor de energia superior aos sinais saudáveis, enquanto nas baixas frequências essa energia é bem inferior, o que facilita a discriminação entre sinais saudáveis e outros com patologia nas pregas vocais.

Figura 12 – Energia no nó A (4,0), que corresponde à faixa de frequência de 0 – 781,25 Hz



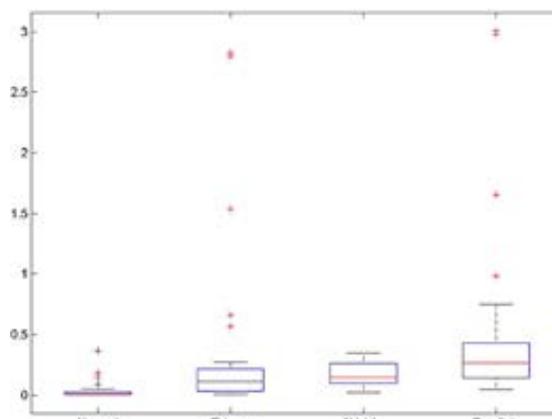
Fonte: Dados da pesquisa

Figura 13 – Energia no nó D (4,5), que corresponde à faixa de frequência de 3906,25 – 4687,5 Hz



Fonte: Dados da pesquisa

Figura 14 – Energia no nó D (4,13), que corresponde à faixa de frequência de 7812,5 – 8593,75 Hz



Fonte: Dados da pesquisa

6 Conclusões

Os resultados da acurácia média se mostraram mais eficientes entre o quarto e o quinto nível de resolução, chegando a atingir valores de até 98,57%. Na maioria dos casos, a entropia forneceu o maior desempenho, sendo adequada para a discriminação entre vozes saudáveis e patológicas. A análise através das *Wavelets* se mostrou eficiente na discriminação entre os sinais de vozes saudáveis e afetadas por patologias, podendo ser usada como uma ferramenta de auxílio ao diagnóstico.

REFERÊNCIAS

AKAY, M. **Time Frequency and Wavelets in Biomedical Signal Processing**, IEEE Press, 1998.

AKBARI, A.; ARJMANDI, M. K. An efficient voice pathology classification scheme based on applying multi-layer linear discriminant analysis to wavelet packet-based features. **Biomedical Signal Processing and Control**, v. 10, 2014.

ALMEIDA, N. C. **Sistema Inteligente para Diagnóstico de Patologias na Laringe utilizando Máquinas de Vetor Suporte**. Tese (Doutorado em Automação e Sistemas)—UFRN, Natal, 2010.

ARJMANDI, K. A.; POOYAN, M. An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine. **Biomedical Signal Processing and Control**, v. 1, n. 7, 2012.

BEHROOZMAND, R.; ALMASGANJ, F. Optimal Selection of wavelet-packet-based features using genetic algorithm in pathological assessment of patient's speech signal with unilateral vocal fold paralysis. **Computers in Biology and Medicine**, v. 37, 2007.

BURRUS, C. S.; GOPINATH, R. A.; GUO, H. **Introduction to Wavelets and Wavelet Transforms**: a primer, 1998.

CASTELANO, C. R. **Estudo Comparativo da Transformada Wavelet no Reconhecimento de Padrões da Íris Humana**. Dissertação (Mestrado em Engenharia Elétrica) – Escola de Engenharia de São Carlos, USP, 2006.

CHOUeiter, G. F.; GLASS, J. R. A wavelet and filter bank framework for phonetic classification, In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS SPEECH AND SIGNAL PROCESSING, 2005. **Proceedings...ICASSP**, 2005.

COIFMAN, R. R.; WICKERHAUSER, M. V. Entropy-based Algorithms for Best Basis Selection. **IEE Transaction on Information Theory**, v. 38, n. 2, 1992.

CORREIA, S. E. N.; COSTA W. C. A.; COSTA S. L. N. C. Detecção Automática de Patologias Laríngeas usando a Transformada Wavelet Discreta. **Anais... do 11th Brazilian Congress on Computational Intelligence (CBIC)**, 2013.

COSTA, S. L. N. C. **Análise acústica, baseada no modelo linear de produção da fala, para discriminação de vozes patológicas**. Tese (Doutorado em Engenharia Elétrica) – Universidade Federal de Campina Grande, 2008.

FONSECA, S. E. Wavelets, Predição Linear e LS-SVM Aplicados na Análise e Classificação de Sinais de Vozes Patológicas. Tese (Doutorado em Engenharia Elétrica) – Universidade de São Paulo, 2008.

GODINO-LLORENTE, J. I.; GÓMEZ-VILDA, P.; BLANCO VELASCO, M. Dimensionality Reduction of a Pathological Voice Quality Assessment System

Based on Gaussian Mixture Models and Short-Term Cepstral Parameters. **IEEE Transactions on Biomedical Engineering**, v. 53, n. 10, 2006.

MAGALHÃES, M. K. *et al.* Percepção Visual da Perturbação do Sinal de Voz por Meio da Transformada de Hilbert. In: XI CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE (CBIS'2008), 2008, Campos do Jordão. **Anais... do XI Congresso Brasileiro de Informática em Saúde**, 2008.

MALLAT, S. **A Wavelet Tour of Signal Processing**, Academic Press, 1999.

PARRAGA, A. **Aplicação da Transformada Wavelet Packet na Análise e Classificação de Sinais de Vozes Patológicas**. Dissertação (Mestrado em Automação e Instrumentação Eletrônica) – UFRS, Porto Alegre, 2002.

PINTO, R. M. da F. **Novas Abordagens ao Estudo de Sinais Biomédicos: Análise em Tempo-Frequência e Transformada de Hilbert-Huang**. Tese (Doutorado em Engenharia Biomédica e Física) – Universidade de Lisboa, 2009.

SANTOS, M. O. *et al.* Avaliações dos distúrbios Vocais em Crianças usando Características baseadas na Transformada Wavelet. In: XXIV CONGRESSO BRASILEIRO DE ENGENHARIA BIOMÉDICA (CBEB 2014), **Anais... 2014**.

SHUCK JR., A.; GUIMARGES, L. V.; WISBE, J. O. Dysphonic Voice Classification using Wavelet Packet Transform and Artificial Neural Network. In: ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY, 2003. **Proceedings of the 25th Annual International Conference of the IEEE**, 2003.

TANG, Y. Y. **Wavelet Theory and Its Applications to Pattern Recognition**, World Scientific, 2000.

TOSI, O. Vibratory Motion. In: **Textbook of phisics for audiologists and speach pathologists**. Lasing, Linko Copies, p.188-208, 1979.