

# Aplicação da Transformada Wavelet Packet para a Classificação de Sinais de Voz Patológicos Usando Rede Neural Artificial

Augusto F. T. da Silva, Samuel R. de Abreu, Silvana Cunha Costa e Suzete E. N. Correia

**Resumo**—Esse trabalho tem como objeto a detecção de patologias laringeas pela análise de sinais de voz. Para tanto, são extraídas medidas de energia e entropia, através da transformada wavelet packet, empregando a família wavelet Daubechies de ordem 4, até o sexto nível de resolução. Para a classificação dos sinais, são utilizadas redes neurais artificiais, obtendo-se uma acurácia média de até 98,57% na discriminação entre vozes saudáveis e vozes afetadas por nódulos nas pregas vocais.

**Palavras-Chave**—Transformada wavelet packet, Detecção de patologias, Redes neurais artificiais.

**Abstract**—The main objective of this work is the detection of laryngeal diseases through the voice signal analysis. Therefore, energy and entropy measures are extracted by wavelet packet transform, using the Daubechies wavelet family of 4 order, up to the sixth level of resolution. For the classification of signals, artificial neural network is used to obtain an average accuracy of up to 98.57% in discriminating between healthy voices and those affected by nodules on the vocal folds.

**Keywords**—Wavelet packet transform, Pathologies detection, Artificial neural networks.

## I. INTRODUÇÃO

Diversos sinais existentes 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. [1].

A voz é o 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. Alterações nos sistemas respiratórios, musculatórios e cerebrais podem causar mudanças no sistema articulatório da fala, causando desordens vocais. Certas alterações provocam mudanças significativas no sinal de voz, ocasionando alterações, tanto na frequência fundamental quanto no espectro do sinal [2]. Patologias como edemas, nódulos e paralisia são bastante recorrentes e influenciam diretamente essas disfunções vocais.

Há algum tempo a ciência vem trabalhando no sentido de fornecer ferramentas aos profissionais da saúde, no intuito de identificar quando esta se comporta de maneira normal ou patológica. Comumente, exames invasivos, tais como a videolaringoscopia e a videoestroboscopia, são empregados para observar as estruturas e características das pregas vocais.

Augusto F. T. da Silva, Samuel R. de Abreu, Silvana Cunha Costa e Suzete E. N. Correia, Unidade Acadêmica de Indústria, Instituto Federal de Educação, Ciências e Tecnologia da Paraíba, João Pessoa-PB, Brasil, E-mails: augustofdc92@gmail.com, abreu.s.ribeiro@gmail.com, silvana@ifpb.edu.br, suzete@ifpb.edu.br.

Devido a uma demanda da população que requer exames rápidos, simples e indolores, técnicas baseadas em análise acústica, por meio de processamento digital de sinais, surgem como uma ferramenta eficiente para auxiliar o diagnóstico médico da laringe a partir da presença de desordens vocais.

A transformada wavelet possibilita a análise simultânea dos sinais tanto no domínio do tempo quanto da frequência, o que permite o seu estudo em faixas de frequências distintas, de modo que em cada faixa de frequência aspectos diferentes dos sinais possam ser observados. Características baseadas na energia e na entropia de Shannon obtidas a partir da análise wavelet têm sido recentemente empregadas na avaliação de patologias laringeas. As wavelets têm sido aplicadas de diversas maneiras, a exemplo da transformada wavelet discreta [3,4] e as wavelets packets [5]. Em trabalho anterior, que empregava a energia normalizada dos coeficientes de detalhes da transformada wavelet discreta, acurácias de 95% e 97% foram obtidas na discriminação entre sinais saudáveis e afetados por edema de Reinke e entre sinais saudáveis e afetados por nódulos vocais, respectivamente. Ambas as classificações considerando a combinação de características em quatro níveis de resolução [4].

O objetivo deste trabalho consiste em empregar a transformada wavelet packet na classificação de sinais de voz, entre saudáveis e patológicos através de uso de redes neurais artificiais. Para tanto foram empregadas medidas de energia e entropia, extraídas dos coeficientes de detalhes e aproximação em seis níveis de resolução, através da família wavelet de Daubechies de ordem 4.

## II. MATERIAIS E MÉTODOS

### A. Base de Dados

Foram analisados 167 sinais 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* [6], sendo deles 53 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.

### B. Transformada wavelet packet

A transformada wavelet packet, proposta por Coifman [7] é uma generalização da transformada wavelet discreta. Na transformada wavelet discreta, o sinal é dividido em

coeficientes de aproximação, resultante da filtragem passa-baixas, e de coeficientes de detalhes, da filtragem passa-altas. Na decomposição apenas os coeficientes de aproximação são novamente subdivididos em coeficientes de aproximação e detalhes. 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 específica [8].

A energia dos coeficientes wavelet packet, para cada sub-banda  $i$  da resolução  $j$ , é obtida por:

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

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 os coeficientes da wavelet packet, segundo a Equação 3. Esta medida irá avaliar o grau de desordem do sinal.

$$H_i = - \sum_{k=1}^n |C_{j,k}^i|^2 \log |C_{j,k}^i|^2. \quad (3)$$

### C. Classificação

As medidas de energia e entropia extraídas são empregadas, individualmente, para a classificação de sinais de vozes saudáveis e afetados 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 dois conjuntos, 90% para teste e os demais para treino. O procedimento foi realizado dez vezes 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 determinado pela média aritmética entre os neurônios de entrada e saída. A função sigmóide foi empregada como função ativação.

### III. RESULTADOS

Para cada sinal de voz, foram calculados os valores de energia e a entropia, para todos os coeficientes wavelet packet até o 6° nível de resolução, empregando a família wavelet Daubechies de ordem 4, através do *software* MATLAB. Quatro casos de classificação foram considerados: sinais saudáveis *versus* patológicos (SDLxPTL), sinais saudáveis *versus* afetados por edema de Reinke (SDLxEDM), sinais saudáveis *versus* afetados por paralisia (SDLxPRL) e sinais saudáveis *versus* afetados por nódulos (SDLxNDL).

As Tabelas I e II mostram os valores de acurácia médio obtidos com a energia e entropia dos coeficientes wavelets, respectivamente. Observa-se que a entropia na maioria dos casos apresentou maior discriminação entre as classes, com exceção da classificação entre sinais saudáveis e nódulos, em que a energia forneceu maior acurácia.

Para a energia dos coeficientes wavelet packet, o quinto nível de resolução foi o mais eficiente, enquanto que para a entropia a discriminação, na maioria dos casos, foi maior no

quarto nível, o que atesta sua eficácia, tendo em vista que no 4° nível de resolução o número de medidas corresponde a metade dos atributos empregados no quinto nível de resolução.

TABELA I. VALORES MÉDIOS DE ACURÁCIA (%) DETERMINADOS PELA ENERGIA DOS SINAIS DE VOZ.

Casos	Níveis de Resolução					
	1°	2°	3°	4°	5°	6°
SDLxPTL	77,36	71,83	80,22	87,90	94,66	87,15
SDLxEDM	60,43	64,11	79,18	85,51	87,51	84,33
SDLxPRL	68,50	78,16	87,54	80,72	91,09	85,77
SDLxNDL	84,40	93,21	94,28	97,14	98,57	97,32

TABELA II. VALORES MÉDIOS DE ACURÁCIA (%) DETERMINADOS PELA ENTROPIA DOS SINAIS DE VOZ.

Casos	Níveis de Resolução					
	1°	2°	3°	4°	5°	6°
SDLxPTL	89,87	93,38	93,38	96,45	94,13	93,34
SDLxEDM	79,95	90,05	88,31	94,64	91,64	82,65
SDLxPRL	86,45	84,45	92,27	93,25	90,45	82,65
SDLxNDL	87,50	86,07	91,60	95,89	97,14	97,32

### IV. CONCLUSÕES

Os resultados da acurácia média se mostraram mais eficientes entre o quarto e o quinto nível de resolução, chegando a atingir valores de acurácia de até 98,57%. Na maioria dos casos a entropia forneceu o maior desempenho, sendo adequada para a discriminação entre vozes saudáveis e patológicas. Em trabalhos futuros será realizada a seleção de características em cada resolução para melhorar a classificação.

### AGRADECIMENTOS

Ao IFPB pelo apoio financeiro.

### REFERÊNCIAS

- [1] M. Akay. *Time Frequency and Wavelets in Biomedical Signal Processing*, IEEE Press, 1998.
- [2] J. I. Godino-Llorente, P. Gómez-Vilda, M. B. Velasco. *Dimensionality Reduction of a Pathological Voice Quality Assessment System Based on Gaussian Mixture Models and Short-Term Cepstral Parameters*, IEEE Trans. on Biomedical Engineering., Vol. 53, No. 10, 2006.
- [3] E. Fonseca, R. Guido, P. Scalassara, C. Maciel e J. Pereira. *Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders*. Computers in Biology and Medicine, v. 37, p. 571-578, 2007.
- [4] S. E .N. Correia, W. C. A. Costa, S. L. N. C. Costa. *Deteção Automática de Patologias Laríngeas usando a Transformada Wavelet Discreta*. Em: Anais do 11th Brazilian Congress on Computational Intelligence (CBIC), 2013.
- [5] R. Behroozmand, F. Almasganj. *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, p. 474-485, 2007.
- [6] KAY ELEMETRICS CORP. *Disordered voice database, model 4337*, 03 Ed, 1994.
- [7] R. R. Coifman, M. V. Wickerhauser. *Entropy-based Algorithms for Best Basis Selection*, IEEE Trans. on Information Theory. v.38, n.2,1992.
- [8] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 3 edition, 2008.