

# AVALIAÇÃO DOS DISTÚRBIOS VOCAIS EM CRIANÇAS USANDO CARACTERÍSTICAS BASEADAS NA TRANSFORMADA WAVELET

M. O. Santos\*, S. C. Costa\*, W. C. de A. Costa\*, S. E. N. Correia\* e L. W. Lopes\*\*

\* Programa de Pós-Graduação em Engenharia Elétrica - IFPB, João Pessoa, Brasil

\*\* Departamento de Fonoaudiologia - UFPB, João Pessoa, Brasil

e-mail: mikaosantos@gmail.com

**Resumo:** A transformada wavelet tem sido usada como uma ferramenta confiável e objetiva para avaliar distúrbios vocais. No entanto, a maioria das pesquisas têm sido empregadas em adultos. No presente estudo, a energia normalizada e a entropia dos coeficientes de detalhe da transformada wavelet discreta, em nove níveis de resolução, são usadas como características para discriminar o grau de severidade do desvio vocal em crianças. Para a classificação, é empregada a análise discriminante quadrática. Os resultados mostram que as medidas de energia e entropia combinadas entre si fornecem acurácia maior que 95% na discriminação entre vozes saudáveis e desviadas.

**Palavras-chave:** Desordens vocais, transformada wavelet, energia e entropia, análise discriminante quadrática.

**Abstract:** Wavelet transform has been used as a reliable and objective tool for assessing voice disorders. However, most of researches have been employed on adults. In the present study, normalized energy and entropy of detail wavelet coefficients obtained by discrete wavelet transform, in nine resolution levels, are used as features to discriminate the severity degree of vocal deviation in children. For classification, a quadratic discriminant analysis is employed. Results show that entropy and energy measures provide accuracy higher than 95% in discriminating healthy voices from deviated ones.

**Keywords:** Voice disorders; wavelet transform; energy and entropy; quadratic discriminant analysis.

## Introdução

Distúrbios da voz podem afetar diferentes faixas etárias. Em crianças, estima-se que a taxa de incidência de disfonias está entre 6% a 23%, aproximadamente [1,2]. Dentre os fatores causadores dos distúrbios vocais, encontram-se o abuso vocal ou uso incorreto da voz, causas orgânicas (alterações estruturais, desvios nasais, patologias da laringe) e causas funcionais (alergias, doenças respiratórias, refluxo gastroesofágico) [1,3,4]. Crianças com distúrbios de voz podem apresentar efeitos negativos no seu desenvolvimento social, educacional e físico [2]. A investigação e o diagnóstico precoce da disфонia infantil permite maior eficácia no tratamento. Entretanto, a avaliação de desordens vocais em crianças apresenta alguns desafios

relacionados às dificuldades de cooperação das mesmas durante os exames tradicionais.

A análise acústica pode ser empregada como um método de apoio ao diagnóstico e tratamento de disfonias, de forma rápida e confortável. As medidas acústicas quantificam informações a respeito do comportamento do sinal de voz. Medidas como *pitch*, *jitter*, *shimmer* e HNR (*Harmonic-to-noise ratio*) tem sido comumente empregadas para avaliar a qualidade vocal [2,3,5]. Devido às diferenças fisiológicas entre crianças e adultos, alterações nas frequências de ressonância e na duração da elocução das vogais são observadas em função da idade [6].

O sucesso do diagnóstico e do tratamento, entretanto, depende da escolha apropriada das características. Disfonias infantis causam alterações na qualidade da voz, sendo a rouquidão e a soproidade as mais frequentes [4], representadas como ruídos em faixas de frequências diferentes no espectro do sinal de voz. A transformada wavelet fornece uma análise dos sinais em diferentes resoluções, de forma que, em cada uma delas, diferentes aspectos dos sinais podem ser observados. Características obtidas a partir da análise wavelet têm sido empregadas na avaliação de desordens vocais em adultos [3,7-10], causadas por patologias laringeas. Para a população pediátrica, no entanto, ainda há poucos trabalhos relacionados [11].

Neste artigo, é realizada a avaliação do grau de severidade do distúrbio vocal em crianças. São usadas como medidas a energia normalizada e a entropia dos coeficientes de detalhes da transformada wavelet, em nove faixas de frequência, usando a família wavelet de Daubechies de ordem 40. Para a classificação é utilizada a análise discriminante quadrática. O desempenho das medidas é avaliado de forma individual e combinada, visando determinar as faixas de frequência que fornecem a maior separação entre os sinais de vozes saudáveis, com desvio leve, moderado e intenso.

## Análise Wavelet

Uma wavelet mãe, representada por  $\psi(t) \in L^2(\mathbf{R})$ , é uma função oscilatória, com média zero e duração finita. A transformada wavelet contínua de uma função  $f(t) \in L^2(\mathbf{R})$  consiste na decomposição de  $f(t)$  em um conjunto de funções elementares  $\psi_{a,b}(t)$ , denominadas wavelets, as quais são geradas através de mudanças de escala e translações da wavelet mãe:

$$\psi_{a,b}(t) \equiv \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad a, b \in \mathbf{R}, \quad (1)$$

em que  $a > 0$  é o fator de escala,  $b$  o fator de translação e  $a^{-1/2}$  o fator de normalização que mantém a mesma energia para todas as wavelets independente da escala utilizada. Quando o fator de escala  $a > 1$  a wavelet encontra-se expandida proporcionando a análise em baixas frequências do sinal. Do contrário, quando  $a < 1$  as wavelets encontram-se comprimidas e permitem uma análise em altas frequências.

A versão discreta da transformada wavelet pode ser definida assumindo  $a = 2^j$  e  $b = k \cdot 2^j$ , em que  $j$  e  $k$  são inteiros. O conjunto discreto de wavelets é representado por:

$$\psi_{j,k}(t) = \sqrt{2^{-j}} \psi(2^{-j}t - k). \quad (2)$$

Na análise multirresolucional, além das wavelets são empregadas versões escalonadas e transladadas de uma função  $\phi(t)$  denominada função escalonamento,

$$\phi_{j,k}(t) = \sqrt{2^{-j}} \phi(2^{-j}t - k). \quad (3)$$

Uma função contínua  $f(t)$  pode ser decomposta na  $j$ -ésima escala ou resolução, em termos das funções base wavelet e escalonamento por:

$$f(t) = \sum_k (c_j(k) \phi_{j,k}(t) + d_j(k) \psi_{j,k}(t)), \quad (4)$$

em que  $c_j(k)$  e  $d_j(k)$  correspondem aos coeficientes de aproximação e detalhe, definidos como:

$$c_j(k) = \sum_m h(m - 2k) c_{j-1}(m), \quad (5)$$

$$d_j(k) = \sum_m g(m - 2k) c_{j-1}(m). \quad (6)$$

As Equações (5) e (6) representam operações de filtragem por meio das respostas ao impulso de filtros de análise passa-baixas  $h(n)$  e passa-altas  $g(n)$ . Para cada nível de resolução  $j$ , o algoritmo da transformada wavelet discreta, proposto por Mallat [12], decompõe o sinal em dois conjuntos de coeficientes: aproximação  $c_j(k)$  e detalhe  $d_j(k)$ . As informações extraídas em uma dada resolução são mantidas nos níveis de resolução superiores. O fator  $2k$ , no índice dos filtros, representa a decimação por um fator 2.

As wavelets de Daubechies são uma família formada por várias funções, que possuem filtros de comprimento diferentes [7]. Tais wavelets são ortogonais e possuem suporte compacto. Segundo Salhi et al [8], as wavelets de Daubechies de ordem 40 são indicadas para análise de desordens vocais.

## Materiais e métodos

Na Figura 1 está ilustrada o diagrama em blocos da metodologia empregada neste trabalho.



Figura 1: Diagrama em blocos do processo de análise e classificação dos desvios vocais infantis.

**Base de dados** – Os sinais de voz empregados neste estudo são provenientes da base de dados fornecida pelo Laboratório de Voz do Departamento de Fonoaudiologia da Universidade Federal da Paraíba. A metodologia de coleta de dados foi aprovada pelo Comitê de Ética em Pesquisa da instituição de origem sob o protocolo número 775/10. A base de dados inclui sinais de vozes da vogal sustentada /ε/ de 93 crianças (48 meninas e 45 meninos) com idades entre 3 e 10 anos.

Os sinais foram gravados com uma taxa de amostragem de 44,1 kHz e duração média de 5 minutos, no tempo máximo de fonação. Foram eliminados 1 segundo inicial e dois segundos finais da elocução, devido às irregularidades do sinal, sendo mantidos um tempo mínimo de 2 segundos em cada emissão.

Foi realizada uma análise perceptivo-auditiva da voz para classificar os sinais de acordo com o seu grau de desvio vocal, com base na escala visual analógica (EAV) [13]. Os sinais foram classificados como grau geral 1 para voz saudável (GG1), grau geral 2 para o desvio leve (GG2), grau geral 3 para o desvio moderado (GG3) e grau geral 4 para o desvio intenso (GG4). Dos 93 sinais analisados, 10 foram considerados saudáveis (sem desvio), 70 com desvio leve e 13 com desvio moderado. Não houve casos de desvio intenso na avaliação da qualidade vocal.

**Extração de Características** – Os sinais de voz da base de dados são analisados por meio da energia normalizada e da entropia dos coeficientes de detalhe da transformada wavelet discreta, obtidas em nove níveis de resolução. A família wavelet empregada na decomposição foi a de Daubechies de ordem 40.

A energia normalizada ( $E$ ) e a entropia de Shannon ( $H$ ) dos coeficientes de detalhe em cada resolução  $j$ , são obtidas através das equações (7) e (8), respectivamente:

$$E_j = \frac{\sum_k |d_j(k)|^2}{\sum_k (|c_j(k)|^2 + |d_j(k)|^2)} \quad (7)$$

$$H_j = -\sum_k p_j(k) \log p_j(k), \quad (8)$$

$$\text{em que } p_j(k) = \frac{|d_j(k)|^2}{\sum_k |d_j(k)|^2}.$$

Usando a energia normalizada dos coeficientes de detalhe como característica, pode-se identificar o quanto a energia do sinal de voz encontra-se distribuída ao longo da frequência [9]. Comumente, a qualidade da voz é alterada na presença de disfonias por meio de parâmetros como aspereza, rouquidão e soproidade. A aspereza ocorre devido a rigidez da mucosa, que causa uma irregularidade vibratória com ruídos nas altas frequências. A rouquidão é proveniente da irregularidade de vibração das dobras vocais que geram ruídos nas baixas frequências. A soproidade indica a presença de ruído de fundo, audível, que corresponde fisiologicamente à fenda glótica [5].

Uma vez que a entropia mede o grau de desordem de um processo, a mesma é influenciada pelas irregularidades e aleatoriedade dos sistemas fisiológicos, a exemplo do sistema de produção vocal [3,9], podendo ser usada como medida na avaliação de desordens vocais.

Como os sinais de voz analisados têm uma frequência de amostragem de 44,1 kHz, de acordo com o critério de Nyquist, a componente de frequência mais alta do sinal é 22,5 kHz. Como nove níveis de resolução são considerados, as medidas de energia e entropia extraídas dos coeficientes de detalhes fornecem informações dos sinais em diferentes faixas de frequência, conforme é apresentado na Tabela 1.

Tabela 1: Níveis de resolução e faixas de frequência para os coeficientes de detalhe da transformada wavelet.

Nível de Resolução	Faixa de Frequência (Hz)
1	12500 a 22500
2	6250 a 12500
3	3125 a 6250
4	1562,5 a 3125
5	781,25 a 1562,5
6	390,625 a 781,25
7	195,32 a 390,635
8	97,66 a 195,32
9	48,83 a 97,66

**Classificação** – As características extraídas, em cada nível de resolução, são utilizadas individualmente e combinadas entre si para a classificação do grau de desvio vocal. Para tanto, empregou-se uma função de análise discriminante quadrática [14]. Os graus de severidade do desvio da voz são avaliados e classificados como saudáveis (GG1), desvio leve

(GG2), desvio moderado (GG3) ou desvio intenso (GG4).

Na classificação foi empregada a validação cruzada considerando 10 subconjuntos. Para cada subconjunto 90% dos sinais foram usados para teste e 10% para treino. A escolha dos sinais se deu de forma aleatória e sem repetição.

## Resultados

Nesta seção estão descritos os resultados obtidos para a classificação, baseados na metodologia apresentada. O processo de classificação dos sinais foi realizado considerando-se quatro casos distintos: GG1 x (GG2 e GG3), GG1 x GG2, GG1 x GG3 e GG2 x GG3. A análise de desempenho foi realizada usando como medida a acurácia que é definida como a relação entre o número de casos corretamente classificados e todos os casos apresentados ao classificador.

Inicialmente, a energia e a entropia dos coeficientes de detalhe das wavelets foram avaliadas individualmente para cada nível de resolução. Em seguida, as características foram agrupadas em um único vetor, contendo 18 medidas (9 de energia e 9 de entropia). Por fim, as características foram combinadas 2 a 2, 3 a 3, ..., 17 a 17 e selecionada a combinação que apresentou a melhor taxa de classificação.

Nas Tabelas 2 a 5 estão apresentados os melhores resultados obtidos para todos os casos de classificação considerados. As Tabelas 2 a 4 incluem os casos de classificação entre os sinais de vozes saudáveis e vozes desviadas. Na Tabela 5, são classificados os desvios vocais quanto a sua intensidade. A combinação das medidas resultou num aumento significativo da acurácia e reduziu o vetor de características de 18 para, no máximo, 8 medidas.

Tabela 2: Classificação GG1 x (GG2 e GG3).

	Medidas	Acurácia (%)
Individual	$E_5$	72,11 ± 3,97
	$H_8$	74,11 ± 2,89
Combinação	Todas	89,22 ± 0,16
	$E_1E_4E_5E_6H_5H_8$	95,66 ± 1,77

Tabela 3: Classificação GG1 x GG2.

	Medidas	Acurácia (%)
Individual	$E_6$	68,75 ± 6,25
	$H_1$	72,50 ± 5,83
Combinação	Todas	86,25 ± 1,25
	$E_1E_4E_5E_6H_5H_8$	95,00 ± 2,04

Tabela 4: Classificação GG1 x GG3.

	Medidas	Acurácia (%)
Individual	$E_1$	81,25 ± 7,93
	$H_2$	75,00 ± 8,70
Combinação	Todas	55,00 ± 2,54
	$E_1E_4H_5H_8$	100,00 ± 0,00

Tabela 5: Classificação GG2 x GG3.

	Medidas	Acurácia (%)
Individual	$E_1$	$76,66 \pm 5,21$
	$H_7$	$54,72 \pm 6,00$
Combinação	Todas	$84,58 \pm 1,48$
	$E_1 E_2 E_3 E_4 E_6 E_9 H_7 H_9$	$90,55 \pm 2,82$

## Discussão

A análise por faixa de frequência por meio da energia normalizada e entropia dos coeficientes de detalhe da transformada wavelet permitiu detectar a presença dos desvios vocais em sinais de vozes infantis. As medidas combinadas apresentaram maior poder discriminatório, reduzindo a dimensionalidade dos dados e aumentando a acurácia. Observa-se que a energia do nível de resolução 1 ( $E_1$ ) aparece em todas as combinações, ressaltando a presença de ruído nas altas frequências nas vozes desviadas. Os níveis de energia 4, 5 e 6, em que situam-se o primeiro e segundo formantes da vogal /  $\varepsilon$  / [6] são predominantes na discriminação, indicando a variabilidade destes parâmetros com o grau do desvio vocal.

Em trabalho anterior [15], foi obtida uma acurácia de  $88,22\% \pm 0,93\%$ , combinando oito medidas de quantificação de recorrência e análise discriminante quadrática na classificação entre GG1 x (GG2 e GG3), empregando a mesma base de dados de vozes infantis. No presente trabalho, manteve-se o mesmo classificador, para fins de comparação. As taxas de classificação acima de 90% sugerem que as medidas analisadas podem ser utilizadas para tratamento de disfonias e avaliação da qualidade vocal. A combinação das medidas conseguiu capturar as diferenças predominantes nas vozes desviadas comparadas a sinais saudáveis, especialmente no caso GG1 x GG3, em que uma acurácia de 100% foi obtida, com apenas quatro medidas.

A entropia e a energia dos coeficientes wavelets já foram empregadas para discriminar patologias laringeas em adultos, usando wavelet packets [3,9]. Neste trabalho, a energia normalizada e a entropia são obtidas em nove níveis de resolução empregando a transformada wavelet discreta, a qual tem um menor custo computacional. Em trabalhos futuros, pretende-se investigar o desempenho das medidas na avaliação da qualidade vocal, bem como outros tipos de classificadores, a exemplo de Redes Neurais Artificiais e Máquinas de Vetor de Suporte.

## Referências

[1] Martins R, Ribeiro C, Mello B, Branco A, Tavares E. Dysphonia in Children. *Journal of Voice*; 2012; 26(5): 674.e17-674.e20.  
 [2] Meredith M, Theis S, McMurray J, Zhang Y, Jiang J. Describing pediatric dysphonia with nonlinear dynamic parameters. *International Journal of Pediatric Otorhinolaryngology*; 2008,72, p.1829-1836.

[3] Arjmandi M, 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* 7; 2012. p. 3– 19.  
 [4] Oliveira R, Teixeira L, Gama A, Medeiros A. Análise perceptivo-auditiva, acústica e autopercepção vocal em crianças. *Jornal da Sociedade Brasileira de Fonoaudiologia*; 2011; 23(2):158-63.  
 [5] Lopes J, Freitas S, Sousa R, Matos J, Abreu F, Ferreira A. A medida HNR: sua relevância na análise acústica da voz e sua estimação precisa. In: *Anais da I Jornadas sobre Tecnologia e Saúde*; 2008.  
 [6] Brod LEM, Seara IC. As vogais orais do português brasileiro na fala infantil e adulta: uma análise comparativa. *Linguagem & Ensino, Pelotas*, v.16, n.1, p. 111-130, jan./jun. 2013.  
 [7] Fonseca E, Guido R, Scalassara P, Maciel C, Pereira J. Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders. *Computers in Biology and Medicine*; 2007; v. 37, p. 571-578, 2007.  
 [8] Salhi L, Talbi M, Cherif A. Voice disorders identification using hybrid approach: wavelet analysis and multilayer neural networks. *World Academy of Science, Engineering and Technology*, 2008,45, pp. 330-339.  
 [9] 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*, 2007, v. 37, pp 474-475.  
 [10] Correia SEN, Costa WCA, Costa SLNC. Detecção Automática de Patologias Laringeas usando a Transformada Wavelet Discreta. Em: *Anais do 11th Brazilian Congress on Computational Intelligence (CBIC)*, 2013.  
 [11] Mahmoudi Z, Rahati S, Ghasemi MM, Asadpour V, Tayarani H, Rajati, M. Classification of voice disorder in children with cochlear implantation and hearing aid using multiple classifier fusion, *BioMedical Engineering OnLine*, 10:3, 2011.  
 [12] Mallat S. *A Wavelet Tour of Signal Processing*, Academic Press, 1999.  
 [13] Yamasaki R, Madazio G, Padovani M, Azevedo R, Behlau M. Correspondence between-Visual Analog Scale and the Numerical Rating Scale Perceptual-Hearing Voices. *XVI Brazilian Congress of Speech*; 2008.  
 [14] Krzanowski W J. *Principles of Multivariate Analysis: A User's Perspective*. New York: Oxford University Press, 1988.  
 [15] Lopes LW, Costa SLNC, Costa WCA, Correia SEN, Vieira JVD. Análise da dinâmica não linear de vozes infantis: nova proposta de avaliação e monitoramento vocal. *Pesquisas em Fonoaudiologia*. 1ed.; Recife: Sociedade Brasileira de Fonoaudiologia, 2013, 1, p. 466-470.