

Detecção Automática de Patologias Laríngeas usando a Transformada Wavelet Discreta

Suzete E. N. Correia, Washington C. de A. Costa, Silvana C. Costa
Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, Campus João Pessoa, PB, Brasil
suzete@ifpb.edu.br, washington@ifpb.edu.br, silvana@ifpb.edu.br

Abstract— O presente trabalho emprega a energia normalizada dos coeficientes de detalhes obtidos por meio da transformada wavelet discreta, para distinguir entre sinais de vozes saudáveis de sinais afetados por edema de Reinke e nódulos nas pregas vocais. A wavelet de Daubechies de ordem 35 é usada para decompor os sinais em oito níveis de resolução. As características extraídas são avaliadas individualmente e de forma combinada, com o intuito de determinar as faixas de frequência que fornecem a melhor discriminação entre as vozes saudáveis e patológicas. Para a classificação, a análise discriminante quadrática é empregada. Os resultados atestam que o quarto nível de resolução provê as melhores taxas de reconhecimento. Uma acurácia de 97% foi obtida na discriminação entre os sinais de vozes saudáveis e os sinais afetados por nódulos vocais.

Palavras-Chave— *Detecção automática de patologias laríngeas; Transformada Wavelet Discreta, Análise Discriminante Quadrática;*

Abstract— The present work uses the normalized energy of the detail coefficients obtained from Discrete Wavelet Transform to distinguish health voices from that affected by Reinke's edema and nodules in vocal folds. The Daubechies wavelet of order 35 is used to decompose the signals at eight resolution levels. The extracted features are evaluated individually and in a combined way in order to determine the frequency bands that provide the best discrimination between healthy and pathological voices. For the classification, quadratic discriminant analysis technique is employed. Results show that the fourth resolution level provides the best recognition rates. An accuracy of 97% is achieved in the discrimination between healthy voices and those affected by vocal nodules.

Keywords— *Automatic laryngeal pathology detection; Discrete Wavelet Transform, Quadratic Discriminant Analysis;*

I. INTRODUÇÃO

A produção da fala é um processo complexo que envolve movimentos musculatórios, respiração, controle do cérebro e um sistema sensor de realimentação auditiva. Alterações estruturais nesses mecanismos podem causar mudanças no processo articulatório da fala, causando distúrbios vocais. Certas alterações laríngeas impedem que as pregas vocais tenham uma vibração glotal harmônica, provocando modificações significativas no sinal de voz [1].

Embora a patologia esteja localizada no sistema vibratório pode, ainda, afetar o movimento articulatório regular durante a produção da fala. Além disso, componentes do sistema de

ressonância podem ser afetados, resultando em mudanças no trato vocal, produzindo irregularidades tanto nas modificações da frequência fundamental quanto na envoltória espectral do sinal [1]. Hartl *et al.* [2] atestam que nos sinais de vozes patológicas, os níveis de energia aumentam nas médias e altas frequências e decrescem nas baixas em relação aos sinais de vozes saudáveis.

A detecção precoce de patologias laríngeas aumenta significativamente a eficácia do tratamento. Para auxiliar o diagnóstico médico, técnicas de processamento digital de sinais podem ser usadas para avaliar a qualidade vocal, através da análise acústica. Estas técnicas podem ser associadas às técnicas convencionais baseadas na observação clínica para melhorar o diagnóstico e o tratamento das distúrbios vocais. Além disso, traz como benefícios um exame menos desconfortável para o paciente, uma avaliação quantitativa dos critérios de avaliação da qualidade da voz e possibilitam o desenvolvimento de sistemas automáticos de auxílio ao diagnóstico por computador com baixo custo [3].

Para a detecção eficiente de uma determinada patologia, torna-se necessário escolher características ou parâmetros capazes de capturar a desordem presente no sinal de voz causada pela patologia. A transformada wavelet permite a análise de sinais em resoluções distintas, de modo que, em cada escala, aspectos diferentes possam ser observados. Como a análise com wavelets fornece informações do sinal tanto no domínio do tempo quanto no domínio da frequência, torna-se uma ferramenta de enorme importância em processamento digital de diversos tipos de sinais, inclusive os de vozes saudáveis ou patológicas [4].

Diversos trabalhos, na literatura recente, empregam as wavelets como extratores de características na discriminação entre vozes saudáveis e patológicas. A maioria utiliza medidas baseadas na energia e na entropia de Shannon dos coeficientes wavelets para a representação dos sinais de voz. As wavelets são aplicadas de diversas formas, tais como a transformada wavelet discreta [4-7], a transformada wavelet contínua [8] e as wavelets packets [3, 9 -12].

Neste artigo, um novo método para a detecção automática de patologias laríngeas baseado na transformada wavelet discreta é apresentado. Como parâmetro, é empregado a energia normalizada dos coeficientes de detalhes obtidos por meio da wavelet Daubechies de ordem 35, em oito níveis de resolução. As características extraídas são avaliadas individualmente, por nível de resolução, e combinadas com o

intuito de determinar as faixas de frequência que fornecem a maior separação entre os sinais de vozes saudáveis e os sinais afetados por edema de Reinke e nódulos nas pregas vocais. Para a classificação é empregada a análise discriminante quadrática.

O artigo está organizado da seguinte maneira: Na Seção II é apresentada uma breve revisão da transformada wavelet discreta e do tipo de wavelet empregada. Na Seção III são apresentadas a base de dados e a metodologia empregada. Os resultados na Seção IV e, por fim, na Seção V, as considerações finais.

II. TRANSFORMADA WAVELET DISCRETA

A transformada wavelet é uma técnica usada para representar, descrever e processar sinais através da combinação linear de funções elementares. Na análise multirresolucional são empregadas, na decomposição, duas funções ortogonais conhecidas como wavelet mãe, $\psi(t)$, e função escalonamento, $\phi(t)$. Todas as funções base são obtidas pela translação e escalonamento dessas funções:

$$\phi_{j,k}(t) = \sqrt{2^{-j}} \phi(2^{-j}t - k), \quad (1)$$

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

em que j e k são inteiros que representam, respectivamente, a escala ou resolução e a translação.

Uma função $f(t)$ pode ser decomposta na j -ésima escala em termos da função wavelet e da função escalonamento por:

$$f(t) = \sum_k (c_j(k)\phi_{j,k}(t) + d_j(k)\psi_{j,k}(t)), \quad (3)$$

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

$$c_j(k) = \sum_m h(m - 2k)c_{j-1}(m), \quad (4)$$

$$d_j(k) = \sum_m g(m - 2k)c_{j-1}(m). \quad (5)$$

As funções $h(n)$ e $g(n)$ são, respectivamente, as respostas ao impulso de filtros de análise passa-baixas e passa-altas. As equações (4) e (5) mostram que os coeficientes wavelets $c_j(k)$ e $d_j(k)$ são obtidos a partir de convoluções discretas com as respostas ao impulso $h(n)$ e $g(n)$. O fator $2k$, no índice dos filtros, representa a decimação por um fator 2.

Para cada nível de resolução j é obtida uma aproximação $c_j(k)$ e um detalhe $d_j(k)$. Os termos aproximação e detalhe correspondem, respectivamente, às componentes de baixas e altas frequências do sinal analisado. As informações extraídas em uma dada resolução são mantidas nos níveis de resolução superiores.

Se o sinal original $c_{j-1}(k)$ possui N amostras, então os sinais $c_j(k)$ e $d_j(k)$ possuem $2^j N$ amostras. Dessa forma, a decomposição wavelet é formada por sequências que possuem o mesmo tamanho do sinal de entrada. A Figura 1 ilustra o algoritmo de decomposição wavelet uni-dimensional, proposto por Mallat [13], conhecido como transformada wavelet rápida.

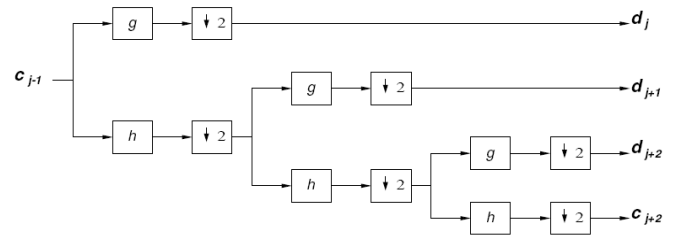


Fig. 1. Algoritmo de decomposição wavelet para três níveis de resolução.

As wavelets de Daubechies são uma família, formada por várias funções que possuem filtros de comprimento diferentes [14]. Os nomes das wavelets são descritos por dbN, em que N é o comprimento do filtro ou ordem da wavelet. Tais wavelets são ortogonais e possuem suporte compacto, ou seja, as respostas ao impulso dos filtros $h(n)$ e $g(n)$ são finitas e os somatórios que definem a transformada wavelet discreta são finitos. Segundo Fonseca *et al.* [4], as wavelets de Daubechies são muito utilizadas para análise de sinais de voz.

III. MATERIAIS E MÉTODOS

A. Base de Dados

Os dados utilizados neste trabalho fazem parte da base de dados gravada pelo *Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab* [15]. Essa base de dados foi construída para auxílio na análise perceptual de vozes desordenadas para aplicações clínicas e de pesquisa e tem sido largamente utilizada em diversos trabalhos relacionados à análise acústica de sinais de voz. A mesma inclui amostras de pacientes com uma grande variedade de desordens da voz provocadas por causas orgânicas, neurológicas, traumáticas e psicogênicas, entre outras. Nesta pesquisa, são utilizados 53 arquivos de vozes saudáveis e 62 sinais de vozes patológicas, dentre as quais 44 são afetadas por edema de Reinke e 18 por nódulos nas pregas vocais.

Dentre os sons produzidos no trato vocal, as vogais sustentadas e algumas consoantes sonoras são as mais facilmente distorcidas na presença da patologia, por isso são usadas para validar sistemas de detecção automática de sinais de vozes patológicas [8]. Todos os sinais utilizados são da vogal sustentada /a/, com duração de 1 a 3 segundos, gravados em um ambiente controlado e amostrados a taxas de 25 kamostras/s, com resolução de 16 bits/amostra. Os valores de amplitude desses sinais encontram-se normalizados no intervalo entre [-1,1].

B. Metodologia

Para cada sinal de voz proveniente da base de dados é aplicado o algoritmo de decomposição da transformada wavelet discreta em oito níveis de resolução. Somente os coeficientes de detalhes são avaliados. Levando-se em consideração que os sinais possuem uma frequência de amostragem de 25000 Hz, então, de acordo com o critério de Nyquist, o maior componente de frequência é 12500 Hz. Sendo assim, devido ao processo de multirresolução da transformada

wavelet, a faixa de frequência do sinal de voz é dividida em sub-bandas. A Tabela I apresenta os níveis de resolução empregados na análise e suas respectivas faixas de frequência.

TABELA I. RESOLUÇÃO E FAIXA DE FREQUÊNCIA ASSOCIADA

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

Os coeficientes de detalhes d_1 , d_2 , d_3 e d_4 foram escolhidos para a extração do ruído em alta frequência, identificado em vozes patológicas. Já os componentes d_5 a d_8 foram selecionados para a análise nas baixas frequências. Na faixa de frequências a partir do nível d_4 estão contidas as frequências de formantes da vogal /a/. Nas faixas abaixo deste nível, de d_5 a d_8 , está contida a região em que se concentram os valores de frequência fundamental, cujos valores médios estão em torno de 200 Hz para vozes femininas e de 100 Hz para vozes masculinas saudáveis [16].

Para a caracterização dos sinais de vozes saudáveis e patológicos, a energia normalizada dos coeficientes de detalhes, em cada resolução, é obtida por:

$$E_j = \frac{\sum_k |d_j(k)|^2}{\sum_k (|c_j(k)|^2 + |d_j(k)|^2)} \quad (6)$$

Esse parâmetro mostra o quanto a energia do sinal de voz está distribuída ao longo das escalas. Em geral, para sons sonoros, sinais de vozes saudáveis apresentam uma periodicidade no tempo, enquanto vozes afetadas por patologias nas pregas vocais apresentam um comportamento irregular tanto das características temporais quanto espectrais [1]. Espera-se, portanto, que a concentração de energia ocorra em faixas de frequências diferentes para os grupos de sinais analisados, de modo que essa informação seja útil para a sua discriminação.

A família wavelet empregada na decomposição foi a de Daubechies de ordem 35, tendo em vista que em trabalho anterior foi a que apresentou melhor potencial de discriminação entre os sinais saudáveis e patológicos [17]. A classificação é feita considerando a energia normalizada dos coeficientes wavelets, em cada nível de resolução individualmente, e pela combinação entre os coeficientes de energia dos níveis observados. Para tanto, empregou-se a análise de discriminante

quadrática, com validação cruzada, em que são utilizados 10 subconjuntos.

IV. RESULTADOS EXPERIMENTAIS

Como forma de avaliação de desempenho na classificação dos sinais é utilizada como medida a acurácia que mede o percentual de sinais reconhecidos ou classificados corretamente pelo sistema. Os resultados são obtidos analisando dois casos: Saudável x Edema de Reinke e Saudável x Nódulo.

A. Saudável x Edema de Reinke

Na Tabela II são apresentados os resultados da classificação obtidos com o uso individual da energia normalizada dos coeficientes de detalhes, em cada nível de resolução analisado. Observa-se que o nível de resolução 4 fornece a maior separação entre as classes analisadas. Neste nível encontram-se as frequências de ressonância do trato vocal, formantes, de primeira e segunda ordem F1 e F2. Estas frequências situam-se em torno de 900 Hz (F1) e de 1500 Hz (F2), para vozes consideradas saudáveis. Para vozes patológicas, esse valores são rebaixados, se situando, em alguns casos, na faixa de frequências do nível 5 [18, 19].

TABELA II. AVALIAÇÃO DE DESEMPENHO DA ENERGIA NORMALIZADA DOS COEFICIENTES DE DETALHES EM CADA NÍVEL DE RESOLUÇÃO – SAUDÁVEL X EDEMA.

Nível de Resolução	Acurácia (%)
1	63,33 ± 14,46
2	68,66 ± 13,77
3	70,00 ± 6,49
4	82,44 ± 10,69
5	62,66 ± 11,48
6	62,22 ± 15,35
7	66,11 ± 13,04
8	76,66 ± 14,80

A Tabela III apresenta os melhores resultados da combinação das energias normalizadas obtidas para cada nível de resolução. As faixas de frequência são combinadas 2-2, 3-3, 4-4, 5-5, 6-6, 7-7 e considerando todos os oito níveis analisados.

Observa-se que o nível de resolução 4 aparece em todas as combinações. O melhor resultado foi obtido com a combinação dos níveis 2-4-5-8. Os níveis mais altos, caracterizam as variações no primeiro e segundo formantes, como também os aspectos ruidosos inseridos pela patologia na produção do som sonoro. O nível 8 caracteriza a diminuição na frequência fundamental provocada pela patologia, que devido à presença de massa (edema), apresenta o padrão vibratório das pregas vocais alterado.

TABELA III. AVALIAÇÃO DE DESEMPENHO DA COMBINAÇÃO DA ENERGIA NORMALIZADA DOS COEFICIENTES DE DETALHES EM VÁRIOS NÍVEIS DE RESOLUÇÃO – SAUDÁVEL X EDEMA

Níveis de Resolução	Acurácia (%)
4-7	90,77 ± 9,08
4-5-8	94,77 ± 7,25
2-4-5-8	95,00 ± 7,07
1-2-4-5-8	94,88 ± 7,16
1-2-4-5-6-8	95,00 ± 5,27
1-2-3-4-5-6-7	94,77 ± 5,51
1-2-3-4-5-6-7-8	93,66 ± 8,92

B. Saudável x Nódulo

Na Tabela IV é apresentada a avaliação de desempenho para a discriminação entre as vozes saudáveis e as vozes afetadas por nódulos vocais. Também neste caso, onde são avaliados os níveis individualmente, o nível 4 de decomposição oferece a maior acurácia. Aspectos semelhantes aos observados nas vozes afetadas por edema.

TABELA IV. AVALIAÇÃO DE DESEMPENHO DA ENERGIA NORMALIZADA DOS COEFICIENTES DE DETALHES EM CADA NÍVEL DE RESOLUÇÃO – SAUDÁVEL X NÓDULO

Nível de Resolução	Acurácia (%)
1	76,25 ± 12,82
2	70,35 ± 15,79
3	80,71 ± 19,06
4	91,60 ± 9,89
5	68,75 ± 23,28
6	32,32 ± 14,83
7	75,00 ± 17,57
8	84,46 ± 10,58

Na Tabela V, estão apresentados os valores de acurácia obtidos para a separação entre as classes saudável e nódulo, mediante a combinação dos níveis de resolução. A maior acurácia é obtida com quatro níveis de resolução, 2-4-5-6, com 97,32% de classificação correta. Observa-se, no entanto, que os níveis 4 e 5 foram determinantes no desempenho do sistema, já que não houve aumento significativo na acurácia, ao se aumentar o número de níveis de resolução nas demais combinações.

V. CONSIDERAÇÕES FINAIS

Neste artigo a energia normalizada dos coeficientes de detalhes obtidos através da wavelet de Daubechies de ordem 35 foi empregada como parâmetro na detecção de patologias

TABELA V. AVALIAÇÃO DE DESEMPENHO DA COMBINAÇÃO DA ENERGIA NORMALIZADA DOS COEFICIENTES DE DETALHES EM VÁRIOS NÍVEL DE RESOLUÇÃO - SAUDÁVEL X NÓDULO

Níveis de Resolução	Acurácia (%)
4-5	97,14 ± 6,02
2-4-8	97,14 ± 6,02
2-4-5-6	97,32 ± 5,66
1-2-4-5-6	97,14 ± 6,02
1-2-3-4-6-7	97,14 ± 6,02
1-2-3-4-5-6-7	97,14 ± 6,02
1-2-3-4-5-6-7-8	95,71 ± 6,9

laríngeas. Oito faixas de frequência foram analisadas. Os resultados obtidos nos dois casos analisados, Saudável x Edema de Reinke e Saudável x Nódulo, atestam que o nível de resolução 4 fornece a maior separação entre as classes analisadas. Na combinação das energias normalizadas em resoluções diferentes, tal nível aparece em todas as combinações. Uma eficácia de 95% foi obtida na discriminação entre os sinais de vozes saudáveis e afetados por edema de Reinke e de 97% na classificação dos sinais entre saudáveis e afetados por nódulos vocais. Os resultados são promissores e interessantes, tendo em vista que uma única medida foi capaz de fornecer taxas superiores a 95% de acurácia, fornecendo uma indicação de que a técnica empregada pode ser efetivamente considerada para ferramenta de apoio ao diagnóstico de patologias laríngeas como nódulos e edema e acompanhamento de terapias vocais.

AGRADECIMENTOS

Os autores agradecem a UFCG pelo fornecimento da base de dados e ao IFPB pelo suporte financeiro.

REFERÊNCIAS

- [1] 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 Transactions on Biomedical Engineering, Vol. 53, No. 10, 2006.
- [2] D. M. Hartl, S. Hans, J. Vaissière, M. Riquet, D. F. Brasnu, "Objective voice quality analysis before and after onset of unilateral vocal fold paralysis", J. Voice, 15, v.3, 1999.
- [3] A. Parraga, Aplicação da Transformada Wavelet Packet na Análise e Classificação de Sinais de Vozes Patológicas. Dissertação (Mestrado), UFRS, Porto Alegre, 2002.
- [4] 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.
- [5] C-W. Jo e D. H. Kim, "Analysis of Disordered Speech Signal using Wavelet Transform" In: Proceedings of 5th International Conference on Spoken Language Processing, 1998.
- [6] J. Nayak, P. S. Bhat, R. Acharya e U. V. Aithal. "Classification and analysis of speech abnormalities". ITBM-RBM, v. 27, 2005.

- [7] L. M. Sousa. Detecção Inteligente de Patologias na Laringe baseada em Máquinas de Vetor de Suporte e na Transformada Wavelet. Dissertação de Mestrado, USP, São Carlos, 2010.
- [8] P. Kulharchik, D. Martynov, I. Kheidorov e O. Kotov. "Vocal Fold Pathology Detection using Modified Wavelet-Like Features and Support Vector Machines", 15th European Signal Processing Conference (EUSIPCO 2007), 2007.
- [9] R. Behroozmand, e 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, 2007
- [10] L. Salhi; M. Talbi and A. Cherif. "Voice disorders identification using hybrid approach: wavelet analysis and multilayer neural networks", *World Academy of Science, Engineering and Technology*, 45, pp. 330-339, 2008.
- [11] P. T. Hosseini e F. Almasganj. "Different mother wavelets and pathological voice". 2nd International Conference on Computer, Control and Communication (ICA 2009), 2009.
- [12] N. C. Almeida. Sistema Inteligente para Diagnóstico de Patologias na Laringe utilizando Máquinas de Vetor Suporte, Dissertação (Mestrado), UFRN, Natal, 2010.
- [13] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1999.
- [14] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial & Applied, 1992.
- [15] KAY ELEMETRICS CORP. *Disordered voice database, model 4337*, 03 Ed, 1994.
- [16] M. Behlau. *A Voz do Especialista*, volume 1. Revinter, Rio de Janeiro, 2001.
- [17] S. E. N. Correia, S. C. Costa, e W. C. A. Costa. "Estudo do Potencial Discriminativo da Transformada Wavelet na Caracterização de Sinais de Vozes Patológicas." In: VI Congresso de Pesquisa e Inovação da Rede Norte Nordeste de Educação Tecnológica (CONNEPI 2011), Natal, 2011.
- [18] S. L. N. C. Costa. "Análise Acústica, Baseada no Modelo Linear de Produção da Fala, para Discriminação de Vozes Patológicas". Tese de Doutorado, UFCG, Campina Grande, 2008.
- [19] I. C. P. Russo, M. Behlau. *Percepção da fala: análise acústica do português brasileiro*. São Paulo: Lovise, 1993.