

Estimação Inteligente do Grau Geral de Desvio Vocal Utilizando Transformada Wavelet Packet

Andriéli H. Barizão*. Victor H. Yoshizumi*.
Danilo H. Spatti**. María E. Dajer***. Rogério A. Flauzino*.

* Departamento de Engenharia Elétrica e de Computação, Universidade de São Paulo, São Carlos, Brasil
(Tel: +55 18 9 8128-2016; e-mail: andrieli.h.b@usp.br); (Tel: +55 44 9 9807-5369; e-mail: yoshizumi@usp.br);
(Tel: +55 16 9 8124-4281; e-mail: raflauzino@usp.br).

** Departamento de Sistemas de Computação Universidade de São Paulo, São Carlos, Brasil
(Tel: +55 16 9 9181-9539; e-mail: spatti@icmc.usp.br)

*** Departamento da Elétrica - Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Brasil
(Tel: +55 43 9 9630-0369; e-mail: medajer@utfpr.edu.br)

Abstract: The current model of auditory-perceptual assessment of the voice is based on visual analogical scales and seeks continuous improvement under numerous aspects, with emphasis on the prescription of a good assessment regarding vocal quality, as well as its diagnosis and monitoring of treatments. Reflecting on the eventualities that degrade the aspects, there is a lack of standardization of the commonly used evaluation protocols. Therefore, the main objectives of this article were to develop computational tools dedicated to the estimation of the subjective parameter, the general degree of vocal deviation. More specifically, the audio file fragmentation technique was developed, the down sample decomposition of these data for level six of the energy coefficients of the wavelet packet transform referring to six different wavelet filter families, which are: daubechies, symlets, coiflets, fejer-korovkin, biorthogonal e reverse biorthogonal, proposing an automated methodology for the implementation of multilayer perceptron neural networks in order to integrate the variables arising from such decomposition. Results with real vocal samples are presented to demonstrate the efficiency of the developed techniques, exposing hit rates greater than 98% for the elaborated method, where it simulates a real environment that can be used by professionals in the field as a quick aid tool in their diagnoses.

Resumo: O atual modelo de avaliação perceptivo-auditiva da voz está alicerçado em escalas analógicas visuais e buscam melhorias contínuas sob inúmeros aspectos, com destaque para a prescrição de uma boa avaliação referente a qualidade vocal, bem como o seu diagnóstico e monitoramento de tratamentos. Refletindo-se sobre as eventualidades que degradam os aspectos, constata-se a não padronização dos protocolos avaliativos comumente utilizados. Assim sendo, os objetivos principais deste artigo consistiram em desenvolver ferramentas computacionais dedicadas à estimação do parâmetro subjetivo, grau geral do desvio vocal. Mais especificamente, desenvolveu-se a técnica de fragmentação dos arquivos de áudios, a decomposição *downsample* desses dados para o nível seis dos coeficientes de energia da transformada wavelet packet referentes a seis famílias de filtros *wavelets* distintos, sendo elas: *daubechies*, *symlets*, *coiflets*, *fejer-korovkin*, *biorthogonal* e *reverse biorthogonal*, propondo uma metodologia automatizada para implementação das redes neurais perceptron multicamadas com a finalidade de integrar as variáveis advindas de tal decomposição. Resultados com amostras vocais reais são apresentados para demonstrar a eficiência das técnicas desenvolvidas, expondo taxas de acertos maiores que 98% para o método elaborado, onde o mesmo simula um ambiente real que poderá ser utilizado pelos profissionais da área como ferramenta de auxílio rápida em seus diagnósticos.

Keywords: Estimation of parameters; auditory-perceptual evaluation; wavelet packet transformed; wavelet filters; artificial neural networks; multilayer perceptron.

Palavras-chaves: Estimação de parâmetros; avaliação perceptivo-auditiva; transformada *wavelet packet*; filtros *wavelets*; redes neurais artificiais; *perceptron* multicamadas.

1. INTRODUÇÃO

A voz humana desempenha um papel vital para interação das pessoas, sendo assim, indiscutível sua relevância nas áreas médicas, mais especificamente no âmbito da fonoaudiologia (VERMEULEN *et al.*, 2020) (SANTOS, 2015). Assim, é

neste contexto da quantificação da qualidade de vozes patológicas que as análises perceptivas-auditivas da voz permanecem como método padrão nos diagnósticos de pacientes com distúrbios vocais. (FUJIMURA *et al.*, 2020). De modo que, a categoria dos professores apresenta alto índice de disfunção vocal, uma vez que formam um grande

grupo de usuários profissionais da voz e acredita-se que correm maior risco relacionados a problemas vocais do que a população em geral (RUSSELL, OATES, GREENWOOD, 1997).

A avaliação perceptivo-auditiva da voz é uma habilidade clínica subjetiva e, portanto, podem variar entre os examinadores. Embora muitas métricas acústicas tenham sido estudadas para potencializar a avaliação objetiva de vozes patológicas, a interpretação das mesmas, em casos individuais se torna difícil, e a técnica não é amplamente utilizada pelos médicos, que acabam confiando em diretrizes e protocolos para ajudá-los a julgar a qualidade da voz de seus pacientes (FUJIMURA *et al.*, 2020).

Os protocolos comumente utilizados incluem o GRBAS (*grade, roughness, breathiness, asthenia, strain*) e o CAPE-V (*Consensus Auditory-Perceptual Evaluation of Voice*). Sendo uma escala ordinal de 4 pontos usada pelo GRBAS e uma escala visual analógica (EVA) utilizada pelo CAPE-V, ambos fornecem o parâmetro para a avaliação da severidade geral (CAPE-V) ou grau (GRBAS) do desvio vocal (COHEN *et al.*, 2018).

Contudo, os diferentes tipos de distúrbios da voz estão classificados como distúrbios orgânicos e funcionais, e foram comparados em termos de pontuação no CAPE-V. O parâmetro do grau geral de desvio vocal apresentou os maiores valores intra e inter-confiabilidade para todos os participantes (ÖZCEBE *et al.*, 2017). Como também, as correlações para os parâmetros entre as escalas GRBAS e CAPE-V, sendo a do parâmetro do grau geral de desvio vocal a maior encontrada (ERTAN-SCHLUTER *et al.*, 2019).

Neste cenário, propõe-se uma abordagem alternativa para os julgamentos referentes a qualidade vocal, mantendo o método não-invasivo abordado pelas avaliações perceptivo-auditivas da voz. Sendo assim, o desenvolvimento de uma ferramenta computacional inteligente que possa extrair aprendizado dos parâmetros dos sinais vocais. Bem como, a implementação da transformada *wavelet packet* (TWP) e as redes neurais artificiais (RNA) que veem se consagrando na área dos estudos biomédicos, tanto para estimar quanto para remover ruídos em eletromiografias (GALIANA-MERINO, RUIZ-FERNANDEZ, MARTINEZ-ESPLA, 2013), detecção automática e classificação cardíaca (ZAIRI *et al.*, 2019) e também para classificação de bio-sinais (LIMA *et al.*, 2019) (BARIZÃO *et al.*, 2018).

Acreditando-se que o desenvolvimento de uma metodologia automatizada para a estimação de parâmetros subjetivos da voz possa auxiliar os profissionais da área, a priori, tem-se como base uma parcela da EVA, Fig.1, que se refere ao primeiro parâmetro observado pelos protocolos mais utilizados pelos avaliadores: grau geral do desvio vocal. Tais julgamentos perceptivos podem ser usados para comparação da magnitude das mudanças sucedidas durante o tratamento realizado pelo paciente, como também para o treinamento da sensibilidade audível dos avaliadores que realizam tais diagnósticos.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Escala Visual Analógica

As ferramentas de medição mais comuns usadas na avaliação perceptivo da qualidade da voz produzem dados ordinais e, portanto, não suportam o estabelecimento de relações matemáticas entre diferentes valores de medição. Isso torna a interpretação deles um desafio. Entre as muitas características desejáveis de qualquer ferramenta de medição psicofísica, está a capacidade de quantificar a diferença entre duas ou mais medidas e a capacidade de interpretação destas medidas de uma maneira que esteja relacionada à experiência do observador. O primeiro permite comparar as medidas usando matemática simples, enquanto o segundo permite que essa comparação seja interpretada de maneira construtiva (EL-HENDAWI, WANG, 2020).



Fig. 1. Exemplo de parcela da escala visual analógica utilizada pelo protocolo CAPE-V.

2.2 Transformada Wavelet Packet

A transformada *wavelet packet*, como pode ser observado em (GALIANA-MERINO, RUIZ-FERNANDEZ, MARTINEZ-ESPLA, 2013) (ZHONG, LI, 2019) e (EL-HENDAWI, WANG, 2020), tem se mostrado promissora na extração de características de sinais em processos de classificação e estimação. Por meio desta, é possível realizar análises tempo-frequência e extrair diferentes níveis de energia do sinal. Pode-se observar as inúmeras aplicações das transformadas *wavelet* em sinais biológicos em (FUJIMURA *et al.*, 2020) (GALIANA-MERINO, RUIZ-FERNANDEZ, MARTINEZ-ESPLA, 2013) (LIMA *et al.*, 2019) (BARIZÃO *et al.*, 2018).

2.3 Perceptron Multicamadas

Por sua vez, em relação ao aprendizado inteligente, tem-se como sugestão, inspirado em (FUJIMURA *et al.*, 2020) (CARVALHO, CAVALCANTE, CORTEZ, 2011) (RABINER, SCHAFFER, 2012), um estimador fundamentado em RNA. Para a proposta deste artigo, a arquitetura utilizada é a *Perceptron Multicamadas* (PMC), por possuir uma das arquiteturas mais versáteis quanto a sua aplicabilidade (Da SILVA, SPATTI, FLAUZINO, 2016).

3. MATERIAIS E MÉTODOS

Serão discutidos nesta seção os materiais e ferramentas necessárias para consolidar a proposta desta pesquisa.

3.1 Banco de Dados

A primeira etapa consiste na construção de um banco de dados composto por gravações acústicas da vogal sustentada /e/ para análises perceptivo-auditivas do grau geral do desvio vocal, como pode ser observado em (ZAMBON, 2012). Tais arquivos foram disponibilizados pela pesquisadora Fabiana Copelli Zambon, não dispendo riscos, tampouco advindos de procedimentos invasivos para as professoras que aceitaram participar deste estudo.

Os arquivos de áudios foram armazenados em extensão .WAV à taxa de 11025Hz e 44100Hz amostras por segundos.

A composição dos arquivos presentes no banco de dados está representada na Tabela 1.

Tabela 1. Composição do banco de dados

Características	Quantidade
Apresentam queixa vocal Buscaram tratamento fonoaudiológico	30
Apresentam queixa vocal Nunca buscaram tratamento fonoaudiológico	30
Não apresentam queixa vocal	30
<i>Total</i>	90

3.2 Pré-processamento digital de sinais

Em seguida, os arquivos de áudios passaram por filtros modelados, com o intuito de remover silêncios e segmentos não vocalizados ou qualquer outro tipo de ruído não relevante ao processo de estimação do grau geral de desvio vocal, restando então 82 arquivos para realização das análises seguintes. Nesta etapa é feito também a segmentação do sinal e janelamento, responsável tanto para agilizar o processo (reduzindo o tempo necessário para a aplicação da TWP e a extração das características) quanto para produzir uma quantidade mais elevada de amostras para o treinamento, aumentando o desempenho e validação das RNA. O janelamento simples é análogo ao conceito de segmentação, onde tem-se um sinal completo e a cada janelamento é dado um passo correspondente ao próprio comprimento da janela.

Antes de iniciar o procedimento, todo o banco de dados é inspecionado, realizando-se a leitura da frequência de amostragem. Feito isso, 7 arquivos foram removidos a partir do critério de Nyquist, onde o mesmo relaciona uma frequência de amostragem ideal para que o sinal possa ser reconstruído apropriadamente, e mantendo assim, uma uniformidade de análise. A Tabela 2 apresenta a quantidade de arquivos não úteis.

Tabela 2. Quantidades de arquivos não úteis

Características	Quantidade
Arquivos de áudios com taxa de amostragem incompatíveis para o estudo	7
Arquivos de áudios com silêncios, segmentos não vocalizados ou qualquer outro tipo de ruído não relevante ao processo de estimação do grau geral do desvio vocal	8
<i>Total</i>	15

Ao final dessas etapas, reservou-se 3 arquivos para testar a generalização da RNA, certificando a confiabilidade do estudo desejado, sendo 1 arquivo pertencente ao grupo 1, o outro ao grupo 2 e o último ao grupo 3. A frequência de amostragem utilizada foi 44100Hz, permitindo gerar 28365 segmentos de 1024 pontos a partir dos 72 arquivos úteis.

3.3 Processamento digital de sinais

O processo referente a extração dos níveis de energia de cada segmento a partir da aplicação da TWP pode ser observado por meio da Fig. 2.

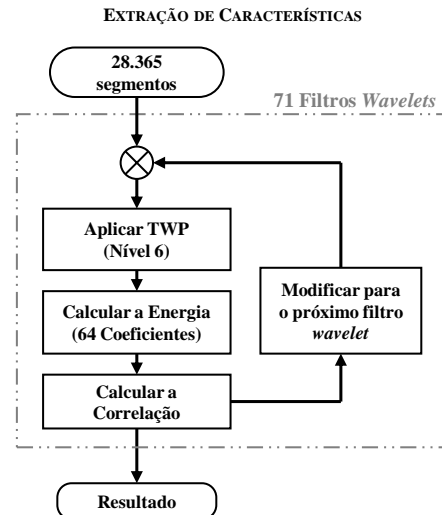


Fig. 2. Procedimento de extração de características dos segmentos.

E adotando-se o nível seis dos coeficientes de energia da TWP, como pode ser visto pela Fig. 3, referentes a seis famílias de filtros wavelets distintos e suas respectivas ordens, apresentados pela Tabela 3, é possível estabelecer o universo amostral de treinamento e validação das RNA.

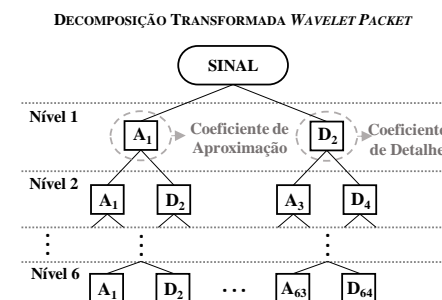


Fig. 3. Decomposição da TWP para o nível 6.

Tabela 3. Filtros wavelets abordados

Famílias Wavelets	Filtros Testados
Symlets	'sym16', 'sym15', 'sym14', 'sym13', 'sym12', 'sym11', 'sym10', 'sym9', 'sym8', 'sym7', 'sym6', 'sym5', 'sym4', 'sym3 e 'sym2'
Reverse Biorthogonal	'rbio6.8', 'rbio5.5', 'rbio4.4', 'rbio3.9',

	'rbio3.7', 'rbio3.5', 'rbio3.3', 'rbio3.1', 'rbio2.8', 'rbio2.6', 'rbio2.4', 'rbio2.2', 'rbio1.5', 'rbio1.3' e 'rbio1.1'
Fejer-Korovkin	'fk22', 'fk14', 'fk8', 'fk6' e 'fk4'
Daubechies	'db16', 'db15', 'db14', 'db13', 'db12', 'db11', 'db10', 'db9', 'db8', 'db7', 'db6', 'db5', 'db4', 'db3', 'db2' e 'db1' or 'haar'
Coiflets	'coif5', 'coif4', 'coif3', 'coif2' e 'coif1'
Biorthogonal	'bior6.8', 'bior5.5', 'bior4.4', 'bior3.9', 'bior3.7', 'bior3.5', 'bior3.3', 'bior3.1', 'bior2.8', 'bior2.6', 'bior2.4', 'bior2.2', 'bior1.5', 'bior1.3' e 'bior1.5'

Para cada segmento de arquivo é aplicada a TWP com uma das 71 configurações de filtros abordados, extraindo o nível de energia dos 64 coeficientes presentes no sexto nível da decomposição. É feito, também, o cálculo e o armazenamento da correlação dos níveis de energia em relação as classes desejadas para utilização na próxima etapa do estudo.

3.4 Características das RNA e processo de treinamento

Visando a maior eficiência na generalização das redes e a redução da complexidade da modelagem, optou-se pela elaboração de uma RNA especialista para a estimação do grau geral de desvio vocal. As características das RNA utilizadas estão descritas na Tabela 4.

Vale ressaltar que a função de ativação escolhida para as camadas escondidas são a *tansig* e para a camada de saída a *purelin*, garantindo robustez para o resultado obtido pela estimação.

Por sua vez, definiu-se um grupo de 29 variações de neurônios para a primeira e segunda camadas escondidas durante o processo de busca pela melhor topologia. Tais variações estão descritas na Tabela 5. Para evitar um descarte de topologia devido a inicialização dos parâmetros aleatórios em um mínimo local da função erro, são realizadas 3 iterações na mesma topologia antes da próxima variação (armazenando apenas o melhor dos três resultados).

Tabela 4. Características das RNAs aplicadas

Arquitetura da rede	Multilayer Perceptron
Tipo de treinamento	Supervised – <i>Off-line</i>
Algoritmo de treinamento	Levenberg-Marquardt (<i>backpropagation</i>)
Taxa de aprendizagem	10^{-2}
Critério de parada (Erro)	10^{-5}
Épocas	2000
Proporção Treinamento / Validação	60% / 40%
Número total de amostras	28.365

Tabela 5. Variações de topologias

1ª Camada escondida (neurônios)	1, 3, 5, 7, 9
2ª Camada escondida (neurônios)	2, 4, 6, 8, 10
Treinamentos por topologia	3 treinamentos

Com objetivo de reduzir o tempo e esforço computacional e, tendo em mãos a análise de correlação de *Pearson* realizada,

foi proposto a criação de 5 grupos distintos de amostras para a RNA. O primeiro deles consiste dos 8 coeficientes mais correlacionados (8C), o segundo dos 16 coeficientes mais correlacionados (16C), o terceiro dos 24 coeficientes mais correlacionados (24C), o quarto dos 32 coeficientes mais correlacionados (32C) e o quinto com todos os coeficientes (64C).

A análise e escolha dos coeficientes é feita individualmente para cada uma das 71 famílias, contudo, para facilitar a visualização, é representada a correlação mediana das famílias na Fig.4.

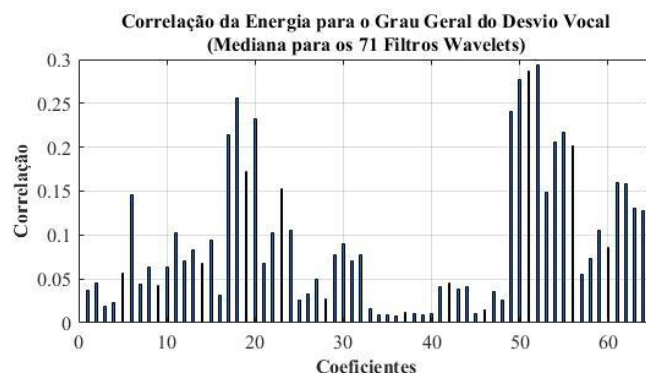


Fig. 4. Correlação mediana da energia para os 64 coeficientes.

Durante a construção do conjunto de amostras apresentado à RNA foram realizados os seguintes procedimentos:

- Normalização dos dados: a presença de níveis de energia muito discrepantes entre cada coeficiente é prejudicial ao processo de treinamento da RNA, logo é realizada a normalização de todo o banco de dados entre -1 e 1.
- Máximos e mínimos: para garantir que durante a etapa de validação todas as amostras apresentadas à rede estejam dentro do universo de treinamento realizado, as amostras que possuíam os valores de máximos ou mínimos estavam sempre presentes no conjunto de treinamento.
- Embaralhamento das amostras: todas as amostras (excluindo máximos e mínimos) são embaralhadas entre o conjunto de treinamento e teste a cada inicialização de um novo treinamento, tornando o processo mais dinâmico e mantendo uma maior variabilidade de amostras apresentadas à rede.

Por fim, os procedimentos descritos nessa seção podem ser visualizados de maneira simplificada por meio do fluxograma que compõe a Figura 5.

Nota-se que a busca pela melhor topologia da PMC capaz de estimar com eficiência o parâmetro subjetivo do grau geral do desvio vocal é resultado da combinação de variações de coeficientes, variações de filtros *wavelets* e também de neurônios por camada, totalizando 30.885 permutações realizadas.

Para que isto se tornasse possível, foi desenvolvido uma rotina de treinamento automatizada responsável por todas as

variações e armazenamento dos resultados. Tais dados obtidos serão analisados ao longo da próxima seção.

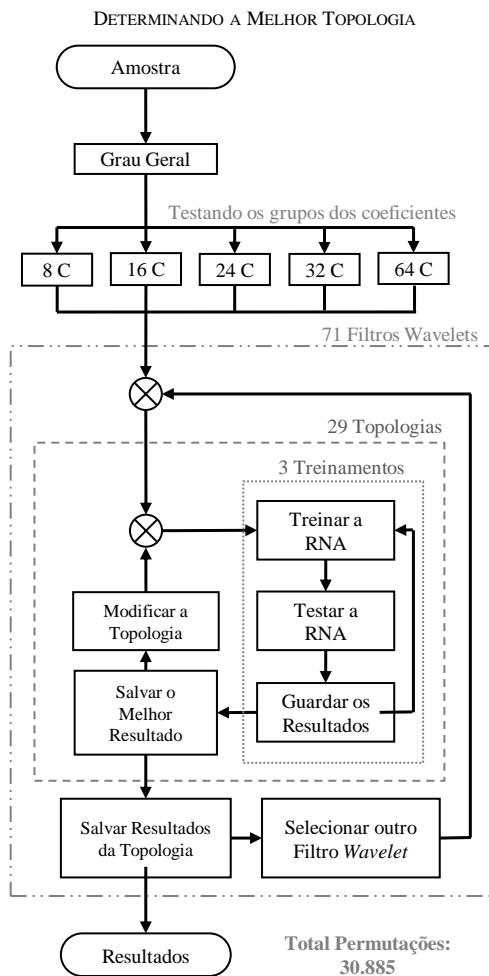


Fig. 5. Procedimento de busca pela melhor topologia abordada.

4. RESULTADOS

Após a finalização da rotina automatizada desenvolvida, obteve-se uma massa de dados referente a todos os treinamentos realizados.

Feita a análise dos dados, os melhores resultados na fase de treinamento/validação para cada grupo de coeficientes mais correlacionados estão representados na Tabela 6.

Tabela 6. Resultados finais treinamento/validação

Rede Neural Artificial	Topologia		C	WPT Filtro	Taxas de acertos (%)
	L1	L2			
Grau Geral do Desvio Vocal	1	2	8	db1	98.13
	1	2	16	db1	98.73
	5	2	24	db1	98.89
	9	4	32	db1	99.32
	7	2	64	db1	99.62

É possível notar que mesmo contendo uma correlação baixa, os resultados para a fase de treinamento/validação foram significativos, sendo assim, as Figuras 6, 7, 8, 9 e 10 apontam o desempenho mediano das outras famílias de filtros *wavelets*

testadas nos grupos das variações das topologias apresentadas.

Após a obtenção da melhor topologia, foi realizada mais uma bateria de testes para assegurar a consistência do resultado obtido. Nesta última etapa, foi acrescentada uma quantidade maior de neurônios por camada com o intuito de averiguar se os erros que ainda existiam na estimação não tinham como origem uma insuficiente quantidade de neurônios (*underfitting*), apresentados na Tabela 7.

Tabela 7. Variações de topologias (overfitting)

1ª Camada escondida (neurônios)	15, 30
2ª Camada escondida (neurônios)	15, 30
Treinamentos por topologia	3 treinamentos

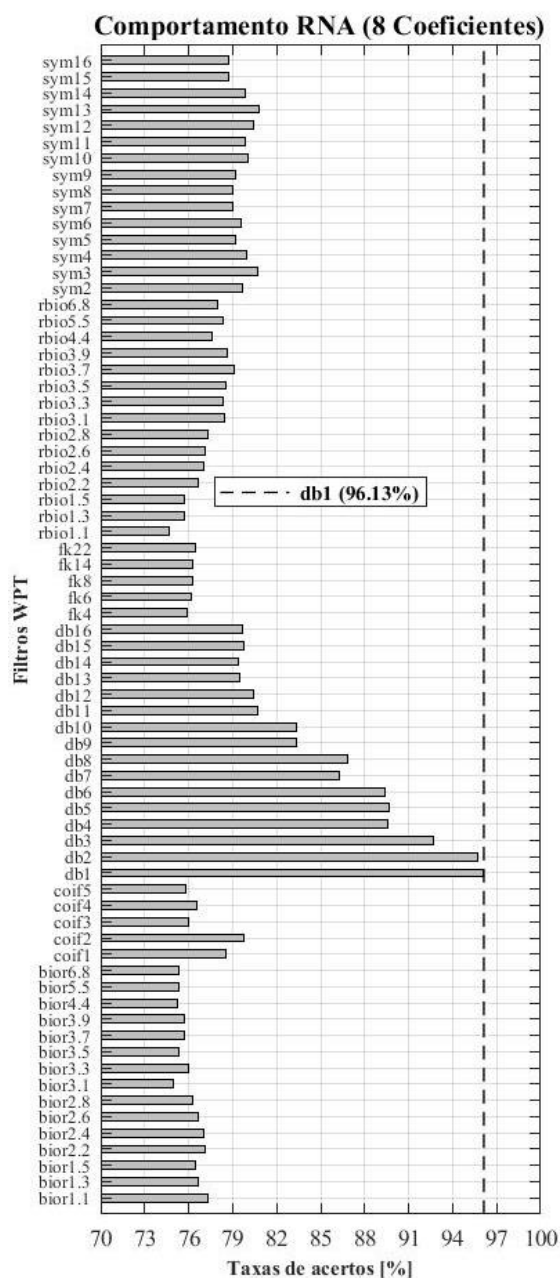


Fig. 6. Comportamento RNA com os 8 coeficientes mais correlacionados.

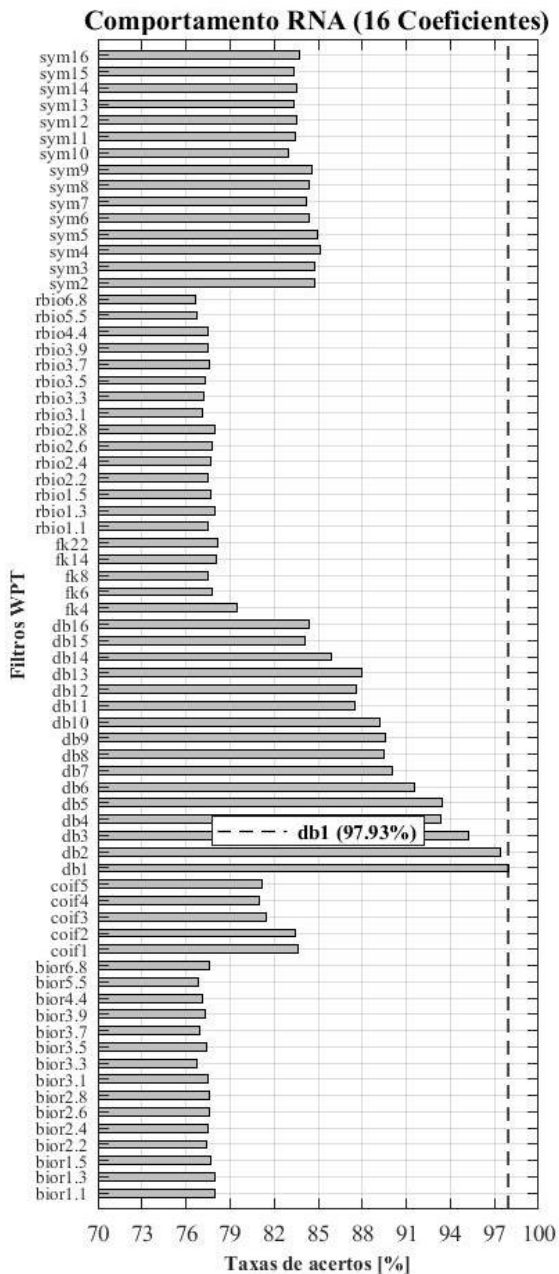


Fig. 7. Comportamento RNA com os 16 coeficientes mais correlacionados.

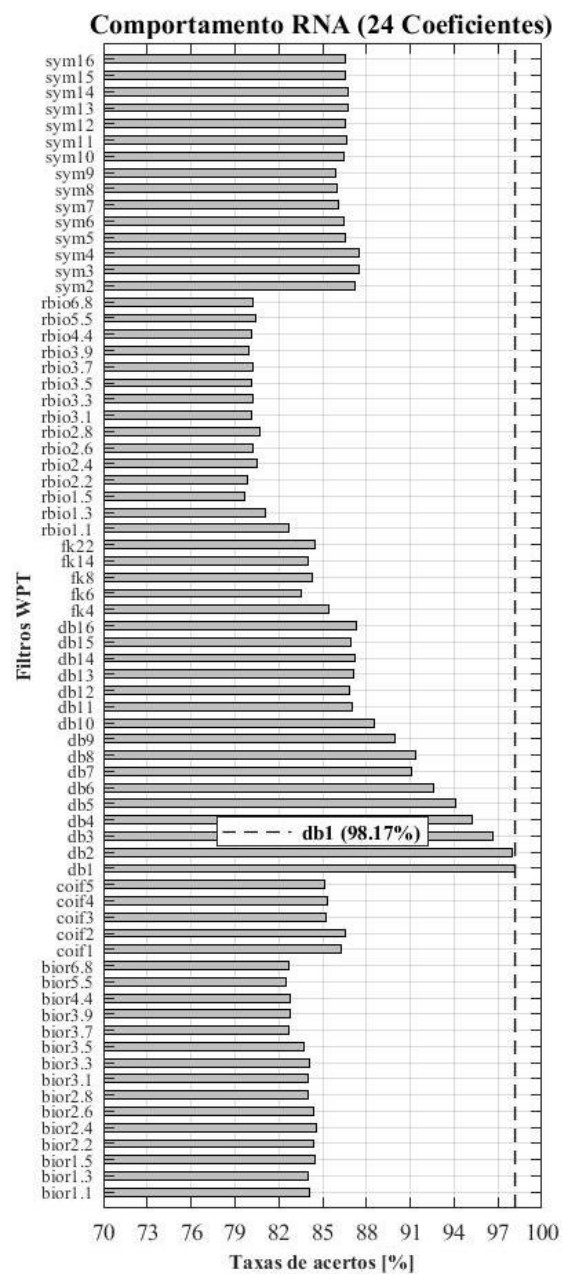


Fig. 8. Comportamento RNA com os 24 coeficientes mais correlacionados.

5. DISCUSSÃO E CONCLUSÕES

O objetivo principal deste trabalho foi propor um método capaz de auxiliar profissionais da voz em suas avaliações subjetivas, agregando maior confiabilidade ao procedimento avaliativo, bem como, facilitando o longo e exaustivo processo envolvido. Foram realizadas mais de 30 mil permutações de configurações das RNA, atingindo elevado nível de eficiência e mantendo taxas de acertos acima dos 98,12%. Nota-se pelos resultados que, mesmo propondo uma análise de correlação para reduzir o conjunto de entradas das RNA, o grupo completo de entradas se mostrou mais eficaz. Deste modo, mesmo os coeficientes de baixa correlação se mostraram essenciais para a estimação

Por fim, cabe salientar que o estudo elaborado não descarta a importância e aplicação dos métodos convencionais de avaliação e diagnóstico, propondo apenas uma ferramenta complementar para auxiliar os profissionais das áreas de saúde vocal de maneira automatizada.

Nota-se que o estimador apresentou índices de acerto consistentes, Tabela 8 e 9. Contudo, a realização de apenas 3 testes (cerca de 4% da população total estudada) não apresenta forte embasamento estatístico para realmente comprovar que as RNA conseguiram adquirir aprendizado em relação a estimação do grau geral do desvio vocal e desenvolver um elevado grau de generalização para operação com arquivos desconhecidos.

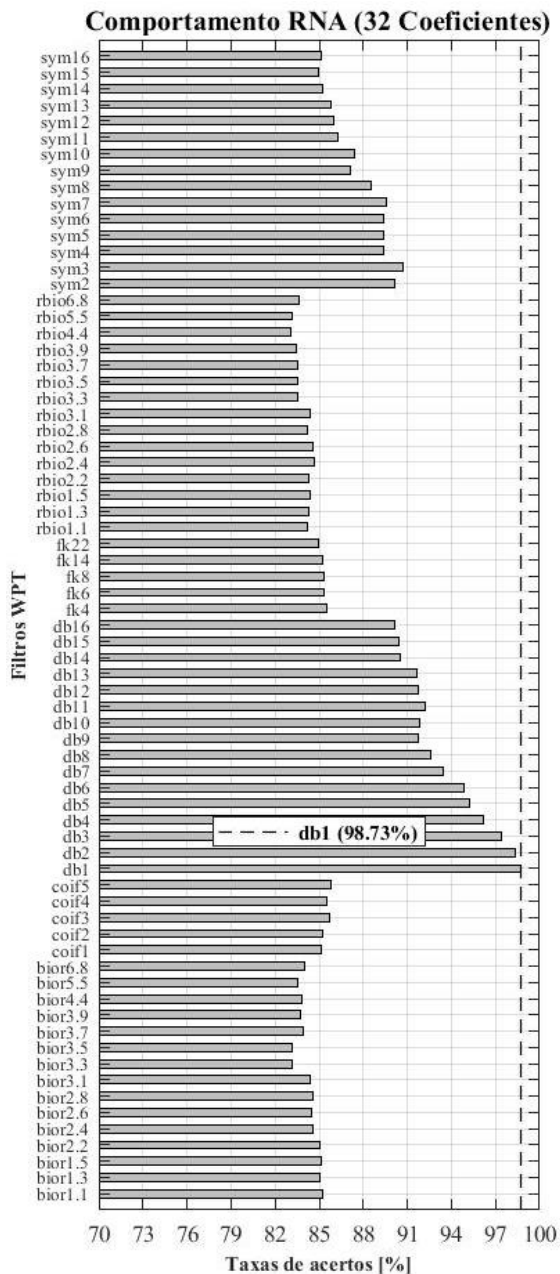


Fig. 9. Comportamento RNA com os 32 coeficientes mais correlacionados.

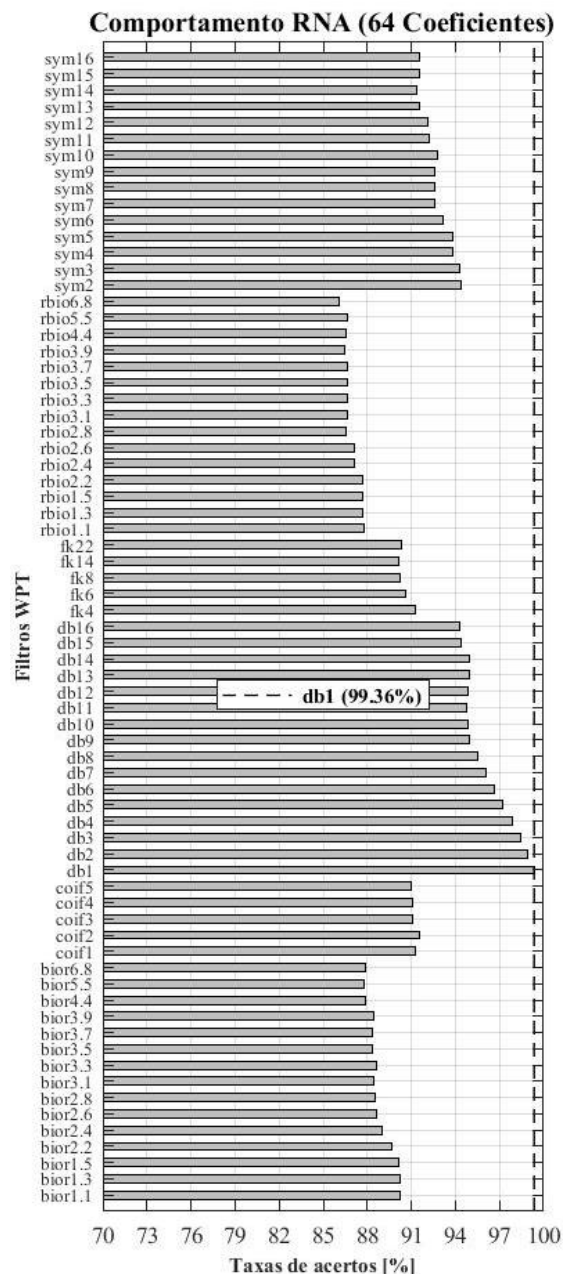


Fig. 10. Comportamento RNA com todos coeficientes.

Tabela 8. Resultados finais generalização

Rede Neural Artificial	Topologia		C	Filtros Wavelets	Taxas de acertos (%)
	L1	L2			
Grau Geral do Desvio Vocal	1	2	16	db1	98,73
	1	2	64	bior1.1	91,43
	1	2	64	rbio1.1	90,38
	1	2	64	sym6	94,41

E por meio dos resultados obtidos, comprova-se a eficácia do método proposto, atingindo elevado nível de confiabilidade.

Deve-se ressaltar que a mesma margem de confiabilidade aplicada aos profissionais fonoaudiólogos que realizam as avaliações (5 pontos para mais ou para menos) foi imposta para os resultados alcançados na fase de generalização das RNA.

Tabela 9. Estimação generalização

Áudios reservados	Filtro wavelets	Estimação percentual	Margem de confiabilidade aceitáveis
[29]	Db1 (16C)	65%	65≤70≤75
[57]		33%	26≤31≤36
[78]		33%	33≤38≤43

Contudo, a RNA db1 (16C) possui quantidade de neurônios artificiais, interconexões e menor grupo de amostras em sua camada de entrada, apresentando assim, geralmente, uma melhor capacidade de generalização, mesmo que seja necessário sacrificar uma pequena parcela de acertos na etapa de treinamento/validação.

AGRADECIMENTOS

Agradeço imensamente as pesquisadoras Mara Behlau e Fabiana Copelli Zambon pela contribuição com a base de dados deste trabalho.

REFERÊNCIAS

- Barizão, A. H. *et al.* (2018). Voice disorder classification using MLP and Wavelet Packet Transform. Piscataway: IEEE World Congress on Computational Intelligence – WCCI. Rio de Janeiro – RJ, Brazil.
- Carvalho, R. T. S. (2012). Transformada Wavelet na Detecção de Patologias da Laringe. 2012. Tese (PhD) – Universidade Federal do Ceará.
- Cohen, W., Lloyd, S., Wynne, D.M., and Townsley, R.B. (2018). Perceptual Evaluation of Voice Disorder in Children Who Have Had Laryngotracheal Reconstruction Surgery and the Relationship Between Clinician Perceptual Rating of Voice Quality and Parent Proxy/Child Self-Report of Voice-Related Quality of Life. The Voice Foundation. Published by Elsevier Inc. *Journal of Voice*, Vol. 33, No. 6, pp. 945.e27945.e35. University of Strathclyde, Graham Hills Building, George Street, United Kingdom.
- da Silva, I.N., Spatti, D.H., and Flauzino, R.A. (2016). *Redes Neurais Artificiais para engenharia e ciências aplicadas*. São Paulo, SP: Artliber.
- Eddins, D.A., Anand, S., Lang, A., and Shrivastav, R. (2019). The Voice Foundation. Published by Elsevier Inc. *Journal of Voice*. Department of Communication Sciences and Disorders, University of South Florida, Tampa, Florida.
- El-Hendawi, M., and Wang, Z. (2020). An ensemble method of full wavelet packet transform and neural network for short term electrical load forecasting, *Electric Power Systems Research* 182, 106265, Elsevier B.V. Doi: 10.1016/j.epr.2020.106265.
- Ertan-Schluter, E., Demirhan, E., Unsal, E.M., and Ozkan, E.T. (2019). The Turkish Version of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V): A Reliability and Validity Study. The Voice Foundation. Published by Elsevier Inc. *Journal of Voice*. Department of Speech and Language Therapy, Kavack Campus, School of Health Sciences, Istanbul Medipol University, Istanbul, Turkey.
- Fujimura, S., Kojima, T., Okanou, Y., Shoji, K., Inoue, M., Omori, K., and Hori, R. (2020). Classification of Voice Disorders Using a One-Dimensional Convolutional Neural Network, *Journal of Voice*. Doi: 10.1016/j.jvoice.2020.02.009.
- Galiana-Merino, J.J., Ruiz-Fernandez, D., and Martinez-Espla, J.J. (2013). Power line interference filtering on surface electromyography based on the stationary wavelet packet transform, *Computer Methods and Programs in Biomedicine*, Volume 111, Issue 2, Pages 338-346. Doi: 10.1016/j.cmpb.2013.04.022.
- Lima, A. A. M.; BARROS, F. K. H.; YOSHIZUMI, V. H.; SPATTI, D. H.; and DAJER, M. E. Optimized Artificial Neural Network for Biosignals Classification Using Genetic Algorithm, *Journal of Control, Automation and Electrical Systems* (2019) 30:371–379. Doi: 10.1007/s40313-019-00454-1.
- Özcebe, E., Aydinli, F.E., Tigrak, T.K., Incebay, Ö., and Yilmaz, T. (2017). Reliability and Validity of the Turkish Version of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). The Voice Foundation. Published by Elsevier Inc. *Journal of Voice*, Vol. 33, No. 3, pp. 382.e1–382.e10. Hacettepe University Faculty of Medicine, Sıhhiye, Ankara, Turkey.
- Rabiner, L.R., and Schafer, R.W. (2012). *Theory and Applications of Digital Speech Processing*. First Edition. New Jersey, US: John Wiley & Sons Inc.
- Russell, A., Oates, J., and Greenwood, K.M. (1998). Prevalence of Voice Problems in Teachers, *Journal of Voice*, Vol. 12, No. 4, pp. 467-479, 26th Annual Symposium: Care of the Professional Voice, Philadelphia, Pennsylvania, U.S.A.
- Santos, M.O. (2015). Análise Acústica de Desvios Vocais Infantis utilizando a Transformada Wavelet. 2015. 80f. Dissertação (Mestrado) – Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, João Pessoa, Brasil.
- Vermeulen, R., Linde, J.V.D., Abdoola, S., Lierde, K.V., and Graham, M.A. (2020). The Effect of Superficial Hydration, With or Without Systemic Hydration, on Voice Quality in Future Female Professional Singers, *Journal of Voice*, Doi: 10.1016/j.jvoice.2020.01.008.
- Zairi, H., Talha, M.K., Meddah, K., and Slimane, S.O. (2019). FPGA-based system for artificial neural network arrhythmia classification. Springer-Verlag London Ltd. *Neural Computing and Applications*. Department of Electronics and Computer Science, University of Sciences and Technology Houari Boumediene, USTHB BP 32 el alia, 16111 Bab Ezzouar, Algiers, Algeria.
- Zambon, F.C. (2012). Estratégias de enfrentamento em professores com queixa de voz. Programa de Pós-graduação (Doutorado) em Distúrbios da Comunicação Humana, Universidade Federal de São Paulo – UNIFESP – São Paulo (SP), Brasil; Centro de Estudos da Voz – CEV – São Paulo (SP), Brasil; Sindicato dos Professores de São Paulo – SINPRO-SP – São Paulo (SP), Brasil.
- Zhong, Z., and Li, H. (2019). Recognition and prediction of ground vibration signal based on machine learning algorithm, *Neural Computing and Applications* (2020) 32:1937–1947. Doi: 10.1007/s00521-019-04496-z.