

Wavelet packet transform and multilayer perceptron to identify voices with a mild degree of vocal deviation

Transformada Wavelet packet y Perceptrón Multicapa para identificación de voces con grado leve de desvío vocal

Mateus Morikawa¹  , Danilo Hernane Spatti²  , María Eugenia Dajer¹  

¹Departamento de Engenharia Elétrica; Universidade Tecnológica Federal do Paraná; Cornélio Procópio; Brasil.

²Departamento de Sistemas de Computação; Universidade de São Paulo; São Carlos; Brasil.



Correspondence

María Eugenia Dajer.
Email: mcdajer@utfpr.edu.br

How to cite

Morikawa, Mateus; Spatti, Danilo Hernane; Dajer, María Eugenia. (2022). Wavelet packet transform and multilayer perceptron to identify voices with a mild degree of vocal deviation. *Revista de Investigación e Innovación en Ciencias de la Salud*. 4(1), 16-25. <https://doi.org/10.46634/riics.126>

Received: 11/11/2021

Revised: 20/01/2022

Accepted: 01/02/2022

Editor

Jorge Mauricio Cuartas Arias, Ph.D. 

Coeditor

Fraidy-Alonso Alzate-Pamplona, MSc. 

Copyright © 2022. Fundación Universitaria María Cano. The *Revista de Investigación e Innovación en Ciencias de la Salud* provides open access to all its content under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International \(CC BY-NC-ND 4.0\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Conflicts of Interest

The authors have declared that no competing interests exist.

Abstract

Introduction. Laryngeal disorders are characterized by a change in the vibratory pattern of the vocal folds. This disorder may have an organic origin described by anatomical fold modification, or a functional origin caused by vocal abuse or misuse. The most common diagnostic methods are performed by invasive imaging features that cause patient discomfort. In addition, mild voice deviations do not stop the individual from using their voices, which makes it difficult to identify the problem and increases the possibility of complications.

Aim. For those reasons, the goal of the present paper was to develop a noninvasive alternative for the identification of voices with a mild degree of vocal deviation applying the Wavelet Packet Transform (WPT) and Multilayer Perceptron (MLP), an Artificial Neural Network (ANN).

Methods. A dataset of 74 audio files were used. Shannon energy and entropy measures were extracted using the Daubechies 2 and Symlet 2 families and then the processing step was performed with the MLP ANN.

Results. The Symlet 2 family was more efficient in its generalization, obtaining 99.75% and 99.56% accuracy by using Shannon energy and entropy measures, respectively. The Daubechies 2 family, however, obtained lower accuracy rates: 91.17% and 70.01%, respectively.

Conclusion. The combination of WPT and MLP presented high accuracy for the identification of voices with a mild degree of vocal deviation.

Keywords

Voice; voice disorder; voice classification; voice deviation; artificial neural network; multilayer perceptron; wavelet packet transform; dysphonia; laryngeal diseases; vocal cords.

Data Availability Statement

All relevant data is in the article and in the appendices. For more detailed information, write to the Corresponding Author.

Funding

None. This research did not receive any specific grants from funding agencies in the public, commercial, or non-profit sectors.

Disclaimer

The content of this article is the sole responsibility of the authors and does not represent an official opinion of their institutions or the Revista de Investigación e Innovación en Ciencias de la Salud.

Author Contributions

Mateus Morikawa: data curation, formal analysis, investigation, methodology, project administration, software, validation, visualization, writing – original draft, writing – review and editing.

Danilo Hernane Spatti: conceptualization, methodology, supervision, writing – review and editing.

María Eugenia Dajer: conceptualization, data curation, methodology, resources, supervision, writing – review and editing.

Resumen

Introducción. Los trastornos laríngeos se caracterizan por un cambio en el patrón vibratorio de los pliegues vocales. Este trastorno puede tener un origen orgánico, descrito como la modificación anatómica de los pliegues vocales, o de origen funcional, provocado por abuso o mal uso de la voz. Los métodos de diagnóstico más comunes se realizan mediante procedimientos invasivos que causan malestar al paciente. Además, los desvíos vocales de grado leve no impiden que el individuo utilice la voz, lo que dificulta la identificación del problema y aumenta la posibilidad de complicaciones futuras.

Objetivo. Por esas razones, el objetivo de esta investigación es desarrollar una herramienta alternativa, no invasiva para la identificación de voces con grado leve de desvío vocal aplicando Transformada Wavelet Packet (WPT) y la red neuronal artificial del tipo Perceptrón Multicapa (PMC).

Métodos. Fue utilizado un banco de datos con 78 voces. Fueron extraídas las medidas de energía y entropía de Shannon usando las familias Daubechies 2 y Symlet 2 para después aplicar la red neuronal PMC.

Resultados. La familia Symlet 2 fue más eficiente en su generalización, obteniendo un 99.75% y un 99.56% de precisión mediante el uso de medidas de energía y entropía de Shannon, respectivamente. La familia Daubechies 2, sin embargo, obtuvo menores índices de precisión: 91.17% y 70.01%, respectivamente.

Conclusión. La combinación de WPT y PMC presentó alta precisión para la identificación de voces con grado leve de desvío vocal.

Palabras clave

Voz; trastorno de la voz; clasificación de voz; desviación de voz; red neuronal artificial; perceptrón multicapas; transformada wavelet packet; afonía; enfermedades laríngeas; cuerdas vocales.

Introduction

The voice is one of the main tools of human communication. According to Imamura, Tsuji and Sennes [1], voice is basically produced by three processes: the movement of the vocal folds interrupting the subglottic airflow, followed by the resonance and articulation of this fundamental sound, which takes place in the supraglottic vocal tract. Any change in this complex mechanism may represent a shift in the vocal quality of a person. As the human voice is essentially an auditory-perceptual signal, any voice disorder is usually recognized as a deviation in vocal quality as stated by Behlau et al. [2].

For Patel and Shrivastav [3], as well as Eadie et al. [4], the auditory-perceptual evaluation is still considered the “gold standard” for traditional evaluation in voice clinics, and it enables the documentation of the severity of voice impairment. Since voice quality is multidimensional, auditory-perceptual evaluations have been performed with structured scales and protocols, suggested by Yamasaki et al. [5] to control the interference factors (the training, task design, type of stimulus, and the listener’s attention and experience). Also, for clinical and research purposes, auditory-perceptual parameters are usually rated using different perceptual scales, such as

the 4-point numerical scale (NS) and the 100 mm visual analog scale (VAS), as proposed by Webb et al. [6], Karnell et al. [7], and Kempster et al. [8].

According to Karnell et al. [7], VAS seems more sensitive to small differences in voice quality deviations than the NS. In Yamasaki et al. [5], boundaries between normal and disordered voices, for Brazilian participants, were found using the VAS. The authors concluded that the 35.5 value corresponds to the cutoff point between normal variation and mild/moderate vocal deviation; the 50.5 value, to the cutoff point between mild/moderate and moderate vocal deviation; and the 90.5, to the cutoff point between moderate and severe deviations. People with a mild voice deviation could not perceive a significant difference in their voice quality or could not identify the problem in the beginning. As a consequence, the individual would continue to use the voice carelessly, which may increase the possibility of complications.

Signal processing tools are widely applied in voice assessment and monitoring as they allow characterizing the state of the voice production system. Since biological signals, e.g. voice, are not stationary, the application of the Fourier Transform does not prove to be an accurate alternative to perform an acoustical analysis. However, the Wavelet Transform theory provides an alternative tool for short-time analysis of quasi-stationary signals such as voice, as emphasized by Tan et al. [9].

The Wavelet Packet Transform (WPT) has been used as an alternative tool, acting as an extractor of signal characteristics as seen in Lima et al. [10]. For Oliveira [11], this tool allows a time-frequency analysis and presents a wide range of applications that enables the unification of a vast number of processing and analysis techniques. The WPT is divided into families and each of them presents a different method for extraction. This is applied in decomposition levels, producing coefficients or nodes for a given dataset, selected at intervals from the data in time, called windowing conforming by Lima et al. [10]. And according to Jiao, Shi and Liu [12], the nodes in the last level of decomposition are called tree leaves or terminal nodes. Lima et al. [10], Ramirez-Villegas and Ramirez-Moreno [13], Zhang et al. [14], Barizão et al. [15] and Alves et al. [16] have used the WPT for the extraction of features from signals within classification processes.

In addition, another tool such as the Artificial Neural Networks (ANNs) can improve the performance of pattern classification in voice signals, as found in Silva, Spatti and Flauzino [17]. According to Haykin [18], ANNs are systems based on the human brain, described as a processing unit consisting of a massive parallel distributed processor, which stores knowledge and makes it available for use. Resembling the “human” brain in two aspects: knowledge, which is acquired by the network from its environment through a learning process, and the connection forces between neurons, called synaptic weights, which are used to store the knowledge that was acquired.

According to Silva, Spatti and Flauzino [17], ANNs are considered adaptive because their internal parameters, called synaptic weights, are adjusted from the presentation of examples related to a particular pattern, so they acquire knowledge (adapt) from experiences. By applying training sessions, the network is able to extract the correlation between information that makes up the application. After the training process, an ANN can generalize patterns and estimate possible solutions.

Multilayer perceptron (MLP) ANNs have, as their main feature, the presence of one or more hidden layers of neurons, and its structure is composed of an input layer, intermediate

or hidden layers, and an output neural layer. According to Silva, Spatti and Flauzino [17], it is considered a powerful and quite versatile tool and can be applied in the solution of problems related to a wide range of areas of knowledge, such as universal approximation of functions, pattern recognition processes, identification and control, prediction of time series, and optimization of systems. ANNs are widely applied in biomedical studies, as in Lima et al. [10], Souzanchi-K, Owhadi-Kareshk and Akbarzadeh-T [19], Baracho et al. [20], Bevilacqua et al. [21], Barizão et al. [15], and Silva et al. [22].

The purpose of this paper is to develop a non-invasive tool for the identification of voices with a mild degree deviation applying the WPT and MLP ANN.

Methods

Database

For this research, the software MATLAB® 2017b [23] (Student License) was used, because it contains the necessary features for the study.

The database was provided by Dr. Fabiana Zambon from SINPRO-SP and it was composed of 90 audio files recorded with the sound of the letter /e/ sustained for an average time of 10 seconds. All the volunteers were female professors between 23 and 66 years old and all of them were assessed and diagnosed either with the presence or absence of some symptoms, e.g. hoarseness, vocal fatigue, discomfort while talking, monotone voice, sore throat, effort while talking, among others. Only 74 audio files were used in this work because 16 samples were damaged. Thus, they were divided into the 3 following groups: 25 audios corresponding normal variation, 29 audios with mild vocal deviation, and 20 audios with moderate voice deviation, according to the cutoff values obtained from the auditory-perceptual analysis proposed by Yamasaki et al. [5]. Further details of the data collection and the classification can be found at Zambon [24].

Since the goal of this paper was to identify voices with a mild degree of deviation, we divided the dataset into two groups: G1 = voices with a mild degree of deviation and G2 = voices without deviation and voices with a moderate degree of deviation.

Procedures

The procedures were composed of the 5 following steps: a) preprocessing, b) segmentation, c) characteristic extraction, d) classification, and e) post-processing.

- a) The preprocessing step consisted of removing any silent parts of the audio files as well as any other sound that was not from the patient, which was considered as noise. It was also necessary to apply the MATLAB function `detrend` to prevent the DC-offset phenomenon from interfering with the recognition of silence. In this sense, in order to ensure the presence of vocal activity, an analysis of 25 milliseconds frames was performed, as suggested by Paliwal, Lyons and Wójcicki [25]. After that, the highest amplitude value of each frame was compared to the 0.03 empirical threshold. As a result, the frames where the highest amplitude was above the threshold were considered as periods with the presence of voice. Thus, by applying the `reshape` function, the signals were rebuilt removing the silence.
- b) In the segmentation step, the objective was to separate the data into a set of training (80%) and a set of testing (20%). For each voice signal, a window of 4096 discretized samples and 50% overlap was applied. Table 1 shows the number of samples for training and testing in group 1 (G1) and group 2 (G2), before and after segmentation.

Table 1. Number of samples for Groups (G1 and G2), pre and post segmentation

Register	Pre-segmentation		Post-segmentation	
	G1	G2	G1	G2
Training	23	36	4402	7723
Testing	6	9	1156	1843

- c) For the characteristic extraction step, the WPT transform was used as it obtains information from both the domain of time and frequency. Moreover, Daubechies 2 family (decomposition level 3) and Symlet 2 families (decomposition level 5) were used as they showed good performance in Lima [26], extracting the Shannon energy and entropy measures from the approximation and detail coefficients.
- d) The processing step was performed by the MLP network with the Levenberg-Marquardt learning algorithm, described by Silva, Spatti and Flauzino [17], using the hyperbolic tangent function in the intermediate layers and a learning rate of 0.2. The topology used is represented by two intermediate layers, which had 1 neuron in the first and 2 neurons in the second layer. Since the MLP uses a supervised learning process, it is necessary to indicate the target values of the answers. Thus, the output has defined the vector [1 -1] for the class Group 1. To the samples of Group 2, the vector [-1 1] was defined. If the result did not fit into either option, the designated vector was [2 2], indicating uncertainty.
- e) Finally, the post-processing step consisted of adjusting the output vectors produced by MLP. Therefore, it has been established a 98% degree of reliability. Thus, each of the two positions of the output vector was compared to the threshold of ± 0.98 . Hence, if the term value was higher than 0.98, this would receive value 1. If the term value was less than -0.98, this would receive -1. For values between -0.98 and 0.98, the term would receive 2. As suggested by Lever, Krzywinski and Altman [27], a confusion matrix was used to evaluate and explain the results.

Results

To prevent the randomization of the initialization of synaptic weights from interfering in the final answer, the network was trained and tested 10 times. Aiming to carry out a more detailed analysis of the classifier, the confusion matrices of each wavelet family were generated from the average of the 10 tests.

According to Tables 2, 3, 4, and 5, it is possible to observe that the proposed classification algorithm obtained an accuracy rate of 99.76% and 99.56% for the Shannon energy and entropy measures using the Symlet 2 family, and 91.17% and 70.01% for the same measures using the Daubechies 2 family.

Table 2. Confusion matrix with accuracy percentage using the Symlet 2 family and energy values

	G1	G2	Uncertainty
G1	99.75 %	0.15%	0.10%
G2	1.14%	97.57%	1.29%

Table 3. Confusion matrix with accuracy percentage using the Symlet 2 family and entropy values

	G1	G2	Uncertainty
G1	99.56 %	0.31%	0.13%
G2	2.19%	96.29%	1.52%

Table 4. Confusion matrix with accuracy percentage using the Daubechies 2 family and energy values

	G1	G2	Uncertainty
G1	91.17 %	3.68%	5.15%
G2	0.50%	98.29%	1.21%

Table 5. Confusion matrix with accuracy percentage using the Daubechies 2 family and entropy values

	G1	G2	Uncertainty
G1	70.01 %	1.97%	28.02%
G2	0.34%	86.75%	12.91%

Discussion

The voice is an important tool for some professionals who use it as a main work instrument. However, when misused, serious vocal disorders may emerge and it becomes a huge problem since the mild deviation does not stop the individual from doing their job. In other words, the initial stage of voice disorders manifests imperceptibly, making it difficult to diagnose, as suggested by Medeiros et al. [28]. Furthermore, as highlighted by Giannini and Ferreira [29] and Cantor-Cutiva et al. [30], professionals from the educational sector are more likely to have voice issues when compared to other occupations, mostly due to the environmental conditions they are in. To make matters worse, the authors also report related disorders that may appear, such as mental and physical disorders, thus emphasizing the importance of the findings herein presented and the relevance in the study of an automated system to aid in the diagnosis.

In agreement with Silva, Spatti and Flauzino [10], as well as Haykin [11], for ANNs or any other artificial intelligence (AI) algorithms, it is crucial to have as much data as possible so that the model will be able to better generalize the issue at hand. In this sense, the size of our dataset presented a challenge for the researchers since it was composed of only 74 audios

and just 29 of which corresponded to the class of interest. In order to solve the problem, we applied the segmentation step, as Lima [26] suggested. In this paper, it was used a window of 4096 discretized samples and 50% overlap for each voice signal.

Tan et al. [9], Lima et al. [10], Barizão et al. [15], and Alves et al. [16] report that WPT is a valuable tool to extract features from non-stationary signals, being a powerful approach when using along with some Artificial Intelligence (AI) algorithm to find patterns. In this sense, the MLP model created in this work was fundamental to showcase new perspectives regarding the usage of the Daubechies 2 family, especially with the Shannon Entropy measure. The training configuration regarding the number of neurons, hidden layers, learning rate, and activation function was kept the same for all those 4 input sets. This may explain why MLP performance decreases when using the Daubechies 2 and Shannon entropy measure, given that for each input set there is a better MLP setting to use.

The results presented in the confusion matrices (Tables 2-5) suggest that Symlet 2 outperformed Daubechies 2, as can be seen from the measures of uncertainties, errors, and successes in identifying the desired class. Although Lima et al. [10] have shown that the Daubechies 2 families and Symlet 2 were efficient for the analysis of vocal signals; for this study the performance of Daubechies 2 using Shannon entropy showed low accuracy percentages. This may be a fair finding once the scope of the above-mentioned work aimed to fit an MLP model capable of categorizing the dataset into the types of dysphonia, not in terms of its severity. Moreover, as Lima [26] indicates, the topology parameters of the MLP are configured empirically in order to achieve its maximum performance, raising the hypothesis that there is a topology that best meets the Daubechies 2 family for this work.

In Table 4, the result of the Daubechies 2 family is less accurate than the Symlet 2 family, and there is an increase in the uncertainty rate. Since this study is about an application of ANN to help identify mild vocal deviations, it becomes more acceptable for the ANN to be uncertain rather than of performing an incorrect classification.

Additionally, it was observed that only 3 neurons in the intermediate layers were enough to perform a good generalization, thus not requiring a great computational performance. It is worth pointing out that besides the fewer numbers of neurons in the hidden layers, it is crucial to consider the learning rate used, and the optimization algorithm Levenberg-Marquardt, which speeds up the learning process.

Limitations

This research has some limitations. When talking about artificial intelligence, there are plenty of algorithms that can be explored to verify their performance at analyzing the voice signal. The MLP was chosen to be used in this work so the findings herein presented may support or contrast other results from the same research group. In addition, future work will explore the use of other wavelet families and the use of larger databases, as well as other types of voice conditions.

Conclusion

This research aimed to train a neural network specialist to recognize voices with a mild degree of deviation. Therefore, the work was grounded in a process chronology that starts from the data treatment and goes all the way until the classifier model. Following this method, it was possible to get outcomes that showed the effectiveness and supported the use of those two WPT families in the vocal signal analysis.

It is concluded that the MLP proved to be robust enough to generate a high rate of correctness in its classification, which, in most cases, surpassed 99% accuracy with 98% reliability.

It was also observed that only 3 neurons in the intermediate layers were enough to perform a good generalization, thus not requiring a great computational performance.

The contribution of this work is the development of a noninvasive computational tool to automatically identify voices with a mild degree of deviation. This tool would be used in clinical settings to assist professionals during screenings, diagnostic process, and for training young professionals to perform auditory-perceptual evaluations.

References

1. Imamura R, Tsuji DH, Sennes LU. Fisiologia da laringe. In Pinho S, Tsuji DH, Bohadana S, editors. *Fundamentos de Laringologia e Voz*. 1st ed. Rio de Janeiro: Revinter Ltda; 2006.
2. Behlau M, Rocha B, Englert M, Madazio G. Validation of the Brazilian Portuguese CAPE-V Instrument—Br CAPE-V for Auditory-Perceptual Analysis. *J Voice*. 2020. doi: <https://doi.org/10.1016/j.jvoice.2020.07.007>
3. Patel S, Shrivastav R. Perception of dysphonic vocal quality: some thoughts and research update. *Perspect Voice Dis*. 2007;17:3–6. doi: <https://doi.org/10.1044/vvd17.2.3>
4. Eadie T, Sroka A, Wright DR, Merati A. Does knowledge of medical diagnosis bias auditory-perceptual judgments of dysphonia? *J Voice*. 2011;25:420–429. doi: <https://doi.org/10.1016/j.jvoice.2009.12.009>
5. Yamasaki R, Madazio G, Leão SHS, Padovani M, Azevedo R, Behlau M. Auditory-perceptual Evaluation of Normal and Dysphonic Voices Using the Voice Deviation Scale. *J Voice*. 2016;31:67–71. doi: <https://doi.org/10.1016/j.jvoice.2016.01.004>
6. Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen N, Wilson JA. The reliability of three perceptual evaluation scales for dysphonia. *Eur Arch Otorhinolaryngol*. 2004;261:429–434. doi: <https://doi.org/10.1007/s00405-003-0707-7>
7. Karnell MP, Melton SD, Childes JM, Coleman T, Dailey S, Hoffman H. Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. *J Voice*. 2007;21:576–590. doi: <https://doi.org/10.1016/j.jvoice.2006.05.001>
8. Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Karemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol*. 2009;18:124–132. doi: [https://doi.org/10.1044/1058-0360\(2008/08-0017\)](https://doi.org/10.1044/1058-0360(2008/08-0017))

9. Tan BT, Fu M, Spray A, Dermody P. The use of wavelet transforms in phoneme recognition. Proceeding of Fourth International Conference on Spoken Language Processing, ICSLP '96; 1996 Oct 3 – Oct 6; Philadelphia, USA. IEEE; 2002. p. 2431-2434. doi: <https://doi.org/10.1109/ICSLP.1996.607300>
10. Lima AAM, Barros FKH, Yoshizumi VH, Spatti DH, Dajer ME. Optimized Artificial Neural Network for Biosignals Classification Using Genetic Algorithm. *J Control Autom Electr.* 2019;30:371-379. doi: <https://doi.org/10.1007/s40313-019-00454-1>
11. Oliveira HM. Análise de Fourier e Wavelets: Sinais Estacionários e não Estacionários. Recife: Editora Universitária, UFPE; 2007.
12. Jiao S, Shi W, Liu Q. Self-adaptative partial discharge denoising based on variation mode decomposition and wavelet packet transform. Chinese automation congress; 2017 Oct 20 – Oct 22; Jinan, China. IEEE; 2018 Jan. p. 6. doi: <https://doi.org/10.3390/en12173242>.
13. Ramirez-Villegas JF, Ramirez-Moreno DF. Wavelet packet Energy, Tsallis entropy and statistical parameterization for support vector-based and neural-based classification of mammographic regions. *J Neurocomputing.* 2012;77(1):82-100. doi: <https://doi.org/10.1016/j.neucom.2011.08.015>.
14. Zhang Y, Dong Z, Wang S, Ji G, Yang J. Preclinical diagnosis of magnetic resonance (MR) brain images via discrete wavelet packet transform with Tsallis entropy and generalized eigenvalue proximal support vector machine (GEPSVM). *J Entropy.* 2015;17(4):1795-1813. doi: <https://doi.org/10.3390/e17041795>
15. Barizão H, Fermino MA, Dajer ME, Liboni LHB, Spatti DH. Voice disorder classification using MLP and wavelet packet transform. 2018 International Joint Conference on Neural Networks (IJCNN); 2018 Jul 8 – Jul 13; Rio de Janeiro, Brazil; IEEE; 2018. p. 8. doi: <https://doi.org/10.1109/IJCNN.2018.8489121>
16. Alves M, Silva G, Bispo BC, Dajer ME, Rodrigues PM. Voice Disorders Detection Through Multiband Cepstral Features of Sustained Vowel. *J Voice.* 2021;35(5):1-10. doi: <https://doi.org/10.1016/j.jvoice.2021.01.018>
17. Silva IND, Spatti DH, Flauzino RA. Redes Neurais Artificiais para engenharia e ciências aplicadas. São Paulo: Artliber; 2010.
18. Haykin S. Redes Neurais: Princípios e Prática. 2nd ed. Hamilton: Bookman; 2001.
19. Souzanchi-K M, Owhadi-Kareshk M, Akbarzadeh-T MR. Control of elastic joint robot based on electromyogram signal by pre-trained Multi-Layer Perceptron. 2016 International Joint Conference on Neural Networks (IJCNN); 2016 Jul 24 – Jul 29; Vancouver, Canada; IEEE; 2016. doi: <https://doi.org/10.1109/IJCNN.2016.7727891>
20. Baracho SF, Pinheiro DJLL, de Melo VV, Coelho RC. A hybrid neural system for the automatic segmentation of the interventricular septum in echocardiographic images. 2016 International Joint Conference on Neural Networks (IJCNN); 2016 Jul 24 – Jul 29; Vancouver, Canada; IEEE; 2016. doi: <https://doi.org/10.1109/IJCNN.2016.7727868>

21. Bevilacqua V, Salatino AA, Di Leo C, Tatolli G, Buongiorno D, Signorile D, et al. Advanced classification of Alzheimer's disease and healthy subjects based on EEG markers. 2015 International Joint Conference on Neural Networks (IJCNN); 2015 Jul 12 – Jul 17; Killarney, Ireland; IEEE; 2015. doi: <https://doi.org/10.1109/IJCNN.2015.7280463>
22. Silva EHD, Morikawa M, Suterio VB, et al. Aplicação De Rede Neural Artificial Especialista Em Reconhecimento De Transtornos Vocais Moderados. In: Dallamuta J, Ajuz Holzman H, organizers. Engenharia Elétrica: Comunicação Integrada no Universo da Energia. 1st ed. Ponta Grossa: Atena Editora; 2021. doi: <https://doi.org/10.22533/at.ed.3732123021>
23. MATLAB. version 9.3 (R2017b). Natick, Massachusetts: The MathWorks Inc.; 2017.
24. Zambon FC. Estratégias de enfrentamento em professores com queixa de voz. [thesis]. São Paulo: Universidade Federal de São Paulo; 2011.
25. Paliwal KK, Lyons JG, Wójcicki KK. Preference for 20 40 ms window duration in speech analysis. 2010 4th International Conference on Signal Processing and Communication Systems; 2010 Dec 13 – Dec 15; Gold Coast, Austrália; IEEE; 2011. doi: <https://doi.org/10.1109/ICSPCS.2010.5709770>
26. Lima AAM. Classificação de Disfonias Utilizando Redes Neurais Artificiais e Transformadas Wavelet Packet. [Bachelor's thesis]. Cornélio Procópio: Universidade Tecnológica Federal do Paraná; 2018.
27. Lever J, Krzywinski M, Altman N. Classification evaluation. *Nat Methods*. 2016;13:603–604. doi: <https://doi.org/10.1038/nmeth.3945>.
28. Medeiros JdaSA, Santos SMM, Teixeira LC, Cortes Gama AC, de Medeiros AM. Sintomas vocais relatados por professoras com disfonia e fatores associados. *J Audiol Commun Res*. 2016;21:1-8. doi: <https://doi.org/10.1590/2317-6431-2015-1553>
29. Giannini SSP, Ferreira LP. Voice disorders in teachers and the International Classification of Functioning, Disability and Health (ICF). *Rev. Investig. Innov. Cienc. Salud* [Internet]. 2021 Aug. 3 [cited 2022 Feb. 5];3(1):33-47. doi: <https://doi.org/10.46634/riics.60>
30. Cantor-Cutiva LC, Cuervo-Diaz DE, Hunter EJ, Moreno-Angarita M. Impairment, disability, and handicap associated with hearing problems and voice disorders among Colombian teachers. *Rev. Investig. Innov. Cienc. Salud* [Internet]. 2021 Aug. 3 [cited 2022 Feb. 5];3(1):4-21. doi: <https://doi.org/10.46634/riics.48>