WAVELET PACKET TRANSFORM AND MULTILAYER PERCEPTRON TO IDENTIFY VOICES WITH A MILD DEGREE DEVIATION

Mateus Morikawa[1], Danilo Hernane Spatti[2], Maria Eugenia Dajer[1]

[1] Department of Electrical Engineering, Federal University of Technology – Paraná (UTFPR-CP), Brazil.

[2] Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, SP, Brazil.

***Keywords*: Voice; Artificial Neural Networks; Wavelet Packet Transform; Multilayer Perceptron.**

INTRODUCTION

The voice is one of the main tools of human communication. According to [1] voice is basically produced by three processes: 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 can mean a change in the vocal quality of a subject. As human voice is essentially an auditory-perceptual signal, any voice disorder is usually recognized as a deviation in vocal quality [2]

Since biological signals, as 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 Packet Transform (TWP) has been used as an alternative tool, acting as an extractor of signal characteristics [3]. In addition, another tool, such as Artificial Neural Networks (ANNs) can improve the performance of pattern classification in voice signals.

The purpose of this paper is to study an alternative way to identify voices with a mild degree deviation, using the Wavelet Packet Transform and Artificial Neural Networks.

METHODS

Database

The database consisted of 74 audio files classified into 3 groups: 25 voices without vocal deviation, 29 voices with mild vocal deviation, and 20 voices with moderate vocal deviation, according to the perceptual-auditory indices observed [4]. The database was provided by Dr. Fabiana Zambóm and further details of the data collection and the classification can be found at [5].

Since the goal of this paper is to identify voices with a mild degree of deviation, we divided the data set 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

For this work, we use MATLAB (student license), and the procedures were composed of the following steps: pre-processing, segmentation, characteristic extraction, classification, and post-processing.

The pre-processing step consisted of removing periods of silence from audio files, as well as any types of sound that are not of the patient, called artifacts.

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 groups 1 and 2, before and after segmentation.

Table 10. Number of samples for Groups (G1 and G2), pre and post segmentation.

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

For the extraction of characteristics step, the Wavelet Packet (TWP) transform was used, since this one obtains information in both, the domain of time and frequency. We used the Daubechies 2 and Symlet 2 families for extracting the energy and Shannon's entropy values from the coefficients of approximation and detail.

The processing step was performed with the Multilayer Perceptron (MLP) network with the Levenberg-Marquardt learning algorithm [6], 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 have 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 desired 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.

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. Therefore, 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.

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 assembled with 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 of 99.76% and 99.56% for 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 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 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 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 Daubechies 2 family and entropy values.

	G1	G2	Uncertainty
G1	70,01 %	1,97%	28,02%
G2	0,34%	86,75%	12,91%

The results presented in the confusion matrices 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 previous work has 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 entropy did not have a good result.

CONCLUSION

This research aimed to train a neural network specialist in recognizing mild voice disorders. 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.

Also, it is observed that only 3 neurons in the intermediate layers have already been enough to perform a good generalization, not requiring thus, a great computational performance.

Future work will explore the use of other wavelet families and the use of larger databases.

ACKNOWLEDGMENTS

Special thanks to Dr. Fabiana Zambón for sharing the voice database.

REFERENCES

[1] Imamura, R. Tsuji, DH; Sennes, LU. Fisiologia da laringe. IN: Pinho, SMR; Tsuji, DH, Bahadana, SC. Fundamentos de Laringologia e Voz. 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, Journal of Voice, 2020.

[3] Lima, AAM.; De Barros, FKH.; Yoshizumi, VH; Spatti, DH; Dajer, ME. Optimized Artificial Neural Network for Biosignals Classification Using Genetic Algorithm. *Journal Of Control, Automation and Electrical Systems*, V.30, pages371–379, 2019.

[4] Yamasaki, R. et al. Auditory-perceptual Evaluation of Normal and Dysphonic Voices Using the Voice Deviation Scale. Journal of Voice, v. 31, n. 1, p. 67-71, 2016.

[5] Zambon, F.C. Estratégias de enfrentamento em professores com queixa de voz. São Paulo, 2011.

[6] Silva, IND; Spatti, DH.; Flauzino, RA. Redes Neurais Artificiais para engenharia e ciências aplicadas. São Paulo, SP: Artliber, 2010.