ISSN 2395-1621

Machine Learning Technique for Voice Disorders Identification Algorithm

[#1]Megha Nagawade, [#2]Prof. Aruna Varma

[1]nagawademegha@gmail.com

[#12]Department of Computer Engineering
Dhole Patil Collage of Engineering, Wagholi, Pune.

ABSTRACT

In this paper we present a new method for voice disorders classification based on machine learning neural network. The processing algorithm is based on a hybrid technique which uses the wavelets energy coefficients as input of the multilayer neural network. The training step uses a speech database of several pathological and normal voices collected from the national hospital "Rabta - Tunis" and was conducted in a supervised mode for discrimination of normal and pathology voices and in a second step classification between neural and vocal pathologies (Parkinson, Alzheimer, laryngeal, dyslexia…). Several simulation results will be presented in function of the disease and will be compared with the clinical diagnosis in order to have an objective evaluation of the developed tool.

Keywords: Speech processing, pathological voices, classification, wavelet transform, neural networks, energy.

ARTICLE INFO

Article History

Received: 18th May 2018

Received in revised form : 18th May 2018

Accepted: 23rd May 2018

Published online : 23rd May 2018

I. INTRODUCTION

Pathological voice recognition has been received a great attention from researchers in the last decade. Speech processing has proved to be an excellent tool for voice disorder detection. Among the most interesting recent works are those concerned with Parkinson's Disease (PD), Multiple Sclerosis (MS) and other diseases which belong to a class of neuro- degenerative diseases that affect patients speech, motor, and cognitive capabilities [3, 11]. The speech production is a complex motor act that implies a big number of muscles, of physiological variables and a neurological control implying different cortical and under cortical regions. We distinguish three systems contributing to the production of the speech: the respiratory system, the laryngeal system and the supra-laryngeal system (the articulators) [14, 15]. The nervous system also controls the prosody. This one schematically covers the variations of height (intonation, melody), the variations of intensity (accentuation) and the temporal progress (pauses, debit, and rhythm).

The analysis of the voice disorder stays essentially clinic [12]. The instrumental measures are spilled little in practice clinic. The most used are the acoustic and aerodynamic measures [13]. The speech analysis is complex and has been disregarded for a long time. A difficulty result is to analyze in the literature the different treatment effect (medical or surgical). Indeed, a many studies don't return specific analysis of the speech. On the other hand, we attend confusion between the modifications of the motivity orofacial and the speech quality that remains the clinic objective [1].

Features assessment of a voice disorder is that the disorder carries on a patient's capacity to communicate are a crucial step to conceive a program of its management. A process of the prosperous assessment allows the pathologist of the speech to diagnose the voice disorder, determine the relative efficiency of several treatment approaches and formulate a prognosis [5]. Physicians often use invasive techniques like endoscopy to diagnose the symptoms of vocal fold disorders. However, it is possible to identify disorders using certain features of speech signals [12, 13].

Different classic techniques are used to extract the vocal parameters and so to make the classification of the pathological voices such as pitch and formant detection: PDA (cepstrum, FFT, spectrogram).

II. DISORDER IDENTIFICATION BY PITCH AND FORMANTS ANALYSIS

2.1. Speech Processing Algorithm

Figure 1 illustrates the algorithm steps for speech processing.

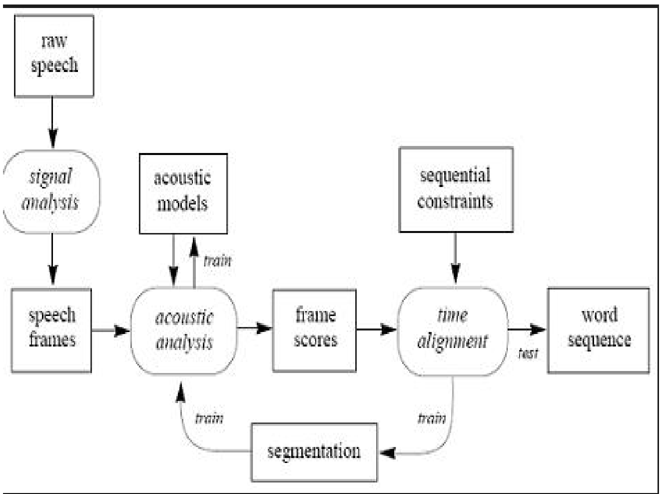
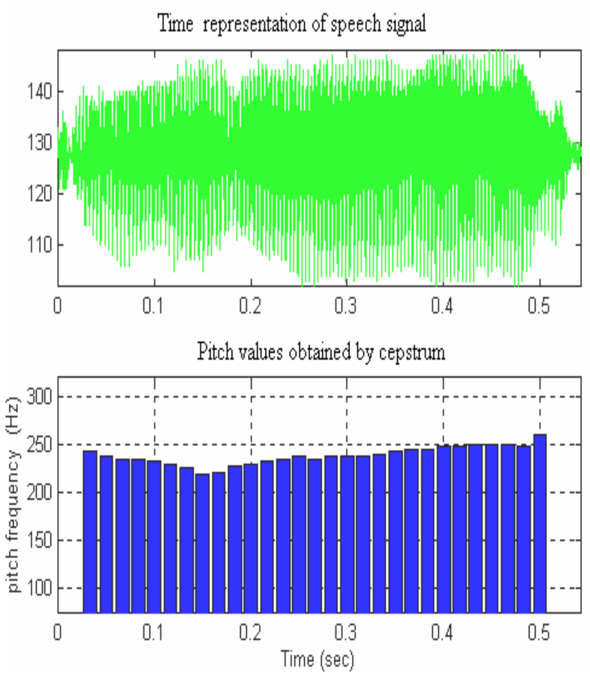


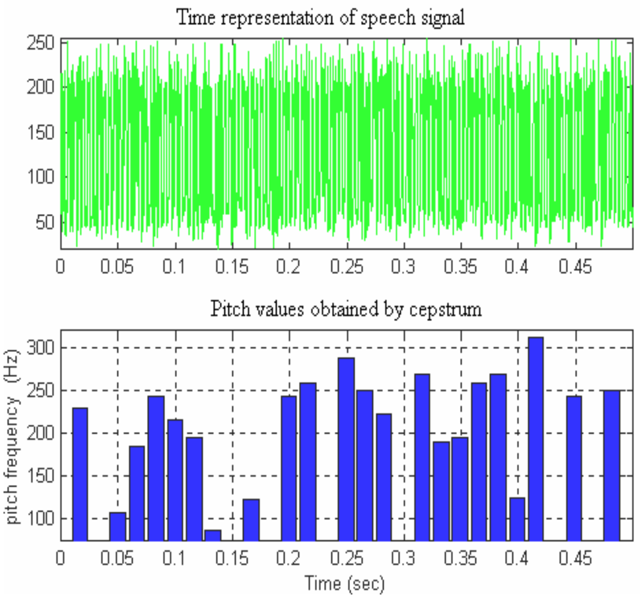
Figure 1. Speech processing algorithm

2.2. Pitch and Formants Analysis Results

Figure 2 illustrates the pitch variation by application of the cepstrum method analysis of normal and pathological female sounds (32 years). The high distortion and the variation of the pitch around the expected value (250 Hz) demonstrate a state of the glottic signal anomaly, resulting of a laryngeal pathology [2, 7, 8].



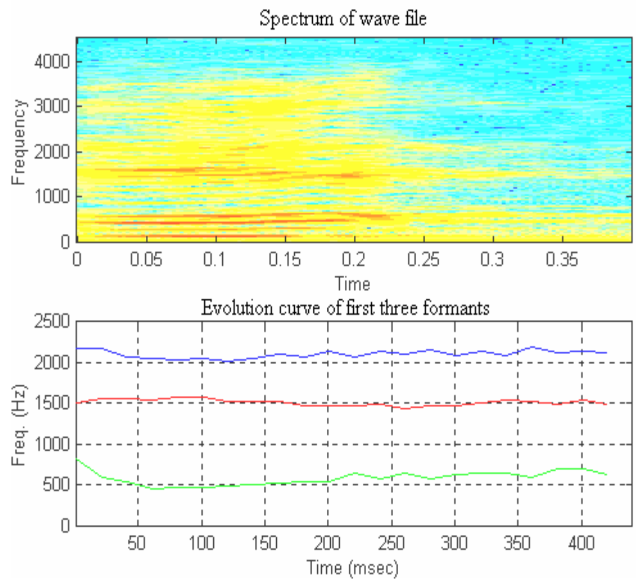
a. Normal female voice.



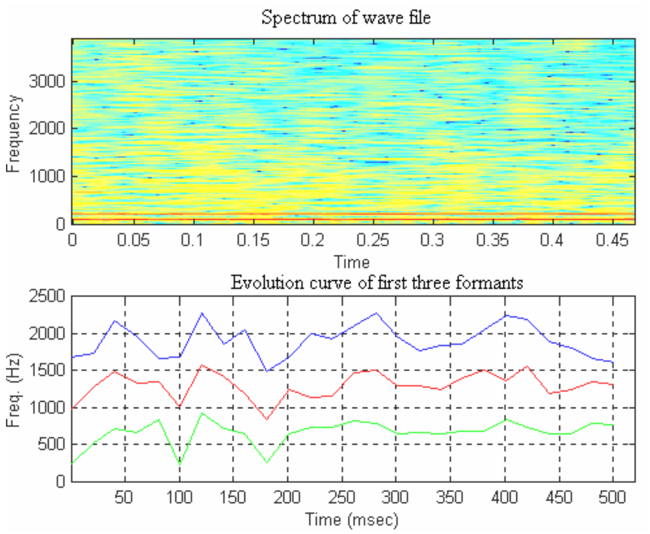
b. Pathological female voice.

Figure 2. Pitch evolution.

Linear Predictive Coding (LPC) method applied on the same voices allows evolution to extract the formants. By comparison to the normal values as shown in Figure 3, the high variations of the formants F_1, F_2 and F_3 of the pathological male sound confirm the conclusions, [2, 7, 8]. Although these methods can help us to distinguish a pathological voice but they remain subjective methods that don't give any quantification values to take the decision. For it, we try to improve this idea by new methods that use wavelet transformed and neural networks.



a. Normal male voice.



b. . Pathological male voice.

Figure 3. Formants evolution

III. NEURAL NETWORKS

Neural networks were chosen as a method of pattern matching for many main reasons. First the Matlab software has a fantastic implementation of several different types of neural networks in its neural network toolbox. The big advantage of the neural networks resides in their automatic training capacity, what permits to solve some problems without requiring to the complex rule writing, while being tolerant to the errors [6].

Neural networks consist of several simple parallel computational units called neurons. These units form a neural network that resembles a biological nervous system. The functioning of a neural network is greatly determined by the ways in which its units connect to other units.

3.1. Proposed System

The Matlab7.0 platform is used for implementation of the neural network formed of three layers, one of input, one of output and a hidden layer as shown in Figure 4. The input layer is formed of the same neurons number that corresponds to the components of the input vector. The input is the feature vector obtained from wavelet decomposition. The hidden layer contains fifteen neurons and the output layer contains only one neuron to give the decision pathological or normal as shown in Figure 4.

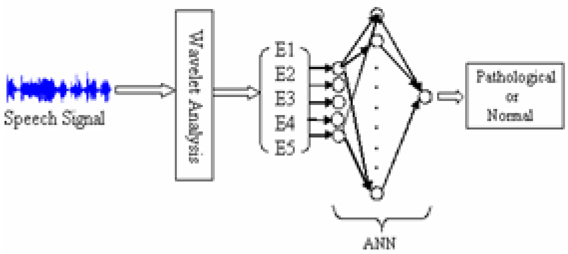


Figure 4. Voices classification model.

3.2. Machine Learning Design

The classic cycle of a neural network development can be separate in seven stages:

- The collection of data.
- The analysis of the data.
- The separation of the data bases.
- The choice of a neural network.
- The formatting of the data.
- The training.
- The testing.

In our survey we use a MNN with only one layer hidden between the input layer and the output layer. Every neuron of the hidden layer is connected to the neurons of the input layer and those of the output layer and there is not a connection between the cells of a same layer. The activation functions used in this type of network are the doorstep or sigmoid functions.

This network follows a supervised training according to the rule of errors correction. The training type used for this network is the supervised fashion. To every well stocked input an answer corresponds waited at the output. So the network is going to alter until it finds the good output.

The speech data base used for the training and for the system validation is formed by normal and pathological voices containing about ten words of every type (words for normal and pathological voices). The data base implies several pathological voices as laryngeal, neurological (Parkinson, Alzheimer, dyslexia, dysphonic…) to come from different mixed speakers (men and women). Thus, the results will be compared to those gotten by the classic methods as the cepstrum, LPC and the spectrogram methods for the extraction of the pitch and formants.

4.4 Feature Extraction

The speech is a highly random signal, and then the classic parameter instability as the pitch, jitter and formants can be common for the two types of voice (pathological and normal). So that, the classification will be efficient and effective one chose to use the normalized energies correspondents to the wavelet transformed coefficients as part of input feature vector for the neural network [10]. A Filter Bank is used to extract the wavelet coefficients.

The energy of every level is normalized against total energy content in the signal.

The lower band scale presents a more dominant periodicity than the in the higher band scale. This periodicity is decreasing in the pathological speech but it is very consistent in the normal speech [10].

IV. RESULTS

We use 60 words on the total, pronounced by different speakers of which 30 are normal and the other present pathologies of vocal or neurological origin. For the training, we use 40 words (20 normal and 20 pathological). After training, the network will be tested with 20 words different from those used for the training (10 normal and 10 pathological).

In order to obtain optimal result, we vary at every time the number of the energy coefficients to the input of the neural network. This procedure requires a variation of the choice of the wavelet filter bank.

V. CONCLUSIONS

The goal of this work is to conceive a tool of help to the clinicians in the Tunisian hospitals. This tool allow to follow up of patients who suffer from illness of vocal and neurological origin.

We presented in this paper a material and software interface of numeric treatment of the patient's vocal signal based on neural networks. Result of the MNN classifier gives the correct classification. The classification rate is between 80% and 100%. We have demonstrated in this study, a feature vector based on wavelet coefficients is useful for classification of normal and pathological speech data. At a preliminary level, the speech data is classified into two classes normal or pathological.

VI. ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable suggestions and comments. The authors also express their thanks to Prof. Aruna Varma for his critical reading, comments, and help.

REFERENCES

[1] Boyanov B. and Hadjitodorov S., "Acoustic Analysis of Pathological Voices: A Voice Analysis System for Screening of Laryngeal Diseases," in Proceedings of IEEE Engineering in Medical and Biology, vol. 16, no. 4, pp. 74- 82. 1997.

[2] Cherif A., "Detection and Formant Extraction of Arabic Speech Processing," Computer Journal of Applied Acoustics, vol. 6, no. 3, pp. 55-58, 2001.

[3] Davis B., "Acoustic Characteristics of Normal and Pathological Voices," in Proceedings of Speech and Language: Advances in Basic Research and Practice, Orland, pp. 271-335, 1979.

[4] Gaouda A., Salama M., Chikhani A., and Sultan M., "Application of Wavelet Analysis for Monitoring Dynamic Performance in Industrial Plants," in Proceedings North American Power Symposium, Laramie, pp. 154-159, 1997.

[5] Jiang J. and Zhang Y., "Nonlinear Dynamic Analysis of Speech from Pathological Subjects," in Proceedings of IEEE Electronics Letters, vol. 38, no. 6, pp. 142-146, 2002.

[6] Kortelainen J. and Noponen K., "Neural Networks," Technical Document, University de Sherbrooke, 2005.

[7] Lotfi S. and Adnene C., "A Speech Processing Interface for Analysis of Pathological Voice," in Proceedings of Communication Technologies from Theory to Applications Conference, Damascus, pp. 250-254, 2006.

[8] Lotfi S., Haythem B., and Adnene C., "Interface d'Analyse Vocale a l'Identification de Certaines Pathologies d'Origine Neurologique et Vocale," in Proceedings of JTM Conference, Tunis, pp. 66-69, 2007.

[9] Mallat S., "A Theory for Multiresolution Signal Decomposition: Wavelet Representation," Computer Journal of IEEE Transaction Pattern Analysis and Machine Intelligence, vol. 11, no. 7, pp. 674-693, 1989.

[10] Nayak J. and Bhat S., "Classification and Analysis of Speech Abnormalities," Computer Journal of ITBM-RBM, vol. 26, no. 5, pp. 319- 327, 2005.

[11] Parsa V. and Jamieson G., "Interactions between Speech Coders and Disordered Speech," Computer Journal of Speech Communication, vol. 40, no. 7, pp. 365-385, 2003.

[12] Plant F., Kessler H., Cheetham B., and Earis J., "Speech Monitoring of Infective Laryngitis," in Proceedings of International Conference on Spoken Language Processing, Philadelphia, pp. 749-752, 1996.

[13] Viera N., McInnes R., and Jack A., "Robust F0 and Jitter Estimation in the Pathological Voices," in Proceedings of International Conference on Spoken Language Processing, Philadelphia, pp. 745-748, 1996.

[14] Wang J. and Cheolwoo J., "Performance of Gaussian Mixture Model as a Classifier for Pathological Voice," in Proceedings of the ASST in Auckland, Australian, pp. 165-169, 2006.

[15] Yu P., Ouaknine M., Revis J., and Giovanni A., "Objective Voice Analysis for Dysphonic Patients: A Multiparametric Protocol Including Acoustic and Aerodynamic Measurements," Computer Journal of Voice, vol. 15, no. 4, pp. 529-542, 2001.