International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:9, 2008

# Voice Disorders Identification Using Hybrid Approach: Wavelet Analysis and Multilayer Neural Networks

L. Salhi, M. Talbi, and A. Cherif

***Abstract*—**This paper presents a new strategy of identification and classification of pathological voices using the hybrid method based on wavelet transform and neural networks. After speech acquisition from a patient, the speech signal is analysed in order to extract the acoustic parameters such as the pitch, the formants, Jitter, and shimmer. Obtained results will be compared to those normal and standard values thanks to a programmable database. Sounds are collected from normal people and patients, and then classified into two different categories. Speech data base is consists of several pathological and normal voices collected from the national hospital "Rabta-Tunis". Speech processing algorithm is conducted in a supervised mode for discrimination of normal and pathology voices and then for 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*—**Formants, Neural Networks, Pathological Voices, Pitch, Wavelet Transform.

## I. INTRODUCTION

THERE are several diseases that adversely affect our human voice which can be organic or neurological. The clinician can use the available apparatus for detection of the kind of the pathology. The diagnosis of pathological voice is an important topic that has been received considerable attention.

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 patient's speech, motor, and cognitive capabilities [1], [2]. We distinguishes three systems contributing to the production of the speech: the respiratory system, the laryngeal system and the supra-laryngeal system (the articulators) [9], [10].

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 [3]. The instrumental measures are spilled little in clinic practice. The most used are the acoustic and aerodynamic measures [4]. The speech analysis is complex and has been disregarded for a long time. The difficult result is to analyze in the literature the different treatment effects (medical or surgical). Indeed, 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 [7].

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 [8]. 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 [3], [4].

In a previous study, we presented three classical techniques to extract the vocal parameters and so to make the classification of pathological voices such as the cepstrum, LPC, spectrogram. In the recent approaches to pathological voice classification, various pattern classification methods have been used. Several researches, such as: application of automatic speaker recognition techniques to pathological voice assessment, [16] identification of voice disorders using speech samples, [5] performance of Gaussian mixture models as a classifier for pathological voice, [10] have recently been applied to various kinds of pathological classification tasks.

Artificial neural network (ANN) can be used in this case because there is no need to compute the details of the data mathematical models and relatively easy to train and has produced a good pathological recognition performance.

In this paper the ANN method was used to classify the mixed voiced data set (pathological and normal voices) into normal and pathological voices.

The multilayer neural network (MNN) with back propagation algorithm was implemented. Feature extraction like pitch, formants and wavelet coefficients with their corresponding energies and their corresponding entropies will be used as input vector to the MNN. The neural network

L. Salhi is with the Signal Processing Laboratory, Physics Departement, Sciences Faculty of Tunis, 1060 Tunis ElManar Tunisia (phone: +216.98.997.885; e-mail: lotfi.salhi@laposte.net).

M. Talbi is with the Signal Processing Laboratory, Physics Departement, Sciences Faculty of Tunis, 1060 Tunis ElManar Tunisia (e-mail: mouradtalbi1969@yahoo.com).

A. Cherif is Professor in Electrical Engineering, Sciences Faculty of Tunis, 1060 Tunis ElManar Tunisia (e-mail: Adnane.cher@fst.rnu.tn).

results using will be illustrated and compared to those of previous study.

## II. Disorder Identification by Pitch and Formants Analysis

### A. Speech Processing Algorithm

The processing speech algorithm can be illustrated by Fig. 1

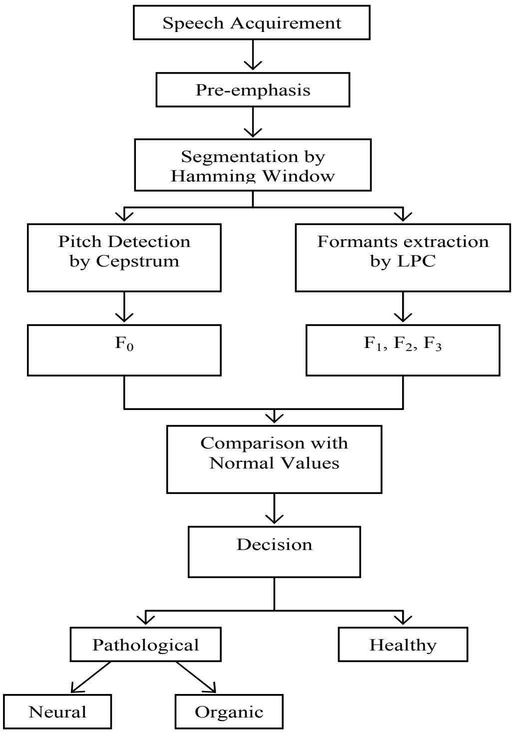


Fig. 1 Speech processing algorithm

### B. Pitch and Formants Analysis Results

Fig. 2 illustrates the pitch variation by application of the cepstrum method analysis of a normal and a pathological female sound (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.

The linear predictive coding (LPC) method applied on the same voices presents the evolution of the three formants $F_1$, $F_2$, and $F_3$. By comparison with the normal values (Fig. 3), the high variations of the formants of the pathological male sound confirm the last conclusions [12], [13].

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. That’s why we will present another method based in an automaic artificial neural network (ANN).

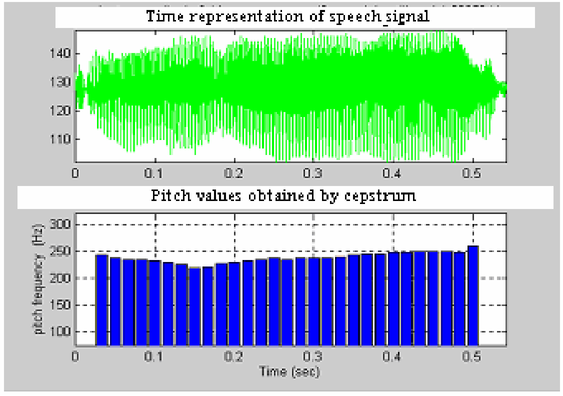


Fig. 2 (a) Pitch evolution of a normal female voice

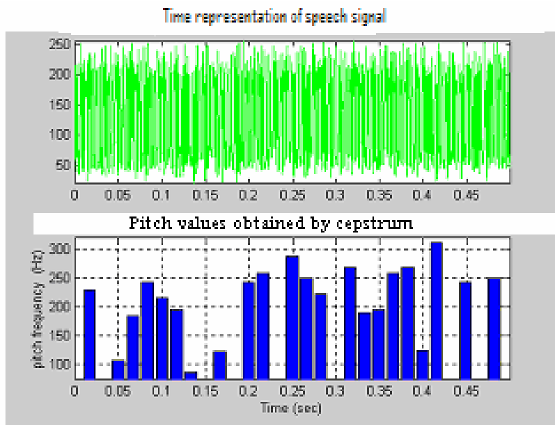


Fig. 2 (b) Pitch evolution of a pathological female voice

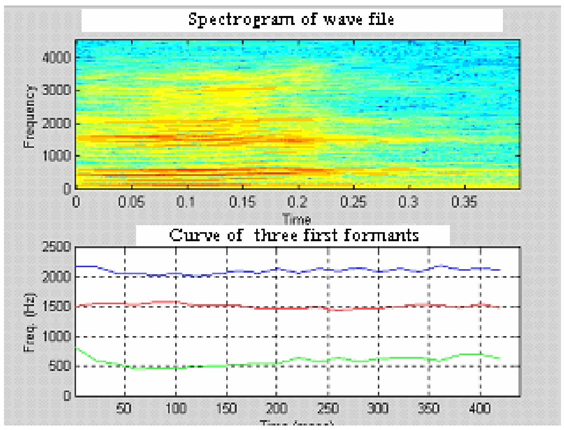


Fig. 3 (a) Formants evolution of a normal male voice

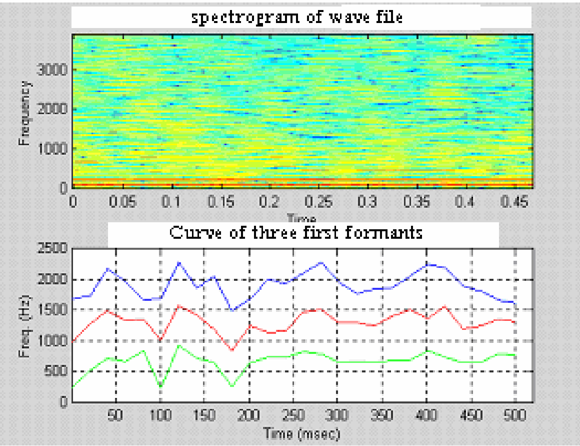


Fig. 3 (b) Formants evolution of a pathological male voice

## III. New Approach for Voice Pathology Classification

This paper propose a technique that uses wavelet analysis to extract a feature vector from speech samples, which is used as input to a Multilayer Neural Network classifier. Wavelet analysis provides a two-dimensional pattern of wavelet coefficients. The wavelet coefficients at various level of scaling with their corresponding normalized energies and entropies are used to formulate a feature vector of speech sample. Attempt is made to use this feature vector as a diagnostic tool to identify pathological disorders in the voice. A three layer feed forward network with sigmoid activation is used for classification. Generalized Back Propagation Algorithm (BPA) is used for training of the network.

### *A. Algorithm of the Hybrid Method*

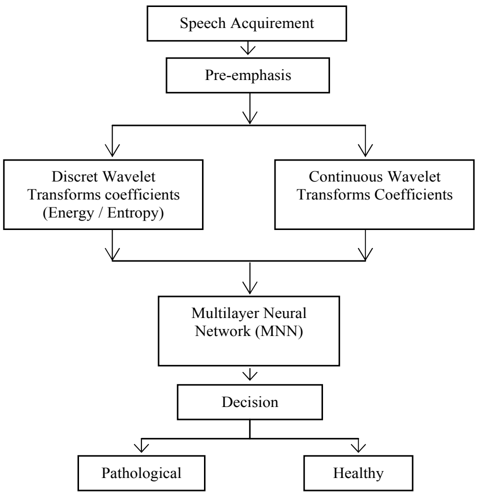


Fig. 4 Hybrid method algorithm

### *B. Wavelet Transforms Analysis*

The wavelet transform can be viewed as transforming the signal from the time domain to the wavelet domain. This new domain contains more complicated basis functions called wavelets, mother wavelets or analysing wavelets.

A wavelet prototype function at a scale s and a spatial displacement u is defined as: [6]

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right) \quad (u \in IR,\ s \in IR_{+}^{*}) \tag{1}$$

This localisation feature, along with wavelets localisation of frequency, makes many functions and operators using wavelets sparse when transformed into the wavelet domain. This sparseness, in turn results in a number of useful applications such as data compression and detecting features in signals.

### *C. Continuous Wavelet Transforms*

The Continuous Wavelet Transform (CWT) is used to decompose a signal into wavelets, small oscillations that are highly localized in time. Whereas the Fourier transform decomposes a signal into infinite length sines and cosines, effectively losing all time-localization information, the CWT's basis functions are scaled and shifted versions of the time-localized mother wavelet. The CWT is used to construct a time-frequency representation of a signal that offers very good time and frequency localization.

The CWT is an excellent tool for mapping the changing properties of non-stationary signals. The CWT is also an ideal tool for determining whether or not a signal is stationary in a global sense. When a signal is judged non-stationary, the CWT can be used to identify stationary sections of the data stream.

Specifically, a Wavelet Transform function f(t) Є $L^2$(R) (defines space of square integrable functions) can be represented as:

$$\begin{aligned} W(f)(u,s) &= \int_{-\infty}^{+\infty} f(t)\psi_{u,s}^{*}(t)dt \\ &= \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{s}}\psi^{*}\left(\frac{t-u}{s}\right)dt \end{aligned} \tag{2}$$

The factor of scale includes an aspect transfer at a time in the time brought by the term u, but also an aspect dilation at a time in time and in amplitude brought by the terms s and $\sqrt{s}$ .

The dilation in amplitude permits to preserve a constant norm for all elements of the basis (wavelet energy). The most important criteria for the choice of a wavelet is to present, for it and its Fourier transformed of the possible weakest oscillations; it is what will permit to assure a good temporal and frequency resolution.

Many different types of mother wavelets are available and for phase evaluation application the most commonly used CWT mother wavelet is the Morlet wavelet. Morlet wavelet can be decomposed in two parts, one for the real part, and the other for the imaginary part. Since our signal is real, we have chosen its real part as defined in Matlab toolbox:

$$\psi(x) = e^{\frac{-x^2}{2}}\cos(5x) \tag{3}$$

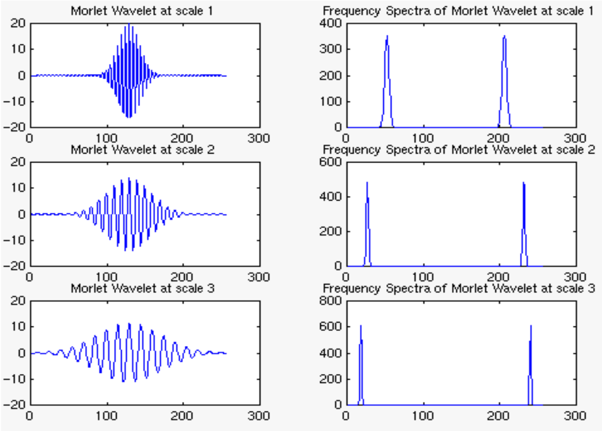


Fig. 5 Morlet wavelets with corresponding spectrum

Fig. 5 shows examples of the Morlet wavelet function at three different scales, along with the corresponding frequency domain information.

*D. Discrete Wavelet Transforms*

The Discrete Wavelet Transform (DWT), which is based on sub-band coding, is found to yield a fast computation of Wavelet Transform. It is easy to implement and reduces the computation time and resources required.

The Discrete Wavelet Transform (DWT) involves choosing scales and positions based on powers of two, so called dyadic scales and positions. The mother wavelet is rescaled or dilated by powers of two and translated by integers. In CWT, the signals are analyzed using a set of basis functions which relate to each other by simple scaling and translation. In the case of DWT, a time-scale representation of the digital signal is obtained using digital filtering techniques. The signal to be analyzed is passed through filters with different cutoff frequencies at different scales.

The criteria for selecting a proper mother wavelet is to have a wavelet function with enough number of vanishing moments in order to represent the salient features of the disturbance. At the same time, this wavelet should provide sharp cut-off frequencies. Furthermore, the selected mother wavelet should be orthonormal.

Daubechies 40 shows the sharper cut-off frequency compared with the others and hence the leakage energy between different resolution levels is reduced [15]. The number of vanishing moments of db40 wavelet is large, and hence it gives a meaningful wavelet spectrum of the analyzed signal. The Daubechies wavelet db40 is selected as a good choice because of its high performance in an informal listening test.

Wavelet that one often uses in the setting of the treatment of the discreet dimensional mono signal is the wavelet of Daubechies. For fast wavelets transformed (DWT), the functions are defined by a game of indications that one designates under the appellation "coefficients of the filters in wavelets" [6].

The Daubechies wavelets with compact support are functions to p hopeless moments, their regularity increases with p.

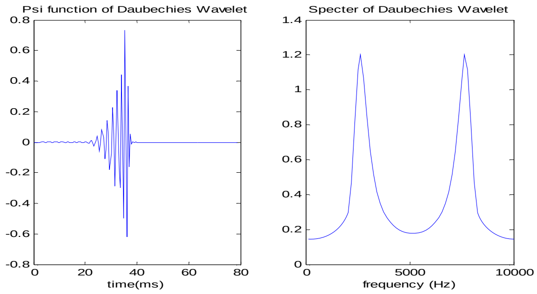


Fig. 6 Wavelet of Daubechies to 40 hopeless moments and its spectrum

Fig. 6 shows the variation of the analysis function of the Daubechies wavelet (db40) and its corresponding spectrum.

*E. DWT and Filter Banks*

Starting with a discrete input signal vector x[n], the first stage of the fast wavelet transform (FWT) algorithm decomposes the signal into two sets of coefficients. These are the approximation coefficients cA1 (low frequency information) and the detail coefficients cD1 (high frequency information). The DWT is computed by successive lowpass and highpass filtering of the discrete time-domain signal as shown in Fig. 7.

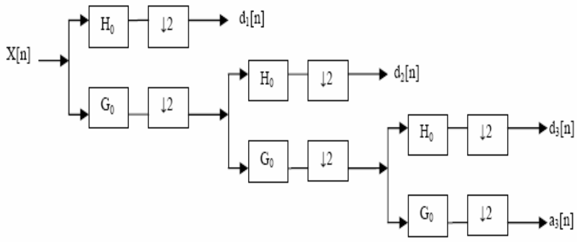


Fig. 7 Three-level wavelet decomposition tree

This is called the Mallat algorithm or Mallat-tree decomposition. Its significance is in the manner it connects the continuous-time mutiresolution to discrete-time filters. In the figure, the signal is denoted by the sequence x[n], where n is an integer.

The low pass filter is denoted by $G_0$ while the high pass filter is denoted by $H_0$. At each level, the high pass filter produces detail information; d[n], while the low pass filter associated with scaling function produces coarse approximations, a[n].

*F. Neural Networks*

Neural network 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 [11].

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.

A neuron (Fig. 8) is an information processing unit, which is an essential part of a neural network. A neuron consists of three main elements: synapses (links), a linear combiner, and an activation function. Each synapse (link) contains a weight factor. Input p(i), which is connected to neurone k, is multiplied by synaptic weight w(k, i).

The linear combiner adds the neurone's weighted inputs together and the activation function F limits the neuron's output. The figure also shows the bias factor b.

Hence, the output of a neuron depends on its inputs and its activation function. There are different types of activation functions that can be used in Matlab. The most commonly used activation functions are hard limit, linear, or sigmoid functions. Naturally, one can also construct hes own activation function.

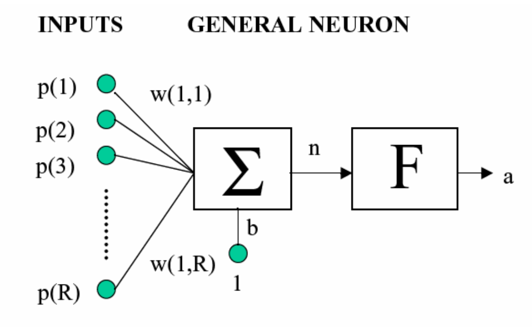


Fig. 8 Model of an artificial neuron

A neural network consists of one or more layers. The neurons are arranged into layers so that the input vector values are fed to the first layer, the output from the first layer is fed to the next and so on, until the output layer is reached. Normally the neurons are completely connected in between layers, so each neuron in layer is connected to every neuron in the next layer.

Each layer has a weight matrix W, a bias vector b, and an output vector a (Fig. 9). The number of neurons usually varies between each layer. In Fig. 9, the number of inputs is R, and the number of neurons in the first layer is S1, while in the second layer it is S2, also the same for other layers. The layers, which are situated between the inputs and the output layer, are called hidden layers. Thus, Fig. 9 shows two hidden layers [11].

A multi-layer neural network can be used to implement an arbitrary Boolean operation, it can be used in formulating boundary surfaces in classification problems, and it can realise almost any arbitrary non-linear function.

A neural network is trained by giving a target output to a certain input group, in which case the term supervised learning is used. Alternatively, a network can be trained through self-guidance, which means that the network parameters adapt according to the input. In both cases, the free parameters in the network, weights and biases, adapt according to the measured data. The training can be gradual (incremental training), which means that the weights and biases are adapted every time that a new training example is fed to the network, or it can be done in batches (batch training), in which case the parameters are not adapted until all the examples have been fed.

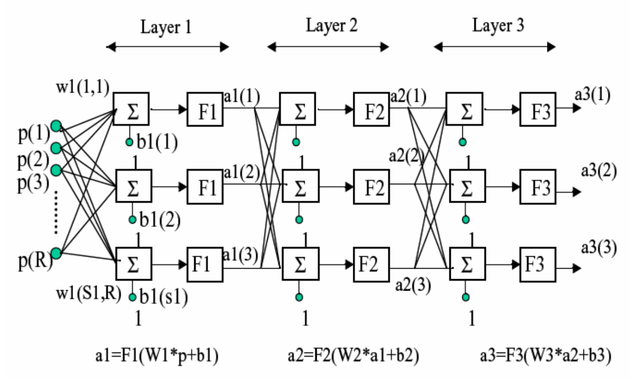


Fig. 9 Multi-layer neural network structure

The backpropagation algorithm is often used in the training of multi-layer neural networks. Backpropagation consists of two stages: forward pass and backward pass through the network. In the forward pass, the input is conveyed layer by layer all the way to the output neuron, which produces the true output of the network. In the backward pass, an error signal is produced by deducting the desired output from the actual output. This error signal is conveyed backwards through the network, layer by layer, simultaneously modifying the values of the network weights, thus bringing the actual output closer to the desired output.

In practice, an activation function is used to form a gradient into the network's weight factor space, and the weights are modified into the opposite direction from the gradient. Thus, the activation function must be constant and derivable.

If sigmoid activation functions are used in the output layer, the outputs of the network are limited to a small range. Also, if a linear activation function is used in the output layer, the network outputs can have any real number values.

## IV. Developed Method

The new method used in this survey is called hybrid since it takes as a basis on a wavelet transformed followed by a neural network. The input vector for the neural network is consists by the normalized coefficients energy corresponding to the coefficients of wavelet transformed of input speech signal. The input signal is the voice recorded of a speaker who can be healthy (normal) or pathological. The pathological voice data base has been prepared with help of the G.E laboratory of UCLA-LosAngeles University and the RABTA hospital of Tunis.

### A. Wavelet Coefficients

Speech is a highly non-stationary signal, and then Fourier Transform is not a very useful tool for analysis. Wavelet Transform approach being a good tool for analysis of non-stationary signals, as it is useful in localizing a symptom both in time and frequency scales. In disordered speech, the non-stationary behaviour of the Pitch can be analyzed using Wavelet Transforms [5].

We apply discreet and continuous wavelet transform coefficients on the same word pronounced by two speakers that are the same sex and probably same age. One of these speakers is healthy whereas the other suffers from Alzheimer's illness. The simulation result of different absolute coefficients is given in Fig. 10 and Fig. 11.

We notice a clear difference between wavelet coefficient evolutions of the two different signals. This analysis method can also provide a visual pattern, which can be of considerable help in diagnosis.

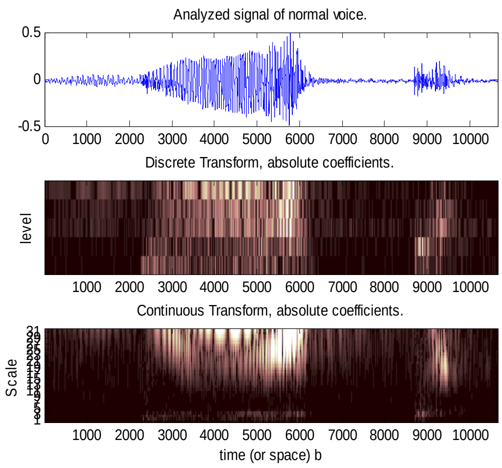


Fig. 10 Wavelet analysis of normal voice

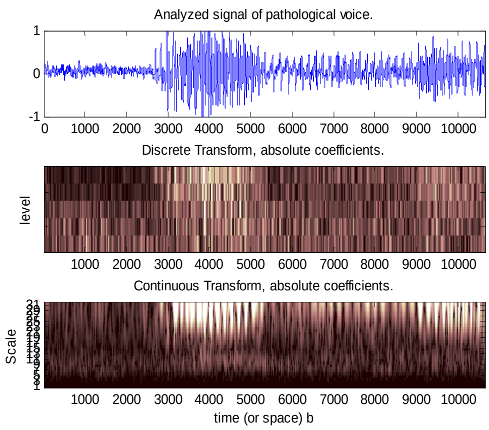


Fig. 11 Wavelet analysis of pathological voice

### B. Proposed System

The Matlab 7.0 platform is used for implementating the neural network formed of three layers, one of input, one of output, and a hidden layer (Fig. 12).

The input layer is formed of five neurons that correspond 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 take the decision, pathological or normal.

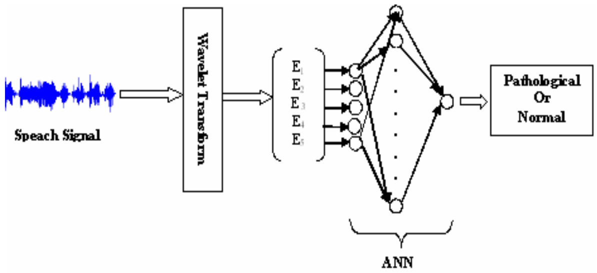


Fig. 12 Voices classification model

### C. Neural Network Design

The classic cycle of a neuron network development can be including in seven stages:

- Collection of data
- Analysis of data
- Separation of data bases
- Choice of a neural network
- Formatting of data
- Training
- Testing

In our survey we use a multilayer neural network (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 the 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 error 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 validation system is formed by normal and pathological voices containing one hundred words of mixed type (50 words for normal voices and 50 words for pathological voices). The data base implies several pathological voices as laryngeal, neurological (Parkinson, Alzheimer, dyslexia, dysphony…) 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.

### D. 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 wavelet coefficients. Continuous wavelet coefficients or wavelet packet decomposition coefficients with their corresponding energy and entropy values will be used as part of input feature vector for the neural network.

A Filter Bank is used to extract the wavelet coefficients. Wavelet packet decomposition at level three was applied to the speech signal using the Daubechies-40 wavelet packet filters.

The lower band scale presents a more dominant periodicty than the higher band scale. This periodicty is decreasing in the pathological speech but it is very consistent in the normal speech [5].

## V. SIMULATION RESULTS

### A. *Discrete Wavelet Transforms Results*

We use 100 words in total, pronounced by different speakers, 50 of whom are normal and the other present pathologies of vocal or neurological origin. For the training, we use 80 words (40 normal and 40 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 results, we vary the nature of the input coefficients vector to the neural network. Different results obtained by normalized wavelet coefficients, also by normalized wavelet energy coefficients and by normalized wavelet entropy coefficients will be presented and compared.

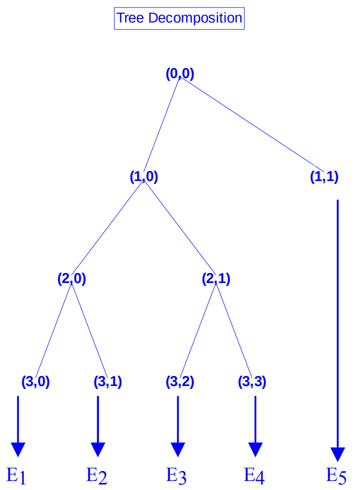


Fig. 13 Discrete wavelet coefficients extraction

- *Approach 1: Wavelet Coefficients As Input Vector To The MNN*

We use a wavelet filter bank with Daubechies-40 wavelet packet filters at level three to extract a five wavelet coefficient then we calculate their corresponding middle values.

The middle value E(s) of the waveforms at the terminal node signals (s) is defined as:

$$E(s) = \frac{1}{N}\sum_{i=1}^{N} s_i \tag{4}$$

Where N is the size of s and i is the coefficients of wavelet packet decomposition.

Tree wavelet decomposition is shown in Fig. 13.

The obtained training curve is given in Fig. 14

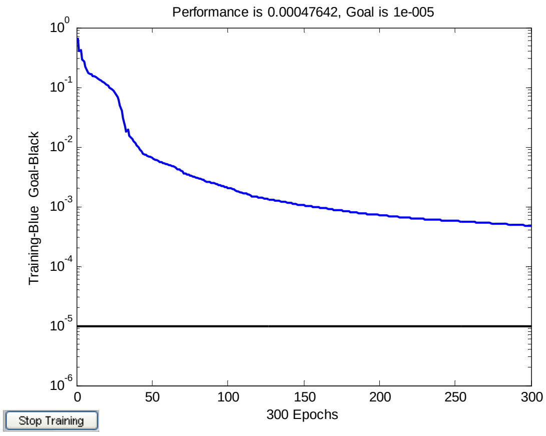


Fig. 14 Training curve with five middle discrete wavelet coefficients

The results of neural network training and testing are regrouped in Table I.

TABLE I
NEURAL NETWORK RESULTS WITH FIVE DISCRETE WAVELET COEFFICIENTS

| Pronounced word | Normal | Pathological |
|---|---|---|
| Training number | 40 | 40 |
| Test number | 10 | 10 |
| Correct classification | 8 | 7 |
| Rate of classification | 80 % | 70% |

- *Approach 2: Wavelet Energy Coefficients As Input Vector To The MNN*

The energy distribution at various levels of scaling can provide information about localized irregularity in vocal fold vibration. The energy content in each scale is extracted and used as feature vector for the ANN classification.

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

$$E_N(i) = \frac{E_i}{E_T} \tag{5}$$

Where i = 1, 2 …
$E_T$ : Total Energy across all the levels.
Ei : Energy at each level.

$$E_T = \sum_i E_i \tag{6}$$

We use a wavelet filter bank to extract a five wavelet coefficient then we calculate their corresponding energy.

Tree wavelet decomposition is shown in Fig. 13.

The obtained training curve is given in Fig. 15.

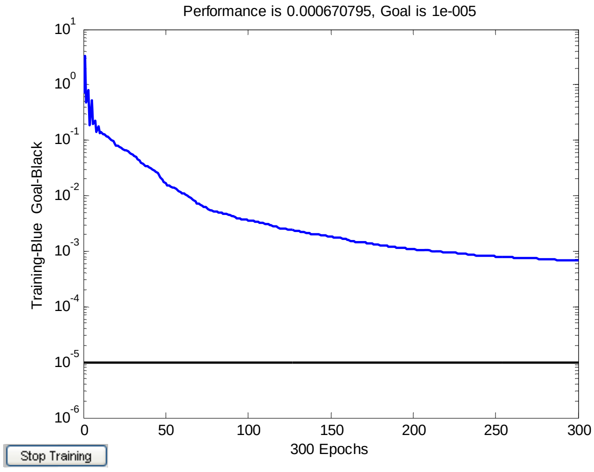


Fig. 15 Training curve with five wavelet energy coefficients

The results of neural network training and testing are regrouped in Table II.

TABLE II
NEURAL NETWORK RESULTS WITH FIVE WAVELET ENERGY COEFFICIENTS

| Pronounced word | Normal | Pathological |
|---|---|---|
| Training number | 40 | 40 |
| Test number | 10 | 10 |
| Correct classification | 9 | 8 |
| Rate of classification | 90 % | 80 % |

- *Approach 3: Wavelet Entropy Coefficients As Input Vector To The MNN*

An entropy-based criterion describes information related properties for an accurate representation of a given signal. Entropy is a common concept in many fields, mainly in signal processing [17]. A method for measuring the entropy appears to be an ideal tool for quantifying the ordering of non-stationary signals. The Shannon entropy (E) is calculated as defined in Equation (6). Where s is the waveforms at the terminal node signals obtained from wavelet decomposition and i in the ($s_i$) is the coefficients of wavelet packet decomposition of s.

$$E(s) = -\sum_i s_i^2 \log(s_i^2) \quad (7)$$

We use a wavelet filter bank to extract a wavelet coefficient then we calculate their corresponding Shannon entropy.

Tree wavelet decomposition is shown in Fig. 13.

The obtained training curve is given in Fig. 16

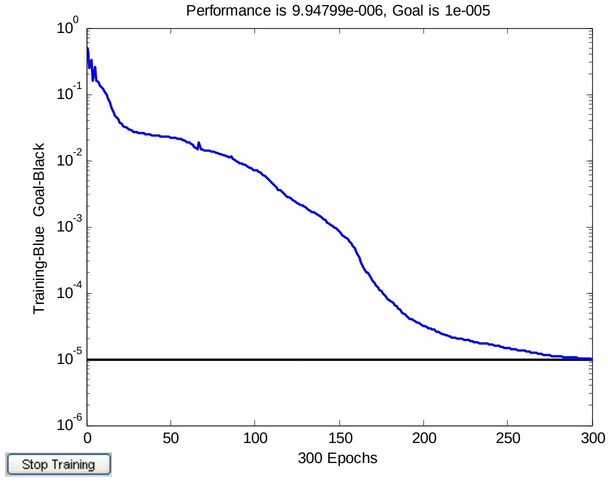


Fig. 16 Training curve with five wavelet entropy coefficients

The results of neural network training and testing are regrouped in Table III.

TABLE III
NEURAL NETWORK RESULTS WITH FIVE WAVELET ENTROPY COEFFICIENTS

| Pronounced word | Normal | Pathological |
|---|---|---|
| Training number | 40 | 40 |
| Test number | 10 | 10 |
| Correct classification | 10 | 9 |
| Rate of classification | 100 % | 90 % |

Then we can summurize these different results in the following diagrams (Fig. 17).

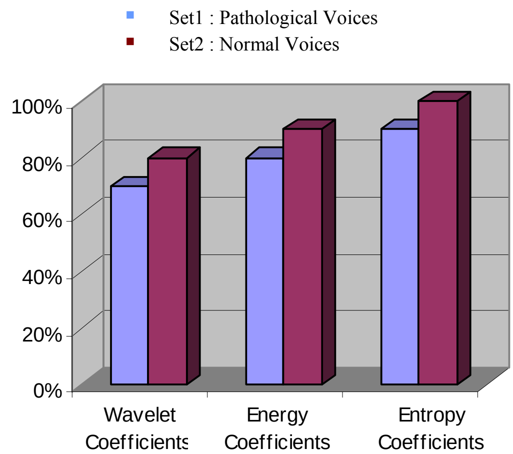


Fig. 17 Voice classification using hybrid method

The stated results show that the proposed method can make an effective interpretation for pathological voices classification. The performance of this hybrid method is given in Table I, II, and III.

### B. Continuous Wavelet Transforms Results

We apply the continuous wavelet transform (CWT) to the level 5 while using the Morlet wavelet. Five coefficients will be recovered that to be used as input vector to the neural network.

The results of neural network training and testing are regrouped in Table IV.

TABLE IV
NEURAL NETWORK RESULTS WITH FIVE CONTINUOUS WAVELET COEFFICIENTS

| Pronounced word | Normal | Pathological |
|---|---|---|
| Training number | 40 | 40 |
| Test number | 10 | 10 |
| Correct classification | 10 | 9 |
| Rate of classification | 100 % | 90 % |

The training curve obtained is given by Fig. 18.

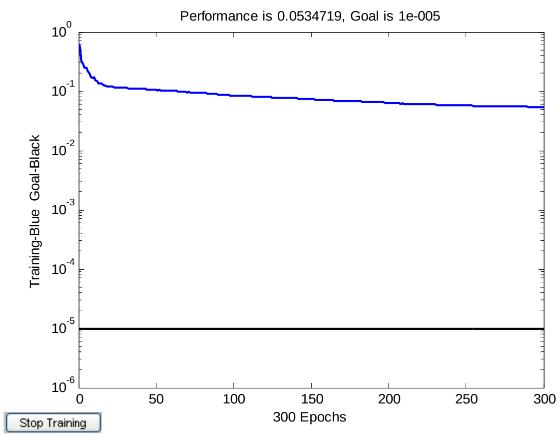


Fig. 18 Training curve with five middle continuous wavelet coefficients

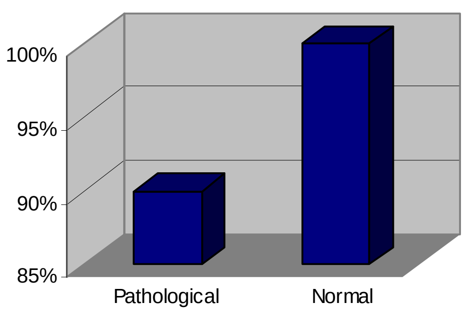


Fig. 19 Classification rate using continuous wavelet transform

### C. Discussion

With wavelets, it is possible to analyze a signal at several levels of resolution, making it possible to capture transient, high-frequency bursts with poor frequency resolution and also slowly varying characteristics with high-frequency resolution.

The DWT has been demonstrated to be an effective tool for extracting information from the speech signals. In comparison with classification techniques, hybrid method has been applied for 5 specific values of wavelet decomposition using Db40 mother wavelet. The decomposition into three levels was quite satisfactory. The classification rate using wavelet entropy was around 90%. Wavelet entropy proved to be a very useful feature for characterizing the speech signals; furthermore the information obtained from the wavelet entropy is related to the energy and consequently to the amplitude of the signal.

The CWT using Morlet wavelet for scale values of 1 to 5, is performed on speech samples (normal, pathological).

Compared to DWT, continuous wavelet transform (CWT) allows more flexibility due to its arbitrary time-scale resolution. These results show that features based on continuous wavelet transforms can serve as one of the basic features for automatic vocal pathology diagnosis.

## VI. CONCLUSION

In this paper we have presented a material and software interface of numeric treatment of the patient's vocal signal based on hybrid approach wavelet transform and neural networks.

The purpose of this work is to conceive a tool to assist the clinicians to follow the evolution of the vocal and neurological pathologies in Tunisian hospitals. The results using the multilayer neural network (MNN) classifier gives the best correct classification. The classification rate is between 70% and 100%. We have demonstrated in this study, a feature vector based on wavelet coefficients that 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. The multilayer neural network (MNN) with back propagation algorithm (BPA) used as a classifier has been proved to be more efficient and more precise than the time-frequency analysis method.

The MNN classifier represents a low cost, accurate, and automatic tool for pathological voice classification using wavelet coefficients. It is presented in this paper as diagnostic tools to aid the physician and clinician in the analysis of speech disease. This work has been validated on a speech pathology database constisting of 8 kinds of pahologies ( 40 sounds collected from national hospitals) in order to increase the reliability of result.

## ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable suggestions and comments. The authors also express their thanks to Professor Saber ABID at ENSET of Tunis for his critical reading, comments, and help.

## REFERENCES

[1] V. Parsa and D. G. Jamieson, “Interactions between speech coders and disordered speech,” Speech Communication, vol. 40, no. 7, pp. 365–385, 2003.
[2] S. B. Davis, “Acoustic characteristics of normal and pathological voices,” Speech and Language: Advances inBasic Research and Practice, vol. 1, pp. 271–335, 1979.
[3] F. Plant, H Kessler, B Cheetham, J Earis, “Speech Monitoring of Infective Laryngitis”, Proceedings of ICSLP96, Philadelphia, pp. 749 – 752 , 1996
[4] M.N. Viera, F.R. McInnes, M.A. Jack “Robust F0 and Jitter estimation in the Pathological voices “, Proceedings of ICSLP96, Philadelphia, pp.745 –748, 1996.
[5] J.Nayak, P.S.Bhat “Classification and analysis of speech abnormalities”, ITBM-RBM 26 (2005) 319-327.
[6] S. Mallat, “A Theory for multiresolution signal decomposition: Wavelet representation”, IEEE Trans. Pattern Analysis and Machine Intelligence. Vol. 11. No. 7 pp674-693 July 1989.
[7] B. Boyanov, S.Hadjitodorov: “Acoustic analysis of pathological voices: a voice analysis system for screening of laryngeal diseases”, Proc. IEEE Engineering in Medical and Biology, (1997), vol. 16, no. 4, 74-82.
[8] J.J. Jiang,Yu Zhang: “Nonlinear dynamic analysis of speech from pathological subjects”, Proc IEEE Electronics Letters, March (2002), vol.38, no.6.
[9] P.Yu, M.Ouaknine, J.Revis, and A.Giovanni, “Objective Voice Analysis for Dysphonic Patients: A Multiparametric Protocol Including Acoustic and Aerodynamic Measurements”, Journal of Voice Vol. 15, No. 4, pp. 529–542, 2001
[10] J.Wang, Jo.Cheolwoo, “Performance of Gaussian Mixture Model as a classifier for Pathological Voice”, proceeding of the ASST in Auckland 2006, pp 165-169.
[11] J.Kortelainen, K.Noponen, « Neural networks », Intelligent Systems 2005
[12] S.Lotfi, C.Adnène, “A Speech Processing Interface for Analysis of Pathological Voices”, in proceeding of ICTTA conference, Damascus 2006.
[13] S.Lotfi, B.Haythem, C.Adnène, “ Interface d’analyse vocale a l’identification de certaines pathologies d’origine neurologique et vocale”, in proceeding of JTM conference, Tunis 2007.
[14] A.chérif, « Pitch detection and formant extraction of Arabic speech processing » Journal of applied acoustics, January 2001
[15] A.M. Gaouda, M. Salama, A. Chikhani, and M. Sultan, “Application of wavelet analysis for monitoring dynamic performance in industrial plants,” North American Power Symposium, Oct. 1997, Laramie, Wyoming.
[16] C.Fredouille, G.Pouchoulin, J-F. Bonastre, M.Azzarello, A.Giovanni, A.Ghio “Application of automatic speaker recognition techniques to pathological voice assessment” in proceeding of international conference on acoustic speech and signal processing (ICASSP 2005)
[17] C. M. Bishop, “Neural Networks for Pattern Recognition”. Oxford: Clarendon Press, 1996.

**L. Salhi** was born the 05-16-1974 at Thala, (center west of Tunisia). He obtained the Bac exp sciences in 1993. He received his bachelor in physics in 1999 from the sciences faculty of Sfax. He has been recruited in computer engineer post in a Tuniso- American company for two years. He received the diploma of Master degree in automatic and signal processing (ATS) in 2004 from the national school of engineers of Tunis (ENIT) with concentration in digital speach signal processing(cochlear filter). Currently, he is registered for doctorate thesis in electrical engineering (signal processing specialty) at the faculty of sciences of Tunis (FST). His thesis subject was the Analysis and Identification of Vocal and Neurological Pathology by digital speech signal processing and classification. He is working a physics teacher at a high school.

**A. Cherif** is a Professor in Electrical Engineering, Coordinator of the Master “Telecommunication Systems”, Director of the Physics Department and of the Signal Processing Laboratory at the Sciences Faculty of Tunis.