Journal of Engineering and Applied Sciences 3 (11): 822-828, 2008
ISSN: 1816-949X
© Medwell Journals, 2008

Gammachirp Wavelet and Neural Network for Identification of Pathological Voices

Lotfi Salhi, Mourad Talbi and Adnene Cherif
Laboratory of Signal Processing, Department of Physics, Faculty of Sciences, Tunis

Abstract: In this study, we present a new method for voice disorders identification based on a gammachirp wavelet transform and Multilayer Neural Network (MNN). The processing algorithm is based on a hybrid technique which uses the gammachirp wavelets coefficients as input of the MNN. 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 identification 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.

Key words: Pathological voices, identification, cochlear filter, gammachirp model, wavelet transforms, neural networks

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 researchers 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 (Parsa and Jamieson, 2003; Davis, 1979). 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 3 systems contributing to the production of the speech: the respiratory system, the laryngeal system and the supra-laryngeal system (Yu *et al*., 2001; Wang and Cheolwoo, 2006). 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, rhythm).

The analysis of the voice disorder stays essentially clinic (Plant *et al*., 1996). The instrumental measures are spilled little in practice clinic. The most used are the acoustic and aerodynamic measures (Viera *et al*., 1996). 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 (Boyanov and Hadjitodorov, 1997). 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 (Jiang and Zhang, 2002). 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 (Plant *et al*., 1996; Vieral *et al*., 1996).

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

MATERIALS AND METHODS

Disorder identification using pitch and formants analysis
Speech processing algorithm: Figure 1 illustrates the speech processing algorithm (Cherif, 2001).

Pitch and formants analysis results: Figure 2 illustrate the pitch variation by application of the cepstrum method analysis of normal and pathological female sounds (32 years) (Salhi and Adnene, 2006; Salhi *et al*., 2006).

Corresponding Author: Lotfi Salhi, Laboratory of Signal Processing, Department of Physics, Faculty of Sciences, Tunis

The Linear Predictive Coding (LPC) method applied on the same voices allows evolution to extract the formants. By comparison to the normal values (Fig. 3), the high variations of the formants F_1, F_2 and F_3 of the pathological male sound confirm the conclusions (Salhi and Adnene, 2006; Salhi *et al.*, 2006).

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 decision. For it, we try to improve this idea by new methods that use wavelet transformed and neural networks.

New approach for voice pathology identification: This study propose a technique that uses gammachirp wavelet analysis to extract a feature vector from speech samples, which is used as input to a Multilayer Neural Network

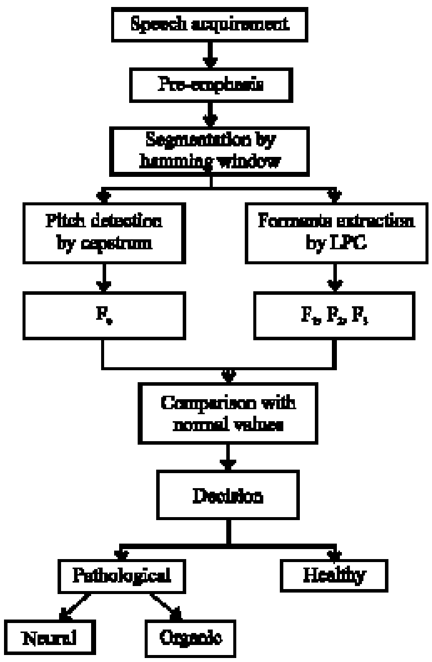

Fig. 1: Speech processing algorithm

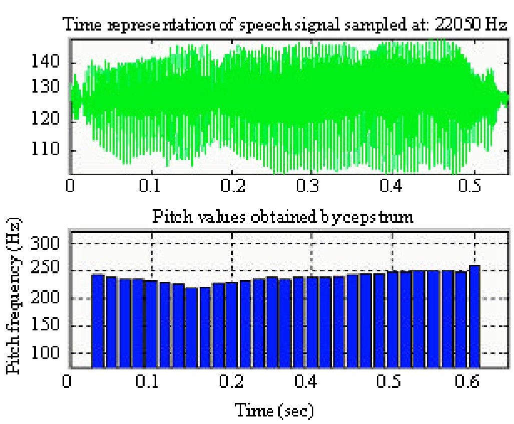

Fig. 2a: Pitch evolution of normal female voice

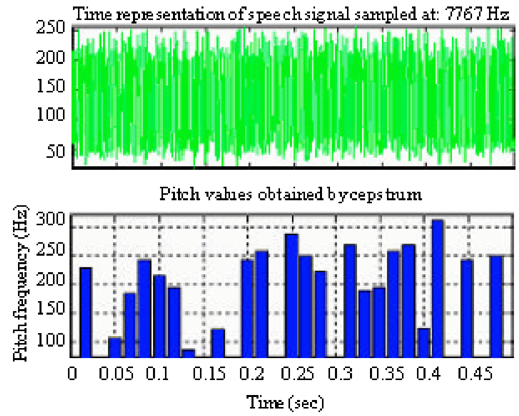

Fig. 2b: Pitch evolution of pathological female voice

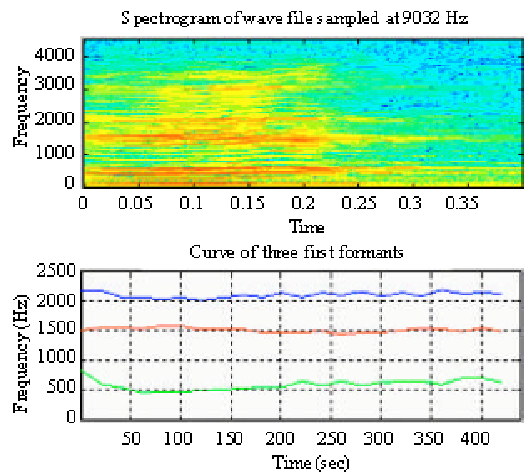

Fig. 3a: Formants evolution of normal male voice

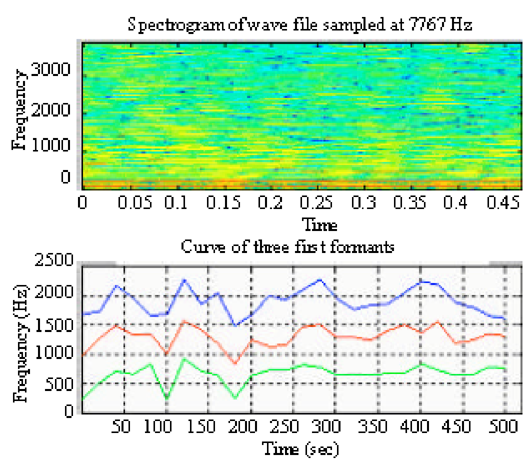

Fig. 3b: Formants evolution of pathological male voice

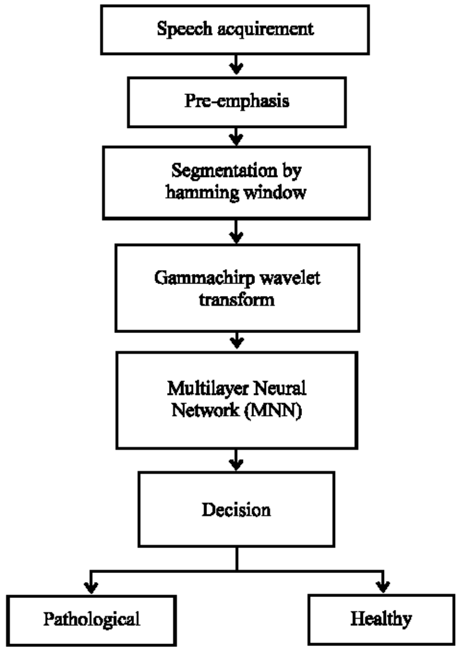

Fig. 4: Hybrid method algorithm

classifier. Wavelet analysis provides a 2-dimensional pattern of wavelet coefficients. The middle values content of wavelet coefficients at various level of scaling is 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 identification. Generalized Back Propagation Algorithm (BPA) is used for training of the network.

Algorithm of the hybrid method: Figure 4 illustrates the hybrid method algorithm (Salhi and Adnene, 2006).

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 (Mallat, 1989):

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

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 $f(t) \in L^2(R)$ (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 s et $\sqrt{s}$.

Gammachirp model of cochlear filter: Several models have been proposed to simulate the working of the cochlear filter. Seen as its temporal-specter properties, the gammachirp filter underwent a good success in the psychoacoustic research. Indeed, globally it answers the requirements and the complexities of the cochlear filter. In addition to its good approximation in the psychoacoustical appareillement, it possesses a temporal-specter optimization of the human auditory filter (Irino and Patterson, 1997).

The notion of the wavelet transform possesses a big importance in the signal treatment domain and in the speech analysis. it comes the idea to exploit the theory of the wavelet transform in the implementation of an auditory model of the cochlear filter proposed by Irino and Patterson (1997) that is the Gammachirp filter.

One is going to be interested to the survey of this filter type and à its implementation as a wavelet under the Matlab flat forms.

The impulsional answer of the gammachirp filter is given by the following function (Irino and Patterson, 1997).

$$g(t) = \lambda_n t^{n-1} e^{-2\pi b ERB(f_0)t} e^{j(2\pi f_0 t + c\ln(t) + \varphi)} \qquad t > 0 \tag{3}$$

n = A whole positive defining the order of the corresponding filter
f_0 = The modulation frequency of the gamma function
φ = The original phase
λ_n = An amplitude normalization parameter
$ERB(f_0)$ = Equivalent Rectangulaire Bandwith

An example of the temporal answer of the gammachirp filter is given by Fig. 5 (Salhi, 2007).

The gammachirp function can be considered like wavelet function and constitute a basis of wavelets. This wavelet possesses the following properties: it is not

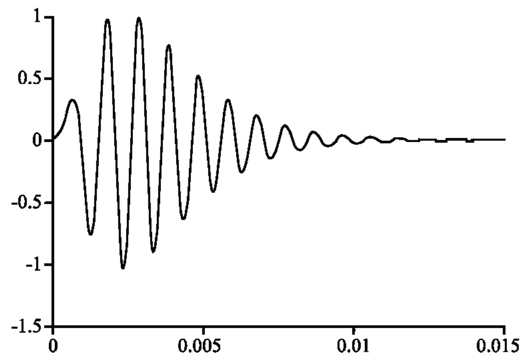

Fig. 5: Temporal answer of a gammachirp function centered on 1000 Hz, with $\lambda_n = 1$, $b = 1$, $c = 0$, $n = 2$ and $\varphi = 0$

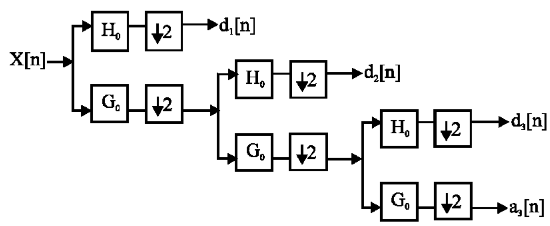

Fig. 6: Three-level wavelet decomposition tree

symmetrical, it is non orthogonal and it doesn't present scale function. This research will be based on the choice of a Gammachirp wavelet centered at the frequency 1000 Hz (Salhi, 2007; Ouni *et al*., 2001).

Wavelet transforms 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 2 sets of coefficients. These are the approximation coefficients cA1 (low frequency information) and the detail coefficients cD1 (high frequency information). The wavelet transform application is computed by successive lowpass and highpass filtering of the discrete time-domain signal as shown in Fig. 6 (Gaouda *et al*., 1997).

In Fig. 7, 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).

Application of the gammachirp wavelet on a speech signal: We applie the implemented Gammachirp wavelt to an input signal that is one of the utilized vowels that is /a/ (F_0 = 100 Hz, F_1 = 730 Hz, F_2 = 1090 Hz, F_3 = 2440 Hz). We get the result that represents the first 5 levels of analysis of the signal by this type of wavelet, as well as the corresponding specters (Salhi, 2007) (Fig. 7).

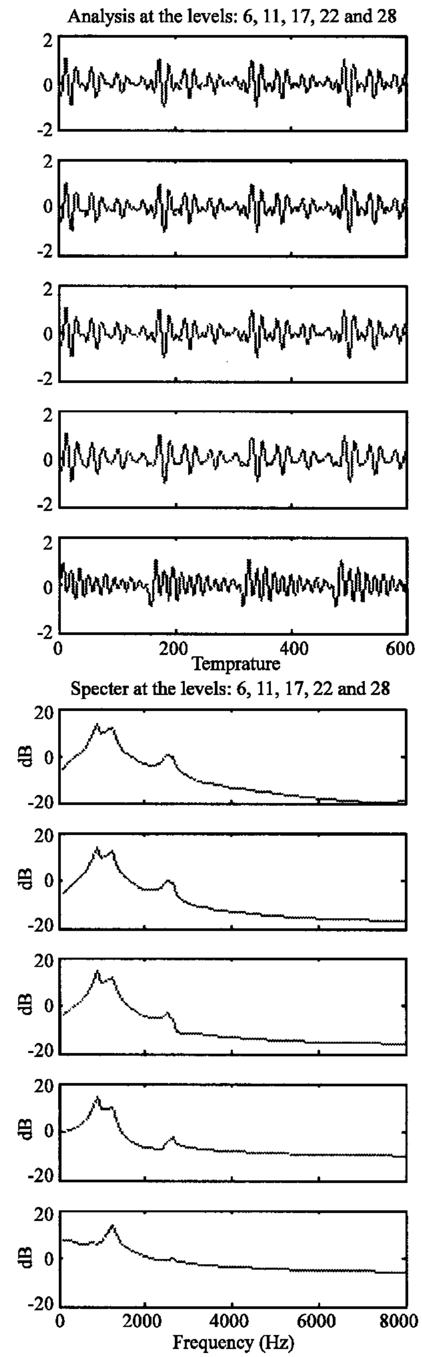

Fig. 7: Analysis of the vowel /a/ by the gammachirp wavelet

The gammachirp filter detects à practically the first level of analysis the three forming of the signal corresponding to the vowel /a/ and it eliminates at every upper level the high frequencies and this as subdividing at every time centers it fréquentiel by s_0^m (with s_0 = 1,13 and m represents the level of analysis).

Neural networks: Neural were chosen as a method of pattern matching for many main reasons. First the Mat Lab 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 (Kortelainen and Noponen, 2005).

A neuron (Fig. 8) is an information processing unit, which is an essential part of a neural network. A neuron consists of 3 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).

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 (Mehrotra *et al*., 1997).

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 S_1, while in the second layer it is S_2, 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 2 hidden layers (Kortelainen and Noponen, 2005).

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.

The backpropagation algorithm is often used in the training of multi-layer neural networks.

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.

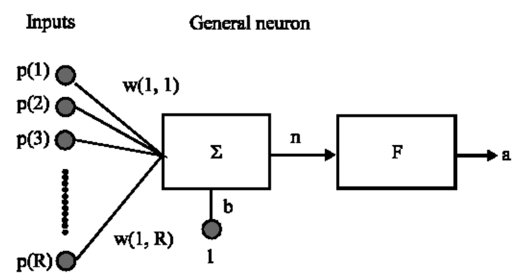

Fig. 8: Model of an artificial neuron

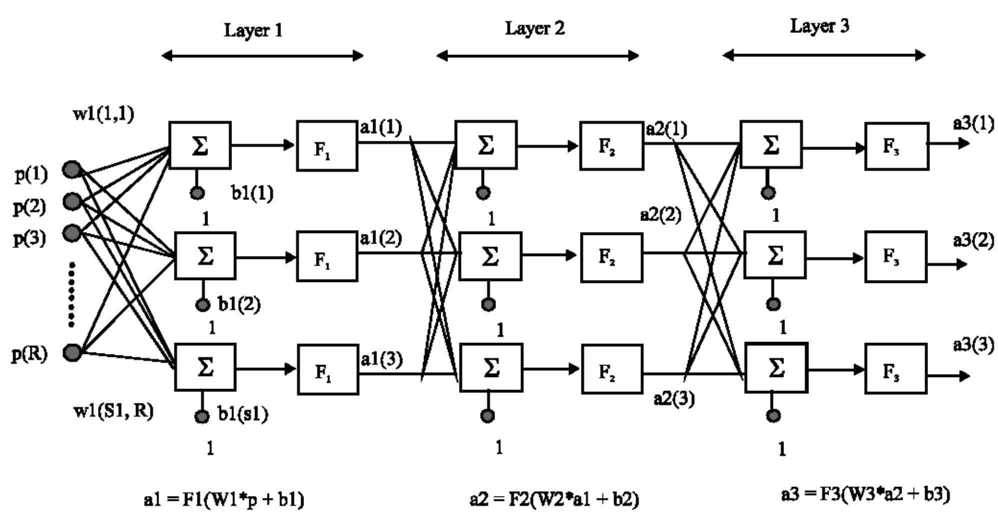

Fig. 9: Multi-layer neural network structure

RESULTS AND DISCUSSION

Proposed system: The Matlab 7.0 platform is used for implementation of the neural network formed of 3 layers, one of input, one of output and a hidden layer (Fig. 10).

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 gammachirp wavelet decomposition.

The hidden layer contains 15 neurons and the output layer contains only one neuron to give the decision pathological or normal (Fig. 10).

Neural network design: In our study, we use a Multilayer Neural Network (MNN) with only one hidden layer 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 (Nayak and Bhat, 2005).

Feature extraction: We apply the Continuous Wavelet Transform (CWT) to the level 5 while using the gammachirp wavelet. Five coefficients will be recovered that to be used as input vector to the neural network. 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.

Neural networks 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 normals and 20 pathological).

After training, the network will be tested with 20 words different from those used for the training (10 normals and 10 pathological).

The obtained training curve is given in Fig. 11. The results of neural network training and testing are regrouped in Table 1.

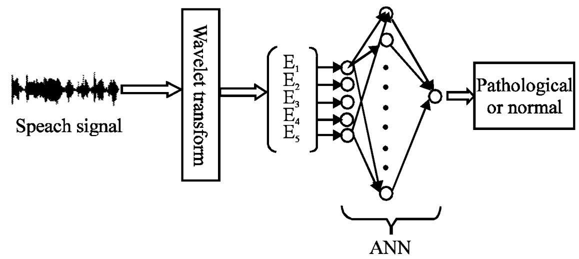

Fig. 10: Voices identification model

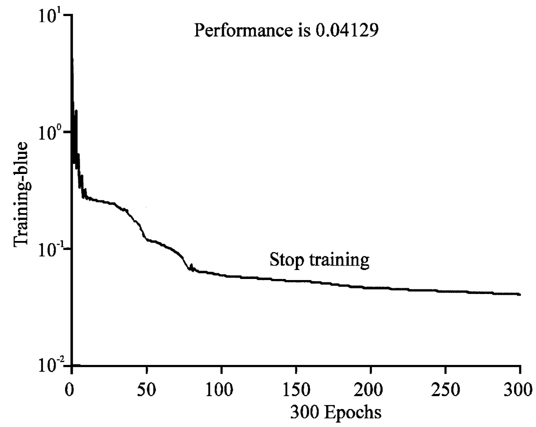

Fig. 11: Training curve with gammachirp wavelet coefficients

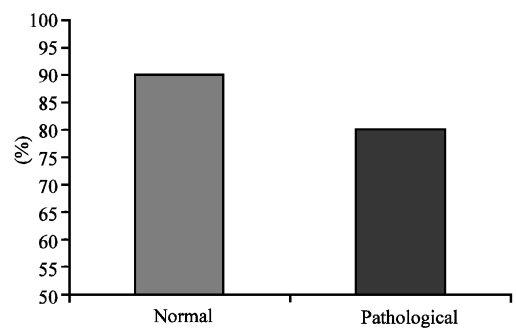

Fig. 12: Voice identification results

Table 1: Neural network results with gammachirp wavelet coefficients

Pronounced word	Normal	Pathological
Training number	20	20
Test number	10	10
Correcte identification	9	8
Rate of identification	90%	80%

Then we can summarize these different results in the following diagrams (Fig. 12).

CONCLUSION

The goal of this research 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 study a material and software interface of numeric treatment of the patient's vocal signal based on gammachirp wavelet transforms and neural networks. Result of the Multilayer Neural Network (MNN) classifier gives the correct identification. The identification rate is between 80 and 90%.

The application of the gammachirp wavelet transforms is then used to analyse and extract coefficients that are used for the MNN input vector. We have demonstrated in this study, a feature vector based on gammachirp wavelet coefficients is useful for identification of normal and pathological speech data. At a preliminary level, the speech data is classified into 2 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.

It is presented in this study, as diagnostic tools to aid the physician and clinician in the analysis of speech disease. Therefore, future work will be focused on the specific recognition of illness type that causes the speech pathology.

Finally, this study has to be validated on a larger speech pathology database to increase the result reliability.

REFERENCES

Boyanov, B. and S. Hadjitodorov, 1997. Acoustic analysis of pathological voices: A voice analysis system for screening of laryngeal diseases. Proc. IEEE. Eng. Med. Biol., 16 (4): 74-82.

Cherif, A., 2001. Pitch detection and formant extraction of Arabic speech processing. J. Applied Acoustics.

Davis, S.B., 1979. Acoustic characteristics of normal and pathological voices. Speech and Language: Advances in Basic Res. Practice, 1: 271-335.

Gaouda, A.M., M. Salama, A. Chikhani and M. Sultan, 1997. Application of wavelet analysis for monitoring dynamic performance in industrial plants. North American Power Symposium, Laramie, Wyoming.

Irino, T. and R.D. Patterson, 1997. A time-domain, level-dependent auditory filter: The gammachirp. JASA, 101 (1): 412-419.

Jiang, J.J. and Yu Zhang, 2002. Nonlinear dynamic analysis of speech from pathological subjects. Proc. IEEE. Elect. Lett., 38 (6).

Kortelainen, J. and K. Noponen, 2005. Neural Networks, Intelligent Systems.

Mallat, S., 1989. A Theory for multiresolution signal decomposition: Wavelet representation. IEEE Trans. Pattern Analysis and Machine Intelligence. 11 (7): 674-693.

Mehrotra, K., K. Mohan and S. Ranka, 1997. Elements of artificial neural networks. Cambridge, Mass. MIT Press.

Nayak, J. and P.S. Bhat, 2005. Classification and analysis of speech abnormalities. ITBM-RBM, 26: 319-327.

Ouni, K., Z. Lachiri and N. Ellouze, 2001. Formant estimation using gammachirp filterbank. In Eurospeech, pp: 2471-2474.

Parsa, V. and D.G. Jamieson, 2003. Interactions between speech coders and disordered speech. Speech Commun., 40 (7): 365-385.

Plant, F., H. Kessler, B. Cheetham and J. Earis, 1996. Speech Monitoring of Infective Laryngitis. Proceedings of ICSLP, Philadelphia, pp: 749-752.

Salhi, L., 2007. Design and implementation of the cochlear filter model based on a wavelet transform as part of speech signals analysis. Res. J. Applied Sci., 2 (4): 512-521.

Salhi, L. and C. Adnene, 2006. A Speech Processing Interface for Analysis of Pathological Voices. In: Proceeding of ICTTA Conference, Damascus.

Salhi, L., B. Haythem and C. Adnene, 2006. Interface d'analyse vocale a l'identification de certaines pathologies d'origine neurologique et vocale. In: Proceeding of JTM Conference, Tunis.

Viera, M.N., F.R. McInnes and M.A. Jack, 1996. Robust F0 andJitter estimation in the Pathological voices. Proc. ICSLP, Philadelphia, pp: 745-748.

Wang, J. and J.O. Cheolwoo, 2006. Performance of Gaussian Mixture Model as a classifier for Pathological Voice. Proceeding of the ASST in Auckland, pp: 165-169.

Yu, P., M. Ouaknine, J. Revis and A. Giovanni, 2001. Objective Voice Analysis for Dysphonic Patients: A Multiparametric Protocol Including Acoustic and Aerodynamic Measurements. J. Voice, 15 (4): 529-542.