Abstract—In this paper we present a new method for voice disorders classification 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 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.

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 [1,2].

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 these 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, rhythm). The analysis of the voice disorder stays essentially clinic [3]. The instrumental measures are spilled little in practice clinic. The most used are the acoustic and aerodynamic measures [4]. 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 [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].

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

A. Speech Processing Algorithm

![Figure 1: Speech processing algorithm](image-url)
B. 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).

![Pitch evolution for a normal female voice](image1)

![Pitch evolution for a pathological female voice](image2)

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

![Formants evolution for a normal male voice](image3)

![Formants evolution for a pathological male voice](image4)

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.

III. NEW APPROACH FOR VOICE PATHOLOGY CLASSIFICATION

This paper 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 classifier. Wavelet analysis provides a two-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 classification. Generalized Back Propagation Algorithm (BPA) is used for training of the network.

A. Algorithm Of The Hybrid Method

![Hybrid method algorithm](image5)

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]

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

(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 \( f(t) \) \((f(t) \in L^2(\mathbb{R}))\) (defines space of square integrable functions) can be represented as:

\[
W(f)(u,s) = \int_{-\infty}^{\infty} f(t) \psi_{s,u}(t) dt
\]

\[W(f)(u,s) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) dt\]

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 \( \sqrt{s} \).

C. 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 [18].

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 that is the Gammachirp filter.

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

The impulsional answer of the gammachirp filter is given by the following function: [18]

\[
g(t) = \lambda_n n^{-1} e^{-2\pi i \text{ERB}(f_0) t} e^{i(2\pi f_0 t + c \ln(t) + \phi)}
\]

\( t > 0 \) \( n \) : a whole positive defining the order of the corresponding filter.

\( f_0 \) : the modulation frequency of the gamma function.

\( \phi \) : the original phase,

\( \lambda_n \) an amplitude normalization parameter.

\( \text{ERB}(f_0) \) : Equivalent Rectangular Bandwith

An example of the temporal answer of the gammachirp filter is given by the following figure.

![Temporal answer of a Gammachirp function centered on 1000 Hz, with \( \lambda_n =1 \), b=1, c=0, n=2 & \( \phi=0 \).](image)

The gammachirp function can be considered like wavelet function and constitute a basis of wavelets. This wavelet possesses the following properties: it is not symmetrical, it is non orthographic and it doesn't present scale function. This work will be based on the choice of a Gammachirp wavelet centered at the frequency 1000 Hz [16] [17].

D. 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 two 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 figure 6.

![Three-level wavelet decomposition tree](image)

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, \( a[n] \).

E. Application of the Gammachirp Wavelet on a Speech Signal

We apply the implemented Gammachirp wavelet to an input signal that is one of the utilized vowels that is /a/ (\( F_0 = 100Hz \), \( F_1 = 730Hz \), \( F_2 = 1090Hz \), \( F_3 = 2440Hz \)). We get the result that represents the first five levels of analysis of the signal by this type of wavelet, as well as the corresponding specters.

![Analysis of the vowel /a/ by the Gammachirp wavelet](image)
The gammachirp filter detects 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).

F. Neural Networks

Neural 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].

A neuron (Figure 9) 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 $w(k, i)$. 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 its own activation function.

Each layer has a weight matrix $W$, a bias vector $b$, and an output vector $a$ (Figure 10). The number of neurons usually varies between each layer. In figure 10, 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, Figure 9 shows two hidden layers [11].

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.

IV. DEVELOPED METHOD

A. 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 (Figure 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 fifteen neurons and the output layer contains only one neuron to give the decision pathological or normal (Figure 10).

B. Neural Network Design

In our survey 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.

C. 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:

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

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

V. 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 figure 11.

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

Table I: Neural Network Results With Gammachirp Wavelet Coefficients

<table>
<thead>
<tr>
<th>Pronounced word</th>
<th>Normal</th>
<th>Pathological</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training number</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Test number</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Correcte classification</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Rate of classification</td>
<td>90 %</td>
<td>80 %</td>
</tr>
</tbody>
</table>

VI. CONCLUSION

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 gammachirp wavelet transforms and neural networks.

Result of the multilayer neural network (MNN) classifier gives the correct classification. The classification 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 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.

It is presented in this paper 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 work has to be validated on a larger speech pathology database to increase the result reliability.

VII. REFERENCES


