ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

Optimization the Parameters for Speech Recognition System Using Genetic Algorithm

M. Ben Nasr[1], S. Saoud[2], A. Cherif[3]

Assistant Professor, Dep. Electronics Faculty of Sciences of Tunis, University of Tunis El-Manar, Tunisia [1]

Assistant Professor, Dep. Electronics Faculty of Sciences of Tunis , University of Tunis El-Manar, Tunisia [2]

Professor, Dep. Electronics Faculty of Sciences of Tunis, University of Tunis El-Manar ,Tunisia [3]

ABSTRACT: This paper adopted Hidden Markov Model (HMM) to recognize Arabic isolated words . The database contain eleven Arabic isolated words. We have repeated each of them twenty five times by mono –locutor. Feature extraction using Bionic Wavelet Transform (BWT) and Mel Frequency Cepstral Coefficient (MFCC) are carried over the speech frame of the input speech. This is followed by Vector Quantization (VQ) and Hidden Markov Modeling. We describe in this paper the use of Genetic Algorithm (GA) to make a global search of optimal HMM parameters. We can find that the performance of speech recognition was improved by the later method. Experimentally it is observed that recognition accuracy is 96.96%.

KEYWORDS: Arabic Speech Recognition; Hidden Markov Model (HMM); Mel Frequency Cepstral Coefficient (MFCC); Bionic Wavelet Transform (B WT); Genetic Algorithm (GA);Vector Quantization (VQ)

I. INTRODUCTION

Speech recognition or more commonly known as automatic speech recognition (ASR) is defined as the process of interpreting human speech in a computer. With the development of ASR technology, we still have problem in speech recognition system.
Various methods were developed for speech recognition and classification. Recurrent Neural Network and Dynamic Time Warping (DTW) [1], Multi-layer Perceptron (MLP) [2] and Hidden Markov Models (HMM) [3]are some common methods used to recognize the speech signal.

HMM is in fact the most popular approach for speech recognition. The hidden Markov models became the perfect solution to speech recognition problem. Indeed, HMMs are rich in mathematical structures and therefore can be used in a large range of applications. In addition, these models give outstanding results when they are properly applied. HMM is a probabilistic finite state automaton. It is constituted of states {nodes}, linked together by transitions {arcs} [4].Our work consists at using our speech database that contains Arabic isolated words that are recorded by a mono-speaker and using a HMM for speech recognition and classification. The optimization of HMM parameters is very important step in Automatic Speech Recognition (ASR) using HMM. Since, obtaining the optimal values of the model parameters will automatically raise the ASRs recognition precision. This procedure is referred to as HMM training ; we have used the speech occurrences to estimate the model parameters. The well known Baum-Welch algorithm is usually utilized to perform HMM training, an initial guess of the parameters is made randomly, then more exact parameters are computed in each iteration until the algorithm converge. [6], however the problem of optimizing model parameters of HMM is of great interest to the researchers.

According to the tests the isolated word recognition system based on Markovian modeling showed the influence of the variation of each model parameter (the state number, the size of dictionary and the coefficients number) on the recognition rate of the system. Thus, it is very hard to find an appropriate estimation for the parameters of HMM.
Genetic Algorithm (GA)[7] has been used in the optimization of HMM in order to attain the optimum model parameters and to train HMM.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

II. FEATURE EXTRACTION

Feature extraction is a critical element in speech recognition since it generates the parameters on which the recognition algorithm is based and reduces the speech signal variability. In this work, we used Bionic Wavelet Transform and Mel Frequency Cepstral Coefficient for feature extraction.

II.1. BIONIC WAVELET TRANSFORM

The variable costs (fuel costs) of thermal unit are usually modeled as a nonlinear function of the unit's power output. Thermal units have a number of steam admission valves that are opened in sequence as the power output is increased. This is particularly emphasized for combined cycle gas turbine. BWT was presented as an adaptive wavelet transform. It is conceived particularly for modeling the human auditory system [8, 9]. The adaptive nature of the Bionic Wavelet Transform (BWT) is assured by substituting the constant factor of the wavelet transform with a variable quality factor. The expression of the mother wavelet $\psi(t)$ is as follows:

expression of the mother wavelet $\psi(t)$ is as follows:

$$\psi(t) = \tilde{\psi}(t)\, e^{j2\pi f_0 t} \tag{1}$$

Where f_0 is the center frequency and $\tilde{\psi}$ is the envelope function of $\tilde{\psi}(t)$. So the function $\tilde{\psi}(t)$ is:

$$\tilde{\psi}(t) = e^{-\left(\frac{t}{T_0}\right)^2} \tag{2}$$

Where T_0 represent the initial-support of the unscaled mother wavelet.
By using a time varying function T, the mother function of the BWT is represented as follows:

$$\psi_T(t) = \frac{1}{T}\tilde{\psi}\left(\frac{t}{T}\right) e^{j2\pi f t} \tag{3}$$

The BWT of a given signal x(t) is expressed as follows:

$$BWT_x(a,\tau) = \frac{1}{\sqrt{|a|}} \int x(t)\tilde{\psi}_T^*\left(\frac{t-\tau}{a}\right) dt$$
$$= \frac{1}{T\sqrt{|a|}} \int x(t)\psi^*\left(\frac{t-\tau}{a\cdot T}\right) e^{-j2\pi f\left(\frac{t-\tau}{a}\right)} dt \tag{4}$$

From where, the adaptive nature of the BWT is taken by a time-varying factor T which represents the scaling of the cochlear filter bank quality at every scale over time . For the human auditory system, Yao and Zhang [8] have used $f_0 = 15165.4\ Hz$. The distinction of the scale variable a is consummated using a pre-defined logarithmic spacing across the desired frequency rang .So; the center frequency at each scale is given as follows [12-14]:

$$f_m = \frac{f_0}{(1.1623)^m} \quad , \quad m = 0,1,2,... \tag{5}$$

For this work, coefficients at 21 scales, $m = 11,12,...30$, are computed through numerical integration of the CWT. The 21 scales answer to center frequencies logarithmically varied from 166.4 Hz to 3369.7 Hz. For every time and scale, the adapting function $T(a,\tau)$ is expressed via the following equation :

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

$$T(a,\tau+\Delta\tau)=\left(1-\tilde{G}_1\frac{BWT_S}{BWT_S+\left|BWT_S(a,s)\right|}\right)^{-1}\times \left(1+\tilde{G}_2\left|\frac{\partial}{\partial\tau}BWT_X(a,\tau)\right|\right)^{-1} \quad (6)$$

where $\tilde{G}_1$ is the active gain factor which represents the outer hair cell resistance function, $\tilde{G}_2$ designates the active gain factor to represent the time-varying compliance of Basilar membrane, BWT_s is a constant representing the time-varying compliance of Basilar membrane, $BWT_x(a,\tau)$ is the Bionic Wavelet Transform at time τ and scale a , and $\Delta\tau$ is time computation step [14]. $\tilde{G}_1$ and $\tilde{G}_2$ are the resolution in time domain and frequency domain which can be increased respectively. In execution, the coefficients of BWT can be readily computed by corresponding coefficients of the CWT by:

$$BWT_x(a,\tau)=K(a,\tau)\cdot CWT_x(a,\tau) \quad (7)$$

Where K is a factor which depends on T. For the Morlet wavelet $\psi(t)=e^{-\left(\frac{t}{T_0}\right)^2}$ that is evenly employed as the mother function. K is given by:

$$K(a,\tau)=\frac{\int_{-\infty}^{+\infty}e^{-t^2}dt}{\sqrt{1+\left(T(a,\tau)/T_0\right)^2}} \quad (8)$$

This is roughly equal to:

$$1.7725\Big/\sqrt{1+\left(T(a,\tau)/T_0\right)^2}$$

In our work, we use $\tilde{G}_1=0.87$, $\tilde{G}_2=45$ $BWT_s=0.8$ and $T_0=0.0005$ that are the same values as in the reference [8-9].Finally, the computation step $\Delta\tau$ is selected to be equal to, $1/f_s$ where f_s is the sampling frequency.

II.2. MEL FREQUENCY CEPSTRAL COEFFICIENT

The MFCC is the most widely used feature parameters for speech recognition. Mel Frequency Cepstral Coefficient method is well known for its performance and relative simplicity. The calculation process of the coefficients can be described as follows [10]:

Step 1: employ Fast Fourier Transform (FFT) to get power spectrum of the speech signal.
Step 2: apply a Mel-space filter-bank to the power spectrum to obtain logarithmic energy value.
Step 3: conduct the discrete cosine transform (DCT) of log filter-bank energies to get MFCC.

III. VECTOR QUANTIZATION

To Feature extraction gives result as a series of vectors characteristic of the time-varying spectral properties of the speech signal. The spectral analysis has significantly reduced the required information rate. It gives the discrete representation of speech sounds. Among the various methods of designing the VQ codebook, binary split algorithm is implemented by the following procedure:

- Conceive a 1-vector codebook; this is the centroïde for the entire set of training vectors.
- Duplicate the size of the codebook by splitting the present codebook.
- Employ the K-means iterative algorithm to obtain the best set of centroids for the divided codebook.
- Repeat second and third steps until a codebook of size M is designed.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

The output of VQ is the index of the codebook vector nearest to the input vector.

IV. HIDDEN MARKOV MODEL

Hidden Markov Model (HMM) is currently the most popular approach to speech recognition. An HMM is characterized by three matrices namely A, B and π .So, the compact notation

λ = {A, B, π } represents a complete parameter set of the model.

Where:

- A= {aij} represents the state transition probability matrix, (NxN) ,
- B ={bij(k)} represents the matrix of observation probability, (NxM),
- $\pi = \{\pi_i\}$ initial state distribution vector(Nx1) ,
- N is the states number in HMM,
- M is the of distinct observations number of symbols per state

To do isolated word recognition the following steps are performed.

1. For each word V in the vocabulary, an HMM model LV must be build: a model parameters (A, B, π) are to be estimated which optimized the likelihood of the training set observations for the V the word.

2. For each unknown word to be recognized, the procedure proven in figure 1 must be effected, that is, measurement of the observation sequence O= (o1, o2,……., oT), through feature analysis of the speech that correspond to the word; followed by calculation of model likelihoods to all possible models P (O/ LV), followed by choice of the word whose model probability is highest – that is.[11]

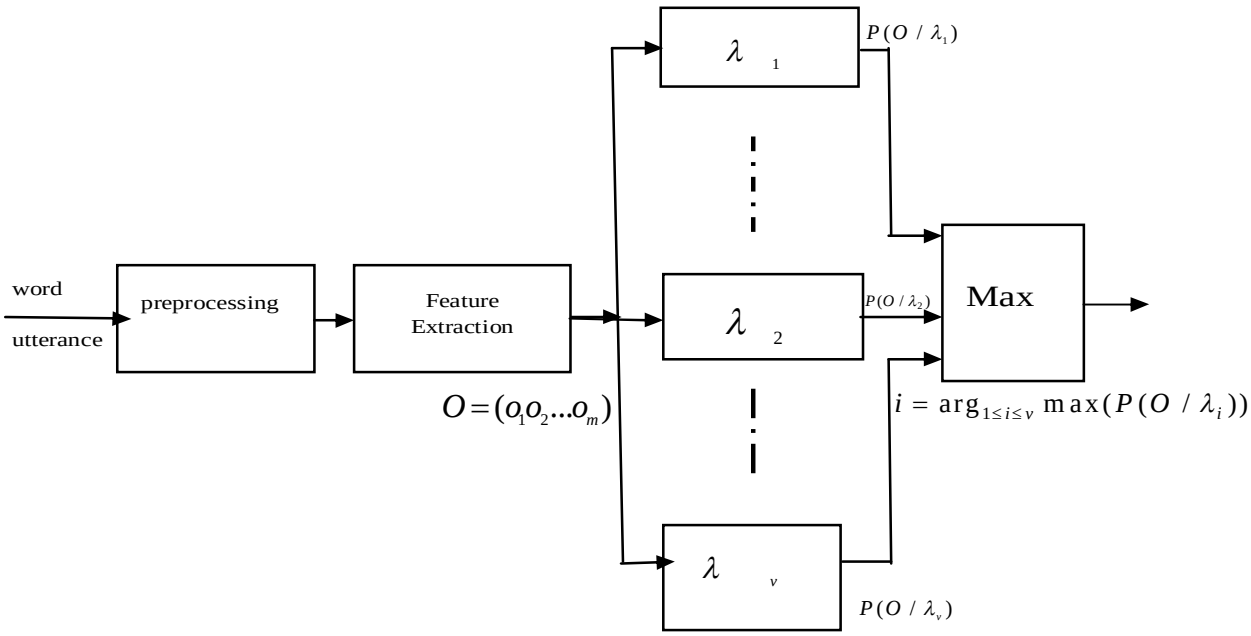

Fig 1: The overall block diagram of an automatic speech recognition system.

For each unknown word to be recognized we calculated the model likelihood for all possible models, and select the model with the highest likelihood.

However, for the moment we still have problem for the selection of proper HMM parameters such as the number of coefficient(the number of MFCC and the number of BWT scales), the number of discrete symbols (or equivalently codebook size) and the number of states to set up the HMM

In the following sections, we will present a Genetic Algorithm method GA for HMM training. By using the global searching capability of GA, we can find the optimal number of coefficients, the optimal number of states and the optimal size of codebook.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

V. THE GA FOR HMM OPTIMIZATION

The genetic algorithm (GA) [10] is a method for solving optimization problems. GA which is based on natural selection can solve a variety of optimization problems such as HMM optimization parameters. Figure 1 shows the flowchart of the genetic algorithm method applied to determine the optimal HMM parameters and to improve the recognition performance.

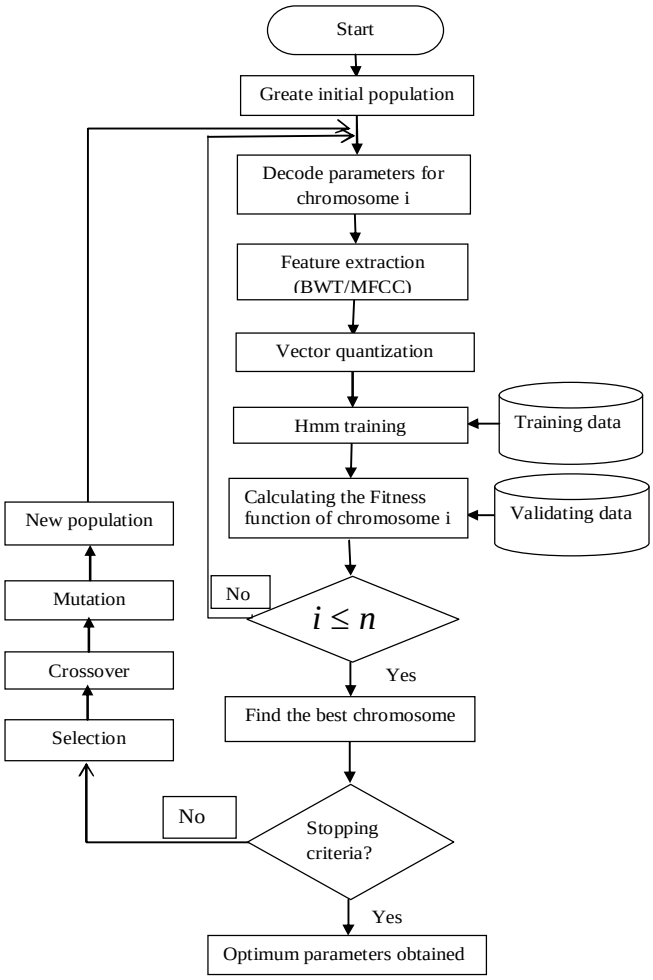

Fig 2: Flowchart of the proposed algorithm

This figure (Fig 2) describes our system by using the GA to optimize HMM parameters and HMM training. All the parameters of the HMM are encoded form a long chromosome and tuned by the GA. Then, as a result of the GA process, the BP algorithm is used to train the network.

The different steps of the adapted GA for optimizing MLP are:

1. Create Initial population of chromosomes

Regarding the genetic algorithm for HMM optimization, we used a population which is composed of n chromosomes to represent the relevant information of the HMM. An HMM can be represented by a direct graph, encode on a chromosome with each parameter (MFCC, BWT, number of states of HMM and codebook size). All these parameters are memorized by a row vector C. First, we have 4 bits for representing MFCCs coefficients. Second, we have 3 bits for representing BWT scales. Third, we have 3 bits for representing the number of states of HMM and finally we have 7 bits for representing the size of codebook. So each chromosome is encoded in 17bit.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

2.Decode parameters for chromosome indice i

- Feature extraction using MFCC and BWT
- Vector quantization
- HMM Training
- Evaluate the Fitness function according to chromosome i. The higher fitness value reflects the chances of the chromosome to be selected in the following generation. The log likelihood represents the probability that the training observation speech have been generated by the present model parameters. It can be expressed as:

$$P_n = (\sum_{k=1}^{M} \log(P(O_k / \lambda_n))) / M \quad (9)$$

- Repeat step 2 until the end of the size of population

3. Find the best chromosome

4. GA stop if generations number is Maximum or fitness =0, then select the optimal parameters.

Otherwise GA repeats the following steps.

- Selection good chromosomes : this operation consists in selecting chromosomes from the population for reproduction based on the relative fitness value of each chromosome. It can be performed by Elitist selection function: The best chromosomes of every generation are warranted to be chosen.
- Crossover: to apply the crossover operator, two chromosomes are randomly selected from the population. Then the two chromosomes are chopped into two parts at the crossover point and they exchange their parts.
- Mutation randomly alters each gene with a small probability, typically less than 10%. This operator introduces innovation into the population and helps prevent premature convergence on a local maximum. It exist many types of mutation, in this work, we used a Flip Bit in which a mutation operator that simply inverts the value of the chosen gene (0 goes to 1 and 1 goes to 0).
- Take the place of the current population with the new population.

Go to 2 the step.

VI. IMPLEMENTATION AND RESULTS

The idea in this work is to generate a speech recognizer for Arabic isolated words by implementing genetic algorithm (GA) with HMM to optimize HMM parameters and to improve the recognition performance. The database used for our approach was an Arabic isolated word spoken by a mono-speaker. Each word was uttered a twenty-five times. Ten occurrences of isolated words were reserved for training purposes and fifteen occurrences were reserved for testing purposes.

The table 1 presents the arabic words used for evaluation the proposed technical .

Table 1: The used vocabulary

Pronunciation	Arabic Writing	English Writing
Khalfa	خلف	backward
Amam	أمام	Forward
Asraa	أسرع	accelerate
Sir	سر	Walk
Istader	إستدر	Turn
Takadam	تقدم	Proceed
Tarajaa	تراجع	Push back
Tawakaf	توقف	Stop
Yamine	يمين	Right
Yassare	يسار	Left
Waraa	وراء	Back

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

The table below reports all the values of the parameters used in the proposed speech recognition system.

Table 2: Parameters used for evolutionary design of HMM

Parameters	**Value**
Coding	Binary
Population size	40
Size of chromosome	18
Maximum number of generations	50
Coefficients MFCC	5-20
Scales BWT	1-7
Fitness Function	100-recognition rate
Selection Technique	Elitist selection.
Probability of selection	0.5
Method of mutation	Random.
Mutation probability	pmut = 1/ Size of chromosome
Method of crossing	One point crossover
Probability of crossover	pcross =0.5
Stopping criterion	Maximum number of generations or fitness =0

- **Results using HMM with GA**

When optimizing the HMM parameters with the proposed architecture in Fig.2 we obtain the results seen in Table 3.

Table 3. reported the obtained results from the different techniques.

Table 3: Recognition rates

Recognition system	**Recognition rate**
MFCC+ MMC	89.7%
MFCC+ΔΔMFCC +MMC	91.5%
(CWT/MFCC+ΔΔMFCC)+MMC	88.4%
(BWT /MFCC+ΔΔMFCC)+MMC	94.4%
[(BWT /MFCC+ΔΔMFCC)+MMC]+AG	96.96%

The values of the GA parameters used in our experiments are given in Table 2. Through these experiments, we found that using the GA and HMM for recognition leads to an important improvement in the accuracy of the recognition rate. HMM is the most amply approach used for speech recognition but any arbitrary estimate of the initial HMM parameters involves difficulties. The performance of the HMM is often related to its parameters. However, the choice of its parameters is very difficult. We can't randomly select the number of coefficient (the number of MFCC and the number of BWT scales), the number of discrete symbols (or equivalently codebook size) and the number of states to set up the HMM. In case of recognition of isolated words as our case, the recognition tests revealed beyond the influence of each parameter variation on the recognition rate and training time. Indeed, the choice of optimum value for each parameter is to maximize the recognition rate and minimizes the prohibitive learning time. Experimental results showed the influence of the variation of each model parameter (the number of state, the dictionary size and the number of coefficients) on the recognition rate of the system. Therefore, the GA is proposed to optimize the HMM parameters because it is worthwhile noting that finding its optimized model parameter is difficult to estimate randomly. So, the GA approach is a good choice for HMM optimization and speed up the convergence time significantly. It's widely known

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

that HMM is the most popular approach for speech recognition but the slow convergence speed of the training HMM and HMM parameters optimization are the great interesting problem. As results, the GA not only overcomes the shortcoming of the slow convergence speed of the training HMM but also helps HMM parameters optimization. The proposed system shown in Table 3 is the best one for the specific problem of Speech Recognition. We can see that HMM trained by GA (with one-point crossover) is the best and HMM plays a key role in recognition performance.
Moreover, from Table 3, it can be seen that the recognition rate increased up to 96.59.

VII.CONCLUSION

Thus In this paper, we describe the utilization of GA and HMM for the problem of Arabic speech recognition.GA is an effective solution used to optimize HMM parameters. Any arbitrary estimate of the initial model parameters such as the following: the number of coefficient (the number of MFCC and the number of BWT scales), the number of discrete symbols (or equivalently codebook size) and the number of states will usually lead to a sub optimal model in practice. Moreover, when the complexity of the problem domain increases and when optimized HMM parameters are desired, manual searching becomes unmanageable and quite difficult.

This paper presents a method based on optimization techniques (GA), which optimizes the HMM parameters and performance. Integrating the GA with Monrovian model can improve speech recognition rate. It can be increasing up to 96.59%. However, speech recognition rate still has room for improvement, where much effort is needed to improve GA method for accelerating the learning process in HMM model and the convergence time significantly. The results were quite encouraging but we realize that our database was mono locator and of limited vocabulary. Our goal is to ultimately design a HMM which would be able to recognize continuous speech recognition on a larger vocabulary. The work is being continued for connected speech recognition

REFERENCES

[1] V. Skorpil and J. Stastny, " Back-Propagation and K-Means Algorithms Comparison", in Proc. of 8 IEEE International Conference on Signal Processing, pp. 16-20, 2006.
[2] Z. Sakka, A. Kachouri and M. Samet "Speech Denoising and Arabic Speaker Recognition System Using Subband Approach", International Review on Computers and Software (IRECOS), Vol. 2. n. 3, pp. 264 – 271, 2007.
[3] Peter Jančovič, Münevver Köküer" Incorporating the voicing information into HMM-based automatic speech recognition in noisy environments" Speech Commu., Volume 51, Issue 5, pp 438–451, 2009.
[4] Lawrence Rabiner, "Tutorial on Hidden Markov Models & selected Applications in speech recognition", Proc. IEEE,
[5] Jelinek F. "Statistical methods for speech recognition",Language, speech and communication: a Bradford book. The MIT Press, Cambridge, MA, January 1998.
[6] S. Kwong, C.W. Chau, K.F. Man, K.T. Ng, "GA-HMM Training for speech recognition", IEEE Int. Conf. on Intelligent Processing System, Australia, pp.502-505, August 1998.
[7] Kwong, S., Chau, C.W., Man, K.F., Tang, K.S., "Optimisation of HMM topology and its model parameters by genetic algorithm", Pattern Recognition, pp 509–522, 2001.
[8] [X. Yuan, "Auditory Model-Based Bionic Wavelet Transform for Speech Enhancement", Master's thesis, Marquette University, Milwaukee, WI, USA, 2003.
[9] J. Yao and Y. T. Zhang, "The application of bionic wavelet transform to speech signal processing in cochlear implants using neural network simulations, IEEE Transactions on Biomedical Engineering, vol.49, no.11, pp.1299-1309, 2002.
[10] M. Ben Nasr, S. Saoud, A. Cherif "Optimization of MLP using Genetic Algorithms applied to Arabic speech recognition", International Review on Computers and Software (I.RE.CO.S.), Vol. 8, n.2, 2013
[11] Negin Najkara,, Farbod Razzazia , Hossein Sametib "A novel approach to HMM-based speech recognition systems using particle swarm optimization Mathematical and Computer Modelling,pp1910-1920, 2010.