International Journal of Digital Information and Wireless Communications (IJDIWC) 1(1): 14-31
The Society of Digital Information and Wireless Communications, 2011 (ISSN 2225-658X)

CONNECTIONIST PROBABILITY ESTIMATORS IN HMM USING GENETIC CLUSTERING APPLICATION FOR SPEECH RECOGNITION AND MEDICAL DIAGNOSIS

Lilia Lazli[1], Boukadoum Mounir[2], Abdennasser Chebira[3], Kurosh Madani[3] and Mohamed Tayeb Laskri[1]
[1] Laboratory of research in Computer Science (LRI/GRIA), Badji Mokhar University, B.P.12 Sidi Amar 23000 Annaba – Algeria.
l_lazli@yahoo.fr, laskri@univ-annaba.org, http://www.univ-annaba.org
[2]Univesité du Québec A Montréal (UQAM), Canada
Boukadoum-mounir@uqam.ca
[3] Images, Signals and Intelligent Systems Laboratory (LISSI / EA 3956)
PARIS XII University, Senart-Fontainebleau Institute of Technology,
Bat.A, Av. Pierre Point, F-77127 Lieusaint, France,
{achebira, kmadani}@univ-paris12.fr, http://www.univ-paris12.fr

ABSTRACT

The main goal of this paper is to compare the performance which can be achieved by five different approaches analyzing their applications' potentiality on real world paradigms. We compare the performance obtained with (1) Multi-network RBF/LVQ structure (2) Discrete Hidden Markov Models (HMM) (3) Hybrid HMM/MLP system using a Multi Layer-Perceptron (MLP) to estimate the HMM emission probabilities and using the K-means algorithm for pattern clustering (4) Hybrid HMM-MLP system using the Fuzzy C-Means (FCM) algorithm for fuzzy pattern clustering and (5) Hybrid HMM-MLP system using the Genetic Algorithm (AG) for genetic clustering. Experimental results on Arabic speech vocabulary and biomedical signals show significant decreases in error rates of hybrid HMM/MLP/AG pattern recognition in comparison to those of other research experiments by integrating three types of features (PLP, log-RASTA PLP, J-RASTA PLP) were used to test the robustness of our hybrid recognizer in the presence of convolution and additive noise.

KEYWORDS

Speech and medical recognition, J-RASTA PLP, fuzzy clustering, Genetic Algorithm, HMM/MLP models.

1 INTRODUCTION

In many target (or pattern) classification problems the availability of multiple looks at an object can substantially improve robustness and reliability in decision making. The use of several aspects is motivated by the difficulty in distinguishing between different classes from a single view at an object [9]. It occurs frequently that returns from two different objects at certain orientations are so similar that they may easily be confused. Consequently, a more reliable decision about the presence and type of an object can be made based upon observations of the received signals or patterns at multiple aspect angles. This allows for more information to accumulate about the size, shape, composition and orientation of the objects, which in turn yields more accurate discrimination. Moreover, when the feature space undergoes changes, owing to different operating and environmental conditions, multi-aspect classification is almost a necessity in order to maintain the

performance of the pattern recognition system.

In this paper we present the Hidden Markov Model (HMM) and apply them to complex pattern recognition problem. We attempt to illustrate some applications of the theory of HMM to real problems to match complex patterns problems as those related to biomedical diagnosis or those linked to social behavior modeling. We introduce the theory and the foundation of Hidden Markov Models (HMM). In the pattern recognition domain, and particularly in speech recognition, HMM techniques hold an important place [7]. There are two reasons why the HMM exists. First the models are very rich in mathematical structure and hence can form the theoretical basis for use in a wide range of applications. Second the models, when applied properly, work very well in practice for several important applications. However, standard HMM require the assumption that adjacent feature vectors are statistically independent and identically distributed. These assumptions can be relaxed by introducing Neural Network (NN) in the HMM framework.

Significant advances have been made in recent years in the area of speaker independent speech recognition. Over the last few years, connectionist models, and Multi Layer-Perceptron (MLP) in particular, have been widely studied as potentially powerful approaches to speech recognition. These neural networks estimate the posterior probabilities used by the HMM. Among these, the hybrid approach using the MLP to estimate HMM emission probabilities has recently been shown to be particularly efficient by example for French speech [1] and American English speech [14].

We then propose a hybrid HMM/MLP model for speech recognition and biomedical diagnosis which makes it possible to join the discriminating capacities, resistance to the noise of MLP and the flexibilities of HMMs in order to obtain better performances than traditional HMM.

We develop also, a method based on concepts of fuzzy logic for clustering and classification of vectors : Fuzzy C-Means (FCM) algorithm, and demonstrate its effectiveness with regard to K-Means (KM) traditional algorithm, owing to the fact that the KM algorithm provides a hard decision (not probabilitized) in connection with the observations in the HMM states.

Regarding the other algorithm of clustering, we specifically, methods involving a supervised clustering by partition and we chose one as the basis for our work. This solution is to make the choice of a measure that we use in our application. This algorithm is a "good" score in a test that measures the quality of a partition. We are thus reduced to an optimization problem. The properties of this algorithm does not guarantee convergence to a global optimum, that is why we are interested in a type heuristics genetic algorithms (GA), less likely to be trapped by the minimum local and now widely used in optimization problems.

Genetic Algorithms (GA) are search procedures based on the mechanics of genetics and natural selection. Numerous techniques and algorithms have been developed to assist the engineer in the creation of optimum development strategies. These procedures quickly converge to optimal solutions after examining only a small fraction of the search space and have been successfully applied to complex engineering optimisation problems.

2 HMM Advantages and Drawbacks

Standard HMM procedures, as defined above, have been very useful for pattern recognition; HMM can deal efficiently with the temporal aspect of pattern (including temporal distortion or time warping) as well as with frequency distortion. There are powerful training and decoding algorithms that permit efficient training on very large databases, and for example recognition of isolated words as well as continuous speech. Given their flexible topology, HMM can easily be extended to include phonological rules (e.g., building word models from phone models) or syntactic rules. For training, only a lexical transcription is necessary (assuming a dictionary of phonological models); explicit segmentation of the training material is not required. An example of a sample HMM is given in figure. 1; this could be the model of a sort assumed to be composed of three stationary states.

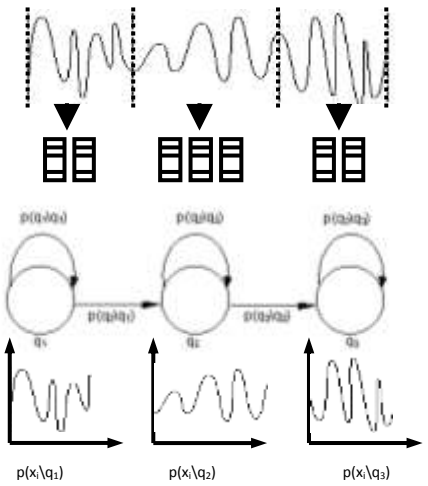

Figure 1. Example of a three-state Hidden Markov Models Where 1) $\{q_1, q_2, q_3\}$: the HMM states. 2) $p(q_i\backslash q_j)$: the transition probability of state q_i to state q_j (i, j = 1..3). 3) $p(x_i\backslash q_j)$: the emission probability of observation x_i from the state q_j (i = 1...number of trames, j = 1..3).

However, the assumptions that permit HMM optimization and improve their efficiency also, practice, limit their generality. As a consequence, although the theory of HMM can accommodate significant extensions (e.g., correlation of acoustic vectors, discriminate training,…), practical considerations such as number of parameters and train-ability limit their implementations to simple systems usually suffering from several drawbacks [2] including:

- Poor discrimination due to training algorithms that maximizes likelihoods instead of *a posteriori* probabilities (i.e., the HMM associated with each pattern unit is trained independently of the other models). Discriminate learning algorithms do exist for HMM but in general they have not scaled well to large problems.
- A priori choice of model topology and statistical distributions, e.g., assuming that the Probability Density Functions (PDF) associated as multivariate Gaussian densities or mixtures of multivariate Gaussian densities, each with a diagonal only covariance matrix (i.e., possible correlation between the components of the acoustic vectors is disregarded).
- Assumption that the state sequences are first-order Markov chains.
- Typically, very limited acoustical context is used, so that possible correlation between successive acoustic vectors is not modeled very well.

Much Artificial Neural Network (ANN) based Automatic pattern recognition research has been motivated by these problems.

3 ESTIMATING HMM LIKELIHOODS WITH ANN

ANN can be used to classify pattern classes such as units. For statistical

recognition systems, the role of the local estimator is to approximate probabilities or PDF. Practically, given the basic HMM equations, we would like to estimate something like $p(x_n \backslash q_k)$, is the value of the probability density function of the observed data vector given the hypothesized HMM state. The ANN in particular the MLP can be trained to produce the posterior probability $p(q_k \backslash x_n)$ of the HMM state give the acoustic data. This can be converted to emission PDF values using Bayes'rule.

Several authors [1-4, 7-10] have shown for speech recognition that ANN can be trained to estimate *a posteriori* probabilities of output classes conditioned on the input pattern. Recently, this property has been successfully used in HMM systems, referred to as hybrid HMM/ANN systems, in which ANN are trained to estimate local probabilities $p(q_k \backslash x_n)$ of HMM states given the acoustic data.

Since the network outputs approximate Bayesian probabilities, $g_k(x_n, \Theta)$ is an estimate of:

$$p(q_k \setminus x_n) = \frac{p(x_n \setminus q_k)\, p(q_k)}{p(x_n)} \qquad (1)$$

which implicitly contains the *a priori* class probability $p(q_k)$. It is thus possible to vary the class priors during classification without retraining, since these probabilities occur only as multiplicative terms in producing the network outputs. As a result, class probabilities can be adjusted during use of a classifier to compensate for training data with class probabilities that are not representative of actual use or test conditions [18], [19].

Thus, scaled likelihoods $p(x_n \backslash q_k)$ for use as emission probabilities in standard HMM can be obtained by dividing the network outputs $g_k(x_n)$ by the training set, which gives us an estimate of :

$$\frac{p(x_n \setminus q_k)}{p(x_n)} \qquad (2)$$

During recognition, the seating factor $p(x_n)$ is a constant for all classes and will not change the classification. It could be argued that, when dividing by the priors, we are using a scaled likelihood, which is no longer a discriminate criterion. However, this needs not the true, since the discriminate training has affected the parametric optimization for the system that is used during recognition. Thus, this permit uses of the standard HMM formalism, while taking advantage of ANN characteristics.

The figure2 shows that an equilibrium that is reached in the ideal case does in fact correspond to the network output g being equal to posterior probability p. Consider an area in feature space around a training pattern x_n, and assume that we consider only two classes: the target class for this pattern, and the class of all patterns that do not belong to the first class. Vectors in the selected area will undergo an upward "force" corresponding to the gradient of the error term for all patterns with a target of 1; the quadratic error term in this case is $\frac{1}{2}(1-g)^2$, with a derivative of $(1-g)$, and this case will occur a fraction of the time given by probability p. Therefore, the upward force (average gradient term) applied to the region is equal to $p(1-g)$. The downward force is similarly defined, and will balance for the equilibrated case. This will only occur when the network output is equal to the posterior probability.

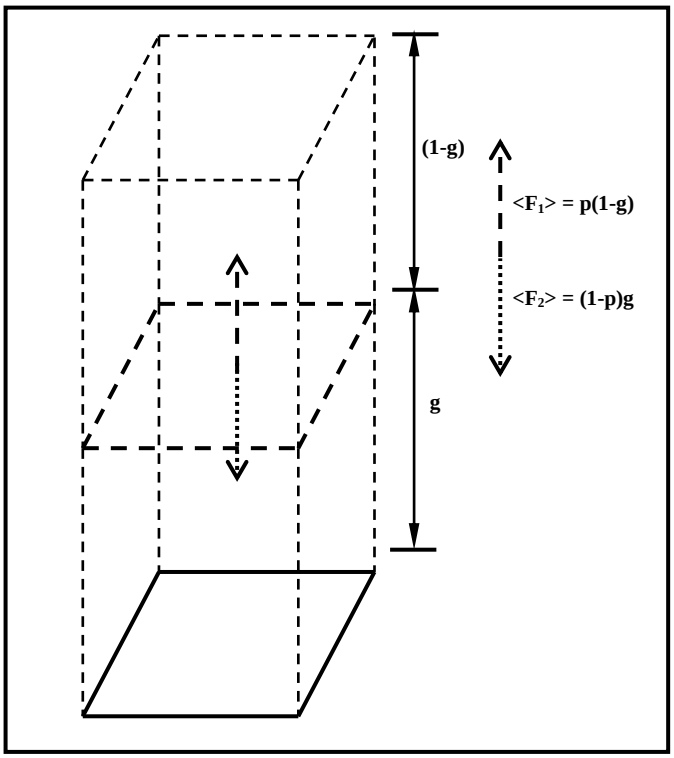

Figure 2. Geometric illustration of proof that ANN will (ideally) produce posterior probabilities. Equilibrium is reached when the upward force, (due to the average of the ANN output from 1 for a target that should be high), is equal to the downward force (due to the average deviation from 0 for a target that should be low).

4 WHY THIS IS GOOD ?

Since we ultimately derive essentially the same probability with a MLP as we would with a conventional (e.g., Gaussian mixture) estimator, what is the point? There are at least two potential advantages that we and others have observed [4, 9, 10, 14]:

1) The standard statistical recognizers require strong assumptions about the statistical character of the input, such as parameterizing the input densities as mixtures of Gaussian densities with no correlation between features, or as the product of discrete densities for different features that are assumed to be statistically independent. This type of assumption is not required with a MLP estimator, which will be an advantage particularly when a mixture of feature types are used, e.g., binary and continuous. Specifically, standard HMM approaches require the assumption that successive acoustic vectors are uncorrelated. For the MLP estimator, multiple inputs can be used from a range of time steps, and the network will learn something about the correlation between the acoustic inputs. Note that the use of such a network will lead to a more general emession probability that may also be used in the global decoding. That is, if $c + d + 1$ frames of acoustic vectors $X_{n-c}^{n+d} = \{x_{n-c},...,x_n,...,x_{n+d}\}$ are used as input to provide contextual information to the network, the output values of the MLP will estimate $p(q_k/X_{n-c}^{n+d}),\ \forall k = 1,..., K.$ This provides a simple mechanism for incorporating acoustic context into the statistical formulation.

2) MLP are a good match to discriminative objective functions (e.g., Mean Squared Error: MSE), and so the probabilities will be optimized to maximize discrimination between sound classes, rather than to most closely match the distributions within each class. It can be argued that such a training is conservative of parameters, since the parameters of the estimate are trained to split the space between classes, rather than to represent the volumes that compose the division of space constituting each class.

5 J-RASTA PLP ANALYSIS

Developing speech recognizer that are robust under realistic acoustic environment has been and still is a major problem for the speech community. Speech distortion can be categorized into two types. There is linear spectral distortion (i.e., convolutional noise), which is typically introduced by microphones and the telephone channel. There is also

additive noise such as stationary or slowly varying background noise.

Three types of features (PLP, Log-RASTA PLP and J-RASTA PLP) were used to test the robustness of our hybrid recognizer in the presence of convolution and additive noise. Perceptual Linear Predictive (PLP) analysis is an extension of linear predictive analysis that takes into account some aspects of human sound perception [20, 21]. Log-RASTA PLP (RelActive SpecTrAl processing) is based on PLP but also aims at reducing the effect of linear spectral distortion. In addidtion of PLP and log-RASTA PLP analysis already used in our past works [22], we used her J-RASTA PLP tries to handle both linear spectral distortion and additive noise simultaneously [23].

J-RASTA-PLP [23] is used as the acoustic pre-processor for both the databases experiments. Each frame of the feature vector represents 25 ms of speech, with 12.5 ms overlap of consecutive frames. J-RASTA PLP was chosen for its robustness to linear spectral distortions in speech signals that are often introduced by communication channels.

J-RASTA-PLP attempts to make recognition more robust and invariant to acoustic environment variables. It in some sense performs speech enhancement but for the purposes of feature extraction. The speech enhancement process added was designed to improve intelligibility of noisy speech for human listeners. Addition of this enhancement may improve recognition scores further.

Speech recordings were sampled over the microphone at 11 kHz. After pre-emphasis (factor 0.95) and application of a Hamming windows, The J-RASTA PLP features were computed every 10 ms on analysis windows of 30 ms.

Each frame is represented by 12 components plus energy (J-RASTA PLP + E). The values of the 13 coefficients are standardized by their standard deviation measured on the frames of training. The features set for our hybrid HMM/MLP system was based on a 26 dimensional vector composed of the cepstral parameters (J-RASTA PLP parameters), the Δ cepstral parameters, the Δ energy and the ΔΔ energy.

RASTA was configured to incorporate high-pass filtering and slight spectral subtraction. A constant J of 1e-6 was used for training. Multiple regressions J mapping was used during testing.

Nine frames of contextual information was used at the input of the MLP (9 frames of context being known as yielding usually the best recognition performance).

6 CLUSTERING PROCEDURE

One of the basic problems that aries in a great variety of fields, including pattern recognition, machine learning and statistics, is the so-called clustering problem [5,13,15,22]. The fundamental data clustering problem may be defined as discovering groups in data or grouping similar objects together. Each of these groups is called a cluster, a region in which the density of objects is locally higher than in other regions.

In this paper, data clustering is viewed as a data partitioning problem. Several approaches to find groups in a given database have been developed [5, 13, 15], but we focus on the K-Means (KM) algorithm as it is one of the most used iterative partitional clustering algorithms and because it may also be used to initialize more expensive clustering algorithms (e.g., the EM algorithm).

6.1 Drawbacks of the K-Means Algorithm

Despite being used in a wide array of applications, the KM algorithm is not exempt from drawbacks. Some of these drawbacks have been extensively reported in literature [5, 13, 15, 22]. The most important are listed below:

- As many clustering methods, the KM algorithm assumes that the number of clusters *k* in the database is known beforehand which, obviously, is not necessarily true in real-world applications.
- As an iterative technique, the KM algorithm is especially sensitive to initial starting conditions (initial clusters and instance order).
- The KM algorithm converges finitely to a local minimum. The running of the algorithm defines a determinist mapping from the initial solution to the final one.
- The vectorial quantization and in particular, the KM algorithm provides a hard and fixed decision not probabilitized which does not transmit enough information on the real observations.

Many clustering algorithms based on fuzzy logic concepts have been motivated for this last problem.

6.2 Fuzzy C-Means Algorithm

In general, and by example for speech context a purely acoustic segmentation of the speech cannot suitably detect the basic units of the vocal signal. One of the causes is that the borders between these units are not acoustically defined. For this reason, we were interested to use the automatic classification methods which are based on fuzzy logic in order to segment the data vectors. Among the adapted algorithms, we have chosen the FCM algorithm which has already been successfully used in various fields and especially in the image processing [13], [15].

FCM algorithm is a method of clustering which allows one piece of data to belong to two or more clusters. The use of the measurement data is used in order to take not of pattern data by considering in spectral domain only. However, this method is applied for searching some general regularity in the collocation of patterns focused on finding a certain class of geometrical shapes favored by the particular objective function [5]. The FCM algorithm is based on minimization of the following objective function, with respect to *U*, a fuzzy c-partition of the data set, and to *V*, a set of *K* prototypes [5]:

$$J_m(U,V) = \sum_{j=1}^{m}\sum_{i=1}^{c} u_{ij}^{m} \left\| X_j - V_i \right\|^2 \qquad (3)$$
$$1 \le 1 \prec \infty$$

Where *m* is any real number greater than 1, u_{ij} is the degree of membership of x_j in the cluster *i*, x_j is the *j* th of d-dimensional measured data, V_i is the d-dimension center of the cluster, and ||*|| is the any norm expressed the similarity between any measured data and the center.

Fuzzy partition is carried out through an iterative optimization of (3) with the update of membership *u* and the cluster centers *V* by [5]:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{ik}}\right)^{2/m-1}} \qquad (4)$$

$$V_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} X_j}{\sum_{j=1}^{n} u_{ij}^{m}} \qquad (5)$$

The criteria in this iteration will stop when $\max_{ij}\left|u_{ij}-\hat{u}_{ij}\right| \prec \varepsilon$ where ε is a termination criterion between 0 and 1.

As a part of our main objective, we aim to find the best and the worst set of initial starting conditions to approach the extremes of the probability distributions of the square-error values. Due to the computational expense of performing an exhaustive search we tackle the problem using Genetic Algorithms.

6.3 Genetic Algorithms

Roughly speaking, we can say that Genetic Algorithms (GA) are kinds of *evolutionary algorithms*, that is, probabilistic search algorithms which simulate natural evolution [24, 25, 26]. GA are used to solve combinatorial optimization problems following the rules of natural selection and natural genetics. They are based upon the survival of the fittest among string structures together with a structured yet randomized information exchange. Working in this way and under certain conditions. GA evolve to the global optima with probability arbitrarily close to 1.

When dealing with GA, the search space of a problem is represented as a collection of *individuals*. The individuals are represented by character strings. Each individual is coding a solution to the problem. In addition, each individual has associated a fitness measure. The part of the space to be examined is called the *population*. The purpose of the use of a GA is to find the individual from the search space with the best "genetic material".

Figure 3 shows the pseudo-code of the GA that we use. First, the initial population is chosen and the fitness of each of its individuals is determined. In our case, the result of FCM clustering is considering as initial population. Next, in every iteration, two parents are selected from the population. This parental couple produces children which, with a probability near zero, are mutated, i.e., hereditary distinctions are changed. After the evaluation of the children, the worst individual of the population is replaced by the fittest of the children. This process is iterated until a convergence criterion is satisfied.

The operators which define the children production process and the mutation process are the *crossover* operator and the *mutation* operator respectively. Both operators are applied with different probabilities and play different roles in the GA. Mutation is needed to explore new areas of the search space and helps the algorithm avoid local optima. Crossover is aimed to increase the average quality of the population. By choosing adequate crossover and mutation operators as well as an appropriate reduction mechanism, the probability that the GA reaches a near-optimal solution in a reasonable number of iterations increases.

```
Begin GA
Choose initial population at random
Evaluate initial population
While not convergence criterion do
Begin
Select two parents from the current
population
Produce children by the selected parents
Mutate the children
Evaluate the children
Replace the worst individual of the population
by the best child
End
Output the best individual found
End GA.
```

Figure 3. The pseudo-code of the Genetic Algorithm

7 VALIDATION ON SPEECH AND BIOMEDICAL SIGNAL CLASSIFICATION PARADIGM

Further assume that for each class in the vocabulary we have a training set of k occurrences (instances) of each class where each instance of the categories constitutes an observation sequence. In order to build our tool, we perform the following operations.

1. For each class v in the vocabulary, we estimate the model parameters λ^v (A, B, π) that optimize the likelihood of the training set for the v^{th} category.

2. For each unknown category to be recognized, the processing of Fig. 4 is carried out: measurement of the observation sequence $O = \{o_1, o_2, \ldots, o_T\}$, via a feature analysis of the signal corresponding to the class; the computation of model likelihoods for all possible models, $P(O/\lambda^v)$, $1 \leq v \leq V$; at the end the selection of the category with the highest likelihood.

7.1 Speech Databases

Three speech databases have been used in this work:

1) The first one referred to as DB1, the isolated digits task has 13 words in the vocabulary: 1, 2, 3, 4, 5, 6, 7, 8, 9, zero, oh, yes, no. They are spoken by 30 speakers, producing a total of 3900 utterances (each word should be marked 10 times). The digits database has about 30,000 frames of training data. This first corpus consists of isolated digits collected over the microphone. This data base was used to conduct extensive testing's of the robustness of various signal processing front ends (PLP, log-RASTA PLP, J-RASTA PLP) in the presence of both convolution and additive noise.

2) The second database, referred to as DB2 contained about 50 speakers saying their last name, first name, the city of birth and the city of residence. Each word should be marked 10 times.

The used training set in the following experiments consists of 2000 sounds.

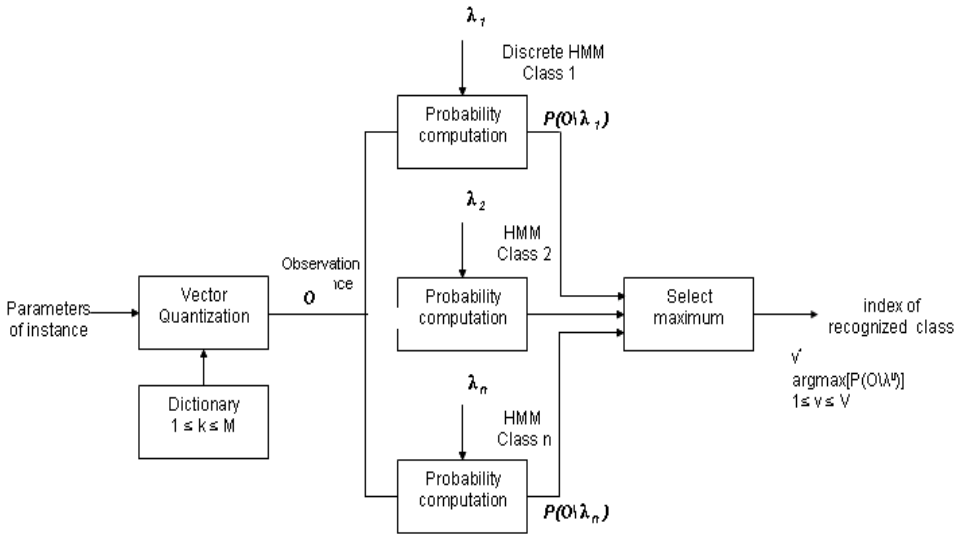

Figure 4. Block diagram of a speech and biomedical database HMM recognizer

3) The third database, referred to as DB3, contained the 13 control words (i.e. View/new, save/save as/save all) so that each speaker pronounces each control word 10 times. The used training set in the following experiments consists of 3900 sounds saying by 30 speakers.

For each databases, 12 speakers were used for training (a non-overlapping subset of these were used for cross-validation used to adapt the learning rate of the MLP), while the remaining 8 speakers were used for testing.

The acoustic feature were quantized into independent codebooks according to the FCM algorithms respectively:

- 128 clusters for the J-RASTA PLP vectors.
- 128 clusters for the first time derivative of cepstral vectors.
- 32 clusters for the first time derivative of energy.
- 32 clusters for the second time derivative of energy.

7.2 Biomedical Database

For task of biomedical database recognition using HMM/MLP model, the object of this survey is the classification of an electric signal coming from a medical test [16], experimental device is described in Fig. 5. The used signals are called Potentials Evoked Auditory (PEA), examples of PEA signals are illustrated in Fig. 6. Indeed, the exploration functional otoneurology possesses a technique permitting the objective survey of the nervous conduction along the auditory ways. The classification of the PEA is a first step in the development of a help tool to the diagnosis. The main difficulty of this classification resides in the resemblance of signals corresponding to different pathologies, but also in the disparity of the signals within a same class. The results of the medical test can be indeed different for two different measures for the same patient.

The PEA signals descended of the examination and their associated pathology are defined in a data base containing the files of 11185 patients.. We chose 3 categories of patients (3 classes) according to the type of their trouble. The categories of patients are:

1) Normal (N): the patients of this category have a normal audition (normal class).

2) Endocochlear (E): these patients suffer from disorders that touches the part of the ear situated before the cochlea (class Endocochlear).

3) Retrocochlear (R): these patients suffer from disorders that touches the part of the ear situated to the level of the cochlea or after the cochlea. (class retrocochlear).

We selected 213 signals (correspondents to patients). So that every process (signal) contains 128 parameters. 92 among the 213 signals belong to the N class, 83, to the class E and 38 to the class R. To construct our basis of training, we chose the signals corresponding to pathologies indicated like being certain by the physician. All PEA signals come from the same experimental system. In order to value the HMM/MLP realized system and for ends of performance comparison, we forced ourselves to apply the same conditions already respected in the work of the group describes in the following article [17] and that uses a multi-network structure heterogeneous set to basis of RBF and LVQ networks and the work describes in the following articles [11], [12] and that uses a discrete HMM. For this reason, the basis of training contains 24 signals, of which 11 correspondent to the class R, 6 to the class E and 7 to the N class.

After the phase of training, when the non learned signals are presented to the HMM, the corresponding class must be designated. For the case in which we wish to use an HMM with a discrete observation symbol density, rather than continuous vectors above, a quantized vector VQ is required to map each continuous observation vector into a discrete codebook index. Once the codebook of vectors has been obtained, the mapping between continuous vectors and codebook indices becomes a simple nearest neighbor computation, i.e., the continuous vector is assigned the index of the nearest codebook vector. Thus the major issue in VQ is the design of an appropriate codebook for quantization.

Codebook sizes of from M = 64 vectors have been used in biomedical database recognition experiments using HMM/MLP model.

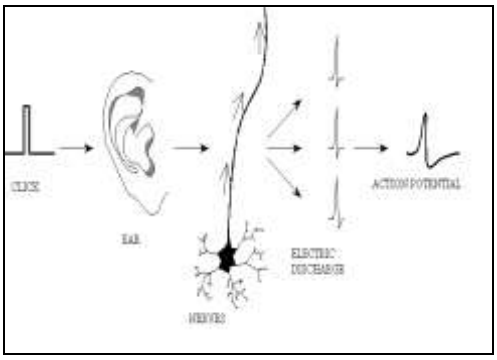

Figure 5. PEA generation experiment

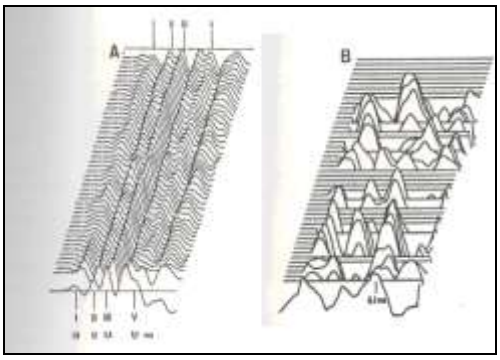

Figure 6. (A) PEA signal for normal patient, (B) Patient with auditory disorder

7.3 Discrete MLP with entries provided by the FCM algorithm

For this case, we compare the performance of the basis hybrid model with that of an hybrid HMM/MLP model using in entry of the network an acoustic vector composed of real values which were obtained by applying the FCM algorithm with 2880 real components corresponding to the various membership degrees of the acoustic vectors to the classes of the "code-book ". We presented each cepstral parameter (J-RASTA PLP, Δ J-RASTA PLP, ΔE, ΔΔE) by a real vector which the components definite the membership degrees of the parameter to the various classes of the "code-book ". For speech databases, for example, we reported the values used for DB2. 10-state, strictly left-to-right, word HMM with emission probabilities computed from an MLP with 9 frames of quantizied acoustic vectors at the input, i.e., the current acoustic vector preceded by the 4 acoustic vectors on the left context and followed by the 4 acoustic vectors on the right context. the entry layer is made up of a real vector with 2880 real components corresponding to the various membership degrees of the acoustic vectors to the classes of the "code-book ".

Thus a MLP with only one hidden layer including 2880 neurons at the entry, 30 neurons for the hidden layer and 10 output neurons was trained. The three-layer MLP is trained by using stochastic gradient descent, and relative entropy as the error criterion. A sigmoid function is applied to the hidden layer units, and softmax (exponential of the unit's weighted sum normalized by the sum of exponentials for the entire layer) is used as the output nonlinearity. The number of neurons of the hidden layer was choosen empirically.

For the PEA signals, a MLP with 64 neurons at the entry, 18 neurons for the hidden layer and 5 output neurons was trained.

7.4 Discrete MLP with entries provided by the GA algorithm

In the case of application of GA, the initial population is chosen and the *fitness* of each of its individuals is determined. In our case, the result of FCM clustering is considering as initial population.

For clarification, we report after, some results of the clusterning of DB1

7.4.1 *Coding of individuals*

The key idea is that the encoding of an individual must enable sampling "effective" in the search space. We chose a representation based on indexing of classes. An individual is therefore encoded by a chain of integers whose ordinal value (= index) represents the number of the object and the cardinal value corresponds to the number of the class to which the object belongs.

Practical consideration**:** The following partition of objects (acoustic vectors) in 10 classes (10-digit).
class 0 = {V_1,V_2,…..V_{21},V_{22}},
class 1 = {V_{23},V_{24},...V_{34},V_{35}},
class 2 = {V_{36},V_{37},...$V_?$…}, …,
class 9 = {…} will be represented by the following chromosome:

000000...	**111111…**	**…**	**888888…**	**999999…**

Figure 7. Example of coding of individuals

7.4.2 *Population size*

It is, a fundamental parameter of the GA, for which there is no general criteria for determining what optimum size should be. We tried to adapt the size of the population the size of the problem to treat.

7.4.3 *Merit function*

An individual or even genotype represents, let us recall, partitioning of *k* observations all classes. We use the function of following representation:

$$g_l = \frac{1}{|C_l|} \sum_{x_i \in C_l} x_i \qquad (6)$$

with g_l the centre of gravity of the class C_l, $l \in [1, M]$, M the number of classes (digit in our case), which allows to infer the M representatives. The test of intra-class criteria, measurement of the quality of a partition whose expression is given in (7) is our cost *objective function* , and we seek to minimize the value. This objective function transformation (criterion) led to the *function of merit* that we use in our algorithm to measure the quality of an individual.

$$w = \sum_{l=1}^{M} \sum_{x_i \in C_l} p_i \, d^2 (x_i , g_l) \qquad (7)$$

where p_i means the weight of the i^{th} object.

7.4.4 *Reproduction*

In this phase, is created on each iteration, a new population, by applying the genetic operators: selection, crossing and mutation. It is to select the most appropriate individuals in the population within the meaning of the function of *merit*, following a method, and reproduce them as what in the next generation.

In our GA we used called *"steady state genetic algorithm"* consisting of variant to replace that a certain percentage of the population each

generation (the size of the population remains constant over the course of the GA). The evolution of the population is ensured through operators in the *selection*, the *crossing* and *mutation* that allow to combine and modify chromosome. We summarize, used genetic operators.

- **Selection:** there are many selection strategies, we present just two briefly. In the selection of *wheel* type the probability of selection of an individual is proportional to its value of adaptation. The selection of type *tournament* is to select a subset of the population and retain the best individual of this sub-group. Seen that in our case, the problem is a minimization of the *fitness* problem. The best individual is one that has the smallest value of fitness

Practical consideration: According to *fitness* of each chromosome of the population, it selects the best individuals.

Table 1. Selection of the best individuals

Parents	fitness	Parents	fitness
1	6.4520	1	6.2424
2	6.8843	2	6.4520
3	7.3374	3	6.4746
4	6.2424	4	6.7076
5	6.7076	5	6.8331
6	7.1452	6	6.8843
7	6.4746	7	7.1452
8	6.8331	8	7.3374

- **Crossing:** it allows to create two individuals (children) by combining genes from both parents obtained in the previous step. This operator allows to increase the quality of a population average. We used the variant to a randomly chosen from possible Cup points crossing point (for a chainof length *l*there are $l - 1$). However, this solution may lead to a child non-*viable* (or even both), this is why it tests eventually to different cut-off points. Can optionally restrict this route by setting a priori, a number of iterations. If at the end of this audit, the child remains unsustainable, we take one of the parents. A child is said to be non-viable, if it does not have the same number of classes as these parents.

Practical consideration: Are the two following parents selected from our application.

Parent 1 Cut-of

00000000…	11110111…	...	33633333…	44744477…
55556655…	66665566…	…	88338822...	99999990…

Parent 2

00000000…	11111115...	…	33333333…	44444444...
55556565…	66666666…	…	88888833…	99999999...

Figure 8. Both parents

It gets 102 cut point. The two children obtained are the following:

Child 1

00000000…	11110111…	…	33633333…	44444444…
55556565…	66666666…	…	88888833…	99999999…

Child 2

00000000…	11111115…	…	33333333…	44744477…
55556655…	66665566…	…	88338822…	99999990…

Figure 9. Two children

In regards to the value of the probability of crossing, we followed the recommendations proposed by Goldberg [24], citing the work of *De jong*, offers a high probability of crossing ($\geq 0)5$).

- **Mutation:** this stage allows to introduce random variation in genotype of the individual (here the children). This operator allows therefore to explore new areas of research space, thus decreasing the risk of converge towards local minima.. Each *gene* is therefore likely to change in a given

probability, it is also possible to choose randomly a portion of the genes. The mutation may lead to an unsustainable individual, this possibility is therefore taken into account at the level of the coding of this operator.

Practical consideration: Be the genes of the enfant1 presented above:

Child before mutation

00000000…	11110111…	…	33633333…	44444444…
55556565…	66666666…	…	88888833…	99999999…

Child after mutation

00000000…	11110111…	…	33633333…	44444444…
55556565…	66666666…	…	88888833…	97999999…

Change of gene

Figure 10. Example of mutation

Here again, we followed recommendations of *Goldberg* [24] for the probability of mutation, by adopting a value inversely proportional to the size of the population.

7.4.5 *Replacement of the new population*

We conducted a replacement of bad solutions. For this purpose, we have classified all solutions (parents and children) according to their adaptation and are kept in our people that the *N* first individuals, i.e. more small values of fitness, obtaining the new population which consists of the *N* best adaptations. If the number of generations is reached, then the best

solution is extracted, and moves to the next step; otherwise, it is a new iteration of reproduction.

Practical consideration: The population of parents and children obtained, is presented in the table on the left. To make a replacement of the worst people, we have classified all the parents and their children according to their fitness growing. It will keep in the population than the *N* first individuals. The new population is presented in the table at right.

Table 2. Best population obtained

Child	fitness	Parents	fitness
1	4.2582	1	3.1490
2	11.5489	2	3.1803
3	7.2761	3	3.2824
4	9.1245	4	3.3785
5	13.5487	5	3.5947
6	5.0214	6	3.8692
7	3.1421	7	3.8761
8	3.8692	8	3.8815
9	3.5103		
10	7.6815		
11	7.2584		
12	3.1436		

New population	fitness
1:Child 7	3.1421
2: Child 12	3.1436
3: Parent 1	3.1490
4: Parent 2	3.1803
5: Parent 3	3.2824
6: Parent 4	3.3785
7: Child 9	3.5103
8: Parent 5	3.5947

7.4.6 *Test of judgment*

There is no criterion of judgment that guarantee the convergence of the GA to an optimal solution. It is customary to fix a priori, the number of iterations. When this number is reached, it takes the best chromosome obtained as the optimal solution. We used this method in our experiments.

For the GA, we set a maximum iteration number that represents the test case.

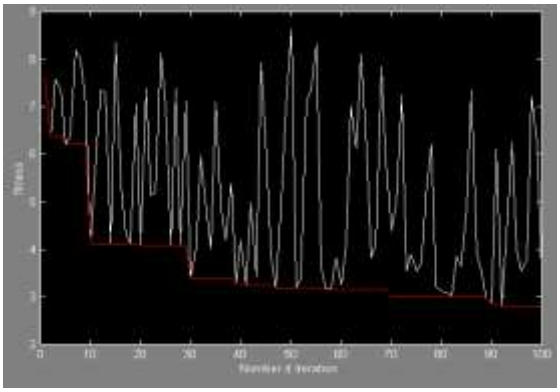

Figure 11. Convergence of the classification process for the base of the Arabic numerals (0-9).

7.5 Results and Discussion

Five types of models were compared within the framework of speech and medical databases recognition. The experimental results are reported in tables 3 and 4. The table 3 summarizes the various results obtained of the models using the acoustic parameters provided by a Log-RASTA PLP and J-RASRA PLP analysis. For the speech databases, all the tests of the five models were carried out on the DB1 in first then on the DB2 and finally on DB3.

For the majority of cases, we can draw a preliminary conclusion from the results reported in tables 3 and 4 for the speech recognition and the PEA signals diagnosis. We reported her, the results describes in [22].

- For the speech databases recorded in real conditions and noisy environment, in majority of cases, the result with application of J-RASTA PLP analysis is better than application of Log-RASTA PLP, this confirmed the theory.
- By comparing the rate with the rate of classification of the system using the multi-network structure (RBF/LVQ) describes in [17], the rate of the discrete HMM is better.
- The hybrid discrete HMM/MLP approaches always outperforms standard discrete HMM.
- The hybrid discrete HMM/MLP system using the GA clustering gave the best results and is more perform than the hybrid discrete HMM/MLP system using the FCM clustering and the rate of th later is beter than hybrid model using KM clustering.

Table 3. Recognition rates for the tree speech databases (DB1, DB2, and DB3) with the five different types of models (M1, M5) and application of acoustic analysis Log-RASTA PLP and J-RASTA PLP. M1: RBF/LVQ, M2: Discrete HMM, M3: Hybrid HMM/MLP with KM entries, M4: Hybrid HMM/MLP with FCM entries and M5: Hybrid HMM/MLP with AG entries.

	DB1		DB2		DB3	
	Log-RASTA PLP	J-RASTA PLP	Log-RASTA PLP	J-RASTA PLP	Log-RASTA PLP	J-RASTA PLP
M1	80.33	79,30	85.12	85,70	72.84	71,78
M2	85.75	87,45	89.68	90,23	75.31	76,48
M3	90.42	92,96	92.37	93,65	75.84	80,35
M4	**97.28**	97,46	96.16	96,89	73.23	**82,62**
M5	96.32	**98,37**	**96.82**	**99,32**	**75.22**	**82,95**

Table 4. Recognition rates for the biomedical databases with the five different type of models (M1,…, M5).

Models	Recognition rates
M1	80.66
M2	84.48
M3	90.12
M4	93.76
M5	**95.23**

8 CONCLUSION AND FUTURE WORK

In this paper, we presented the test of five types of models in the framework of speech recognition and medical diagnosis. A discriminate training algorithm for hybrid HMM/MLP system based on the genetic clustering are described. Our results on isolated speech and biomedical signals recognition tasks show an increase in the estimates of the posterior probilities of the correct class after training, and significant decreases in error rates in comparison to the five systems: 1) Multi-network RBF/LVQ structure, 2) Discrete HMM, 3) HMM/MLP approach with KM clustering, 4) HMM/MLP approach

with FCM clustering and 5) HMM/MLP approach with GA clustering.

These preliminary experiments have set a baseline performance for our hybrid GA/HMM/MLP system. Better recognition rates were observed. From the effectiveness view point of the models, it seems obvious that the hybrid models are more powerful than discrete HMM or multi-network RBF/LVQ structure for the PEA signals diagnosis and speech databases.

We thus envisage improving the performance of the suggested system with the following points:

- It would be important to use other techniques of speech parameters extraction and to compare the recognition rate of the system with that using the log-RASTA PLP and J-RASTA PLP analysis. We think of using the LDA (Linear Discriminate Analysis) and CMS (Cepstral Mean Subtraction) owing to the fact that these representations are currently considered among most powerful in ASR.

- It appears also interesting to use the continuous HMM with a multi-Gaussian distribution and to compare the performance of the system with that of the discrete HMM.

- In addition, for an extended speech vocabulary, it is interesting to use the phonemes models instead of words, which facilitates the training with relatively small bases.

- For the PEA signals recognition, the main idea is to define a fusion scheme: cooperation of HMM with the multi-network RBF/LVQ structure in order to succeed to a hybrid model and compared the perfomance with the HMM/MLP/GA model proposed in this paper.

9 REFERENCES

1. O. DEROO, C. RIIS, F. MALFRERE, H.LEICH, S. DUPONT, V. FONTAINE AND J.M. BOÎTE. «Hybrid HMM/ANN system for speaker independent continuous speech recognition in French ". Thesis, Faculté polytechnique de Mons – TCTS, BELGIUM, (1997).
2. H. BOURLARD, S. DUPONT. "Sub-band-based speech recognition". In Proc, IEEE International, Conf, Acoustic, Speech and Signal Process, pp.1251-1254, MUNICH, (1997).
3. F. BERTHOMMIER, H. GLOTIN. "A new SNR-feature mapping for robust multi-stream speech recognition". In Berkeley University of California, editor, Proceeding. International. Congress on Phonetic Sciences (ICPhS), Vol1 of XIV, pp. 711-715, SANFRANCISCO, (1999).
4. J-M. BOITE, H. BOURLARD, B. D'HOORE, S. ACCAINO, J.VANTIEGHEM. "Task independent and dependent training: performance comparison of HMM and hybrid HMM/MLP approaches". IEEE, vol.I, pp.617-620, (1994).
5. J.C. BEZDEK, J. KELLER, R. KRISHNAPWAM and N.R. PAL. "Fuzzy models and algorithms for pattern recognition and image processing". KLUWER, Boston, LONDON, (1999).
6. H. HERMANSKY, N. MORGAN. "RASTA Processing of speech". IEEE Trans. On Speech and Audio Processing, vol.2, no.4, pp. 578-589, (1994).
7. A. HAGEN, A. MORRIS. "Comparison of HMM experts with MLP experts in the full combination multi-band approach to robust ASR". To appear in International Conference on Spoken Language Processing, BEIJING, (2000).
8. A. HAGEN, A. MORRIS. "From multi-band full combination to multi-stream full combination processing in robust ASR" to appear in ISCA Tutorial Research Workshop ASR2000, Paris, FRANCE, (2000).
9. L. LAZLI, M. SELLAMI. "Hybrid HMM-MLP system based on fuzzy logic for arabic speech recognition". PRIS2003, The third international workshop on Pattern Recognition in

Information Systems, pp. 150-155, April 22-23, Angers, FRANCE, (2003).

10. L. LAZLI, M. SELLAMI. "Connectionist Probability Estimators in HMM Speech Recognition using Fuzzy Logic". MLDM 2003: the 3rd international conference on Machine Learning & Data Mining in pattern recognition, LNAI 2734, Springer-verlag, pp.379-388, July 5-7, Leipzig, GERMANY, (2003).
11. L. LAZLI, A-N. CHEBIRA, K. MADANI. "Hidden Markov Models for Complex Pattern Classification". Ninth International Conference on Pattern Recognition and Information Processing, PRIP'07.http://uiip.bas-net.by/conf/prip2007/prip2007.php-id=200.htm, may 22-24, Minsk, BELARUS, (2007).
12. L. LAZLI, A-N. CHEBIRA, M-T. LASKRI, K. MADANI. "Using hidden Markov models for classification of potentials evoked auditory". Conférence maghrébine sur les technologies de l'information, MCSEAI'08, pp. 477-480, avril 28-30, USTMB, Oran, ALGERIA, (2008).
13. D-L. PHAM, J-L. PRINCE. "An Adaptive Fuzzy C-means algorithm for Image Segmentation in the presence of Intensity In homogeneities". Pattern Recognition Letters. 20(1), pp. 57-68, (1999).
14. S-K. RIIS, A. KROGH . "Hidden Neural Networks: A framework for HMM-NN hybrids". IEEE 1997, to appear in Proc. ICASSP-97, Apr 21-24, Munich, GERMANY, (1997).
15. H. TIMM . "Fuzzy Cluster Analysis of Classified Data". IFSA/Nafips, VANCOUVER, (2001).
16. J-F. MOTSH. "La dynamique temporelle du trons cérébral: Recueil, extraction et analyse optimale des potentiels évoqués auditifs du tronc cérébral". Thesis, University of "Créteil", PARIS XII, (1987).