Research Journal of Applied Sciences 2 (4): 458-467, 2007

Genetic Supervised Classification of Standard Arabic Fricative Consonants for the Automatic Speech Recognition

M. Aissiou and M. Guerti
Algiers National Polytechnique School, BP 182, 16200 El-Harrach, Algiers, Algeria

Abstract: The purpose of this study is the application of the Genetic Algorithms (GAs) to the supervised classification level, in order to recognize Standard Arabic (SA) fricative consonants of continuous, naturally spoken, speech. We have used GAs because of their advantages in resolving complicated optimization problems, where analytic methods fail. For that, we have analyzed a corpus that contains several sentences composed of the thirteen types of fricative consonants in the initial, medium and final positions, recorded by several male Jordanian speakers. Nearly all the world's languages contain at least one fricative sound. The SA language occupies a rather exceptional position in that nearly half of it's consonants is fricatives and nearly half of fricative inventory is situated far back in the uvular, pharyngeal and glottal areas. We have used Mel Frequency Cepstrum Coefficients (MFCCs) method to extract vocal tract coefficients from the speech signal. To represents temporal variations in the speech signal, the first and second derivatives of both MFCCs and energy are added to the set of static parameters. The acoustic segments classification and the GAs have been explored. Among a set of classifiers like Bayesian, likelihood and distance classifier, we have used the distance one. It is based on the classification measure criterion. So, we formulate the supervised classification as a function optimization problem and we have used the decision rule Mahalanobis distance as the fitness function for the GA evaluation. We report promising results with a classification recognition accuracy of 82%.

Key words: Supervised classification, genetic algorithms, standard arabic fricative consonants, automatic speech recognition

INTRODUCTION

The Automatic Speech Recognition (ASR) has been intensively researched for more than four decades. In the last one, significant improvements and successes were achieved. However, there are too few studies about the application of evolutionary algorithms to the automatic speech domain. A typical ASR system consists of two components: feature analysis which includes parameters extraction and pattern classification.

The ASR deals with mathematical and technical aspects of classifying different acoustic segments through their observable information. The extracted parameters of the speech signal are classified by using a certain type of measures, such as Mahalanobis or Euclidean distance, likelihood and Bayesian, over class models. In our case, we have used the Mahalanobis distance as a criterion measure or a decision rule in the supervised classification of the fricative consonants of SA.

Since the number of features of acoustic segments of the vocal continuum is high, Linear Discriminant Analysis (LDA) is a technique that we have used to extract relevant features from speech signals. It is plausible that classes are not linearly separable, but nonlinearly separable in the original space. In this case, it is possible to transform the descent data space to an even higher dimensional space where the classes are linearly separable. Kernel version of LDA (KDA) enables this transformation without having too much extra computation.

In this study, we have used GAs as a function optimization problem to evolve structures representing sets of acoustic segments rules (Hansohm, 2000; Aissiou and Guerti, 2004; Greene, 2003).

The GAs are search methods for good solutions in a large population of candidate solutions. GAs use extensive search of the current candidate space, to find the currently best approximations to the unknown solution. Their advantages are as follow:

- They work with both continuous and discrete parameters, execute simultaneous searches over several regions of the search space and work with a population instead of an unique point;

Correspoding Author: M. Guerti, Algiers National Polytechnique School, BP 182, 16200 El-Harrach, Algiers, Algeria

- They optimize a large number of parameters, have successfully found global minimum even on very complex and complicated objective function and their computer implementations are portable and modular.

Also, they are tolerant to incomplete and noise data (Miller and Goldberg, 1995).

For the purposes of this study, the phonetic framework is that of the International Phonetic Alphabet (IPA) as used within the California University Phonological Segment Inventory Database (UPSID).

STANDARD ARABIC SOUNDS DESCRIPTION

Standard Arabic is composed of the following segments:

- Twenty eight phonetically distinct consonant phonemes. Generally, consonants have less energy than vowels. The characteristics that form a vowel are relatively more prominent and stable than those of the consonants. The consonant vary individually, making it easier to deal with them in group
- Three short vowels, (a, u, i), which contrast phonemically with their long counterparts,([a :, u:, i]. Throughout the text, phonemic length is indicated by writing the vowels symbol twice
- The 6 correspondent variants of the short and the long vowels, in emphatic context
- The silence, called (suku:n).

The vowels were recorded in the following ways: in isolation, within the Consonant-Vowel (CV) sequences. All the Arabic vowels are oral and fully voiced. But, they can be nasalized in nasal consonants context (assimilation).The difference between short and long ones is approximately double or more. The relative duration of the short vowels is from 100-150 msec. With the long ones, it is from 225-350 msec (Al Ani, 1970).

The SA phonemes classification is based on physiological speech parameters that consist of both horizontal and vertical places and various manners of articulation. They are produced by a closure or a narrow constriction in the vocal tract. The tongue position in the vocal cavity and the lips form definite the articulation place which serves to classify the consonants (Al Ani, 1970).

The 28 consonants of SA are classified physiologically as follow :

- 13 fricatives;
- 08 stops: [?], [k], [q], [t], [d], [b], [t*], [d*] ;
- 02 nasals: [n], [m];
- 03 sonorants: [y], [w], [r] ;
- 01 trill: [l];
- 01 affricate [dZ].

The relative duration of the consonants depends upon whether they occur initially, medially or finally. It also depends on whether they are aspirated or unaspirated, voiced or unvoiced and single or geminated.

Fricative consonants of standard arabic: The fricatives form the largest set of consonants in the SA language which has thirteen fricative consonants. They are produced in the vocal cavity by a narrow constriction that causes the airflow to be consistently turbulent (Table 1).

The voiced fricatives are opposite to their voiceless ones. Acoustically, voiceless ones usually possess a high random noise and voiced ones usually possess weak resonance structure. Relatively to the articulation place, the fricatives classification is as follows:

- The labiodental (f): This phoneme is voiceless and appears as a random noise
- The 3 interdentals ([D], [θ], [T])
- The 3 dento-alveolars ([s], [z], [S])
- The palatal [ʃ]: This phoneme is voiceless and appears as a random noise
- The 5 back consonants which are the voiced uvular [ɣ], the voiceless velar [X], the two pharyngeals ([h], [ε]) and the voiceless glottal ([h]).

Particularities of standard arabic sounds: The SA is characterized by three phonetic phenomena which are the presence of the emphatic consonants, the geminate ones and by the presence of glottal, pharyngeal, velar and uvular ones called back consonants. The SA possesses eight back phonemes.

Emphasis: The emphasis is a complex phenomenon that possesses certain characteristics descriped as follow (Bonnot, 1979) :

- The tongue root positioned towards back of the mouth
- Pharyngalization
- Velarization

There are 4 emphatic consonants of SA whose two consonants are fricatives (Table 1).

Table 1: Fricative consonants of standard arabic with their various transcription

International Phonetic Alphabet (I.P.A)	Arabic-speaker transcription	Arabic fricative consonants
Unvoiced consonants		
(θ)	(±)	(ث)
(ħ)	(Ÿ)	(ح)
(X)	(¡)	(خ)
(T)	(†)	(ذ)
(s)	(s)	(س)
(ʃ)	(ŝ)	(ش)
(S)	(β)	(ص)
(f)	(f)	(ف)
Voice consonants		
(z)	(z)	(ز)
(D)*	(¼)	(ظ)
(ε)	(ε)	(ع)
(γ)	(‰)	(غ)
(h)	(h)	(ه)

*: Emphatic consonants

On the neighbouring emphatic consonants, all the Arabic vowels are strongly influenced. So we obtain variants: Emphasized vowels in opposition to the not emphasized ones. From acoustic point of view, the unique element that characterize the emphatic consonant in comparison to it's opposite non emphatic one, is the variation of the second formant transition noted F_2 in the CV context (Cantineau, 1960).

Gemination: The gemination corresponds to the consonant production with intensive energy concentration. In phonetic terms, the distinction between the geminated and ingeminated segments is predicated on the fact that the hold phase in the production of the consonant is lengthened to approximately double the length of the ingeminated one. At a phonological level, it is important to note that in Arabic the geminated segments can never occur word-initially. The entire Arabic consonant can be geminate except the glottal stop consonant [?] (Bonnot, 1979).

SPEECH FRONT END PROCESSING

Front-end processing is the first component in ASR, therefore the quality of the front-end processing will greatly determine the quality of the later other components. Speech signal changes continuously due to the movements of vocal system and it is intrinsically non-stationary. Nonetheless, in short segments, typically 20 to 40 ms, speech could be regarded as pseudo-stationary signal. Speech analysis is generally carried out in frequency domain with short segments across which the speech signal is assumed to be stationary and it is often called short-term spectral analysis.

Signal acquisition and conditioning: Signal acquisition and conditioning involves sampling the analogous signal and performing some preliminary processing to produce a digital one (16 bits). Voiced segments of this digital signal exhibit a negative spectral slope. Pre-emphasis is the process by which this natural slope is offset, equalising the dynamic range across the entire frequency band. Generally preemphasis is performed by filtering the speech signal with the first order impulse response filter, which has the form as follow :

$$H_1(z) = 1 - 0,97(z)^{-1} \quad (1)$$

From an engineering point of view, the speech signal can be seen as the output of a quasi-stationary random process. Assuming that this is the case, the pre-emphasised stream of digital data is analysed in frames of 20 ms, at intervals of 10 ms. The discontinuities at the ends of each frame cause excessive spectral leakage. The distortion is reduced by weighting the samples with a tapered window. The most common tapered window is the Hamming window which is based on the function shown as follow:

$$W(n) = \begin{cases} 0.54 + 0.46 \cos(\pi n(N-1)) & n \in [0, N\text{-}1] \\ 00 & \text{Elsewhere} \end{cases} \quad (2)$$

Feature extraction: As we know feature extraction influences the recognition rate greatly, it is too important for any recognition/ classification systems. Feature extraction is to convert an observed speech signal to some type of parametric representation for further analysis and processing. It transforms the high-dimensional speech signal space to a relatively low-dimensional features subspace while preserving the speech discriminative information to application. Several techniques can be used to extract vocal tract coefficients from the speech signal. These include filter bank analysis, Linear Predictive Coding (LPC) and cepstral analysis (Vergin *et al*., 1999).

Short-term mel-frequency cepstrum: Until now, MFCCs are the best known and most commonly used features for ASR. They are extensions of the cepstral which are used to better represent human auditory models. The approximation of Mel from frequency can be expressed as:

$$\text{mel}(f) = 2595.\log_{10}(1 + f/700) \quad (3)$$

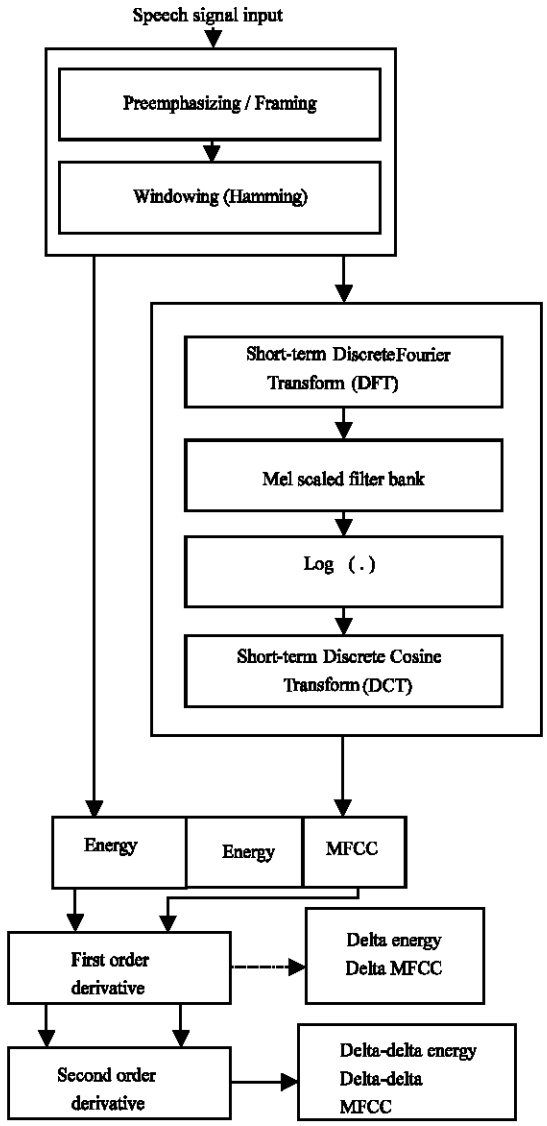

Fig. 1: Computation steps of MFCCs (Campbell, 1997)

Where f denotes the real frequency and mel (f) denotes the perceived frequency. The Mel-frequency Warping is normally realized by Filter banks. Filter banks can be implemented in both time
domain and frequency domain. For the purpose of MFCC processor, filter banks are implemented in frequency domain before the logarithm and the inverse Discrete Fourier Transform (DFT). The most commonly used filter shaper is triangular. The last step before getting MFCCs is the inverse DFT (iDFT). Normally the Discrete Cosine Transform (DCT) will be performed instead of iDFT (Fig. 1).

The $M_c$cepstral coefficients noted C_n are calculated as followed (Vergin *et al*., 1999):

$$C_n = \sum_{k=1}^{n} x_k \cos\left(n - \frac{\pi(j-0.5)}{20}\right) \quad (4)$$
$$\forall \quad 0 \le n \le M_c - 1$$

It will results the MFCCs vector noted V_j and expressed as:

$$V_j = (c_1, c_{2,\ldots}, c_{M_c}) \quad (5)$$

Just the first coefficients of the cepstral sequence are interesting for smoothing the spectrum and minimising the pitch influence. The first thirteen MFCCs of the consonants sounds describe its unique properties sufficiently to differentiate them each other (Vergin *et al*., 1999).

Short-time energy: It is common to append an energy coefficient to the cepstrum feature vector. It is computed as the logarithm of the accumulated frame energy. Differences in energy among phonemes show that it is a good feature to distinguish between them. The normalized log of the raw signal one is usually used as the energy coefficient. It is computed as the logarithm of the signal one.

Differential features: Temporal changes, in speech spectra, play an important role in perception. This information is captured in the form of velocity coefficients and acceleration coefficients (collectively referred to as differential or dynamic features). No time evolution information is included in MFCCs. But it is often included in the feature set by cepstral derivatives. The first order derivative of MFCCs is called Delta coefficients and their second order derivative is called Delta-Delta coefficients. The first ones tell us somehow the speech rate and the second ones give us something similar to the speech acceleration. The delta coefficients are computed using linear regression :

$$\Delta x(m) = \frac{\sum_{i=1}^{j} (i)\left[x(m+i) - x(m-1)\right]}{2\sum_{i=1}^{j} i^2} \quad (6)$$

Where, 2_j+1 is the regression window size and x denotes the cepstrum.

The second order derivatives are computed using the same linear regression applied to a window of delta coefficients.

Dimension reduction of the features space: Speech recognition can be cast as a pattern classification problem where we would like to classify an input acoustic signal into one of all possible phonemes. However, the features number of acoustic segments of the vocal continuum is high, so that it is unreasonable to solve the problem as a regular classification one.

The each m acoustic segments parameter vector noted $V_{nm}$is constituted of the first thirteen MFCCs, the energy and their first and second time derivatives. The description of this vector is as follow :

$$V_{nm} = (c_1, c_{2,...}, c_{M_c}, ..., c_n) \quad (7)$$

For this reason, it is reasonable to borrow pattern classification techniques to help in speech recognition. LDA is one such technique that can be used to extract relevant features from speech signals that discriminate well between sub-unit classes. The problem is usually cast as a dimension reduction problem where many candidate features are pulled together and a rectangular linear transformation is found where the resultant feature vector has small dimensions, yet it carries the most discriminative information to separate classes well.

Kernel-based machine learning and pattern classification techniques have achieved considerable success recently. Among those are Support Vector Machine (SVM) and Kernel version LDA. Already, kernel version of LDA is very popular among pattern classification community. These methods enable to extract features or decision rules that have nonlinear boundaries by projecting parameter vector onto a higher-dimensional feature space through a non-linear Kernel function which is often defined as the dot product between a parameter vector and a reference vector (Muller *et al*., 2001).

$$k(x, \mu_i) = (x.\mu_i) \quad (8)$$

In this higher-dimensional feature space, non-linear class boundaries may become linear. Then the decision plane is pursuit in the feature space. The projection of decision plane in the parameter space is a non-linear decision boundary (Cristianini and Shawe, 2000; Kocsor and Paczolay, 2001).

It result of a set of reduced acoustic parameters vectors whose description is as follows :

$$Vkda_{im} = (Kda_{m1}, Kda_{m2}, ..., Kda_{mn}) \quad (9)$$

SUPERVISED CLASSIFICATION

Classification of objects is an important area of pattern recognition and artificial intelligence. Classifier design can be performed with labelled or unlabeled data. Classification operation usually uses supervised learning methods that induce a classification model from a database. Using a supervised learning method, the computer is given a set of objects with known classification and is asked to classify an unknown object based on the information acquired by it during the training phase (Celeux *et al*., 1989).

We assign an input feature vector to one of K existing classes based on a classification measure. Conventional classification measures include Mahalanobis or Euclidean distance, likelihood and Bayesian a posteriori probability. These methods measures lead to linear classification methods (the decision boundaries they generate are linear). Linear methods, however, have the limitation that they have little computational flexibility and are enable to handle complex nonlinear decision boundaries.

Consequently, acoustic classification is the elaboration of a decision rule that assigns acoustic feature vectors of phonemes to one of existent classes (Fig. 2).

Classification criterion is also called decision rule. The most widely used classification criteria are distances, Bayesian decision rule and likelihood. A brief summary of these criteria is given in the following :

- Distance criterion is the simplest and most direct one. The basic idea of distance classification criterion is

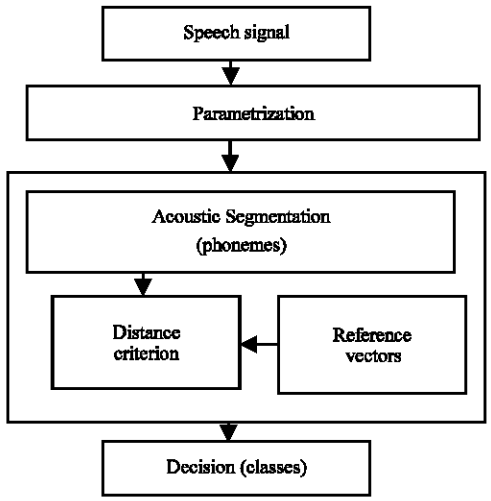

Fig. 2: Acoustic classification procedure

that a data is classified to a class that is closest to it. Euclidean and Mahalanobis distances are the two most common forms. Suppose we have k classes, let $(\mu_i, \sum_i)$be the known parameter set of class i, where μ is the reference vector of class i. $(\sum_i)$ is the covariance. The square form of the Euclidean distance of an observation vector x from class i is :

$$d_i(x) = \|x - \mu_i\|^2 \quad (10)$$

The square form of Mahalanobis distance of x from class i is :

$$d_i(x) = (x - \mu_i^K)\sum_i^{-1}(x - \mu_i) \quad (11)$$

Euclidean distance is in fact a special case of Mahalanobis distance:

- Bayesian decision rule is based on the assumption that the classification problems are posed in probabilistic terms and all of the relevant probabilities are known. It will assign an observation vector to the class that has the largest a posteriori probability
- Likelihood criterion is a special case of Bayesian classification one. It assumes that all of the priori probabilities are equal and the distribution of classes is normal

Based on the classification criterion used in discriminate functions, classifiers can be grouped into Bayesian, likelihood or distance classifiers.

GENETIC CLASSIFICATION

Gas have been used broadly in search and optimization. The optimization is the search for a better solution to solve a given problem. It consists of trying several solutions and to use information collected in this process in order to improve the quality of the solutions.

Our GA consists on looking for, among the various acoustic segments of a vocal sequence, structures which correspond to those of the representative reference vectors of every class. To associate entities to be recognized in classes, it is enough to find an analytical correspondence between the two groups of the acoustic vectors respectively vocal sequences and entities to be recognized of reference. The vectors of the initial population represent the search space for potential solutions. The GA will have for task to make on these initial vectors some modifications, reorganizations, mutations, to produce the best vector as the potential solution (Gilleron and Tommasi, 2000; Aissiou and Guerti, 2006).

Genetic algorithm steps: Before applying the GA to the problem, the user designs an artificial chromosome of a certain fixed size and then defines a mapping (encoding) between the points in the search space of the problem and the instances of the artificial chromosome. Once the reproduction and the fitness function have been properly defined, a GA is evolved. It starts by generating an initial population of chromosomes. Then, the GA loops over an iteration process to make the population evolve. The GA transforms a population of individual objects, each one with an associated fitness value, into a new generation of the population using the Darwinian principle of reproduction and survival of the fittest and analogous of naturally occurring genetic operations such as crossover (sexual recombination) and mutation. Each individual in the population represents a possible solution to a given problem. Each iteration consists of the following steps :

Selection: The first step consists in selecting individuals for reproduction. This selection is done randomly with a probability depending on the relative fitness of the individuals so that best ones are more often chosen for reproduction than poor ones.

Reproduction: In the second step, offspring are bred by the selected individuals. For generating new chromosomes, the algorithm can use both recombination and mutation.

Evaluation: Then the fitness of the new chromosomes is evaluated.

Replacement: During the last step, individuals from the old population are killed and replaced by the new ones.

Genetic classifier initial population: GA starts by generating an initial population of chromosomes. This first population must offer a wide diversity of genetic materials. The gene pool should be as large as possible so that any solution of the search space can be engendered. Generally, the initial population is generated randomly. We have chosen the initial class number, noted N, equal to SA alphabet phonemes number. The SA alphabet is composed of 28 consonants, 06 vowels and their 06 vocalic variants in emphatic context.

With the introduction of the concept of classifier initial population in the ASR context, we must definite the two followed sets:

- The T set of M speech segments. The duration of each segment is equals to 10 ms. This set has the following form :

$$T = \{S_1,...,S_m,...,S_M\} \quad (12)$$

The dimension of the corresponding parameter vector of each acoustic segment is fixed to n. Where the position n°i contains the coefficient Kda_{im} of the acoustic segment n° m. Each acoustic segment is represented by a parameter vector noted $Vkda_{mn}$and has the following form:

$$Vkda_{im} = (Kda_{m1},Kda_{m2},...,Kda_{mn}) \quad (13)$$

- The set noted C constituted of all the fricative consonants It has the following form :

$$C = \{C_1,...,C_k,...,C_{13}\} \quad (14)$$

The effect of genetic research is the dispatching of the acoustic segments among a certain number of classes. The number of these classes is equal to the number of fricative consonants of SA.

Reference acoustic data: The choice of the type and the indices reliability and the relevant acoustic parameters of fricative consonants to be recognized are very important. These last ones, as acoustic vectors form, constitute the set of reference data for the AG. The AG refers to this set, during the evaluation phase as potential solutions, (Aissiou and Gueti, 2006).
Each fricative of class k consonant is represented by a parameters vector noted μ_{kj}:

$$\mu_{Kj} = (Kda_{K1},Kda_{K2},...,Kda_{Kj}) \quad (15)$$

The combination of static and dynamic features provides additional discrimination for speech recognition. These acoustic vectors which are reference are obtained during the learning phase. Each of the thirteen fricatives is represented by an acoustic reference vector of a specific order (Table 2).

Table 2: Reference acoustic vectors of standard Arabic fricative consonants

Consonnants	μ_{kj}	Vector order (vo_K)
(±)	(Kda_{11},..., Kda_{a1tt})	tt
….	……	…
(Ÿ)	(Kda_{k1},..., Kda_{khh})	hh
…..	……..	…
(h)	(Kda_{131},...., Kda_{13h})	h

Encoding of the problem: Each solution is represented through a chromosome, which is just an abstract representation. The choice of an efficient representation is one of the most important issues in designing a GA. We use a fixed length chromosome for the GA, where each chromosome contains, as genes, the coefficients $Kda_{im}$of several segments noted S_m of the vocalic continuum of the chosen corpus (Goldberg, 1989).

The individual is elaborated relatively to the consonant to be classified. For example, if we want to classify the fricative [h], we proceed to the segmentation of the vocalic continuum into a set of acoustic vectors according to the corresponding individual that will be definite as the chromosome. This acoustic vector noted $V_{m\,vok}$ is composed of the coefficients of a certain number of successive concatenated acoustic segments and its size is equal to the order noted $vo_k$of the consonant reference vector.

$$V_{m\ vo_k} = \begin{pmatrix} kda_{m1},Kda_{m2},...,Kda_{(m+1)1},..., \\ Kda_{(2m)n},..., Kda_{rn},...,Kda_{vo_K} \end{pmatrix} \quad (16)$$

Where: $r = \frac{vo_K}{n}$

and r is the number of acoustic segments S_m concatenated to form the acoustic vector of the whole individual.

It result that the initial population noted pop will be composed of all possible acoustic vectors $V_{m\,vok}$when we vary the values of m and vo_K. The different values of m are as follows:

$$m_i = 1,...,M \quad (17)$$

So, for each value of m we have an individual.

Evaluation function: The Evaluation Function (FA) has to be most discriminating possible. This discriminating character concerns two additional aspects: The internal cohesion of the classification which is the degree of affinity of segments inside every class and the differentiation of the classification which corresponds to the dissimilarity of the classes among them (Jang and Spears, 1991).

The FA can be represented rigorously by the Mahalanobis distance between the vectors of the acoustic segments and the vectors of reference of every class of fricatives consonants. It can be expressed as follows (Koza, 1994; Wright 1991).

$$F_K = d_k\left(V_{m\ vo_K}\right) = \left(V_{m\ vo_K} - \mu_k^{13}\right) \sum_k^{-1}\left(V_{m\ vo_K} - \mu_k\right) \quad (18)$$

To estimate potential solutions of our problem, we minimize the FA to decide of the type of membership classes of the acoustic segment or the sets of these segments. So, we just maximize the inverse function of FA noted F and looking for the extrema. This inverse function is as follows:

$$F = 1/F_K \quad (19)$$

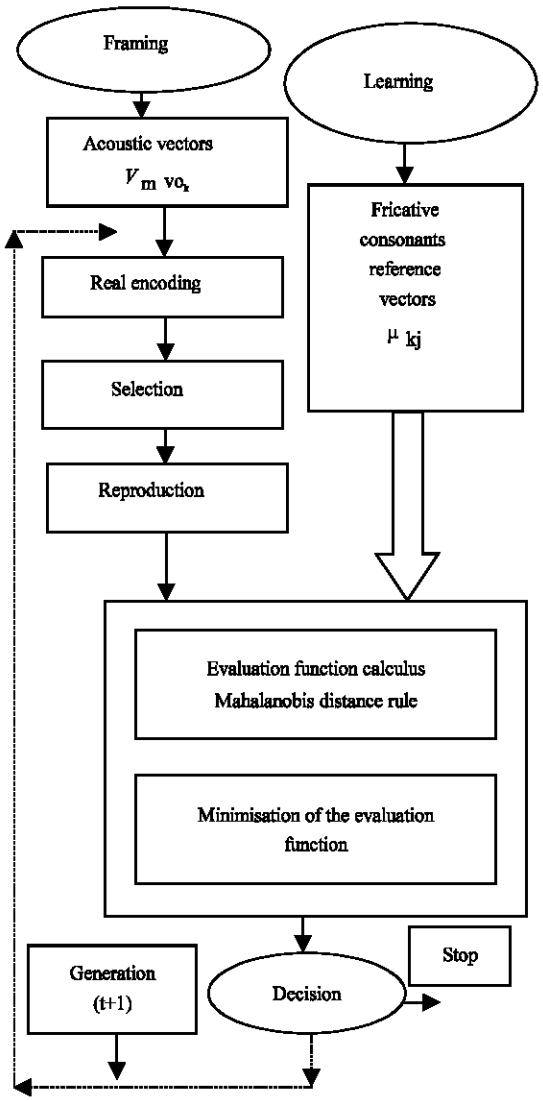

Fig. 3: Flowchart of classification genetic algorithm

The search for this extrêma can give two types of results: The global maximum means the existence of search for the same type of consonant in the rest of the vocal continuum and the end of the vocal continuum means that there is no fricative consonant of class k in the vocal continuum. In that case, we repeat the research for the other types of fricatives (Fig. 3).

Selection operator: Selection is supposed to be able to compare each individual in the population. It is done by using a fitness function. Each chromosome has an associated value corresponding to the fitness of the solution it represents. The fitness should correspond to an evaluation of how good the candidate solution is. The optimal solution is the one which maximizes the fitness function. There are many ways to select individuals. The higher the fitness function, the more chance an individual has to be selected. The selection pressure is defined as the degree to which the better individuals are favoured. The higher the selection pressured, the more the better individuals are favoured. This selection pressure drives the GA to improve the population fitness over the successive generations (Bridges and Goldberg, 1991; Holland, 1975).

Roulette wheel selection is the classic and more popular fitness-proportionate selection. It simply assigns to each solution a sector of a roulette wheel whose size is proportional to the appropriate fitness measure and then chooses a random position on the wheel and the solution to which that position was assigned.

Individuals replacement methods: Once offspring are produced, a method must determine which of the current members of the population, if any, should be replaced by the new solutions. Basically, there are two kinds of methods for maintaining the population; generational updates and steady state updates. The basic generational update scheme consists in producing N children from a population of size N to form the population at the next time step (generation) and this new population of children completely replaces the parent selection.

In a steady state update, new individuals are inserted in the population as soon as they are created, as opposed to the generational update where an entire new generation is produced at each time step. The insertion of a new individual usually necessitates the replacement of another population member. The individual to be deleted can be chosen as the worst member of the population, or as the oldest member of the population, but those methods are quite radical. A subtle alternative is to replace the most similar member in the existing population.

Genetic algorithms convergence: GAs are supposed to converge naturally around optima as they are discovered, but unfortunately these optima are not guaranteed to be the global optimal solution of the problem. A common failure mode of GAs is usually called premature convergence, by which is normally meant that the population converged to an unacceptably poor solution. To tackle this, various techniques could be developed, such as increasing the mutation rate, or even better. Typically, mutation rate should be higher at the end of a run to avoid loss of diversity.

A larger population size or a lower selection pressure can also help avoid premature convergence. In fact, only a good compromise between selection pressure, population size and mutation rate and so on, can lead to an adequate algorithm that finds good solutions in a relatively short among of time (Davis, 1996; Ingber and Rosen, 1992).

Stopping criteria: The GAs are stopped when the population converges toward the optimal solution. Unfortunately, evolution often looks like a never-ending process. GAs cannot be expected to stop spontaneously on the global optimum solution as if by magic. GAs are not even guaranteed to find the global optimum solution. So evolution has to be stopped at some point according to a pre-determined criterion. In fact there are many ways to decide to stop the algorithm. The simplest one is to stop evolution after a fixed number of iterations. A better solution consists of continuing the iteration for as long as any improvements are noticeable. Finally we can also wait until most or all of the members of the population are similar or identical.

RESULTS AND DISCUSSION

We have developed a genetic algorithm for acoustic unit classification task. Our GA uses an elitist generational population model without any overlap of the populations. Fixed length chromosomes were used relatively to the order of reference vector of each fricative consonant. The GAs are very advantageous compared to the deterministic algorithms, in the sense that further improvements can be made in various ways. Also, they are more suitable for large-sized problems.

In order to maximize the reproducibility and minimize the ambiguity in the GA implementation, the population size was chosen to be dependant on the size chosen corpus. Uniform random selection was used to reduce the selection pressure since roulette selection resulted in premature convergence (Goldberg, 1989).

Our simple, two point, crossover operator was used with crossover probability noted $p_c$equals to 0.8 and the probability uniform mutation noted p_m was fixed to 0.05. This means that too poor attribute values were modified in each recombination. These values are chosen empirically to improve our GA by performing multiple executions and then choosing the best solution (Jong and Spears, 1991).

The experiment was conducted using a medium-sized corpus composed of 20 naturally spoken sentences of Standard Arabic continuous speech from Arabic normalized database. This corpus is spoken by several male Jordanian speakers in low noisy environment. It contains more then one hundred fricatives which occurred in the three possible positions: initially, medially and finally. The speech continuum segmentation was performed manually by using version 1.9 of Emu Speech Data System software (Emu Segmenter).

The results analysis shows that the GAs application to the speech classification domain is very interesting, since we have obtained good results. This evaluation was conducted in terms of two criteria: Classification accuracy and computation time. Fricatives consonants were extracted and classified successfully with a 82% global accuracy (9% substitution, 5% insertion and 4% deletion).

The unclassified fricatives are essentially due to the segmentation difficulties and to the coarticulation phenomenon problems. In natural speech, there are no marked boundaries between acoustic data segments. Word and phoneme boundaries are non-existent. Even with expert labelling of the acoustic data, it is very difficult to establish a hard boundary between the phonemes and words that form an utterance. Speech is a stream of phonemes that sounds very smooth. This smoothness results from the coordination of the articulators' movements by the brain. The movements of these articulators-lips, tongue, jaw, velum and larynx- are coordinated so that movements needed for adjacent phonemes are simultaneous and overlapping.

The results show that the classification accuracy of fricatives which occur in the initial and final positions of words is higher than the accuracy of the medium positions ones. The medium fricatives are strongly influenced by the preceding and the following sounds. Approximately, there is similar classification accuracy between the thirteen types of SA fricatives. It means that we have used a good front-end signal processing used in our system to characterize, acoustically, the fricative consonants.

Also, we can say that the unclassified fricatives accuracy is due to the fact that the GA has many inherent variations and parameters that need to be handled properly in the implementation stage, in order for reasonable or good performance to be obtained.

In order to improve the performance of our GA, we propose the diversification of the recombination techniques and the introduction variations to its parameters values. We have to test other structure modification operators and to use a corpus of high-quality recordings of read continuous speech of Standard Arabic.

CONCLUSION

In this study, we have shown a novel speech segments supervised classification method based on the Genetic Algorithms in order to recognize, automatically, Standard Arabic fricative consonants. We have used a medium-sized corpus composed of 20 naturally spoken sentences of continuous speech from Arabic corpus recorded by several male Jordanian speakers in low noisy environment.

We formulate the supervised classification problem as a function optimization problem. And we have used the decision rule Mahalanobis distance as the GA evaluation function or fitness function. We used two structure modification operators: two point crossover and uniform mutation. Uniform random selection was used to reduce the selection pressure.

Our promising results show that the GA is capable of accurately classify almost all instances. We have compared these results with other promising machine learning approaches and we have found that our classification model GA is approximately as performing as the other known approaches like artificial neural networks.

REFERENCES

Aissiou, M. and M. Guerti, 2006. Classification Génétique des Voyelles de l'Arabe Standard. International Conference on Control, Modelling and Diagnostic, Annaba, Algeria, pp: 75.

Aissiou, M. and M. Guerti, 2004. Genetic Algorithms Application to the Standard Arab vocalic Recognition. Fifth International Arab Conference on Information Technology, Constantine, Session E1: Speech Processing, pp: 452-456.

Al Ani, S.H., 1970. Arabic Phonology an acoustical and physiological investigation, Edition Mouton.

Bonnot, J.F., 1979. Etude expérimentale de certains aspects de la gémination et de l'emphase en Arabe. Actes du 9eme ICPS Copenhague, Vol.1.

Bridges, C.L. and D.E. Goldberg, 1991. An analysis of multipoint crossover, the Foundation of Genetic Algorithms. Foga.

Cantineau, J., 1960. Cours de Phonétique Arabe. Klincksiek, Paris.

Campbell, J.P., 1997. Speaker Recognition: A Tutorial. Proc. IEEE., pp: 1437-1462.

Cristianini, N. and J. Shawe-Taylor, 2000. An Introduction to Support Vector Machines and other kernel-based learning methods. Cambridge University Press.

Celeux, G., E. Diday, G. Govart, Y. Lechevallier and H. Ralambondrainy, 1989. Classification automatique des données. Edition Dunod.

De Jong, K.A. and W.M. Spears, 1991. Learning concept classification rules using genetic algorithms. Int. Joint Conf. Artificial Intell., pp: 651-656.

Davis, L., 1991. Handbook of Genetic Algorithms. Van Nostrand Reinhold, New York.

Greene, W., 2003. Unsupervised Hierarchical Clustering via a Genetic Algorithm. Press, I. (Ed.), the Congress on Evolutionary Computation, Canberra, Australia, pp: 998-1005.

Gilleron, R. and M. Tommasi, 2000. Découverte de Connaissances à partir de données. Technical report, Grappa, université de Lille 3, France.

Goldberg, D.E., 1989. Genetic Algorithms in Search, optimisation and Machine Learning Reading, Madison Wesley.

Hansohm, J., 2000. Two-mode Clustering with Genetic Algorithms, Classification, Automation and New Media. Twenty fourth Annual Conference of the Gesellschaft Fur Klassifikation, E.V., pp: 87-94.

Holland, J.H., 1975. Adaptation in Natural and Artificial Systems, University of Michigan Press.

Ingber, L. and B. Rosen, 1992. Genetic Algorithms and very fast simulated re-annealing. Mathematical Computer Modeling, pp: 87-100.

Kocsor, A., L. Toth and D. Paczolay, 2001. A Non linearized Discriminant Analysis and its Application to Speech Impediment Therapy, in V. Matousek *et al.* (Eds.). Springer Verlag, LNAI Series, 2166: 249-257.

Koza, J.R., 1994. Genetic Programming II. the MIT Press.

Miller, B.L. and D.E. Goldberg, 1995. Genetic Algorithms, Selection Schemes and the Varying Effect of Noise. Illigal Report 95009.

Muller, K.R., S. Mika, G. Ratsch, K. Tsuda and B. Scholkopf, 2001. An introduction to kernel-based learning algorithms. IEEE. Trans. Neural Net., 12: 181-202.

Vergin,R., D.O. Shaughnessy and A. Farhat, 1999. Generalized mel frequency cepstral coefficients for large-vocabulary speaker-independent continuous-speech recognition. IEEE. Trans. Speech and Audio Proc., pp: 525-532.

Wright, A.H., 1991. Genetic Algorithms for real parameter optimization. The Foundation of Genetic Algorithms. Foga.