*Proceedings of the 2012 International Conference on Industrial Engineering and Operations Management*
*Istanbul, Turkey, July 3 – 6, 2012*

# Genetic Algorithms in Speech Recognition Systems

**Zahira Benkhellat and Ali Belmehdi**
**Department of Informatics**
**Bejaia University, Algeria**

## Abstract

In speech recognition, the training process plays an important role. When a good training model for a speech pattern is obtained, this not only enhances the speed of recognition tremendously .but also improves the quality of the overall performance in recognizing the speech utterance. In general, there are two classic approaches for this development, namely Dynamic Time Warping (DTW) and Hidden Markov Model (HMM). In this article, Genetic Algorithm (GA) is applied to solve involved nonlinear, discrete and constrained problems for DTW .Because of the intrinsic properties of GA, the associated non trival K-best paths of DTW can be identified without extra computational cost. The obtained results show the important contribution of the genetic algorithms in temporal alignment through the increasingly small factor of distortion.

## Keywords
Genetic Algorithms, artificial intelligence, speech recognition, signal processing.

## 1. Introduction

The early work on speech recognition goes back to 50 years (system of recognition of digits). The introduction of numerical methods and computer technology has increased system capacity [1]. Among the systems most known of speech recognition systems, the Via Voice systems and Dragon Naturally Speaking system. These systems are based on, concerning the recognition of their modules a probabilistic statistical approach uses the hidden Markov model HMM (Hidden Markov Model) [2, 3]. Another approach used in the systems of speech recognition is the Dynamic Time Warping (DTW) method or dynamic temporal alignment [4]. It is to define a dissimilarity index between two words by putting in optimal correspondence time scales of the two words. This technique is faced with two problems: the best alignment path and constraints on the standardization factors used in a path of deformation. Genetic algorithms constitute a large family of statistical algorithms developed by Holland [5] and further developed by Goldberg [6]. In this work, it is to apply these algorithms on one hand, and to optimize temporal alignment path used in the DTW minimizing the distortion between two pronunciations of words on the other hand, to proceed to relaxation of the normalization factor.

## 2. Generalities on Speech Recognition Systems

The speech recognition system is able to transcribe a human voice into digital information, understandable and recognizable by the computer.


Figure 1: Block diagram of a speech recognition system

The speech signal is first digitized and modeled, is usually some a frequency form. The next module in the chain of acoustic (Figure 1), which is extracted from the relevant parameters for speech recognition. These parameters are sent to a recognition module to identify the present sounds in the signal. Among the methods used in the module for speech recognition, the alignment method of temporal dynamic DTW (Dynamic Time Warping).

### 2.1. Structure of a recognition system based on DTW method:

The structure of a speech recognition based on this method is represented by Figure 2. In this system there are two phases:

- A learning phase to form the reference dictionary (R1.. RN).
- A recognition phase during which the system will search the reference closest to the word to be identified by calculating the distance between the unknown word and the words of reference.

If we represent the extracted vectors of the parameterization module of a test word X = (x1, x2, .. xi ... .. xN) on the horizontal axis and the reference vectors of the word Y = (y1, y2, ..., yj ... yM). on the vertical axis.

- The comparison between an unknown word and a word of reference may be represented by a temporel alignment path C = c (1), c (2), ... .., c (k ),.... c (K) . The sequence C achieves a match between the word elements of the unknown and the reference.

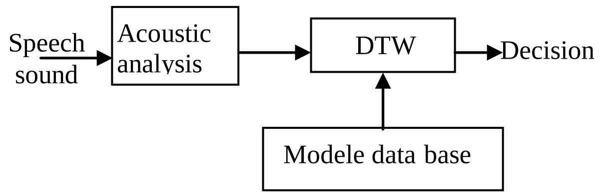


Figure 2: System for speech recognition based on DTW

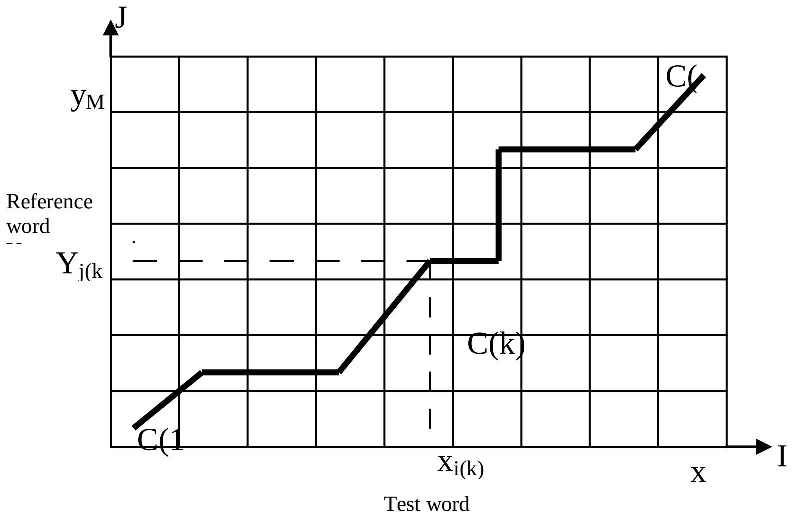


Figure 3: The optimal path between two words

**2.1.1. Dissimilarity between two words**

The basic principle of DTW is to search among all possible paths, the path that minimizes the dissimilarity between the two words given by:

$$D_c(X,Y) = Min_c \frac{\sum_{k=1}^{k} d(c(k)).w(k)}{N(w)} \quad (1)$$

Where

d: distance local.

w(k): weight function.

N (k): normalization factor.

The minimization of the dissimilarity D between two words is determined by the dynamic programming based on the choice of weighting function w (k) in a way that the normalization factor is constant.

$$D_c(X, Y) = \frac{1}{N} Min_C \sum_{k=1}^{K} d(c(k)).w(k) \quad (2)$$

**2.1.2. Weighting function:**

Several types of weighting functions have been proposed in the literature for the constant normalization factor.

- .Symmetrical form

$$w(k)= i(k)-i(k-1) +j(k)- j(k-1) \quad (3)$$

The normalization coefficient is:

$$N(w) =\sum_k w(k) = N+M \quad (4)$$

- Asymmetrical

$$w(k)= i(k)-i (k-1) \quad (5)$$

$$N(w) =\sum_k w(k) = N \quad (6)$$

$$w(k)= j(k)-j(k-1) \quad (7)$$

$$N(w) =\sum_k w(k) = M \quad (8)$$

**2.1.3. Constraints:**

To correspond to a physical reality, the temporal functions ik and jk should be growing and respect certain conditions of continuity.

- Condition on the initial and final points.

  The Correspondence between extrimites two words to compare

$$i (1)=1 \; i(K) =N \quad (9)$$

$$j(1)= 1 \; j(K)=M \quad (10)$$

- Monotonicity:

$$i(k+1) \geq i(k) \quad (11)$$

$$j(k+1) \geq j(k) \quad (12)$$

- Continuity:

$$i(k+1)-i(k) \leq 1 \quad (13)$$

$$j(k+1)-j(k) \leq 1 \quad (14)$$

For the purpose of maintaining the complete information

- Allowable Regions

$$| i(k)-j(k)|<r \quad (15)$$

Intended to reject a way which implies an excessive distortion of time axis.

**2.1.4. Organizational structure of the DTW:**

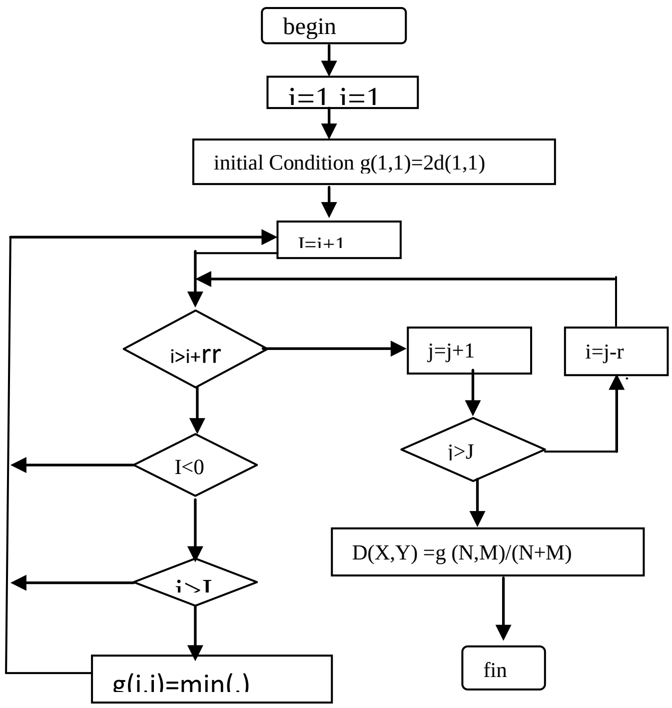


Figure 4: Structure of DTW

**2.1.5. The disadvantages of the DTW method**

- The first problem concerns the standardization factors used in a way that alignment must be constant.
- The second consists of the optimization of path alignment.

To remedy these inconveniences, we introduced another optimization technique such as genetic algorithms which consist of a broad family of stochastic algorithms developed in the 70s. The genetic algorithms were inspired by the Darwinian Theory on the evolution of living beings, based on mechanisms of natural and genetic selection.

## 3. Operating Principle of a Genetic Algorithm

The operation of a genetic algorithm is based on the following phases

### 3.1. Organization of the GA

- An initial population arbitrarily chosen.
- An assessment of individuals of this population is done using an evaluation function called fitness which allows the assessment of their relative performance.
- Based on this performance we create a new population by using evolutionary operators such as selection, crossover and mutation.

- **Crossover:** The crossover is the process allows which permits the combination of genotypes of two parent individuals to get two new descendants.

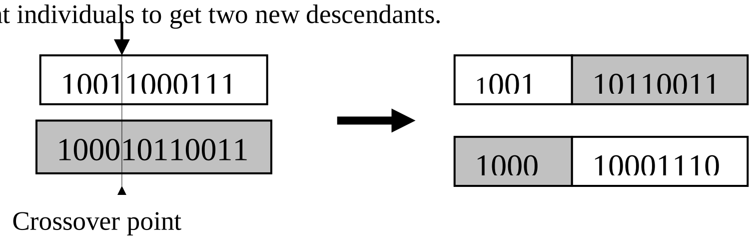


Figure 5: Schematic representation of the crossover

- **Mutation** the process of mutation consists alters the coding of a chromosome.

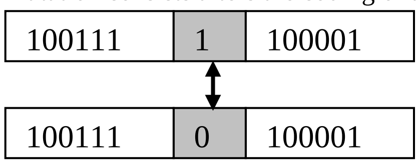


Figure 6: Schematic representation of mutation in a chromosome

### 3.2. Criteria for stopping the algorithm

- Stop the algorithm after a certain number of generations.
- Stop the algorithm when the best individual has not been improved since a number of generations.
- Stop the algorithm when there is loss of genetic diversity.

## 4. Association of genetic algorithms to DTW

### 4.1. The method (GTW)

- coding chromosome (path alignment C) C=c(1)c(2)…c(K)=(i(1),j(1))(i(2),j(2))...(i(K),j(K))
- Creation of an initial population submissive to the constraints of (monotony, continuity, allowable regions).
- choice of a fitness function that will support the measure of distortion for each chromosome (dcn = dn (X, Y)).
- Application of genetic operators (crossover and mutation).

### 4.2. The method GTW-RSW

The GTW combined with the relaxation of the weighting function differs from the GTW by:

- The arbitrary choice of the values of the weighting function w (k) (relaxation).
- The normalization variable factor.

### 4.3. Hybrid crossover with GTW

The hybrid crossover operator whose procedure is listed as follows:

- Randomly select two chromosomes A and B and perform the normal crossover operation .An offspring C is reproduced.
- Swap chromosomes A and B and performs the crossover procedures again. Another offspring D is reproduced.
- Discrimination treatment is carried out, knowing that the best chromosome among A, B, C, D will be in the population in place of the weakest among A and B.

## 5. Application

### 5.1. Presentation of the database

The application is made on isolate words, pronounced by only one speaker representing the word of Berber language. Another version for the test of recognition, such as words are the same tests as those used in the reference but pronounced differently.

### 5.2. Comparative Study

A comparative study between the different methods (DTW, GTW, GTW-RSW, hybrid-GTW) is performed. The comparison between different methods for word recognition testing database is represented by the following histogram:

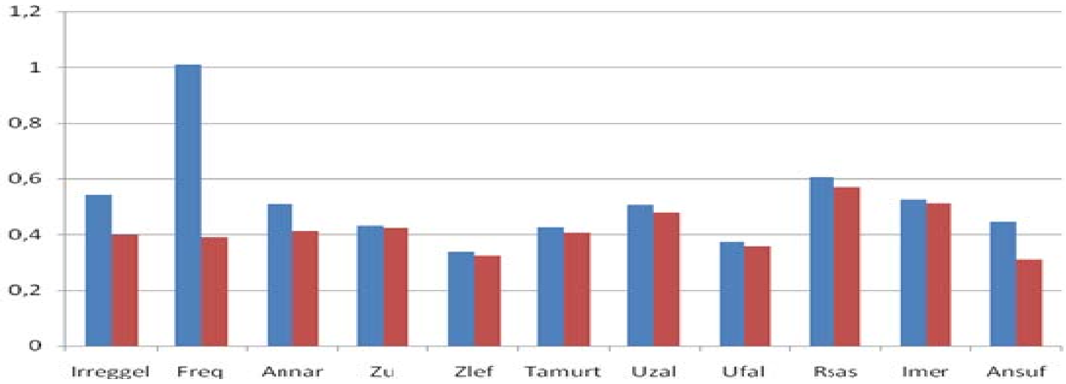


Figure 7: Histogram comparison

We note that the distortions produced by the DTW method is very large compared to the GTW and hybridizations made to the GTW method gives better results.

### 5.3. Influence of the parameters

The last approach we have studied is the influence of different parameters of genetic algorithm on the optimal path and minimal distortion.

- **In terms of population size**

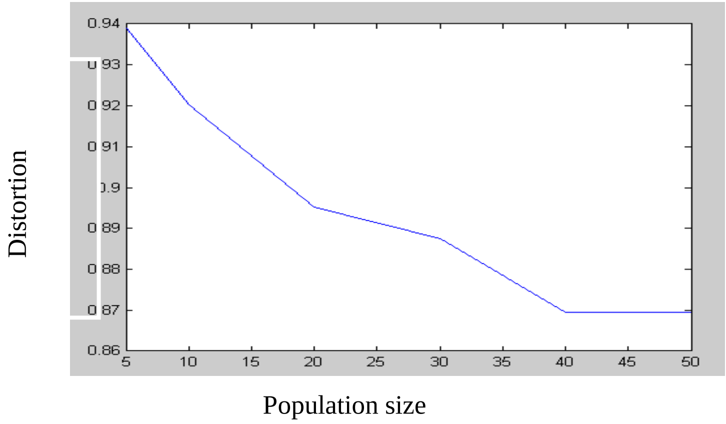


Figure 8: Influence of population size on the distortion

Figure 8 shows the increased number of chromosomes in the population improves the performance of the time alignment path.

- **In the number of generations**

We note that the distortion decreases with increasing the number of generations and it is noted that no improvement is made from the 40th generation.

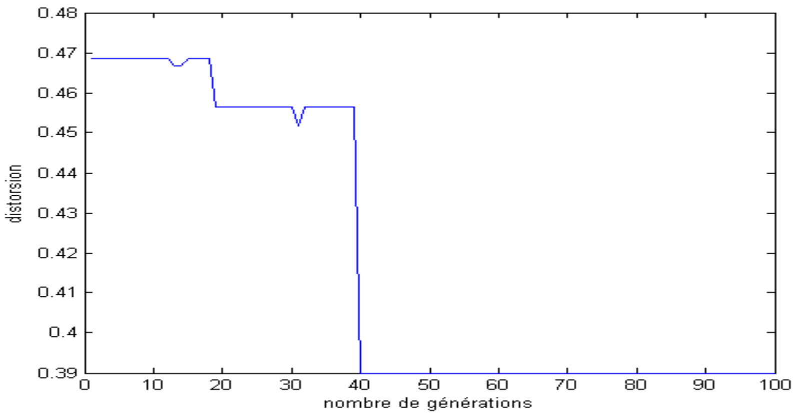


Figure 9: Influence of generation number on the distortion

## 6. Conclusion

The introduction of genetic algorithms has helped to alleviate the problems encountered in the DTW method. They give flexibility in relation to factors of standardization and a better alignment with DTW to. We are limited to the approach and the DTW recognition of isolated words monolocuteur. From our point of view, genetic algorithms can also be used for the optimization of systems for recognition of continuous speech. It can also solve these problems by genetic multi objectives approaches.

## References

1. Haton, J. P., Reconnaissance automatique de la parole, Dunod Edition, Paris, 1991.
2. Rabiner, L.R., A tutorial on hidden markov models and selected application in speech recognition, *in proceedings of IEEE*, vol.77, no 2, pp.257-286, 1989.
3. Berras, C., Reconnaissance de la parole continue, adaptation au locuteur et contrôle temporel dans la méthode de Markov cachés, *PHD Thesis*, Paris University VI, 1996.
4. Schiba, H. S., Dynamic programming algorithm optimisation for spoken words recognition, *IEEE Trans. on Acoustic, Speech and Signal Processing*, vol. ASSP-26, Feburary, pp. 43-49, 1978.
5. Goldberg, D.E., Computer-aided gas pipeline operation using genetic algorithms and rule learning, *PHD Thesis*, Michigan University, 1983.
6. Spalanzani, A., Algorithmes évolutionnaires pour l'étude de la robustesse des systèmes de reconnaissance automatique de la parole, *PHD Thesis* , Joseph Fourrier Grenoble University, 1999.