Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.778012

A VECTOR QUANTIZATION OF THE SPEECH SIGNAL BY A NEW ROBUST ALGORITHM FOR ITS COMPRESSION

Abdelghani Rouini[1], Messaouda Larbi[1]
[1]Applied Automation and Industrial Diagnostics Laboratory (LAADI), Faculty of Science and Technology University of Ziane Achour of Djelfa, 17000, Algeria

Abstract- Our study consists in conceiving a compression of system of the speech signal by the vector quantization. The speech signal is analyzed by the method of Prediction Linear LPC which allows to converter this signal in acoustic vectors. Our system functions in two phases; the phase of training and the phase of coding. In the first phase, a dictionary of prototype vectors is generated by new robust algorithm. During the phase of coding, the speech signal undergoes the same acoustic treatment. Each acoustic vector is quantified and coded by the index of the prototype vector nearest to the vector of entry within the meaning of metric selected. Then, a binary code is attributed to each index. Thus, our system transforms the speech signal into a sequence of index and strongly contributes to its compression.
Keywords— Prediction linear LPC. A dictionary of prototype. Speech signal.

I. INTRODUCTION

Speech processing is today a fundamental component of engineering sciences, located at the intersection of digital signal processing and language processing, this scientific discipline has experienced since the 1960s a dazzling expansion, related to the development of telecommunications resources and techniques. [1-3]

The particular importance of speech processing in this more general context is explained by the privileged position of speech as a vector of information in our human society.

The extraordinary singularity of this science, which fundamentally differentiates it from other components of information processing, is undoubtedly due to the fascinating role that the human brain plays both in the production and understanding of speech and the extent of the functions that it unconsciously implements to achieve it practically instantaneously.

The continuous development of communication techniques with the machine has helped researchers to better understand phonation in order to better model it in order to meet the main objectives pursued in speech processing, the latter presents some distinct techniques, which are as follows: [4-5]

- Recognition of the word.
- Effective voice signal encoding for transmission or recording that extends from MIC encoding to complex algorithms that eliminate redundancy.
- Some medical applications and the study of languages.
- Identification or verification of the speaker.

Our main objective is to study and implement effective algorithm for a Vector quantization of the speech signal and its compression.

1. VECTOR QUANTIFICATION

A spectrum can be objectively represented by different sets of parameters, in particular the spectrum of the autoregressive AR model, which corresponds very well to the envelope of the vocal spectrum, characterized by linear prediction coefficients.
An ordered set of parameters characterizing a vocal spectrum is called an acoustic vector or spectral vector. [6-7]
Signals are therefore interpreted as vectors in a vector space. To calculate the distances between spectral vectors, it is first important to set a metric in the space generated by the types of parameters. To reduce the number of distances to be calculated, we often proceed to a partition of the space of the acoustic vectors, each class is represented by a particular vector called centroid or prototype: we speak of Vector Quantification (QV). [8-10]
Vector Quantification is an operation that generalizes scalar quantification. It concerns the representation of a real vector x=[x1,x2,… ,xk] whose k components are at continuous real values $x \in R^k$ by a vector belonging to a finite set of M vectors.
Thus, to build a vector quantification system is to operate a partition of RK in Ci classes; in each of them, there is a particular vector yi called centroid; Each vector x of Ci will be represented by the centroid yi associated with Ci. The set of M centroids constitutes a dictionary (code-book).
The Q.V. must be organized to minimize the average distortion (average of quantization distortions) for a given M-size dictionary.
Applications of vector quantification are mainly low-rate encoding (less than 8 K bits/s) and recognition.
In fact, for transmission or recording, a ui code is assigned to each yi vector, ie a symbol (word code) belonging to a finite set U. Most often, we use a binary code with B bits; so, we have M=2B
The objective is to reduce the transmission rate of the speech signal by data compression.
To build a vector quantification system, one needs:

1. A large set of acoustic vectors x1, x2…, xL that form the learning set. This set is used to create the optimal set of dictionary vectors to represent the spectral variability observed in the learning set. If the dictionary is of size $M=2^B$ then it is preferable, to create an optimal dictionary, that L must be at least 10M.
2. A measure of similarity or distance between two acoustic vectors in order to partition all learning vectors.
3. A procedure for calculating the centroid of each class, the acoustic vectors of a given partition class must be closer to their centroid than to any other centroid.
4. A procedure for the classification of arbitrary acoustic vectors, which chooses the centroid closest to the input vector and uses the index corresponds to this centroid as the resulting spectral representation. Otherwise, the classification procedure is a quantifier that accepts as input an acoustic vector and that gives as output the index of the centroid closest to the input vector in the sense of the chosen metric.

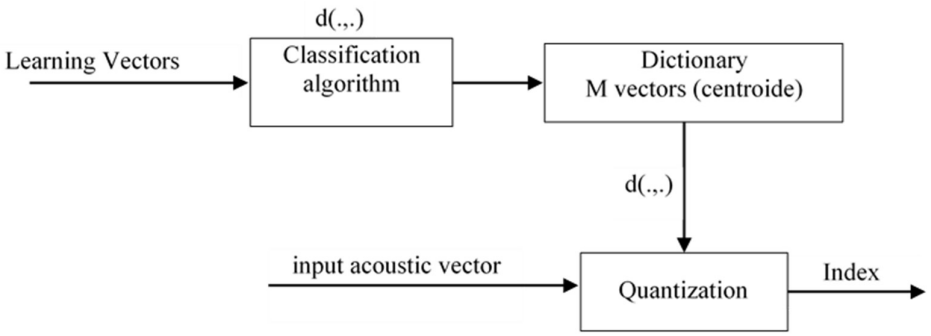


Figure. 1 Basic structure of Vector Quantification

2. DESIGN AND IMPLEMENTATION

The implementation of the vector quantification method for a speech compression system involves solving some problems related to the different stages of the realization of this system. Such a system consists mainly of three stages. The first is the signal processing step where we are interested in the extraction of parameters on each analysis frame. What kind of parameters and analysis methods will be used? The second is the construction of the prototype dictionary. Which QV method will we choose? The third step is the coding of the index representing the prototype closest to the input acoustic vector.

More other problems that require appropriate solutions. Indeed, the choice of a solution or a method to solve one of these problems can influence the accuracy and the computation time, as well as the flow of information.

Our compression system consists of two programs:

- An apprenticeship program;
- A coding program.

The objective of the learning phase is to build the prototype dictionary. While the objective of the coding phase is to quantify each input vector and to assign it a suitable code in the sense of an index of dissimilarity.

A. Software Description

We have composed the following modules for our software (Fig.2):

- A voice signal acquisition and recording module.
- A module of LPC acoustic analysis for the extraction of acoustic vectors.
- A learning module to build the prototype dictionary.
- A vector quantification module to encode each input acoustic vector by replacing it with the nearest prototype index belonging to the quantizer dictionary.

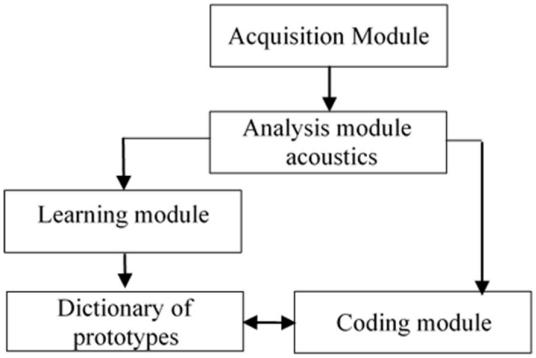


Figure. 2 Overall Software Structure

B. Description of system phases

Our work is divided into two phases:

- Learning phase.
- Coding phase.

Figure 3 shows schematically the structure of the system.

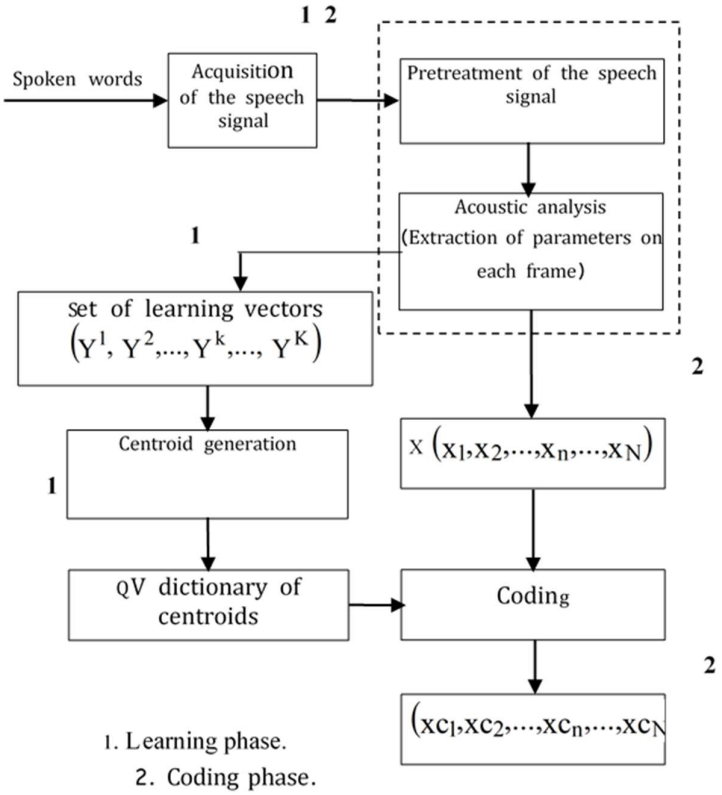


Figure. 3 System Structure

- **Learning phase**

This phase aims to build the prototype dictionary. The learning phase consists of the following modules: the acquisition module, the acoustic analysis module, the prototype dictionary generation module.

a. Acquisition Module

It consists of recording word sounds (in Arabic language), using a microphone, a sound card associated with an interface software.
The voice signal of the words was recorded in the following format: sampled at a rate of 8 KHz and scanned on 16 bits, on the mono channel, then stored on the hard disk in.wav extension files.

b. Acoustic analysis module

After the recording, the acoustic analysis is performed. The objective of this module is to reduce the redundancy of the speech signal, by keeping among all available data only a set of relevant parameters in order to reduce the amount of computation and storage during learning and coding processing.
The parameters chosen are the LPC linear prediction coefficients which are extracted by the autocorrelation method solved by the LIVINSON-DURBIN algorithm. [11-13]
LPC analysis goes through the following steps:

- The calculation and subtraction of the average value of the signal;
- The pre-acceptance of the sampled signal by a transmittance filter (1-0.9375 Z^{-1});
- Splitting the signal into frames with a 50% overlap.
- The weighting of each frame by the Hamming window;
- Application of the autocorrelation method on each frame.

The acoustic analysis module is provided with the following data:

- The recorded signal;
- The frequency of sampling
- The duration of an analysis frame;
- the order of prediction p.

At the output of this module, the signal is transformed into a sequence of LPC vectors $x = \{a_0, a_1, a_2, \dots a_p\}$ where p is the prediction order. All vectors are stored in an *.mat.

c. Learning module

The learning phase consists in generating the prototype dictionary. A vector quantification algorithm is previously applied to all learning vectors.

- **Coding phase**

In this phase, each word undergoes the same treatment as the words forming the learning set except for the construction of the prototype dictionary.
A coding module accepting, as input an acoustic vector, provides, as output, the index of the centroid closest to this input.

Presentation of the coding algorithm

The steps in this algorithm are:

- Calculate the distance (ITAKURA) between all word vectors with all centroids
- We determine for each frame the prototype which presents the smallest distance by contribution to the other centroids with the frame considered.
- We substitute each acoustic vector (LPC coefficients) by the index of its optimal centroid of the QV dictionary.

3. RESULTS AND DISCUTIONS

This section describes the tests performed and the results obtained.
The experimental conditions

The recording is done on a sound card (SOUND BLASTER AUDIOPCI 128). The latter works with the interface software (Creative Wave studio) which allowed us to record, process and visualize the digitized speech signal on the screen.
The learning corpus consists of 20 words pronounced per speaker in Arabic language. Thus, our system is classified in single speaker mode. The choice of words in this corpus is not arbitrary, indeed, we have taken as an application, a subset of commands from the Arabic version of Windows Explorer. They are shown in order in Table 1.

Table 1. Words from the learning and testing corpus.

Fille	**New**	**End**	**Zoom**	**Transitio n**	**Searc h**	**Tool s**	**Run**	**Remov e**	**Creatio n**
Canc el	**Feature s**	**Prin t**	**Copie s**	**Show**	**Paste**	**Cut**	**Accep t**	**Cancel**	**Ignore**

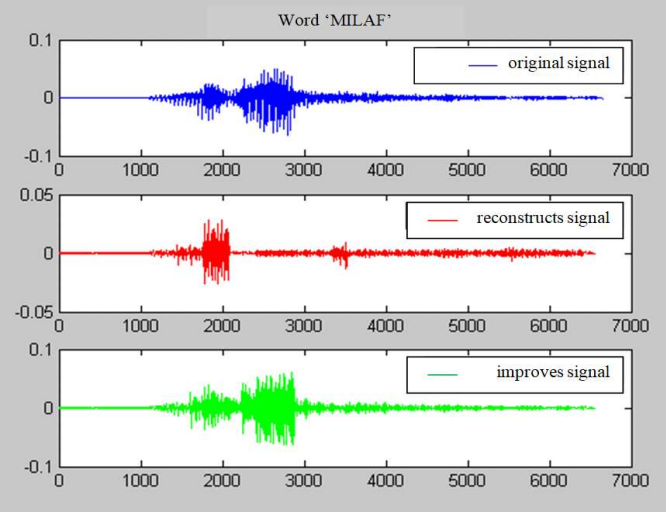
The corpus is recorded with a sampling rate of 8 KHz on a size of 16 bits.
During the analysis phase, the extraction of the acoustic parameters is performed on frames of about 20 ms, with an overlap of 15 ms, fixing the linear prediction order to 11.
The system is implemented using the MATLAB language version 15.

Tests

a. Experience 1

The purpose of this experiment is to evaluate the performance of our system according to the size of the dictionary by setting the following parameters:

- The sampling rate is equal to 8Khz.



- The order of prediction is equal 11.
- The duration of each analysis segment is equal to 20ms in order to obtain the stationarity of the speech signal.

The original signal is reconstructed for different values of M.

To evaluate performance, the distance between the original signal and the reconstructed signal is calculated

We can improve the performance of the system, we add information characterizing the volume of sound on each segment, so each segment will be coded by two coefficients (index of the prototype and the value corresponding to the volume).

The volume information is defined by the maximum value of the segment.

As an example, we took two sounds of the words Millaf (Fille) and Lassek (Paste) , and we found the results presented in the table.2:

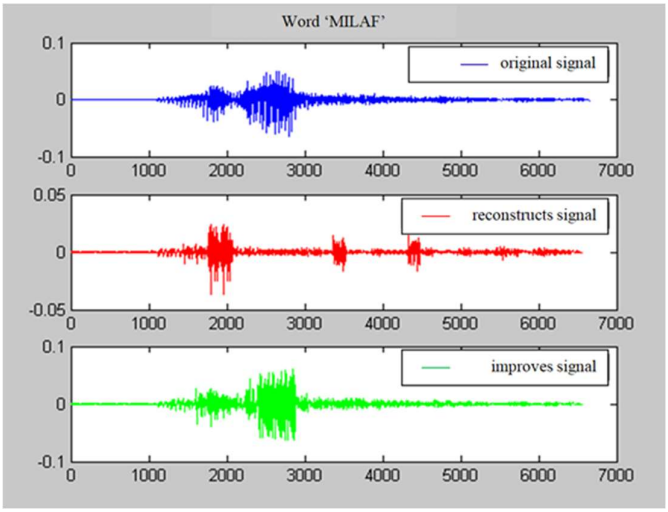
The original and reconstructed signals for different values of M are illustrated by the following figures:

Table 2. Influence of Dictionary Size on System Performance

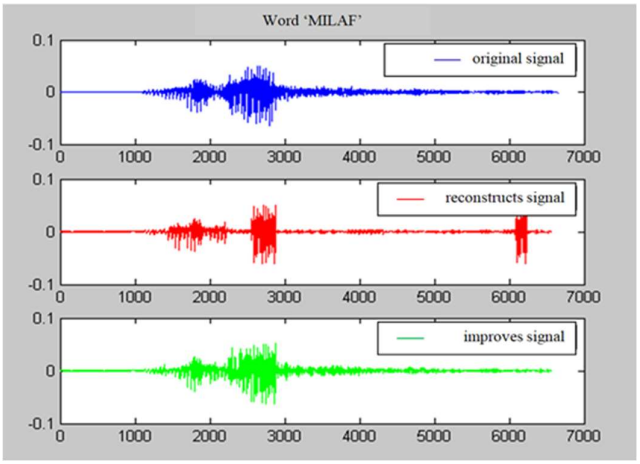
256	128	64	32	M	Word
0.0977	0.1273	0.1565	0.1812	The average distance	MILLAF (Fille)
2.750	1.430	0.820	0.550	Execution time(s)	
0.1181	0.1532	0.1730	0.2095	The average distance	LASSEK (Paste)
2.910	1.540	0.930	0.600	Execution time(s)	

M=32

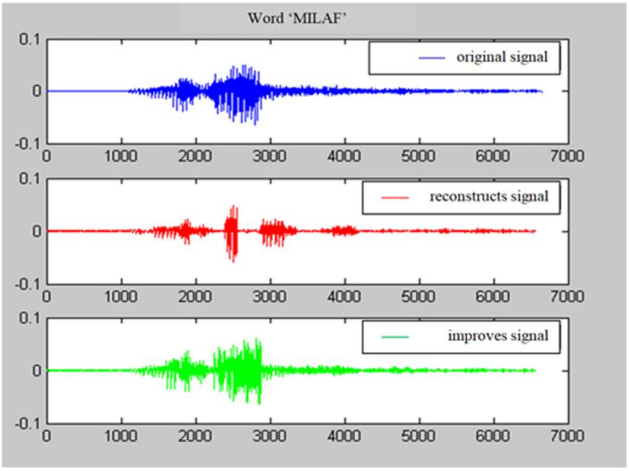
M=64



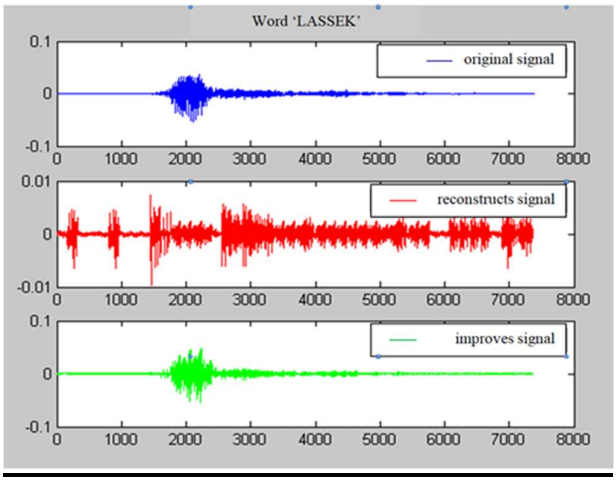
M=128



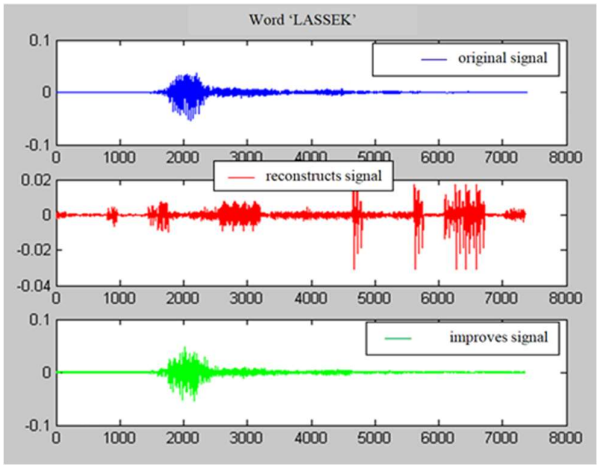
M=256



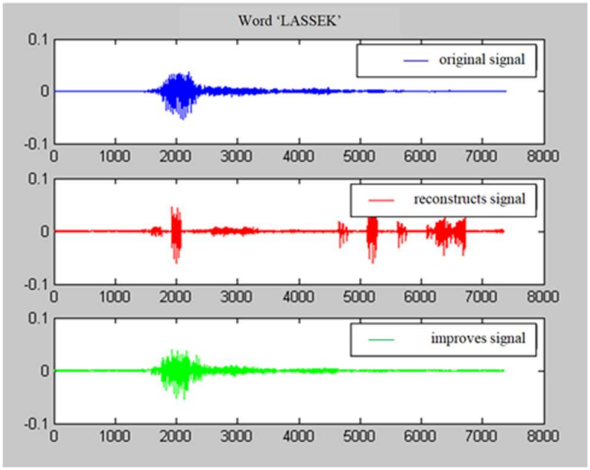
M=32



M=64



M=128



M=256

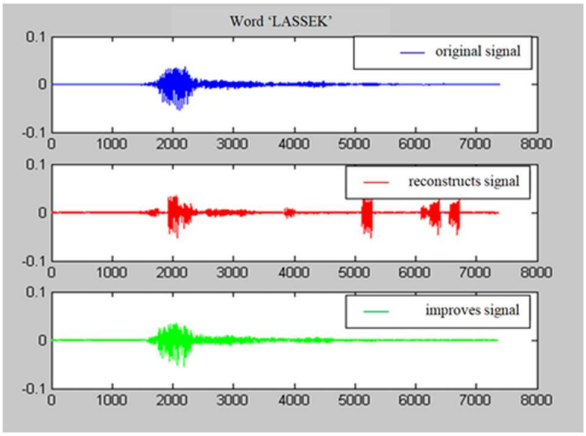


Figure. 4 Curves of the original and reconstructed signal of the words MILLAF and LASSEK

From Table 2, it can be seen that:
• The average distance decreases when the size M of the dictionary increases.
• The execution time increases progressively with the size of the dictionary.
According to the curves of figure. 4, the following points can be noticed:
• The shape of the reconstructed signal is better for the case of signal coding with enhancement.
• The more the value of M increases, the better the resemblance between the original signal and the signal reconstituted with improvement.
Thus, it is necessary to reach a compromise between the execution time, the precision and the memory space, so that the system is efficient.

b. **Experience** 2

The second experiment carried out concerns the choice of the size of the dictionary of prototypes. As we have seen previously, the goal of vector quantization is to minimize the distortion (global distance) which depends on the size of the dictionary. Figure. 5 shows the variations of the average global distance D/L of L learning vectors as a function of the size M of the dictionary.
These results are obtained from twenty words spoken by a speaker contributing to the construction of the corpus, under the following conditions:

- Fs=8KHz.
- The duration of an analysis frame is 20 ms.
- N = 160 (80 recoveries).
- Prediction order 11.

Figure. 5 shows that the average global distance decreases when the size of the dictionary M increases, and that the decrease is small when the size M is greater than 64.
This figure clearly shows, for our case, that there are no big differences between the global distances corresponding to 64, 128 and 256.

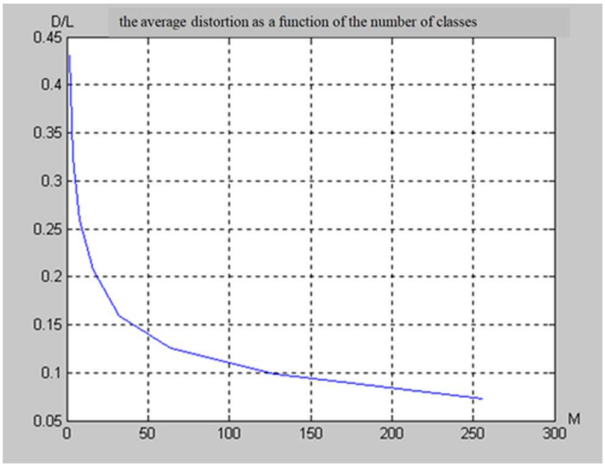


Figure. 5 The effect of QV dictionary size on overall distance

According to the curve above, one can fix the size M to 128 In order to obtain a compromise between the resolution and the execution time at the level of the vector quantization of the signal.

c. **Experience** 3

This experiment allows us to study the influence of the duration of the segment on the precision and the compression rate. The following parameters are set:

- The sampling frequency is 8Khz.
- The prediction order is 11.
- The size of dictionary M is 128.

The analysis frame duration is varied from 10, 20, 30 ms.

Then we proceed to the calculation of the compression ratio, we have two cases:

a. Compression without improvement of the reconstructed signal

The signal x is converted into a sequence of acoustic vectors. Each vector is represented by the index of the suitable prototype. What it does, the signal x is converted into a sequence of indices. If $M=2^B$, then each index is coded on B bits.

The compression ratio Tc is calculated by the following relationship:

$$Tc = \frac{Fs \times b1 \times dt}{b2} \qquad (1)$$

Fs: The sampling frequency

dt: The duration of each frame.

*b*1: The number of bits of a sample before compression equal to 16.

b2: The number of bits after compression.

b. Compression with improvement of the reconstructed signal

Each improved signal segment is represented by two pieces of information:

- The prototype numbers.
- The original signal MAX value in this segment.

So, we do the same calculation as the previous case, but we add to the relation (1) the number of bits of information which represents the MAX value of the original signal.
The following table is filled in for different duration values:

Table 3. the influence of duration on the performance of the MILLAF word system

30	20	10	The duration of the segment (ms)
548.571	365.714	182.8571	Compression ratio (without improvement)
166.956	111.3043	55.6521	Compression ratio (with improvement)
0.1018	0.1273	0.1666	The average distance
0.990	1.530	2.690	Execution time (s)

According to Table 3, we notice that if the duration increases, we will have a decrease in the execution time and the average distance and an increase in the compression ratio.

4. CONCLUSION

The work that we have studied throughout this thesis is based on the compression of the speech signal by vector quantization, using a new algorithm.
For this purpose we analyzed the speech signals by the linear prediction method which can contribute to the reduction of the bit rate of the information.
In order to evaluate the performance of the system, we carried out three test experiments based on the effect of the size of the prototype dictionary, the duration of the analysis segment and the type of information to be encoded. These experiments told us that vector quantization can give good results.
Finally, we recommend that people who are passionate about this area of research use the vector quantization method because it is easy to implement and gives satisfactory results.

Reference

1. M. Arjona Ramírez and M. Minami, "Low bit rate speech coding," in Wiley Encyclopedia of Telecommunications, J. G. Proakis, Ed., New York: Wiley, 2003, vol. 3, pp. 1299-1308.
2. Sayood, K., (1996), “Introduction to data compression”, Morgan Kaufmann Publishers, San Francisco.
3. Vaseghi, Saeed V., (2007), “Multimedia Signal Processing Theory and Applications in Speech, Music and Communications”, John Wiley & Sons Ltd.
4. Bellamy, John C., (2000), “Digital Telephony”, John Wiley & Sons,Inc, Wiley Series in Telecommunications and Signal Processing.
5. Chu, Wai C., (2003), “Speech coding algorithms: foundation and evaluation of standardized coders”, Wiley-IEEE.

6. Makhoul, J., Roucos, S. and Gish, H., (1985), "Vector quantization in speech coding", Proc. IEEE. Vol. 73, pp. 1551-1588, November.
7. Paliwal. K. K. and Atal, B. S., (1991), “Efficient vector quantization of LPC parameters at 24 bits/frame”, in: Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, pp. 661–664
8. Gupta, Shipra (May 2016). "Application of MFCC in Text Independent Speaker Recognition" (PDF). International Journal of Advanced Research in Computer Science and Software Engineering. 6 (5): 805-810 (806). ISSN 2277-128X. S2CID 212485331
9. Ram,M. Satya Sai, Siddaiah, P., and Latha, M. Madhavi, (2008), ”Multi Switched Split vector quantizer”, International Journal of Computer, Information, and Systems Science, and Engineering, Winter,
10. Kleijn. W. B., (1995), “An Introduction to Speech Coding, in Speech Coding and Synthesis”, Elsevier, pp.1-47.
11. B.S. Atal (2006). "The history of linear prediction". IEEE Signal ProcessingMagazine. 23 (2):154161. Bibcode:2006ISPM...23..154A. doi:10.1109/MSP.2006.1598091. S2CID 15601493.
12. Subramaniam, A.D. and Rao, B.D., (2003), “PDF optimized parametric vector quantization of speech line spectral frequencies”, IEEE Trans. Speech Audio Process, vol. 11, Issue 2, pp.130–142, March.
13. Gardner, W. R, and Rao, B. D., (1995), “Theoretical analysis of the highrate vector quantization of LPC parameters”, IEEE Trans. Speech Audio Process, vol. 3, Issue 5, pp.367–381. September.