Application of HMM and Substitution Phonemic

Bouamama Réda Sadouki
Electronics Department,
University of Sidi Bel Abbès
Telecom Laboratory LTT Sidi Bel Abbès, Algeria

Abed Ahcéne
Telecom Laboratory LTT
ENP.10, Hassen Badi Street
BP16200, Algeria

Mouhamed Djebbouri
Electronics Department,
University of
Sidi Bel Abbès Telecom
Laboratory LTT Sidi Bel Abbès,
Algeria

ABSTRACT
The performances of an automatic system of recognition of word are, generally, directly related to quality, the type and the quantity of the data of training. This article shows the effect like the speaker on the performances of a system of recognition of the words isolated directed towards the problem from pathology from the spoken Arabic, in particular, substitution of the spoken Arabic. The system suggested is based on the models of markov hidden (HMM) whose exit is modeled by a density multigaussiennes.

Keywords
Automatic Speech Recognition; Language Pathology; Phonemic Substitution; HMM; Arabic spoke.

1. INTRODUCTION
The automatic treatment of the signal of word is a technology among the techniques most successful with the field of the man-machine communication. It gathers all the tasks relating to the automatic recognition of the speaker and the automatic recognition of word. The automatic speech recognition can be defined like any decision-making process consisting in using the characteristics of the signal of word to extract the linguistic message contained in this signal. Generally, one finds the recognition automatic of the words isolated and the automatic recognition from word continuous. The majority of the systems of automatic recognition of word rests on the model of markov hidden (HMM) [8], he is regarded as the state of the art of these systems. The applications of these techniques are too vast go from the legal applications, of the financial applications but also of the educational applications. Among these applications, the pathology of the spoken language is found. The work in progress is generated in the field of the pathology of the spoken Arabic, in particular, the substitution of the spoken Arabic. This article is organized as follows:

The second section represents the Arab language and one of its pathology, in particular, the substitution of the spoken Arabic. The third section is booked for the problem of the automatic recognition of word. The system suggested and its various tasks are decree in results got with some discussions. And finally a conclusion is represented with the sixth section.

2. PATHOLOGY OF THE SPOKEN ARABIC

2.1 It Arab Standard and Modern
The Arab language is a Semitic language. The Semitic languages are characterized by two classes of distinctive and pompous phonemes [1] [6]. From use point of view modern Arabic is the first language in the Arab world: Algeria, Tunisia, Morocco, etc In the Arab world standard and modern Arabic (MSA) is the official language of the newspapers, the scientific demonstrations, education national, the university researches and the various disciplines of telecommunication.

One of works phonetic Arabic standard and modern (MSA) is consisted 30 phonemes, of which six are vowels and 24 are consonants [1]. The phoneme is the smallest element of the units of the word; it indicates the difference in the direction, the word and the sentence. One distinguishes three long and three short vowels [6]. The concatenation of the phonemes between them produces what calls the syllable. The syllables allowed in the Arab language are: CV, CVC where V indicates a long or short vowel although C indicates a consonant. All the Arab syllables can only start with a consonant [1]. The training of the Arab language for a child Arabic-speaking person passes by several periods. One quotes for examples the prelinguistic period and the period linguistics. In each one of these periods the child acquired certain competences. Generally, with old five years and more the child uses this language in the natural life, it uses it to communicate and express its ideas and its needs with the others. However, certain numbers of child encounter problems of training, these problems can be seen like disorders of the spoken language. Among these disorders one finds substitution phonemic.

2.2 Substitution of Spoken Arabic
The substitution of spoken Arabic is produced when the child replaces a phoneme not yet acquired a very close phoneme articulatory plane [7]. For example, the phoneme / ر / is usually replaced by the phoneme / ب / so the word / راجب / becomes / لاجب / and / ثروحة / becomes / ثلوحة /.

Generally, the substitution is produced when the pivot point is forward or moves backward. When the child replaces the phoneme / ج / the phoneme / ج / eg جوافة / to / جوافة / so the pivot point is advance that the phoneme / ج / is pronounced in the middle of the tongue and the phoneme / د / is pronounced on the side of the tongue. However if the child overrides the phoneme / ق / the phoneme / ء / eg / ئمر / to / تمر / so the pivot point moves back by what the phoneme / ت / is pronounced in extreme language and the phoneme / ء / is pronounced in extreme pharynx.

3. AUTOMATIC SPEECH RECOGNITION
By definition, the automatic speech recognition process is decision to use the characteristics of the speech signal to extract the linguistic message in a voice conversation. It includes two large classes, mention the recognition of isolated words and continuous speech recognition. Practically we can build a system for automatic speech recognition using the following steps:

- Choose the set of acoustic parameters and associated treatments that better represents the properties of the

acoustic signal;

Each of these steps involves many choices and a significant effort in research and development.

3.1 The Acoustic Coefficients

The speech signal is a non-stationary signal and therefore cannot use it directly. So, the speech signal is first transformed to a vector representation this parametric representation must be more compact and suitable for statistical modeling. Various combinations of the acoustic coefficients, articulatory, and were used for hearing thereof [10]. Generally, there are two types of acoustic coefficients: the coefficients (LPC) coefficients (MFCC) and their derivatives.The speech signal first passed through a pre-processing module that segments the speech signal to a sequence of frames (30 to 60 ms).

$$H(n)=\begin{cases}0.54-0.46\cos\left(\frac{2\pi n}{N}\right) & 0\le n\le N-1\\ 0\end{cases} \quad (1)$$

If the sampling frequency is then the window size is used:

$$N=(F_e * t)/1000 \qquad 20<t<40 \quad (2)$$

In order to obtain the short-term spectrum of each frame, the spectral parameters obtained are usually processed to provide a frame representation weakly correlated and reduced dimension. The scan settings are stored in acoustic vector coefficients. The acoustic parameter vector is usually increased by adding other terms such as energy and dynamic coefficients.

$$X=\{x_1,x_2,...,x_T\} \quad (3)$$

At this level, two main methods exist for spectral analysis in speech recognition: filter bank and linear prediction. Linear prediction has been conventionally used several reasons [13 However, these have been the main tool of analysis in recent years as they show better results in the presence of noise [12]. In any case, the spectrum obtained is generally converted into cepstrum.

3.2 Test and Evaluation of Performance

Before marketing any automatic speech recognition, the manufacturer must improve and show its performance. It is important to evaluate the performance of speech recognition system, reliably, based on sets of independent expressions to the training corpus. Typically the error tau words or phrases error tau is for measuring the performance evaluation system. In the current application the error is measured tau words. The error tau words known on the name MCR (misclassification rates) can be defined as the number of words that are incorrectly classified on the total number of words tested.

$$MRC=\frac{Number\ of\ incorrect\ \ words}{Total\ number\ \ of\ words\ tested} \quad (4)$$

Is defined: incorrectly recognized

The rate of substitution:

$$Sub=\frac{Number\ of\ incorrect\ (\text{substitution})\ words}{Total\ number\ \ of\ words\ tested} \quad (5)$$

4. METHODS AND MATERIALS

4.1 Presentation System

Figure (1) shows the proposed for the problem of substitution of spoken Arabic system. As an example, we chose the word / كلام / who contain the phoneme / ك / usually substitute the phoneme / ت /. shows the possible pronunciations of the word / كتاب /.

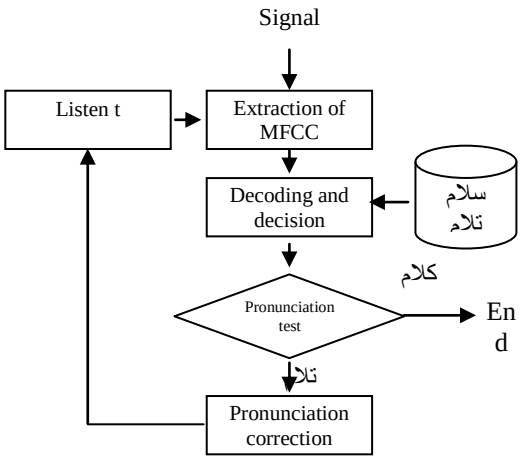

Fig 1: Substitution of Spoken Arabic

4.2 Speech Corpus

A lack of couse database for the problem of pathology of spoken Arabic. We record 20 records of each phoneme references using a set of 12 children age 4. The total number of words used is 250 words, 80 words are used for learning each word references and the rest are used for performance evaluation. During the learning phase of each reference word is represented by a statistical model.

4.3 Hidden Markov Model

Each word is represented by a reference HMM. The data set is denoted by, and connected with the GMM by the equation:

$$p\left(\wedge N|U\right)U \quad (7)$$

including:

N: Represents the acoustic observations .

c_i : Represents the weight vector.

Θ : Represents the GMM model given by:

$$\Theta=\{c_1,...,c_T,U_1,...,U_T\} \quad (8)$$

with:

$$U_i=\{\mu_i,\Sigma_i\} \quad (9)$$

Usually a HMM is given by the vector.

$$\lambda=\{F,O,P\} \quad (10)$$

including:

F : Represents the transition matrix including all statements q_i .

O : Represents the observation matrix including GMM parameters.

P : Represents the initial model of the HMM.

This algorithm is a modified version of the EM [4] algorithm.

The idea of this algorithm is to find the maximum iteratively. This algorithm takes into account all possible paths. However, there is another method called the Viterbi algorithm [5] [14], if the new model provides an improvement in the fit of the data. The initial model is given by:

$$\tilde{a}_{ij} = \frac{\sum_{n=1}^{N-1} \xi_{ij}(n)}{\sum_{n=1}^{N} \gamma_{ij}(n)} \quad (12)$$

Including:

$\sum_{n=1}^{N-1} \xi_{ij}(n)$: represents the number of transitions from state q_i obverse state q_j at iteration n .

$\sum_{n=1}^{N} \gamma_{ij}(n)$: represents the number of outgoing transitions from state q_i to iteration n .

GMM parameters are estimated , mean , covariance matrix and the degree of participation in the global model to the state q_i , they are rated as follows:

$$\tilde{c}_{il} = \frac{\sum_{n=1}^{N} \gamma_{il}(n)}{\sum_{n=1}^{N} \gamma_{i}(n)} \quad (13)$$

with :

l : Represents the 2nd Gaussian observation to the state q_i vectors.

5. RESULTS AND DISCUSSION

The main objective of the work necks and show the effect of gender of the speaker in the learning phase on the performance of a system addresses the problem of substitution of spoken Arabic. As a first experiment, we studied the performance of the proposed system distinguishes between signals records using the male children that signals records with female children. The results obtained are shown in Figure (3).

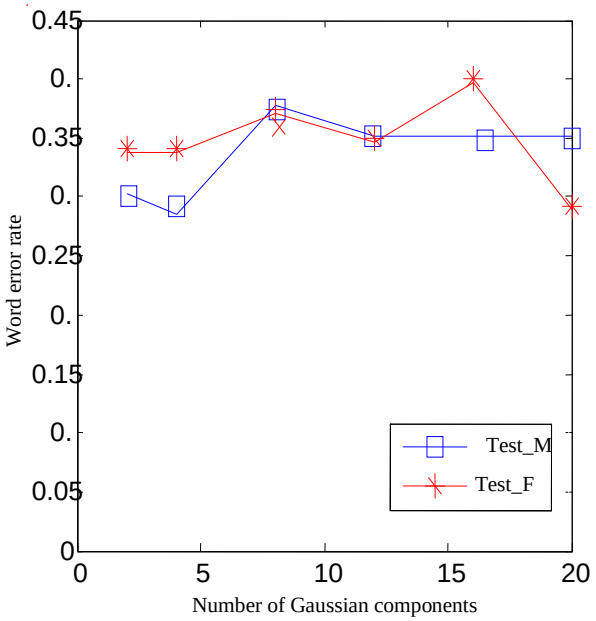

Fig 3: Error Rate Word based on the Number of Gaussian Components Case 1

The results give the tau error of words based on the number of Gaussian components whose training corpus takes only male speakers (Test_M) where the female speakers (Test_F). From these results we see a disturbance on system performance for both tests; performance improves with time and decreased at other times. with the male speakers. The results obtained are shown in Figure (4).

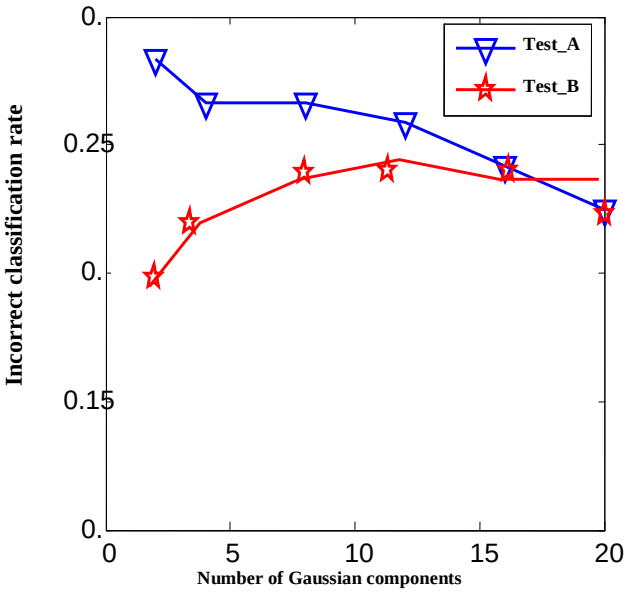

Fig 4: Error Rate Word based on the Number of Gaussian Components Case 2

The two curves show the tau error words depending on the number of Gaussian components. From these results we notice a remarkable improvement on the performance of the system where the error words tau decreases at least 0.23.

5.1 Number of Gaussian Densities

In This part we analyze the influence of the number of Gaussian components included in the GMM model performance evaluation for the substitution of spoken Arabic. For both types tested (sub_ini and sub_fin) the rate of substitution (Sub) is used as a performance criterion. Each word is represented by a HMM of three states or the output is modeled by a multi-Gaussians densities (GMM) number of components ranging from 3, 6, 9, 12, 18 and 22. The size of the acoustic coefficients is 12 including the energy, the first and second Drive ES final resulting in a vector of size 39. The figure shows the results obtained for these two types of tests. It gives the substitution rate based on the number of Gaussian components.

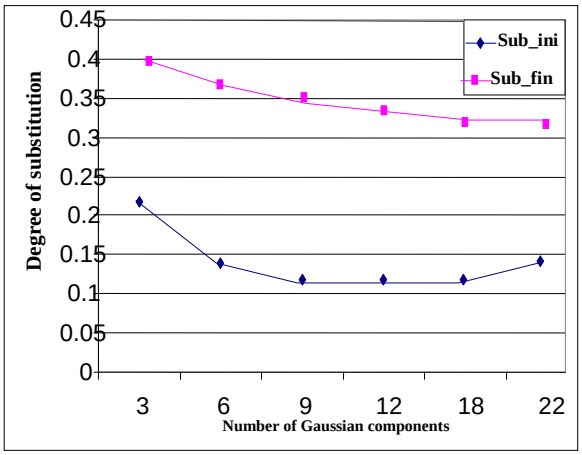

Fig 5: Degree of Substitution on the Number of Gaussian Components

For the first type of test (sub_ini), we note that the rate of substitution cleared rapidly when the number of Gaussian components varies between 2 and 8 when the number of Gaussian components varies between 8 and 16, we observe that the performance of the system stabilize, this means that the minimum number of Gaussian components to obtain adequate performance is 8. For the second type of test (sub_fin), we observe that the stabilization system performance requires a number of larger than in the first type of test Gaussian components, it needs 16 Gaussian components. In another, the performance for this second type of test is less efficient than the performance obtained for the first type of test.

5.2 Number of States

This part is to study the influence of the number of states for the HMM model performance evaluation for the substitution of spoken Arabic. The substitution rate is calculated for the two types of tests: initial substitution (sub_ini) and final substitution (sub_fin

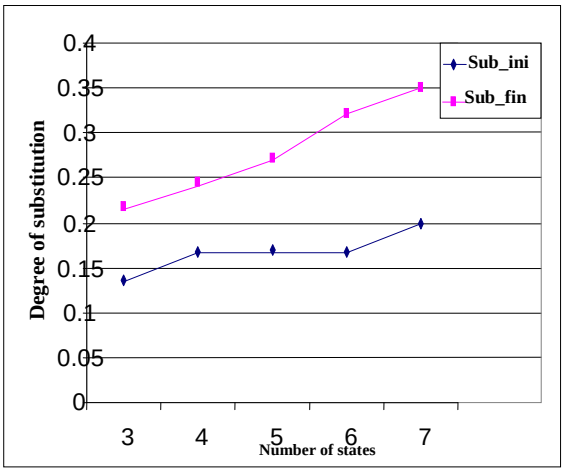

Fig 6:Degree of substitution on the number of states

The results obtained allow us to see that the rate of substitution for the two types of tests (sub_ini and sub_fin) increases when the number of states augments 3 to 6 and are given the best performance when the number of states equals to 3 these two types of tests. In other performance data for the first type of test is still better than the performance data for the second type of test. From these results we can conclude that the model with 3 states work best way for this type of application and given.

5.3 Coefficients MFCC

This part is to study the influence of the size of the vectors of acoustic coefficient on performance evaluation for the substitution of spoken Arabic. The substitution rate is calculated for both types of tests (and sub_ini sub_fin).

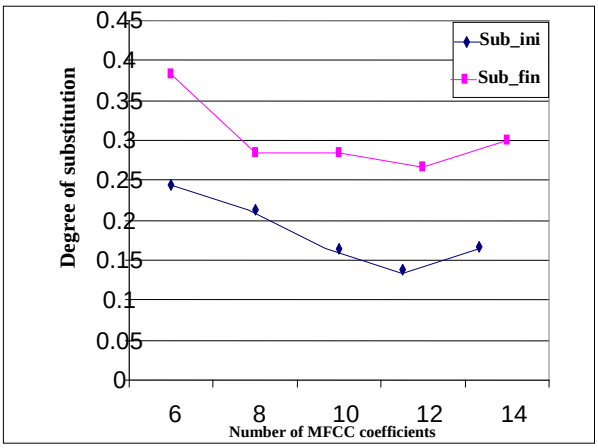

Fig 7 :Degree of substitution on the number of MFCC Coefficients

From these results, it can be seen that the degree of substitution decreases rapidly when the size of the vector coefficients of acoustic exchange 6-12 it is noted for the two types of tests. However, when the size of the vector of acoustic coefficients increases to 14 the degree of substitution for the two types of tests is increasing accordingly the performance is reduced. In another, the performance for (sub_ini) is better than the performance obtained (sub_fin).

5.4 Quantity Learning

This part is to study the influence of the amount of learning on performance evaluation for the substitution of Arabic spoken data. And as the first experiments, the rate of substitution of the two types of tests (and sub_ini sub_fin) is measured.

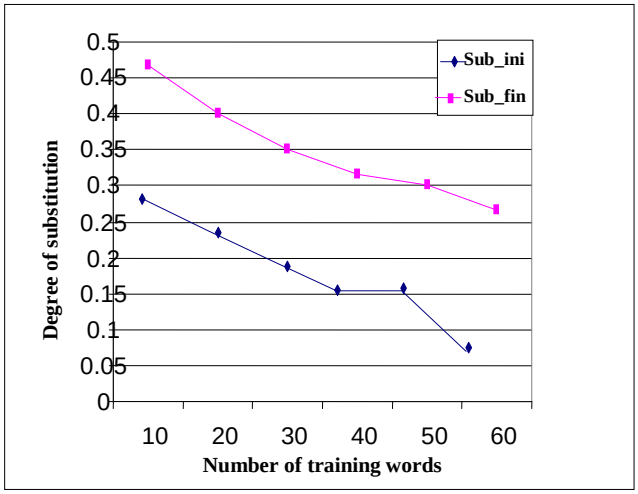

Fig 8 :Degree of substitution on the number of training words

6. CONCLUSION

In this work, we presented a deal with the problem of substitution of Arabic spoken with the automatic speech recognition system. The study system pronunciation of the word / سلام / can be pronounced as / تلام /. The main goal is to even the

effect of child gender on the performance of the system where the error in the final tau reaches 0,34. The results obtained encourage us to use this system for other types of pathology of spoken Arabic is mentioned adding and study the performance based on other parameters such as cepstral coefficients and the number of states ... etc.

7. REFERENCES

[1] Kim Y., Chan Y., Evermann G., Gales F., Mrva D., Sim C., and Woodland C., "Development of the CUHTK 2004 Broadcast News Transcription Systems," in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, USA, pp. 861864, 2005.

[2] L. E. Baum, "An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes," Inequalities, vol. 3, pp. 1-8, 1972.

[3] Nahar K., Elshafei M., AlKhatib G., AlMuhtaseb H., and Alghamdi M., "Statistical Analysis of Arabic Phonemes for Continuous Arabic Speech Recognition," International Journal of Computer and Information Technology, vol. 1, no. 2, pp. 4961, 2012.

[4] Z.A. Benselama, M. Guerti and M.A. Bencherif, "Arabic Speech Pathology Therapy Computer Aided System," Journal of Computer Science 3 (9), pp. 685-692, 2007.

[5] J.A. Bilmes, "A gentle Tutorial of the EM Algorithem and its applications to Parametre Estimation for Gaussian Mixture and Hidden Markov Models, " Technical report, ICSI-TR-97-021, 1998.

[6] M. Elshafei, " Toward an Arabic Text-to -Speech System', The Arabian Journal for Science and Engineering, " 16(4B): , ,pp.565-83, Oct 1991.

[7] A. Eshajhse, "Articulation and speech disorders: types, treatment and diagnostic, " Saudi Arabia, limited golden papers. 1997.

[8] G. D.Forney, "The Viterbi algorithm, " IEEE Proceedings,vol. 61, pp. 268-278, March 1973.

[9] Soltau H, Saon G et al (2007) The IBM 2006 Gale Arabic ASR system. IEEE international conference on acoustics,speech and signal processing, 2007. ICASSP 2007 Stanford Log-linear Part-Of-Speech Tagger, 2011.

[10] L. B. Jackson, " Digital filters and signal processing," New York: Kluwer Academic Publishers, 1996.

[11] Vergyri D, Kirchhoff K, Duh K, Stolcke A (2004) Morphology-based language modeling for Arabic speech recognition. International conference on speech andlanguage processing. Jeju Island, pp 1252–1255

[12] P. Lockwood, C. Baillargeat, J. Gillot, J. Boudy and G. Faucon, " Noise reduction for speech enhancement in cars: non linear spectral subtraction," Proc. Eurospeech 91. Genova, Italia, 1991.

[13] L. Rabiner and B. Juang, " Fundamentals of Speech Recognition," Prentice-Hall, 1993.

[14] A. J. Viterbi, " Error bounds for convolutional codes and an asymptotically optimal decoding Aalgorithm, " IEEE Transactions on Information Theory, vol. IT-13, pp. 260-269, April 1967.