Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 7 (18): 2560-2567, 2007
ISSN 1812-5654

Contribution to the Vocal Print Recognition in Arabic Language

Abdennaceur Kachouri, Tarak Hdiji, Zied Sakka and Mounir Samet
Laboratoire d'Electronique et des Technologies de l'Information 'LETI', ENIS,
Université de Sfax-B.P.W, 3038 Sfax, Tunisia

Abstract: The study presents a new database dedicated to speaker recognition applications. The main characteristics of this Arabic database are spoken by native and non-native speakers, a single session of sentence reading and relatively extensive speech samples suitable for learning person specific speech characteristics. This speech database is dedicated to the modelling and the representation of speakers. The representation consists in extracting parameters (MFCC: Mel Frequency Cepstral Coefficients or LPCC: Linear Prediction Cepstral Coefficients) that characterize the voice or a speaker's vocal print from isolated words either linked from the Arabic database prepared for this work. The technique used in the phase of recognition adapted to this type of data and that showed more performance is the one of HMM (Hidden Markov Models). One tidy preparation of the training database and a good choice of entrance parameters permits to finish an effective model.

Key words: Speaker recognition, MFCC, LPCC, Isolated words, HMM, Arabic database

INTRODUCTION

In many applications (control of access, criminology, banking transactions, GSM...etc.), it is indispensable to characterize an individual by a print in order to differentiate it of others without no ambiguousness; this print is a key coding a person's identity without redundancy or variability.

The creation of the vocal Arabic print is one of the delicate topics in the domain of Arabic word recognition. Due to its morphological, syntactic, phonetic and phonologic properties, the Arabic language is considered to be one of the most difficult languages for written and spoken language processing. Research on Arabic speech processing has made significant progress due to more improved signal processing technologies and to recent advances in the knowledge of the prosodic and the segmental characteristics of Arabic and the acoustic modelling of Arab schemes (Barkat, 2005; Gravier, 2000; O'Shaughnessy, 2000).

In this study, Arabic words were investigated from the speech recognition problem point of view. A probabilistic approach of Hidden Markov Models based speech recognition system was designed and tested with automatic Arabic word recognition. In particular the study deals with the following issues: (1) investigation of the spelling and other problems of Arab language in automatic speech recognition; (2) the development of a new Arabic database to be used in the recognition process; (3) presentation of the Arabic system identification including speech pre-processing, feature extraction and classification recognition phase and (4) an objective evaluation of the performance of the process experimental recognition system.

THE ARAB LANGUAGE IN AUTOMATIC SPEECH RECOGNITION

Introduction: The originality of Arab phonetics is based, mainly on the relevance of the duration in the vowel system and on the presence of emphatic consonants and of the feature of gemination. These particular aspects play a fundamental role in the nominal and verbal morphological development (Selouani and Caelen, 1996). The Arabic language is a very rich language on the level graphematic, phonetic, phonological, morphological and syntactic. In the same way all these aspects are interlaced. It is imperatively necessary to attack all these aspects adequately if one wants to automate this language correctly (Bayeh *et al*., 2004; El-Imam, 2004).

The emphasis: The emphasis is a phonetic feature characterizing 4 consonants, (02) plosives:/T/,/d/and (02) fricatives: /δ/, /S/. These consonants are articulated in the former part of the oral cavity, the root of the language is deferred behind against the posterior pharyngeal wall and a digging of the language is observed. Indeed, in a context VCV (Vowel Consonant Vowel), it is very difficult to keep the character emphatic of /D/ and generally it is its opposite by this feature, /D/, which is carried out.

Corresponding Author: Abdennaceur Kachouri, Laboratoire d'Electronique et des Technologies de l'Information 'LETI', ENIS, Université de Sfax - B.P.W, 3038 Sfax, Tunisia Tel: (+216) 98 438 956 Fax: (+216) 74 275 595

Long vowel and problem of the duration: The parameter lasted is very significant in the Arab language. It characterizes not only the geminated vowels, but also consonants. This characteristic compensates for the poverty of the Arab vowel system. As well at the grammatical level as at the semantic level, this parameter is fundamental. Concerning this feature, a double problem arises in automatic recognition of Arabic: It is necessary to detect the lengthened phonemes all while making sure that this prolongation is relevant; i.e., by distinguishing it from lengthening due to the flow from elocution, with a particular accent of speaker.

For example, two words: /jamal /(camel-جمل) and /jamaal /(beauty-جمال) differs only by lengthening from the final vowel. One requires system of recognition to detect the 2 vowels without deteriorating the temporal property. A temporal alignment, on the contrary, would penalize this detection.

The gemination: The school traditionalist of the Arab grammarians considers that the feature of gemination is an unfolding of the consonant (to pronounce a consonant in a way supported induced this feature on it).

The gemination is defined as being the succession of two marked identical consonants consecutively. In Arabic, the gemination expressed using the symbol " ” " (الشدة, alchadda). This symbol plays a significant role in the definition and the direction of certain words. For the Arab language the parameter of duration is very significant as well at the semantic level as at the grammatical level. It characterizes not only the geminated vowels, but also consonants. This characteristic compensates for the poverty of the Arab vowel system. Concerning this feature, a double problem arises in automatic recognition of Arabic: it is necessary to detect the lengthened phonemes all while making sure that this prolongation is relevant, i.e., by distinguishing it from lengthening due to the flow of elocution, with a particular accent of the speaker.

Indeed if one observes the example of the word صلى/salla: /to request derived from the root صلا which is not opposed that by the gemination of the consonant ل to the word صلى/salted /(to roast) derived from the root صلى, we measure at the same time the importance and the difficulty of an automatic system of detecting this feature.

The gemination appears by the reinforcement of the articulation and a prolongation of the closing of the plosive or continuing other consonants. There too, the school traditionalist is essential by the fact that it regards the gemination as a simple unfolding of the consonant (Barkat, 2005). It is obvious that there is a difference in duration notable between the geminated consonant and its simple counterpart.

Several studies similar to this work were presented for other languages, where the gemination is regarded as a remarkable feature, in particular that for Italian (Giovanardi and Di Benedetto, 1998), the Greek and the Indian. For standard Arabic no study is made to our knowledge within the framework of the spontaneous word. Work of (Jomaa, 1993), is interested by the effect of the quantity of contrast for sequences VVC (succession of Vowel-Vowel-Consonant) and CCV (Consonant-Consonant-Vowel) and to the vocalic quantity for the dialectical Arabic.

In Khouja *et al.* (2005), the results proved that the duration of the geminated consonant was appreciably double among that simple, as well as the duration of the vowel preceding this consonant. Indeed, the duration of the consonant and the vowel which proceeds it can be a source of decidability for the gemination. To be able to solve the problem of the gemination, in a system of continuous automatic speech recognition for the Arab language, it would be necessary to differentiate between a gemination and a simple consonant followed by a long vowel. That passes by a good acoustic automatic approach for the determination of the borders between consonant and vowel.

Morphology of the standard Arab language: The Arabic language morphology represents a special kind of morphological systems (Tahir *et al.*, 2004). Indeed, it is a concatenate morphology in some cases, for instance, the case of regular masculine plural, regular feminine plural, dual and the conjugation affixes. But in other cases, it is a non-concatenate morphology characterized by the manipulation of the two essential factors "root" and "scheme". The use of these two factors makes the majority of morphological rules perfectly regular (except for the case of the "weak" roots).

Ambiguity of syntactic decision: In Arabic language, there are some cases in which morphological information can't provide the right syntactic information for the words. Here are some examples that we can mention in this way:

Some exceptions about the agreement:

- The agreement in number between the verb and the subject depends on the position of the subject in relation to the verb; indeed we have the following rules:
- Subject + verb? Agreement

Example: التلاميذ يلعبون the students are playing.

- Verb + subject? No agreement

Example: يلعب التلاميذ the students are playing.

- In some special cases with which we have to make, one chooses or not the agreement in the kind between verb and subject.

Example:

لا يَنفَعُ الخَاسِرِينَ نَدَمًا مِنْهُم /lâ yanfa'u-l-jâsiRîna nadâmatuhum/: no agreement
لا يَنفَعُ الخَاسِرِينَ نَدَمًا مِنْهُم /lâ tanfa'u-l-jâsiRîna nadâmatuhum/: agreement

- The implicit pronouns that can play a main role in the sentence's syntactic structure, although they are not present in the sentence.

Morphological Irregularity: In Arabic language, there are some morphological irregularities; we can mention some examples of them:

- We can in some special cases suppress the last syllable of a name that comes after a particle of call.

For example, we can say: أفاطِمَة مهلا /'afâTimu mahlan/" instead of saying: أفاطِمَة مهلا /'afâTimata mahlan/". This operation is called الترخيم (we note that this operation is optional).

- While concatenating particles, some of them lose their initial shape.
- There are some exceptional cases where the schemes can not provide deterministic linguistic information of the word.

Because of the particularity of the Arabic language morphology we suggest to construct a particular database that is adapted to the characteristics of this language (Tahir *et al*., 2004).

ARABIC DATA BASE

Acoustic features: With regard to the acoustic features, they carry in particular on:

- The nature of the signal: word, music, either noise...
- Sources of registration and their natures: number of microphones, type of microphone...
- The transmission channel type: narrow strip (telephone), large strip (studio)...
- The speaker: the identity, the kind, the emotional state, the pathological state.

So, the problems met at the time of the recognition (Automatic Recognition Speaker's or Automatic Recognition Speech: "ARS") come of the big variability of the content of a data base. Two categories of problems are to underline: problems owed to conditions of registration on the one hand and problems bound to the nature of the word on the other hand.

Conditions of registration: Conditions of registration influence the resonant document content. Information bound to the transmission channel and the material of registration are elements coming to disrupt (to distort and to degrade) the signal. According to the used transmission channel, some more important information losses are measured. The telephonic registrations, for example, are degraded further that registrations in studio. With the massive arrival of the mobile telephony of new phenomena as losses of information and the compression of data bring new difficulties. The material of registration (in particular microphones) has the own technical specificities that influence the signal. It has been shown that these conditions of registration have an impact on performances of systems of ARSs (Van Vuuren, 1996).

Documents can contain a registration coming from several sources. The nature of these sources is generally the word, but also of music or noises bound to the environment.

Speech nature: Although the signal of word varies according to the individual (what permits to differentiate individuals between them), variations intra speaker gives back problematic the tries of recognition:

- Experiences show that it exist variations in the signal for one same speaker pronouncing several times the same word or sentence; the speaker is incapable to reproduce to the identical the same word or the same sentence.
- Of variations short-term intra-speaker is as present. They are bound in pathological and emotional state mainly (tiredness, cold, stress...) (Alghazi, 1993; Bänziger *et al*., 2000; Karlsson *et al*., 1998).
- Some long-term variations also exists, they come in particular from the ageing of the individual.

Very sensible, other problems can intervene, for example speakers who change their voice voluntarily or that imitate another speaker. The spontaneous word, in opposition to the word either prepared read, is a factor that gives back more difficult the detection of speakers. This type of word encourages recoveries of voice between speakers (takes a speaker's word whereas interlocutor didn't finish to express) and watch a bigger variability.

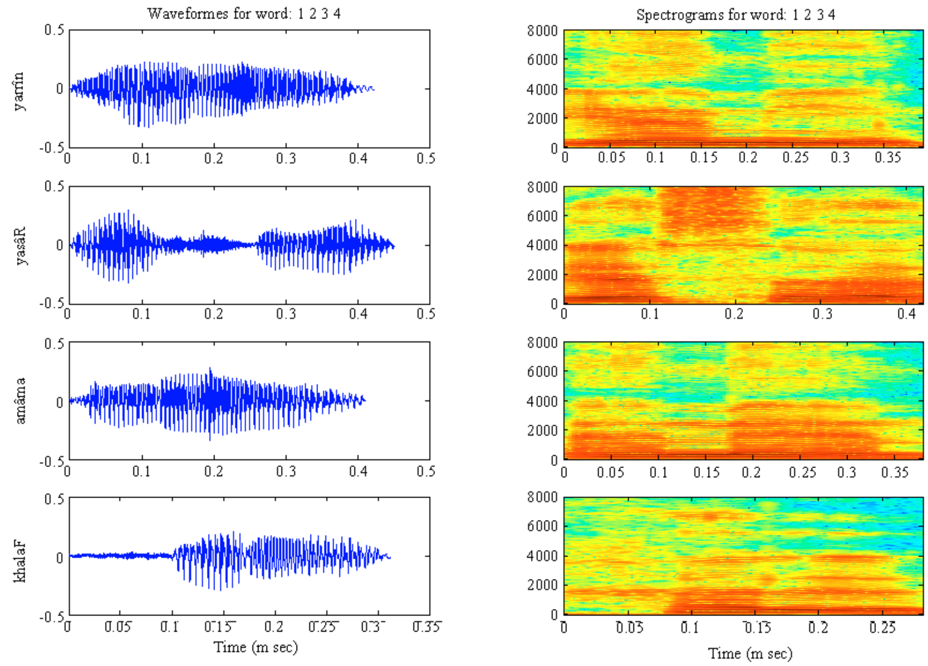

Fig. 1: Waveforms and spectrograms of all Arabic words

Adopted data base: Strategies of the recognition of the word and methodologies for languages multi-syllabic (for example English, French) underwent a research and a substantial development in the two last decades. In the ten past years, techniques of the recognition of the word have also been applied in spoken Chinese language that tries to solve the problem of entered of characters Chinese in systems data processing. The recognition of the applied word on the Arabic language is important and useful like all other language of the same way. It can be applied to solve the problem owed to the wealth of the Arabic language while replacing either increasing other methodologies. The different methodologies accomplished some auspicious results.

That and considering the absence of a corpus annotated in Arab language, we chose the implementation of a system of voice recognition on a basis of data in Arab language worked out by our team of the research laboratory. The done work concerned:

The registration of a data base including 4 words and 2 sentences in Arabic language pronounced by 23 different sex speakers and ages (1800 words and sentences). These data are amplified meadow and sampled with a frequency of 16 KHz spoke following.

Since every speaker pronounces the same word or sentence in succession several times, one been brought to cut up words and sentences to eliminate silences and the intermediate interferences (slip, breathing, hesitation…etc.).

The registration of words and sentences has been achieved in a studio of registration (no noised, not of redundancy, not of bottom noise…) in order to already apply methods of recognition applied on older data base (Example: the TIMIT base) and spoke following to be able to compare results gotten by these bases with our results. Figure 1 show temporal vocal signal and spectrogram for word: يمين/yamîn/, يسار/yasâR/, امام/amâma/, خلف/khalaF/.

In the following of this study, our gait consists in studying and to develop different phases of the recognition system (acquirement, feature extraction and classification).

SPEAKER IDENTIFICATION SYSTEM

Recognition system: The basic elements of a speaker recognition system are shown in Fig. 2. It consists of three main sections:

- A front-end section for extraction of a set of useful discriminative speech features from the time-domain speech samples.
- A middle section that consists of acoustic speech models, a language model, a speech network and speaker adaptation.
- A speech decoder that outputs the most likely word sequence, given the speech feature vectors and a lattice network of acoustic word.

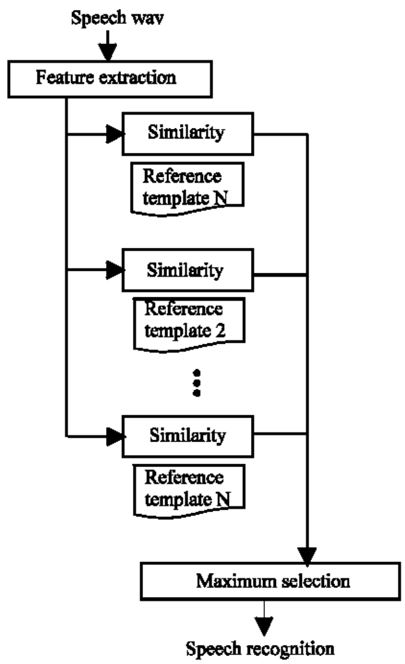

Fig. 2: The outline of a speech recognition system

Front-end feature extraction: The feature extraction subsystem converts time-domain raw speech samples into a compact and efficient sequence of spectral-temporal feature vectors that retain the phonemic information but discard some of the variations due to speaker variability and noise. The most widely used features for speech recognition are cepstral feature vectors which are obtained from a discrete cosinus transform function of the logarithm of magnitude spectrum of speech.

Acoustic speech models: For speech recognition, an efficient set of acoustic models is needed to capture the mean and variance of the spectral-temporal trajectory of speech sounds and to discriminate between different speech sounds. In selecting a speech model we have choosing the Hidden Markov Models (HMMs) described in the next chapter and used for medium to large vocabulary speech recognition systems.

Speech features and models resolutions: The resolution of a speech recognition system depends on the following factors:

- Acoustic feature resolution, i.e., spectral-temporal resolution.
- Acoustic model resolution.
- Context resolution.

The spectral and temporal resolution of speech features are determined by the following factors:

- The speech signal window size for feature extraction (typically 25 ms).
- The rate at which speech features are sampled (usually every 5 to 10 ms).
- Speech feature vector dimensions; typically 13 cepstral features plus 13 first difference and 13 sec difference cepstral features.

Model resolution is determined by:

- The number of models.
- The number of states per model.
- The number of sub-state models per state. For example, when using hidden Markov models, each HMM has N states (typically N = 3-5) and the distribution of feature vectors within each state is modelled by a mixture of M multi-variate Gaussian densities. Therefore the model for each phoneme has N×M Gaussian distributions with each Gaussian density parameterised by a mean feature vector and a covariance matrix.

Hidden Markov model: Since their introduction in speech treatment, the models of hidden Markov took a considerable importance, to the point that the quasi-totality of the automatic speech pattern recognition systems uses this modelling (Young *et al*., 2001). The models of hidden Markov suppose that the phenomenon modelised is an uncertain process and unobservable that appear by broadcasts themselves uncertain. These two levels give to the Markov approach a flexibility that is appealing for represent a phenomenon as complex that the speech production.

Modelling of the speech signal: In the goal to achieve isolated words recognition, we suppose that a machine of Markov represent a word; in a more general case, these models of words are themselves constructed by concatenation of acoustic basis units. The states can be interpreted like configurations of the phonatory device and the observations given out at the time of the arrival in a state correspond to the acoustic plots. These plots are ordinarily represented by vectors of continuous parameters and it is necessary either to amount to the case of the discreet broadcast symbols by vectorial quantification, either of modelised the probabilities of broadcast by continuous densities probability.

Classification by HMM method: The difficulties encountered for the development of Automatic systems speech recognition come from the variability of the speech signal. Among the developed methods, the statistical approach by Hidden Markov Model (HMM) seems the most efficient. We constituted a system of isolated words recognition as the Fig. 2.

The HMM model used is a right left model (two cases are selected: tri-states and five states). Every state is defined by an average and a variance, the algorithm of Baum-Welsh to determine a new average and a new variance using the basis of training every time. The decoding uses the algorithm of Viterbi in order to find the sequence of states likeliest correspondent to the parameters observed in a composite model in order to deduct the corresponding word. The result of the decoding is compared to the reference labels by dynamic alignment in order to count the identified labels, omitted and substituted by another, to insert and to calculate the rate recognition.

Pre-processing and feature extraction: In order to define parameters more the adapted to the Speaker's Automatic Verification (SAV), we undertook to record a big data base of word in Arabic language understanding a big number of speakers.

The phase of preprocessing is important because the choice for one or the other of the possible methods permits to get results of recognition besides or less good quality (Zheng and Zhang, 2000). This stage consists in cutting the useful signal into frames of 256 samples corresponding to 30 ms at 8 KHz sampling frequency. For every frame we apply a Mel filter-bank of order 8.

In this study, we tackled the problem of parameterization by the analysis of the spectre of various speakers of the recorded Arab base. This analysis was based on two types of parameters in order to extract most robust and most effective for our recognition system tested on the Arab data base. These two types of parameters are the LPCC and the MFCC since they showed their evidence in several work of SAR (Speech Automatic Recognition) (Gabzili *et al*., 2003; Hdiji *et al*., 2005).

RESULTS AND DISCUSSION

Experiments were carried out on Arabic speech data base. The system is trained to recognition four words, which are: يمين /yamîn/, يسار /yasâR/, امام /amâma/, خلف/khalaF/, with the two methods of parameterization MFCC and LPCC and under two distinct conditions from recognition: mode multi-speaker and speaker-independent mode. Each word is pronounced by different speakers (male and female). Trained and testing data contained more than 1000 wav files each one represents one word. The recognition system used throughout this study is based on a speech representation by temporal and cepstral parameters and on the modelling of words by HMMs. A pre-emphasis as well as hamming window is applied. The function requires the following parameters: signal, sampling frequency, window type and number of coefficients. Default values are shown in Table 1.

For the mode Multi-Speaker (MS), the base of training is made of 192 words, the first three repetitions of each word pronounced by 16 speakers (16 speakers ×3 repetitions ×4 words). In the phase of test, we use all the 800 words (20 speakers ×10 repetitions ×4 words). That implies that the whole of data of training is a subset of the whole of data of test.

The protocol followed in the speaker-independent mode (SI) is as follows: 10 repetitions of three speakers (speaker 22, speaker 23, speaker 5) used for the training (the training base is made of 120 words (3 speakers ×10 repetitions ×4 words)), the tests are made with the twenty remaining speakers (speaker 1 to 21). All the results obtained with the various combinations [Mode (MS or SI), parameters MFCC or LPCC, a number of HMM states (3 or 5 states)] are presented in the confusion matrix as shown in Table 2-7.

In general, for the multi-speaker mode, the total performance of the system was 93.25% by using LPCC coefficients and five states HMM model; this rate is reasonably high.

The worst rate is obtained in the case of word امام with a rate of 70% in a system of recognition using LPCC coefficients and three states HMM model. The best rate is obtained with the same model with the word يسار with a rate of 100%.

For the speaker-independent mode, the total performance of the system is 93, 13% for a system using MFCC coefficients and five states HMM model witch is acceptable if we consider the acoustic and phonetic resemblance existing between the words in the data base. Table 8 shows the recognition rates of present system for these various combinations.

Table 1: Defaults parameters

Parameters	Default value
Sampling frequency	16 KHZ, 16 bits
Data base	4 isolated words
Speakers	23
Repetitions	10
Pre-emphasized	1-0.95 Z^{-1}
Window	Hamming
LPC order	12
LPCC/MFCC Coef	12/12

Table 2: Confusion matrix relating to the recognition system of Arabic words using 12 MFCC coefficients and three states HMM model (Multi-Speakers mode)

MFCC12/3S/MS	خلف	امام	يسار	يمين	Recognition rate (%)
يمين	1	2	1	196	98.00
يسار	42	0	158	0	79.00
امام	29	164	2	5	82.00
خلف	194	6	0	0	97.00
Total					89.00

Table 3: Confusion matrix relating to the recognition system of Arabic words using 12 MFCC coefficients and five states HMM model (Multi-Speakers mode)

MFCC12/5S/MS	خلف	امام	يسار	يمين	Recognition rate (%)
يمين	1	1	0	198	99.00
يسار	37	0	163	0	81.50
امام	0	196	0	4	98.00
خلف	185	15	0	0	92.50
Total					92.75

Table 4: Confusion matrix relating to the recognition system of Arabic words using 12 LPCC coefficients and five states HMM model (Multi-Speakers mode)

LPCC12/5S/MS	خلف	امام	يسار	يمين	Recognition rate (%)
يمين	1	4	1	194	97.00
يسار	1	0	199	0	99.50
امام	1	172	3	24	86.00
خلف	181	14	5	0	90.50
Total					93.25

Table 5: Confusion matrix relating to the recognition system of Arabic words using 12 LPCC coefficients and three states HMM model (Multi-Speakers mode)

LPCC12/3S/MS	خلف	امام	يسار	يمين	Recognition Rate (%)
يمين	0	6	0	194	97.00
يسار	0	0	200	0	100.00
امام	36	140	22	2	70.00
خلف	182	13	5	0	91.00
Total					89.50

Table 6: Confusion matrix relating to the recognition system of Arabic words using 12 MFCC coefficients and five states HMM model (Speakers Independent mode)

MFCC12/5S/SI	خلف	امام	يسار	يمين	Recognition rate (%)
يمين	2	0	3	195	97.50
يسار	4	0	196	0	98.00
امام	14	161	11	14	80.50
خلف	193	6	1	0	96.50
Total					93.13

Table 7: Confusion matrix relating to the recognition system of Arabic words using 12 LPCC coefficients and three states HMM model (Speakers Independent mode)

LPCC12/5S/SI	خلف	امام	يسار	يمين	Recognition rate (%)
يمين	3	26	2	169	84.50
يسار	0	0	200	0	100.00
امام	19	119	22	40	59.50
خلف	97	3	100	0	48.50
Total					73.13

These results show the influence of the phase of parameterization, the model of the HMM and the speaker mode on the results of recognition and we note so the influence of the phonetic of each word in the recognition rate.

Table 8: Recognition rate evaluation

Parameters	Recognized words	Substituted words	Total No. of words	Percentage
MFCC12/3S/MS	712	88	800	89.00
MFCC12/5S/MS	742	58	800	92.75
LPCC12/5S/MS	746	54	800	93.25
LPCC12/3S/MS	716	84	800	89.50
MFCC12/5S/SI	745	55	800	93.13
LPCC12/5S/SI	585	215	800	73.13

By comparing the rates obtained with other work (Gabzili *et al*., 2003) completed on the basis of data TIMIT, we can notice that they are worse and that is due:

- We don't take into account the features characteristics of Arabic language which are the emphasis, the gemination and duration (see the example of the word امام).
- With the great difference between volume of data of our base and that of TIMIT or other bases.
- In the conditions of recordings of the Arab data base and with cutting of the words.

CONCLUSIONS

The results gotten by the statistical approach by Hidden Markov Model are already satisfactory. But, we consider that this work could be pursued with other models for the speaker's automatic identification. The choice of such approach resides in the hypothesis that is more robust than only one speaker using the acoustic space in totality of the speakers working independently on different acoustic parameter subsets. In this study, we have tested two technique of parametrisation MFCC and LPCC with the classification by the HMM.

Although the rate of recognition is relatively middle, a very tidy construction of the training basis permits to finalize a robust and effective model for the vocal recognition. The increase of the vocabulary drags the multiplication of credentials in the dictionary, therefore of capacities of storage and the number of necessary calculations. Besides it appears the risk to have words that are acoustically near.

REFERENCES

Alghazi, V. *et al*., 1993. Transform representation of the spectra of acoustic speech segments with applications-II: Speech analysis, synthesis and coding, IEEE Trans. Speech Audio Processing, 1: 277-286.

Bänziger, T., G. Klasmeyer, T. Johnstone, T. Kamceva and K.R. Scherer, 2000. To improve the systems of automatic checking of the speaker by integrating emotional variability: Methods and first data. Acts des XXIIIèmes Journées d'Etudes sur la Parole (JEP), Aussois, France, pp: 341-344.

Barkat, M., 2005. Vowel Backing and Vowel Lowering in Arabic Vernaculars, in Encyclopedia of Arabic Language and Linguistics. Versteegh, K. (Ed.), Leiden, Brill.

Bayeh, R., S. Lin, G. Chollet and C. Mokbel, 2004. Recognition of the Spoken Arabic starting from acoustic models about the French. JEP-TALN, Fès, 19-22 avril.

El-Imam, Y.A., 2004. Phonetization of Arabic: Rules and algorithms. Computer Speech and Language, 18: 339-373.

Gabzili, H., Z. Lachiri and N. Ellouze, 2003. A comparative study of parameterization methods providing by HTK for a word recognition system. 2nd International Conference on Signals Systems Decision and Information Technology, SSD'03, Tunisia.

Giovanardi, M. and M-G. Di Benedetto, 1998. Analysis of Singleton and Geminate Fricative in Italien, WEB-SLS. European J. Language Speech, pp: 1-13. http://wrangler.essex.ac.uk/web-sls.

Gravier, G., 2000. Analyze statistical with two dimensions for modeling segmentale of the word signal, Application to the recognition. Ph.D Thesis, ENST, Paris.

Hdiji, T., Z. Sakka, A. Kachouri and M. Samet, 2005. Using hidden markov models for arabic word in vocal print recognition. International Symposium On Computational Intelligence And intelligent Informatics; ICIII'05. Carthage TUNISIA.

Jomaa, M., 1993. Effect of quantity contrasts on the temporal regulation of mandibular movements in Arabic. Acts du colloque du 2ème congrés: Langue arabe et technologies informatiques avancées, pp: 141-169.

Karlsson, I., T. Banziger, J. Dankovicov, T. Johnstone, J. Lindberg, H. Melin, F. Nolan and K. Scherer, 1998. Speaker verification with elicited speaking-styles in the verivox project'. Workshop on Speaker Recognition and its Commercial and Forensic Applications (RLA2C), pp: 207-210.

Khouja, M., M. Zrigui and M. Ben Ahmed, 2005. Acoustic study of the duration of the gemination for the Arab word. SETIT, Sousse-Tunisia.

O'Shaughnessy, D., 2000. Speech Communication: Human and Machine. 2nd Edn., IEEE Press, New York.

Selouani, S. and J. Caelen, 1996. Validation of phonetic features by a standard Arabic system recognition. Proc. des XXI JEP, Avignon, pp: 347-350.

Tahir, Y., N. Chenfour and M. Harti, 2004. Modeling with object of a morphological data base for Arab language, JEP-TALN. Traitement Automatique de 1' Arabe, Fès, 20 avril.

Van Vuuren, S., 1996. Comparison of text-independent speaker recognition methods on telephone speech with acoustic mismatch. Proceedings of International Conference on Spoken Language Processing (ICSLP 96), pp: 1788-1791.

Young, S.J., P.C. Woodland and W.J. Byrne, 2001. HTK Reference Manual, for HTK version 3.1, December.

Zheng, F. and G. Zhang, 2000. The energy information into MFCC. International Conference on Spoken Language Processing, Beijing, pp: I-389~292, Oct. 16-20, Beijing.