Asian Journal of Information Technology 5(8):795-799, 2006
© Medwell Online, 2006

A New Architecture of Isolated Arabicwords Recognition System

Hocine Bourouba, Mouldi Bedda and Rafik Djemili
Automatic and Signals Laboratory of Annaba, Badji Mokhtar University, Algeria

Abstract: In this study, we present a new architecture system of isolated spoken word recognition using HMM model. This study is an alternative study used in speech recognition using sex dependent hidden Markov models. The new study is introduced, evaluated and compared with traditional study GHMM for isolated word recognition system. Both these studyes apply the same principles of feature extraction and time-sequence modeling; the principal difference lies in the the architecture used for training and recognition phases.

Key words: Speech recognition, HMM(hidden Markov model), speech analysis

INTRODUCTION

Our study concerns the Arabic pronunciation Speech is one of the most used ways of communication. Since many years, researchers attempt to conceive devices allowing vocal man-machine dialogue. Their hope is to control many electronics systems by speech, or to control a robot. The computer development has allowed a direct communication man-machine and many great project of speech recognition have been developed in the last years. Today, the speech recognition is in full rapid expansion and we observe the multiplication of the applications domain (robot control, security systems, rolling and moving seats control for handicapped motor bodies, vocal phone number, etc.). Using the speech recognition device, we can do many vocal control operations.

The speech recognition is treated by many methods as the statistics methods (the Hidden Markov Model, the hybrid Model, etc.)[1,2]. The conception of an automatic speech pattern recognition system is made difficult by the complexity of the speech signal as the variability inter and intra speaker. The main reason of the inter-speakers differences is a physiological nature. The speech is produced mainly thanks to the vocal cords that which generate a sound at a base fundamental frequency. This base frequency will be different from one individual to another and more generally from one kind to another, a voice of man being more serious than a voice of woman, the frequency of fundamental being weaker[3]. Currently, the most effective operators of RAP rest on the statistical methods. The Hidden Markov Models (GHMMs) are widely used in number of practical applications and especially suitable in speech recognition because of their ability to handle the variability of the speech signal. Many ways are still open for the improvement of their quality. Among those, the reinforcement of discrimination between models appears us one of most promising.

The main solutions proposed to compensate the lack of discrimination in Markov models intervene in the training phase of the models and in recognition phase[4]. An alternative consists in locally introducing discrimination into the definition of the models. Among the considered methods, the use of neural networks as discriminating estimator of probability[5] proved to be relatively efficient although, equally, expensive and complex to put of it.

The solution, suggested in this study, is the bursting of the models by the parallel use of two Markov model for each word one for the men and other for the woman to decrease intra speaker variability will and thus to increase the models discrimination. Our experiments show that a clear improvement of the rate recognition in traditional system HMM is observed when is intended for only one sex (man or woman). For that we developed a system to two Markov models.

The objective of this study is a recognition device of isolated words. The applied strategy is the Markov model combining two dependent sex models.

GHMM recognition system: Figure-1 illustrates the implementation of the developed HMM recognition system (baseline system). It is an isolated word recognition system based on Continuous hidden Markov model with mixture Gaussian[6,7]. Two steps could be distinguished:

- Training step (Fig. 1), which permits the model parameter estimation.

Corresponding Author: Hocine Bourouba, Automatic and Signals Laboratory of Annaba, Badji Mokhtar University, Algeria

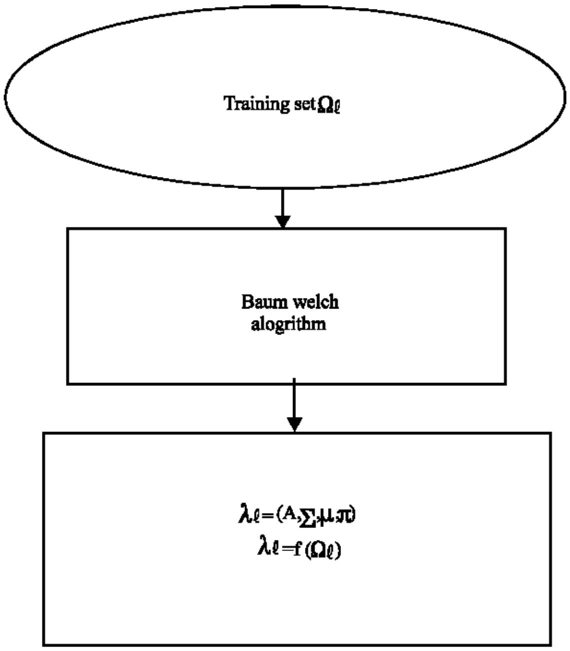

Fig. 1: Training step in baseline isolated word recognition system

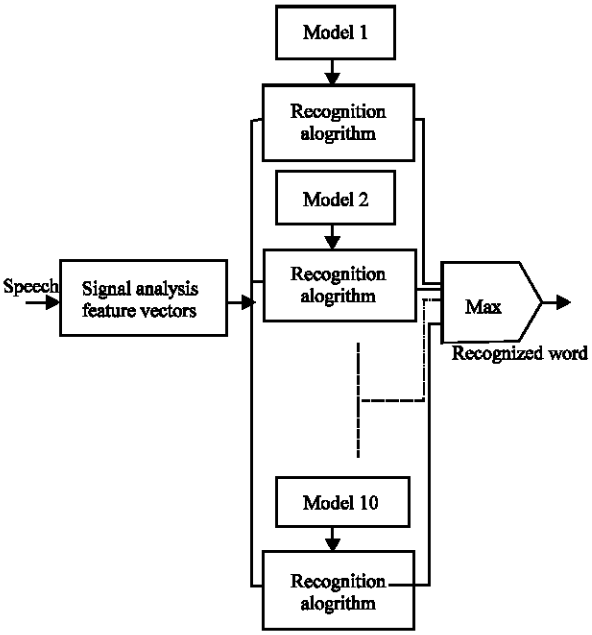

Fig. 2: Recognition step in baseline isolated word recognition system

- Recognition step (Fig. 2), in which the system selects the most probable model to issue the unknown word.

NEW recognition system: In the new system approach same steps are used as baseline system. The full training

diagonal. The models' topologies are displayed in Fig. 1[9].

We assume that L finite sets Ω_ℓ are given, with N_ℓ patterns each (repetitions of the same word). The set of training data is

$$\Omega = \Omega_1 + \Omega_2 = \bigcup_{\ell_1=1}^{L} \Omega_{\ell_1} + \bigcup_{\ell_2=1}^{L} \Omega_{\ell_2} = \bigcup_{\ell=1}^{L} \Omega_\ell \qquad (1)$$

Where Ω_ℓ = {yℓ, 1, yℓ, 2,......yℓ, Nℓ} L is the number of words in vocabulary of the speech recognition system. yℓ is a pattern representing the i^{th} repetition of the ℓ_{th} word. The training set is divided into two sets one of word pronounced by male speaker Ω_1 and the other by female speaker Ω_2 in order to trained two HMM Models for each word. A flow diagram of the training hybrid system step is given in Fig. 3.

Training step: The digitized speech signal is pre-emphases by a first-order digital network in order to spectrally flatten the signal S(n) = S(n)- μS(n-1), the factor μ = 0.95. The signal is fragmented into frames by using a 25.6 ms Hamming window with 10ms shifting. For each frame a Mel Frequency Cepstral Coefficient (MFCC)[10], the log energy E are computed. Each frame is then represented by a feature vector y_t (for training Markov model) as follow:

$$y_t = \{MFCC(12), E\} \qquad (2)$$
$$S = \{y_1, y_2, ...yN_S\}$$

feature vectors S = {y1, y2,.........y_{NS}} of training data $\Omega\ell$ for parameter estimation Markov models. The estimation of model's parameter is done by the Baum-Welch algorithm. Where $\Omega\ell$ = {yℓ,1,yℓ,2,......yℓ,N_ℓ}, L is the number

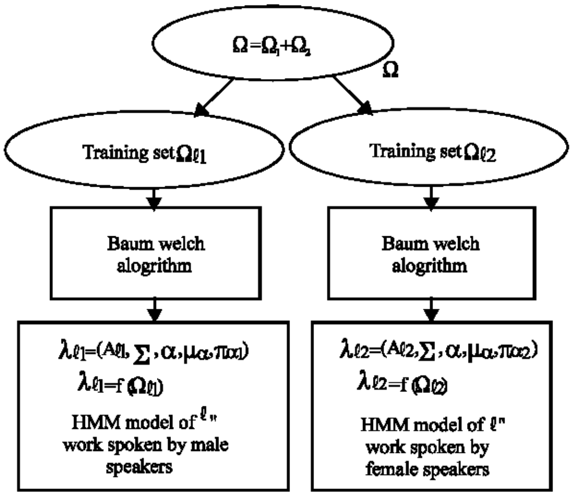
$\Omega=\Omega_1+\Omega_2$
Ω
Training set $\Omega\ell 1$
Training set $\Omega\ell 2$
Baum welch alogrithm
Baum welch alogrithm
$\lambda\ell 1$=(Aℓ1, ∑ , α, μα, πα1)
$\lambda\ell 1$=f ($\Omega\ell 1$)
HMM model of ℓ" work spoken by male speakers
$\lambda\ell 2$=(Aℓ2, ∑ , α, μα, πα2)
$\lambda\ell 2$=f ($\Omega\ell 2$)
HMM model of ℓ" work spoken by female speakers

Fig. 3: Training step in the New recognition system

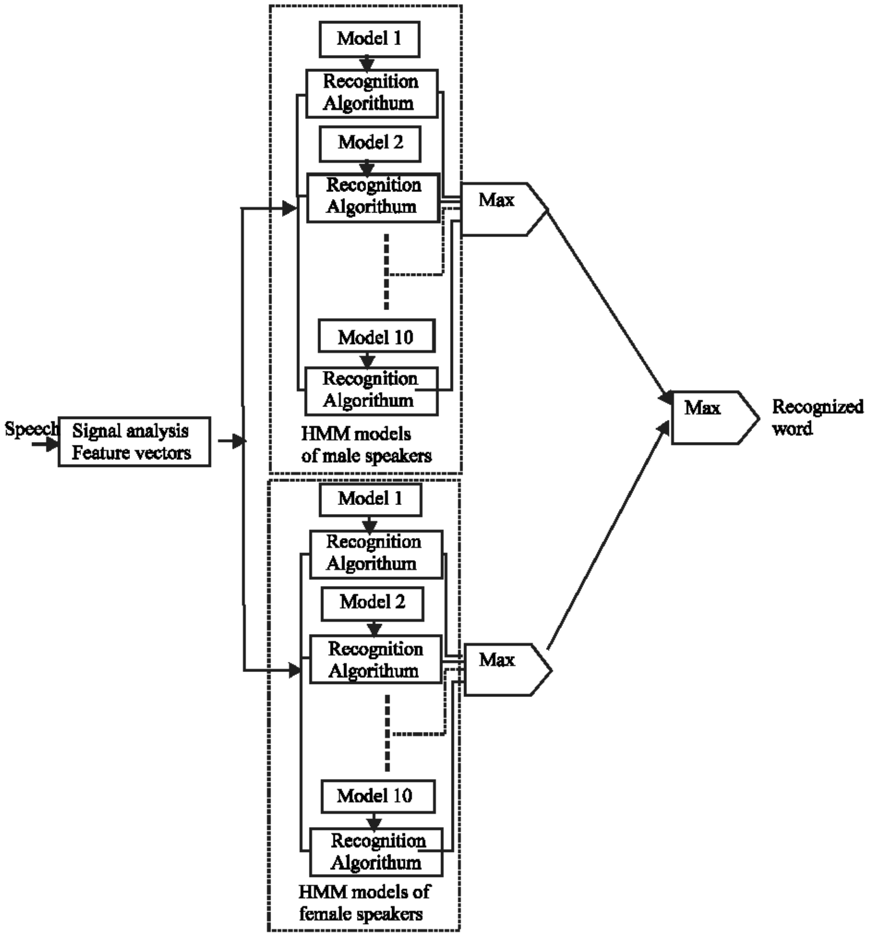

Fig. 4: Recognition step in the New isolated word recognition system

of words in vocabulary of the speech recognition system. $y_{\ell,i}$ is a pattern(sequence vectors) representing the i^{th} repetition of the ℓ^{th} word

Recognition step: For each X(unknown word) we compute the probability P(X,λ) of generating the sequence by the model , we have used the Viterbi algorithm particularly the logarithm of the maximum likelihood. The model with the highest likelihood is selected to issue the unknown word (Fig. 4)[11].

Database preparation: The corpus is constituted of 92 speakers (46 male and 46 female) pronounce each word in Arabic language recorded with a sampling frequency of 11.025 kHz in a size of 16 bits. The speech corpus, divided in a training and test set. this division is displayed in Table 1. The isolated words pronounced are: sifr, wahed, ithnani, thalatha, arbaa, khamsa, setta, sabea, thamania, tessea. The speech files were processed to achieve a MFCC+E (Mel Frequency Cepstrum Coefficients + energie) representation, with a frame length of 23.22 msec. and a frame overlap of 11.61 msec.

Table 1: Subset used in training and recognition set

S1	10 male speakers	S5	S1+S2
S2	10 female speakers	S6	S3+S4
S3	20 male speakers	S7	46 male speakers
S4	20 female speakers	S8	46 female speakers

RESULTS

The results obtained are compared to the baseline system. In the first experiment, the rate recognition of two baseline system depends on the sex (same sex speakers) is evaluated and compared with baseline system in independent of the sex (male and female speakers). In the second experiment the hybrid system is evaluated and compared with the baseline system independent of the sex.

First experiments: In this experiment, the training set consisting of five occurrences of each digit by 10 speaker.

Table 2: Recognition rate of baseline system based sex dependent model (HMM.M, HMM.F) and sex independent model(HMM.MF)

	HMM.M	HMM.F	HMM.MF
Training set	S1	S2	S5
Test set	S7	S8	S7+S8
G=1	93.15	91.61	88.96
2	92.39	93.22	91.78
3	91.89	93.24	92.27

Table 3: Recognition rate of baseline system based sex dependent model (HMM.M, HMM.F) and sex independent model(HMM.MF)

	HMM.M	HMM.F
Training set	S1	S2
Test set	S8	S7
G=1	63.41	42.70
2	53.39	51.20
3	47.96	53.24

Table 2 and 3 show the results obtained with baseline system for male, female and mix speakers respectively for M=1,2,3 (number of Gaussian mixture) and MFCC+E features vector.

HMM.M: hmm models based male speakers

HMM.F: hmm models based female speakers

HMM.ML: hmm models based male/female speakers

Second experiment: In this experiment, the training set consisting of eigth occurrences of each digit by 20 speaker. Table 3 and 4 show the results obtained with baseline system for male, female and mix speakers respectively for M=1,2,3 (number of Gaussian mixture) and MFCC+E features vector.

Table 2 and 4 shows the results obtained with the baseline GHMM system . It can be seen that for the all groups, the best result was provided by the sex dependent models compared to sex independent models. The dependents models gives a more significant rate of recognition compared to the independents model .

Table 4: Recognition rate of baseline system based sex dependent model (HMM.M, HMM.F) and sex independent model(HMM.MF)

	HMM.M	HMM.F	HMM.MF
Training set	S3	S4	S3
Test set	S7	S8	S7+S8
G=1	93.80	91.98	91.08
2	93.89	95.61	93.86
3	93.83	96.65	94.41

Table 5: Recognition rate of baseline system based sex dependent model (HMM.M, HMM.F) and sex independent model(HMM.MF)

	HMM.M	HMM.F
Training set	S1	S2
Test set	S7	S8
G=1	67.54	44.15
2	59.67	58.50
3	63.52	62.39

In another hand, we have observe also (Table 3 et 5) that the recognition of the words pronounced by a different sex speaker that used in training set to obtened HMM models (male speaker by female models or female speaker by male models) presents a an important recognition rate.

Third second experiments: In this expriment, the sex dependent models of male speaker and sex dependent models of femare speaker are combined in a single system (M-G/F-G) (Table 6 and 7).

MF-G: traditional system (mix speaker) with G gaussien

M-G: traditional system (male speaker) with G gaussien

F-G: traditional system (female speaker) with G gaussien

M-G/F-G: New system (combining models)

These results allow us to make the following conclusions:

- The use The sex dependent models in the recognition system presents a big importance compared to the sex independent models.
- The new hybrid system gives a more significant rate recognition compared to the baseline system.

Table 6: Recognition rate of baseline system based sex independent model(MF-G) compared to the new system based sex dependent model (usind depend sex model of the first expriment)

M-1	F-1	MF-1	M-1/F-1
93.15	91.61	88.96	92.33
M-2	F-2	MF-2	M-2/F-2
92.39	93.22	91.78	92.62
M-3	F-3	MF-3	M-3/F-3
91.89	93.24	92.27	92.47
M-1	F-3	MF-3	M-1/F-3
93.15	93.24	92.27	93.18
M-1	F-1	MF-1	M-1/F-1
93.80	91.98	92.08	92.97

Table 7: Recognition rate of baseline system based sex independent model(MF-G) compared to the new system based sex dependent model (usind depend sex model of the second expriment)

M-2	F-2	MF-2	M-2/F-2
93.89	95.61	93.86	94.75
M-3	F-3	MF-3	M-3/F-3
93.83	96.65	94.41	95.28
M-1	F-3	MF-3	M-1/F-3
93.80	96.65	94.41	95.23
M-2	F-3	MF-3	M-2/F-3
93.98	96.65	94.41	95.25

CONCLUSION

In this study, we presented a new architecture system for the isolated spoken words. This study is a contribution to the Arabic speech recognition domain. We have presented search experiments done in order to improve HMM word recognition system. Experimental results shown that the use sex dependent models is important to use sex independent models.

The new architecture presents the property of discrimination in relation to the classic architecture. This system uniting the capacity of GHMM modelling and sex dependent model. The results show that the performances of our new system is better compared to the baseline system.

In the future the many block of the system are study to increase their performance as the HMM model (number of state, number of Gaussian) or to modify the architecture system.

REFERENCES

1. Djemili, R., M. Bedda and H. Bourouba 2004. Recognition Of Spoken Arabic Digits Using Neural Predictive Hidden Markov Models" International Arab J. Inform. Tech. IAJIT, 1: 226-233.
2. Dempster, A.P., N.M. Laird and D.B. Rubin, 1977. Maximum likelihood from incomplete data via the EM Algorithm. J. Royal Statistical Soc., 39: 185-197.
3. Jouvet, D., 1995. Modèle de Markov pour la reconnaissance de la parole Ecole thématique: Fondements et Perspectives en traitement automatique de la parole, Marseille, pp: 99-108.
4. Baudat, G. and F. Anouar. Generalized discriminant analysis using a kernel approach.
5. Renals, S., N. Morgan, H. Bourlard, M. Cohen and H. Franco, 1994. Connectionist probability estimators in HMM speech recognition, IEEE Transactions on Speech and Audio Processing, 2: 161-174.
6. Rabiner, L.R., 1989. A tutorial on hidden Markov Models and Selected applications in speech recognition. Proc, IEEE Trans Speech Process, pp: 77.
7. Huang, X.D., H.W. Hon, M.Y. Hwang and K.F. Lee, 1993. A comparative study of discrete, semi continuous and continuous hidden Markov models. Computer Speech and Language, 7: 359-368.
8. Rabiner, L. and B. Juang, 1999. Fundamentals of speech recognition. Englewood Cliffs, NJ: Printice-Hall.
9. Jelenik, F., 1997. Statistical methods for speech recognition. MIT Press.
10. Zhu, Q. and A. Alwan, 2000. On the use of variable frame rate analysis in speech recognition. Proc. IEEE ICASSP, Turkey, III: 1783-1786.
11. Bochitti, C. and L.P. Ricotti, 1999. Speech recognition. John Wiley, England.