ISCA Archive
http://www.isca-speech.org/archive

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

Phoneme Recognition System based on HMM with Distributed VQ Codebook

***Mohamed Debyeche* [1], *Mohamed Afify* [2], *Jean Paul Haton* [2]**

[1:] Institut d'Electronique, Laboratoire Traitement du Signal
USTHB BP 32 EL-Alia Alger Algérie
Phone: + 213 2 24 71 87 Fax: + 213 2 24 71 87

[2:] CRIN / CNRS / INRIA Lorraine
BP 239 54506 Vandœuvre Lès-Nancy France

Abstract

In this paper a new variant of HMM named distributed VQ HMM (DVQHMM) is presented. Its main characteristic is the use of a code books distributed on HMM states with a new manner of HMM parameters estimation. Procedures for training and HMM evaluation of each recognition unit are described. Comparative results on an isolated phoneme recognition system are shown, between DVQHMM and conventional VQ HMM. The use of the proposed method improves the performance of the system.

1. Introduction

HMM (Hidden Markov Model) is a model which ensures a probabilistic description of different sequence of speech (Word, syllable, phoneme,......). This model is characterized by a discrete output probabilities or continuous output density functions and a topology defined by a number of states, state transition probabilities and initial state distribution. In discrete Hidden Markov Model a conventional vector quantization is used[ ], in this methode the codewords are designed to minimize the average quantization distorsion for all training vectors. In order to improve the discrimination ability of HMM's by minimizing the loss information of signal in the coventional VQ procedure, we propose a new vector quantization method where, the VQ codewords are extracted from the HMM state segment of recognition units. Hence, the codebooks are distributed on HMM states. The recognition accuracy of the HMM based recognition system using the proposed VQ method is evaluated and compared to the case using the conventional codebook. The results indicate a better recognition rate with the new VQ HMM.

2. Conventionnal VQ HMM

The conventionnal VQ codebook design is often associated to LBG or MKM algorithm[1].

HMM parameters estimation

Let the states of HMM be $S=\{S_1, S_2,.........,S_N\}$ and the output symbols $Y = \{Y_1, Y_2,.............,Y_M\}$ respectively. Then, an first order HMM is define by aset of states transition probability $A = \{a_{ij}\}$, a set of output symbol observation probability $B = \{b_j(k)\}$ and initial states probability $\Pi = \{\pi_i\}$.

Given an output symbol sequence $Y_1^T = \{y_1, y_2, ...y_T\}$ for training and λ an HMM model, HMM parameter a_{ij} and $b_j(k)$ can be reestimated by Baum-Welch algorithm[2].

To describe the procedure for reestimation HMM parameter, we first define define $\varepsilon_t(i,j)$, the probability of being in state i at time t, and state j at time t+1, given the model and the observation sequence, i.e.

$$\varepsilon_t(i,j) = \Pr(S_t = i, S_{t+1} = j | Y_1^T, \lambda)$$

From the definition of the forward $\alpha_t(i)$ and backword $\beta_t(i)$ variables [2], we can write $\varepsilon_t(i,j)$ in the form

$$\varepsilon_t(i,j) = \frac{\alpha_t(i)a_{ij}b_j(Y_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(Y_{t+1})\beta_{t+1}(j)}$$

Using the above formulas, we can give a method for reestimation of the parameters of an HMM. A set of reasonable reestimation formulas for A and B is:

$$a_{ij} = \frac{\sum_{t=0}^{T-1}\alpha_t(i)a_{ij}b_j(y_{t+1})\beta_{t+1}(j)}{\sum_{t=0}^{T-1}\alpha_t(i)\beta_t(i)}$$

where

$$\gamma_t(i) = \sum_{j=1}^{N}\varepsilon_t(i,j)$$

$$b_j(k) = \frac{\sum_{t=1,O_t=Y_k}^{T-1}\gamma_t(j)}{\sum_{t=1}^{T-1}\gamma_t(j)}$$

3. Distributed VQ HMM

From the idea that the discrimination ability of the HMM model can be improved if the codewords of codebooks are observed with concentration in certain state, we have built a seperate codebooks distributed on HMM states. The steps which permit the codebooks construction are as follows:

- vector quantize all the training data in a single codebook by using LBG algorithm and estimate, HMM parameters of each recognition unit.
- divide each recognition unit into state segments by the Viterbi algorithm (backtracking).
- group all the segments of similar state and then build a codebook for each state.

HMM parameters estimation

It is known that the effect of the output symbol observation probability on the performance of the HMM is greater than that of a state transition probability. So, we have reestimated in our work, only the discrete output probabilities. The restimation formula for $b_j(k)$ is

given by:

$$b_j(k) = \frac{N_k^{(j)}}{N_j} \qquad (1)$$

where:

N_k: :number of vector in the class represented by the centroïde k.

N_j : total number of vector in state j.

4. Recognition System

The main parts of the recognition system based on DVQHMM are:

Signal Processing

The overall observation vector used is the 12 MFCC coefficient and the weighted corresponding delta cepstrum coefficients.

Structure of HMM

Recognition units are the four arabic emphatic consonants /s/, /d/, /∂/ et /t/. The model used for eachrecognition unit is first order, left-to-right, Markov model with five states (Bakis Model).

Generating Phoneme Reference Model

In order to generate phoneme models from a training data set of labeled speech obtained by using the APHAK software[8], first the conventional VQ and Baum-Welch algorthm are used to estimate HMM's. Secondly, each recognition unit is divided into state segments by using Viterbi algorithm, and grouped per state, hence a codebook for each state are generated. HMM parameters, especially $b_j(k)$, are estimated by using equation (1).

Recognition

In the recognition, we first perform vector quantization for each frame using all codebooks of all states. Thus,

for a frame sequence x_1^T of lengh T frames, we have the symbol sequence $\{ O_t^{(j)}, 1 \le t \le T, 1 \le j \le N \}^{(2)}$. We, then, perform a conventional Viterbi search but using symbol in sequence given by (2).

4. Primilary Experiments and Results

Evaluation of the recognition system is done by the calculation of the rate of the phonetic recognition. This rate is obtained by comparing the system output to the hand labelled sentences. The results in both multispeakers and speaker independant are summurized in table-1 and table-2.

Phoneme	Present	Recognition rate (%)	
		CVQ	DVQ
/d/	118	70	79
/t/	172	65	70
/s/	224	81	84
/∂/	56	63	69

Table-1: multispeakers results

Phoneme	Present	Recognition rate (%)	
		CVQ	DVQ
/d/	118	65	78
/t/	172	60	66
/s/	224	70	78
/∂/	56	58	62

Table-2: speaker independant results

5. Conclusion

This work is contributes to speech recognition based on HMM techniques. We have presented a new variant of VQ quantization and parameters estimation for HMM system recognition. The first experiments carried out for consonants, specific to arabic language,have shown that the proposed method has contributed to improve the recognition rate of the system.

References

[1] : Y. Linde , A. Buzo et R.M. Gray "An algorithm for Vector Quantizer" IEEE Trans on Comm , Vol 28 , N°1, 1980

[2] : L.R. Rabiner "A tutorial on hidden Markov Models and Selcted applications in speech recognition" Proc , IEEE Trans Speech Process , Vol 77 , N° 2 , 1989.

[3] : S.E Levinson, L.R. Rabiner, and M. Sondhi " An introduction to the application of the theorie probabilistic functions of a Markov process to automatic speech recognition, " The Bell Syst. Tech. J, Vol 62 pp 1035-1074, 1989

[4] : L.R. Rabiner , J.G. Wilpon et F.K. Soong " Hight Performance Connected Digit Recognition Using Hidden Markov Models" IEEE Trans on Acoust Speech , and Signal Processing , Vol 37 , N°8, 1989.

[5] : F. Brugnara , D. Falavigna et M. Omologo "Automatic Segmentation and Labeling of Speech Based on Hidden Markov Models" Speech Communication 12 pp 357-370 , 1993.

[6] D. Jouvet, "Modèle de Markov pour la reconnaissance de la parole " Ecole thématique: fondements et Perspectives en traitement automatique de la parole, Marseille, pp99-108, 17-25 juillet 1995.

[7]: M. Debyeche,A. Houacine et J.P Haton "A Knowledge-based approach for Arabic Amphatic consonant Identification Based On Speech Spectrogram Reading " 30th Southeastern Symposium on System Theory 5SST98 West Virginia, USA, pp325-328, 1998.

[8]: M.Debyeche S. Khelifi.& J.P Haton, "APHAK: Un Logiciel Interactif d’Analyse du Signal de Parole," JTEA 97.