HIDDEN MARKOV MODELS IN SPEECH RECOGNITION

J. Krajčovič [1), M. Hrnčár [2), Ľ. Muzikářová [2)

[1) Siemens Program and System Engineering s.r.o,
Bytčická 2, 010 01 Žilina, tel.: +421 907 898 215, mail: jan.krajcovic@siemens.com
[2) Katedra riadiacich a informačných systémov, Elektrotechnická fakulta ŽU v Žiline
Univerzitná 1, 010 26 Žilina, tel.: +421 41 513 2260, mail: martin.hrncar@fel.uniza.sk, ludmila.muzikarova@fel.uniza.sk

Summary: Voice control is one of the perspective areas of interdisciplinary field called Human Machine Interface (HMI). By the voice control of machines, we usually deal with the recognition of commands from previously defined set. There are many different techniques of speech recognition. At presence, Hidden Markov Models (HMM) are among the most frequently used. This article includes basic terms as well as required mathematical apparatus from stated area.

1. INTRODUCTION

Generally systems for voice recognition assume that processed signal is coded message. In case that we want to make a reverse operation original sequence has to be analog speech signal which is converted to a sequence of vectors of discrete parameters. Basic assumption is that the speech signal is stabile. In such a case the speech signal should be represented by the sequence of the vectors of the parameters. Sequences are sufficiently described by classical statistic methods, moments of first and second degree of power spectrum for approximation. In the work are the most frequently used spectrums or linear predicting coefficients, eventually derived representations.

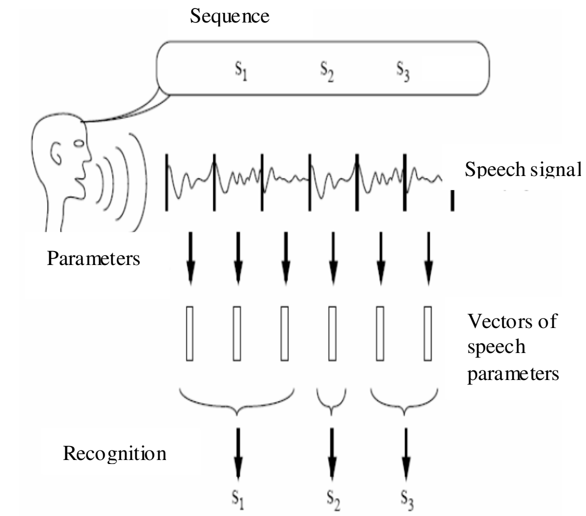

Fig. 1 Process of speech recognition

The task of a recognizer is to create an assignment among the sequences of the speech vectors and basic sequences of symbols which we want to recognize. Described transaction is complicated by two factors [1]:

- The assignment of the speech symbols is clearly defined because similar speeches can be generated by different symbols.
- It is not possible to clearly establish a border among these symbols hence the speech signal can not be directly represented by a sequence of chained static data.

2. HIDDEN MARKOV MODEL

Considering huge variability of the speech signals we need to choose enough robust instruments for process of reorganization. Hidden Markov Model (HMM) meets this condition. HMM connects attributes of homogenous Markov chain therefore for its complete description is necessarily to know the vector of initial allocations π_0 with the length N statuses which describes probability on the beginning of the modeling in chosen status as well as the transition matrix A. The transition matrix has dimension N x N and the item a_{ij} reflects to the probability of the transition from status i to status j. For calculation of the probability of the occurrence of arbitrary status in time t are determined Chapman-Kolmogorov's formulas [3]:

$$P(X_n = j) = \sum_{i=1}^{N} P(X_n = j | X_0 = i) \cdot P(X_0 = i) = \\ = \sum_{i=1}^{N} A_{ij}^n \cdot P(X_0 = i) = \pi_0 \cdot A_{ij}^n \quad (1)$$

In this way we can identify and describe stationary sectors which can be identified by vectors of speech attributes and also can be described by Probability Density Function (PDF). Generally for each hidden status exists different PDF (analog modeling) or more precisely, probability of generated vector of attributes (discreet modeling) [5]. For each word or other basic entity (phoneme) structure of the model is designed (count of hidden statuses and transitions among them) in advance and on the basis on training we can determine the exact probability of transitions among statuses and equivalent PDF. By this is the model determined and also we can assert that with recognition is understood finding of an already trained model which has the highest priority to generate a sequence of attributes corresponded to searched signal. For problems with recognition of speech is used method of model linkage one after another. Moreover with definition of a suitable noise model we do not need to use detecting algorithm [2]. For huge values of SNR (Signal to noise ratio) that significantly improves process of recognition.

For discreet models is used exact probability. This model has to have restricted number of all possible vectors of attributes in uniquely determined collection with size L vectors. Quantization of vectors which uses correlation among vector items is used for this purpose. It divides the vector space into zones according to acoustical scales of distance to the vectors of code book (so called centroid) [3]. It is worthy to remark that this method does not have analytic solution and has to be done by iterations. Except the matrix A and vector V which were mentioned above yet is a discrete model described by the matrix B with dimension N x L (where N is number of status and L is number of centroids). Its item b_{ij} is also described like $b_i(O_j)$ and indicates the probability of generating of input vector j in status i.

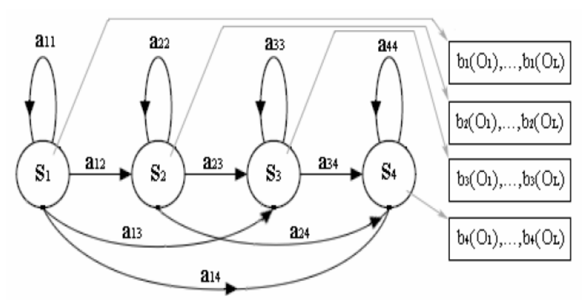

Fig. 2 Four statuses discreet HMM

For analog HMM the probability of occurrence of input vector is modeled by probability density function [4]. Thus we lost duty of using of vector quantization because each observation can theoretically have infinite number of values. For density function is used Gaussian distribution

$$f(x,\mu,U)=\frac{1}{\sqrt{(2\pi)^d\cdot|U|}}e^{-\frac{1}{2}(x-\mu)U^{-1}(x-\mu)}, \quad (2)$$

where d is size of the input vector, x is vector of average values for each item of the vector of attributes and U is covariance matrix. In practice is often assumed that covariance matrix is diagonal what causes significant decreasing of data and decreasing of computing severity. For better modeling is frequently used the mix of the Gaussian functions what better describes environments, type of speakers and so one. The probability of occurrence observed vector is then

$$b_j(x|\lambda)=\sum_{i=1}^{N}c_{ji}\cdot f(x,\mu_{ji},U_{ji}), \quad (3)$$

where c_{ji} are weights of components.

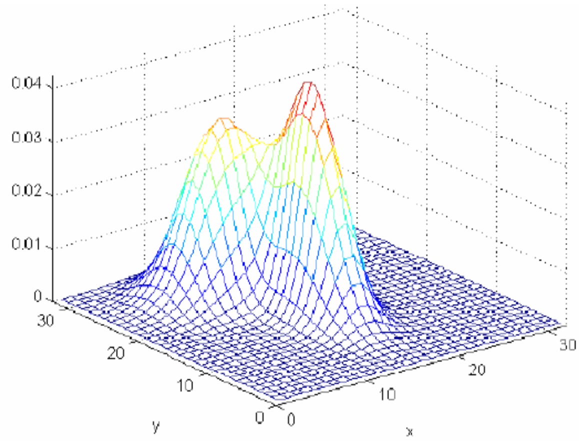

Fig. 3 The mix of two two-dimensional Gaussian PDFs

During of the modeling with HMM are solved three problems:

- How to calculate probability of occurrence of input sequence of vectors $O=\{o_1,o_2..\}$ for model λ.
- How to find out parameters of the model.
- How to decode sequence of hidden statuses $S=\{s_1, s_2..\}$ which the best describes the input sequence O.

3. ADAPTIVE HMM ALGORITHM

In the next chapter, two HMM schemes will be described. The first is represented as HMM-EKF (Hidden Markov Model – Extended Kalman Filter) and contains only non-linear items for the information signal model [3]. The second scheme consists of HMM-KF (Hidden Markov Model – Kalman Filter) items and is made of simplified assumptions of the KF results for the channel approximation bound with the HMM filter for approximation of the signal state.

3.1 Adaptive HMM with EKF

First, let's define $\underline{X}_k=(\alpha_k,x_k)^T$. From this equation we can derive the following formulas:

$$K_k=\sum k|k-1h_k\left(H_k\sum k|k-1H_k+R_k\right)^{-1}, \quad (4)$$

$$\sum k|k=\sum k|k-1-$$

$$-\sum k|k-1H_k\left(H_k\sum k|k-1H_k+R_k\right)^{-1}H_k^T\sum k|k-1, \quad (5)$$

$$\sum k+1|k = F_k \sum k|kF_k + G_k Q_k G_k^T \,, \qquad (6)$$

Equation (4) describes the Kalman's gain and equations (5) and (6) are Riccati's equations.

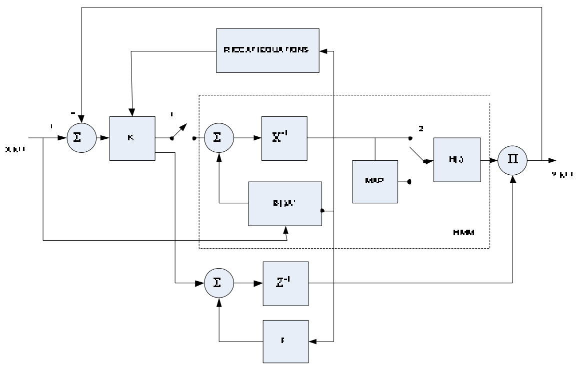

Fig. 4 The EKF-HMM scheme for the adaptive HMM filter

Figure 4 displays a block diagram for the adaptive HMM diagram if switch 1 is closed and switch 2 is in its upper state. If the switch 1 was in the open state, the diagram would work as a HMM-KM filter. Next, let's think of a state for a maximum priority approximation of α_k which can be acquired by turning the switch 2 in the lower state. This approach will be similar to that of using matching filter which work on the principles of using the similarity of message symbols.

3.2 Adaptive HMM-Filters with HMM-KF schemas

The next diagram may seem to be derived from the HMM-EKF diagram assuming that the Kalman's gain is zero. The logical meaning of this statement is confirmed by an example when the parameters describing the transmission channel are constant and asymptotically drop to zero. If the changes in the channel are slow, we expect the items of the Kalman's filter to be asymptotically low. The final diagram may look as a composition of HMM filter with a Kalman's filter.

Figure 5 displays the block diagram. In fact, hybrid versions may be derived by entering low values of the input parameters.

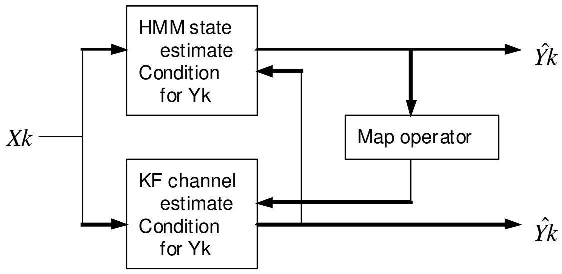

Fig. 5 HMM-KF adaptive diagram with a MAP approximation

4. CONCLUSION

This article offers basic information from HMM theory. Matlab software seems to be the most appropriate way of simulation of HMM. Results of simulation of utilization of HMM for the purpose of automatic speech recognition will be published in a forthcoming article.

Acknowledgement

This paper is supported by cultural and educational grant agency within project KEGA K-057-06-00 „Innovation of laboratory education methodology for modeling and simulation in Matlab and its combination with educational models by e-learning".

REFERENCES

[1] Steve Young, Gunnar Evermann, Thomas Hain, Dan Kershaw, Gareth Moore, Julian Odell, Dave Ollason, Dan Povey, Valtcho Valtchev, Phil Woodland: The HTK Book, Cambridge University Engineering Department, 2002.

[2] J. Kotuliaková, G. Rozinaj: Číslicové spracovanie signálov 1, FABER Bratislava, 1996, ISBN 80-967503-1

[3] E. Keller: Fundamentals of speech synthesis ans speech recognition, John Wiley and sons, 1994, ISBN 0-13-015157-2

[4] Conings, 1. B., V. Krishnamurthy and J. B. Moore: Recursive prediction error techniques for adaptive estimation of hidden Markov models, Proc. 12th IFA C World Congr., Sydney, Jrdy 1993

[5] Rabiner, L. R.: A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE, 1983, ISBN 257-285