DISCRETE HMMs FOR CLASSIFICATION OF MIXTURES OF SIGNALS

*Christophe. Couvreur** *Vincent Fontaine* *Henri Leich*

TCTS, Faculté Polytechnique de Mons

Boulevard Dolez 31, B-7000 Mons, Belgium

Tel: +32 65 374042; Fax: +32 65 374129

e-mail: {couvreur,fontaine,leich}@tcts.fpms.ac.be

ABSTRACT

The concept of mixtures of discrete HMMs (MDHMM) is introduced. The application of MDHMMs to the classification of mixtures of signals is described. The optimal decision rule is presented. Alternative algorithms with reduced computational load are proposed: a simplified decision statistic is defined and sub-optimal search methods are discussed. The performance of the various algorithms are analyzed on Monte-Carlo simulations.

1 INTRODUCTION

Discrete hidden Markov models (DHMMs) have been widely used for signal classification purposes. In these applications, the signal to be classified x_τ is transformed into a sequence of discrete symbols $\{o_t\}$ by a pre-processor; the sequence is then fed to a DHMM classifier (Fig. 1). The classifier chooses the DHMM that best "matches" the sequence $\{o_t\}$ among the possible DHMMs. This "matching" is generally made in a probabilistic sense, that is, the sequence is assigned the DHMM with the maximum *a posteriori* probability. In speech recognition, the pre-processor usually performs some spectral transform (e.g., LPC, cepstrum, filter bank, ...) of a sliding window of the signal followed by a vector quantization. In frequency tracking, the pre-processor is a set of harmonic signal detectors [1].

In many situations, we observe not a single signal x_τ of interest, but a combination of such. For example, in environmental sound recognition [4], multiple sound sources (e.g., cars, trucks, airplanes, animals, ...) can be present and their contributions sum up in the signal recorded at the microphone. Let us assume that there are c simultaneous signals $x_{i,\tau}$, $i = 1, \ldots, c$, each of which has to be classified, and that a family of possible DHMMs for the signals is available. If the signals were observed separately, it would be possible to use the classification scheme of Fig. 1 on each signal: pre-process each $x_{i,\tau}$ to get sequences of symbols $\{o_{i,t}\}$ and classify these sequences according to their "match" with the DHMMs. In practice, however, we only have access to the sum $y_\tau = \sum_{i=1}^{c} x_{i,\tau}$ of these signals (Fig. 2). Therefore, the set of sequences $\{o_{i,t}\}$ is not available and has to be estimated from y_τ. Pre-processing schemes for that purpose have been proposed previously, e.g., in [5] and [7] for the "simultaneous VQ" of multiple AR signals, in [1] for the simultaneous detection of multiple sinusoids, and [2] for auditory scene analysis. In this paper, we propose extensions of the standard discrete HMM approach for the classification of the mixtures of symbols obtained at the output of these pre-processors.

* C. Couvreur is a Research Assistant of the Belgian National Fund for Scientific Research (F.N.R.S.) under a Belgacom fellowship.

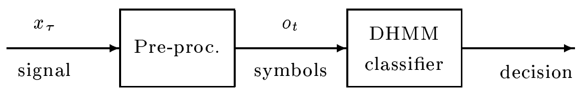

Figure 1: Classification of signal with a discrete HMM.

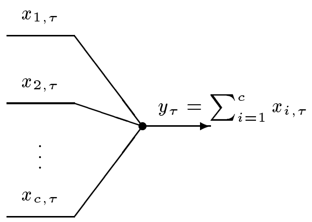

Figure 2: A mixture of c signals.

2 MIXTURES OF DHMMs

In this section the concept of *mixture of DHMMs* (MDHMM) is introduced. As usual, a DHMM is defined as a probabilistic function of a Markov chain, where the observations are conditionally independent given the hidden state of the chain. Let $\{\lambda_1, \ldots, \lambda_c\}$ denote a set of c DHMMs and let $q_{i,t} \in \mathcal{S}_i$ denote the state of the i-th DHMM at time t, where $\mathcal{S}_i = \{\sigma_{i,1}, \ldots, \sigma_{i,N_i}\}$ is the set of N_i individual state of i-th hidden Markov chain. The evolution of $q_{i,t}$ is characterized by the $N_i \times N_i$ transition matrix $\mathbf{A}_i = (a_{i,mn})$, $a_{i,mn} = P[q_{i,t+1} = \sigma_{i,n} | q_{i,t} = \sigma_{i,m}]$, $1 \leq m, n \leq N_i$. We assume that all Markov chains are ergodic and stationary, i.e., that the initial state distribution $\boldsymbol{\pi}_i$ on $\mathcal{S}_i$ is the unique solution of $\boldsymbol{\pi}'_i \mathbf{A}_i = \boldsymbol{\pi}'_i$, where $\boldsymbol{\pi}'_i$ denotes the transpose of $\boldsymbol{\pi}_i$. Let $\{o_{i,t}\} \in \mathcal{O}_i$ denote the sequence of discrete observations of the i-th hidden Markov chain, where $\mathcal{O}_i = \{\omega_{i,1}, \ldots, \omega_{i,M_i}\}$ is the set of M_i possible observations. The state-conditional emission probabilities are defined by the stochastic matrix $\mathbf{B}_i = (b_{i,mn})$, $b_{i,mn} = P[o_{i,t} = \omega_{i,n} | q_{i,t} = \sigma_{i,m}]$, $1 \leq m \leq N_i$, $1 \leq n \leq M_i$.

Consider the Cartesian DHMM or CDHMM [3] obtained from $\lambda_1, \ldots, \lambda_c$, denoted $\bar{\lambda}$. The states and observations sequences of $\bar{\lambda}$ are formed from the states and observations of the components DHMMs by $\bar{q}_t = (q_{1,t}, \ldots, q_{c,t})'$ and $\bar{o}_t = (o_{1,t}, \ldots, o_{c,t})'$. The state space $\bar{\mathcal{S}}$ of $\bar{\lambda}$ is thus the Cartesian product of the state spaces of the component DHMMs λ_i, i.e., $\bar{\mathcal{S}} = \bigotimes_{i=1}^{c} \mathcal{S}_i$, and the observation space

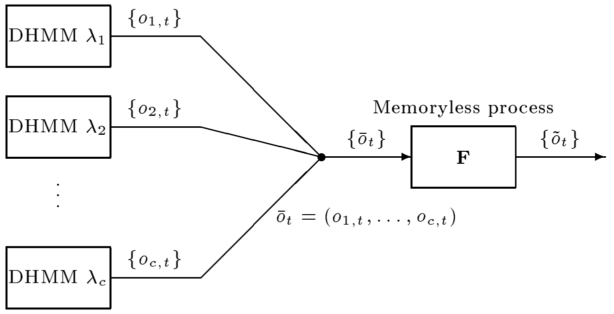

Figure 3: A mixture of c DHMMs.

of $\bar{\lambda}$ is the Cartesian product of the observation spaces, i.e., $\bar{\mathcal{O}} = \bigotimes_{i=1}^{c} \mathcal{O}_i$. The CDHMM $\bar{\lambda}$ is equivalent to a "standard" discrete HMM with state space $\bar{\mathcal{S}} = \{\bar{\sigma}_1, \ldots, \bar{\sigma}_{\bar{N}}\}$, $\bar{N} = \prod_{i=1}^{c} N_i$, and observation space $\bar{\mathcal{O}} = \{\omega_1, \ldots, \omega_{\bar{M}}\}$, $\bar{M} = \prod_{i=1}^{c} M_i$, where

$$\begin{aligned} \bar{\sigma}_n &= (\sigma_{1,\ell_1}, \ldots, \sigma_{c,\ell_c}), \quad 1 \le \ell_i \le N_i, \\ n &= \sum_{i=1}^{c-1} \left[(\ell_i - 1) \prod_{j=i+1}^{c} N_j \right] + \ell_c, \\ \bar{\omega}_m &= (\omega_{1,\ell_1}, \ldots, \omega_{c,\ell_c}), \quad 1 \le \ell_i \le M_i, \\ m &= \sum_{i=1}^{c-1} \left[(\ell_i - 1) \prod_{j=i+1}^{c} M_j \right] + \ell_c. \end{aligned}$$

We assume that the components DHMMs λ_i are independent, hence the transition and emission matrices of $\bar{\lambda}$ are given by the Kronecker products

$$\bar{\mathbf{A}} = \bigotimes_{i=1}^{c} \mathbf{A}_i \quad \text{and} \quad \bar{\mathbf{B}} = \bigotimes_{i=1}^{c} \mathbf{B}_i, \tag{1}$$

where $\bar{\mathbf{A}}$ is a $\bar{N} \times \bar{N}$ stochastic matrix and $\bar{\mathbf{B}}$ is a $\bar{N} \times \bar{M}$ stochastic matrix. It follows directly that $\bar{\lambda}$ is ergodic and that the initial stationary distribution on $\bar{\mathcal{S}}$ for $\bar{\lambda}$ is given by $\bar{\boldsymbol{\pi}} = \bigotimes_{i=1}^{c} \boldsymbol{\pi}_i$.

For signal processing reasons that will appear more clearly in Sec. 3.1, let us assume that $\{\bar{o}_t\}$ is not observed directly, but through a memoryless process $f : \bar{\mathcal{O}} \to \tilde{\mathcal{O}}$. That is, we only have access to a probabilistic function of $\bar{o}_t$, $\tilde{o}_t = f(\bar{o}_t)$; the states $\bar{q}_t$ are doubly hidden. Let $\tilde{M}$ be the cardinal of $\tilde{\mathcal{O}}$, the probabilistic mapping f can be characterized by a $\bar{M} \times \tilde{M}$ stochastic matrix $\mathbf{F} = (f_{mn})$, $f_{mn} = P[\tilde{o}_t = \tilde{\omega}_n | \bar{o}_t = \bar{\omega}_m]$, $1 \le m \le \bar{M}$, $1 \le n \le \tilde{M}$. It is not difficult to show that $\{\tilde{o}_t\}$ can be viewed as a DHMM with state space $\bar{\mathcal{S}}$ and observation space $\tilde{\mathcal{O}}$, with transition matrix $\bar{\mathbf{A}}$ and $\bar{N} \times \tilde{M}$ emission matrix

$$\tilde{\mathbf{B}} = \bar{\mathbf{B}}\mathbf{F}. \tag{2}$$

Figure 3 summarizes the definition of a mixture of DHMMs.

Denote by $\tilde{\lambda}$ the DHMM equivalent to a mixture of DHMM. This DHMM is ergodic and stationary, if started with the initial stationary distribution $\bar{\boldsymbol{\pi}}$. It is completely defined by the matrices $\bar{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ which can be easily computed from the component DHMMs parameters and the matrix $\mathbf{F}$ by (1) and (2). All the standard computational methods available for DHMMs can be applied to $\tilde{\lambda}$. For example, if $\tilde{O}_1^T = \{\tilde{o}_1, \ldots, \tilde{o}_T\}$ is a sample of length T of the process $\{\tilde{o}_t\}$, its likelihood $P[\tilde{O}_1^T | \tilde{\lambda}]$ can be computed efficiently by the usual forward-backward procedure.

Note that the computational load required by mixtures of DHMMs is generally high. However, if the matrix $\mathbf{F}$ has a particular structure, e.g., if $\mathbf{F}$ is diagonal, or block-diagonal, there can be consequent simplifications or factorizations of the algorithms. Further details on mixtures of DHMMs can be found in [7].

3 CLASSIFICATION OF MIXTURES OF SIGNALS

Let $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_K\}$ be a dictionary of DHMMs. A single signal x_τ will be classified as corresponding to DHMMs $\hat{\lambda}$ if $\hat{\lambda} = \arg\max_{\lambda_i} P[\lambda_i | O_1^T]$, where $O_1^T = \{o_1, \ldots, o_T\}$ is a sample of length T of the output of the pre-processor (Fig. 1) to which x_τ is fed, and $P[\lambda_i | O_1^T]$ is the *a posteriori* probability of DHMMs λ_i given O_1^T. The mixtures of DHMMs introduced in the previous section can be used to extend this classification method to mixtures of signals of the type of Fig. 2.

3.1 Mixture Pre-Processing

Consider a mixture of signals y_τ like that of Fig. 2. If each component signal $x_{i,\tau}$ could be pre-processed separately, the resulting sequences of symbols $O_{i,1}^T = \{o_{i,1}, \ldots, o_{i,T}\}$ could be used to form a vector sequence $\bar{O}_1^T = \{\bar{o}_1, \ldots, \bar{o}_T\}$, $\bar{o}_t = (o_{1,t}, \ldots, o_{c,t})'$. If DHMMs are used as models for the component sequences $\{o_{i,t}\}$, the vector sequence can be modeled as a CDHMM. Given access to y_τ only, different pre-processors have to be used that will perform both a decomposition of the mixture and the "pre-processing" conversion to a sequence of discrete symbols $\tilde{O}_1^T = \{\tilde{o}_1, \ldots, \tilde{o}_T\}$, This sequence $\tilde{O}_1^T$ can be viewed as an estimate of $\bar{O}_1^T$. For example, pre-processors for mixtures of signals have been proposed for detection/tracking of sinusoids in [1], for an extension of the LPC-VQ method to mixture of signals in [5] (see also the companion paper [6]), or for auditory-model-based signal decomposition in [2].

The effect of these "mixture" pre-processors on y_τ can be modeled as an application of the "single" pre-processor to each of the $x_{i,\tau}$ followed by a memoryless process that maps the sequences $\{\bar{o}_t\}$ onto $\{\tilde{o}_t\}$ and accounts for the possible detection errors in the pre-processor. This memoryless prossess can be characterized by a stochastic matrix whose row index corresponds to all the possible "Cartesian" vectors obtained by combination of DHMMs in Λ, and whose column index corresponds to all the possible outputs of the pre-processor. Let $\mathbf{F}_\Lambda$ be this matrix. It will possess some particular structure due to the properties of the pre-processor: rows that correspond to similar combinations of symbols up to a permutation are equal, rows that contain repeated symbols cannot be differentiated. The following frequency tracking example illustrates these properties. Assume that the pre-processor makes a decision on the presence or absence of a sinusoid in series of frequency bins, and assume that two sources are present. If two bins are "on," the pre-processor cannot decide which sinusoid belongs to which source: detection is performed up to a permutation of symbols. If the two sinusoids are in the same frequency bin, this bin will be the only bin "on," and there will be no difference with the detection of a single sinusoid in this bin: detection is performed up to a repetition of symbols.

3.2 Problem Statement

Let $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_K\}$ be a dictionary of stationary ergodic DHMMs. Let $\xi = \{i_1, i_2, \ldots, i_c\}$, $i_m \ne i_n$ if $m \ne n$, $1 \le i_n \le K$, be an index set for the components in Λ. There are 2^K possible different index sets ξ. To each ξ corresponds a series of CDHMMs obtained by Cartesian products of the ordering of the λ_i, $i \in \xi$. These CDHMMs are identical, up to a permutation of the states. So, we will consider that they form a class of equivalence and we will speak of *the* CDHMM corresponding to an index set ξ, which will be denoted $\bar{\lambda}_\xi$. In practice, any of the member of this class can be used for the computations. In accordance with the considerations on the pre-processor of Sec. 3.1), the memoryless process f_ξ is such that it yields the same MDHMM $\tilde{\lambda}_\xi$ when acting on any CDHMM of the equivalence class $\bar{\lambda}_\xi$. The MDHMM classification problem can then be stated as: given a sample $\tilde{O}_1^T$, find the index set of the DHMM components that are present in $\{\tilde{o}_t\}$.

In our development, we will make the following hypotheses. All the DHMMs in Λ share the same output set $\mathcal{O} = \{\omega_1, \ldots, \omega_M\}$. Hence, the CDHMM $\bar{\lambda}_\xi$ corresponding to ξ will have state set $\bar{\mathcal{S}}_\xi = \bigotimes_{i\in\xi} \mathcal{S}_i$ (up to a permutation of the coordinates) and observation set $\bar{\mathcal{O}}_\xi = \mathcal{O}^{\otimes c}$. Clearly, $\bar{N}_\xi = \#\bar{\mathcal{S}}_\xi = \prod_{i\in\xi} N_i$ and $\bar{M}_\xi = \#\bar{\mathcal{O}}_\xi = M^c$. The output set $\tilde{\mathcal{O}}$ is the set of all subsets of $\mathcal{O}$, hence, $\tilde{M} = 2^N$ for all ξ. The matrix $\mathbf{F}_\xi$ can be obtained by selecting the appropriate rows of $\mathbf{F}_\Lambda$. The parameters of the MDHMM $\tilde{\lambda}_\xi$ can be obtained from

$$\bar{\mathbf{A}}_\xi = \bigotimes_{i\in\xi} \mathbf{A}_i, \quad \bar{\mathbf{B}}_\xi = \bigotimes_{i\in\xi} \mathbf{B}_i, \quad \tilde{\mathbf{B}}_\xi = \bar{\mathbf{B}}_\xi \mathbf{F}_\xi. \tag{3}$$

We will also assume that the dictionary Λ is *identifiable*, i.e., that

$$d(\tilde{\lambda}_\xi, \tilde{\lambda}_{\xi'}) > 0, \quad \forall\, \xi \neq \xi', \tag{4}$$

where $d(\cdot,\cdot)$ is a probabilistic distance for DHMMs, e.g., the Kullback-Leibler distance introduced in [8].

3.3 Bayes Decision Rule

The Bayes decision rule with the minimum probability of error is simply

$$\begin{aligned} \hat{\xi} &= \arg\max_\xi P[\tilde{\lambda}_\xi | O_1^T] \\ &= \arg\max_\xi P[O_1^T | \tilde{\lambda}_\xi] P[\xi], \end{aligned} \tag{5}$$

where $P[\xi]$ is the *a priori* probability of ξ (or, equivalently, of $\tilde{\lambda}_\xi$). That is, the best estimate of the components that are present in the signal is given by the index set (and corresponding MDHMM) that has the highest *a posteriori* probability given the data O_1^T.

A simple prior for ξ can be obtained by assuming that each of the components λ_i is "on" with probability P_i and is "off" with probability $(1 - P_i)$. We have then

$$P[\xi] = \prod_{i\in\xi} P_i \prod_{i\notin\xi} (1 - P_i).$$

3.4 Sub-optimal Methods

The maximization of (5) requires the computation of $P[O_1^T|\tilde{\lambda}_\xi]$ *for all* ξ, i.e., for the 2^K possible values of ξ. For a given xi, the computation of $P[O_1^T|\tilde{\lambda}_\xi]$ requires in general $O(\tilde{M}_\xi^2 T)$ operations with the forward method or any equivalent. This computational load can rapidly overcome the potential of even the most powerful workstations. The complexity of the combinatorial problem can be reduced in two ways: simplified heuristic search strategies can be used instead of the complete combinatorial exploration of all the possibilities for ξ, and approximations of (5) can be used instead of the full posterior probability.

3.4.1 Alternative Decision Statistic

Instead of the posterior probability (5), simplified decision statistics with reduced computational load can be used. One such statistic based on the ergodicity and stationarity properties of MDHMMs will now be introduced.

An ergodic and stationary MDHMM $\tilde{\lambda}_\xi$ induces a stationary distribution on $\tilde{\mathcal{O}}$. Let $\boldsymbol{\mu}_\xi$ be this stationary distribution: $\boldsymbol{\mu}_\xi = (\mu_{\xi,1}, \mu_{\xi,2}, \ldots, \mu_{\xi,\tilde{M}})'$, $\mu_{\xi,i} = P[\tilde{o}_t = \tilde{\omega}_i | \tilde{\lambda}_\xi]$. This marginal distribution is related to the parameters of $\tilde{\lambda}_\xi$ by

$$\boldsymbol{\mu}_\xi = \mathbf{B}'_\xi \boldsymbol{\pi}_\xi. \tag{6}$$

If $\{\tilde{o}_t\}$ is generated by a stationary ergodic $\tilde{\lambda}_\xi$, the empirical distribution $\boldsymbol{\mu}(O_1^T)$ must converge to $\boldsymbol{\mu}_\xi$ when T increases. That is,

$$\lim_{T\to\infty} \mu_i(O_1^T) = \mu_{\xi,i} \quad \text{a.s.}, \tag{7}$$

with

$$\mu_i(O_1^T) = \frac{1}{T} \sum_{t=1}^T 1_{\{\tilde{o}_t = \tilde{\omega}_i\}}, \tag{8}$$

where 1_E is the indicator function for the event E. This suggests a decision statistic: compare the empirical distribution of O_1^T to the distributions corresponding to the various ξ and select the closest one according to a probabilistic distance. For example, with the Kullback-Leibler distance

$$D(\boldsymbol{\mu}(O_1^T) \| \boldsymbol{\mu}_\xi) = \sum_{i=1}^{\tilde{M}} \mu_i(O_1^T) \log \frac{\mu_i(O_1^T)}{\mu_{\xi,i}}, \tag{9}$$

we get

$$\breve{\xi} = \arg\min_\xi D(\boldsymbol{\mu}(O_1^T) \| \boldsymbol{\mu}_\xi). \tag{10}$$

The computational complexity of (9) is $O(\tilde{M})$. Note that the minimizer $\breve{\xi}$ need not be unique even if the identifiablity condition (4) is fulfilled. Ties can be broken with the posterior probability (5).

More complex schemes based on the combination of a simplified decision statistic and the posterior probability are also possible. For example, the simplified decision statistic (9) can be used to select L-best candidate ξs, and the final decision among these L hypotheses can be made with (5).

3.4.2 Simplified Combinatorial Exploration

Instead of considering all 2^K possible ξs, we can evaluate the decision statistic (posterior probability or other) for some of them only. Sub-optimal search methods similar to feature selection in pattern recognition can be applied (see, e.g., Chap. 5 of [9]). We will not give further details on heuristic search strategies here, but will suggest a very simple way to reduce the search space.

The pre-processor can provide an estimate of the number of DHMM components present. The observed variable $\tilde{o}_t$ belongs to the set of all subsets of $\mathcal{O}$. If the pre-processor behaves properly, it can be expected that the number of elements in $\tilde{o}_t$, viewed as a subset of $\mathcal{O}$, are "close" to c. Let $\tilde{c}_t = \#\tilde{o}_t$ be this number of elements. The relation between $\tilde{c}_t$ and c can be analyzed precisely for given Λ and $\mathbf{F}$, and proper choices for the selection of a size for the index sets ξ to be considered based on $\tilde{c}_t$ can then be made [7]. One simple choice is to take the index sets ξ such that

$$\inf_t \tilde{c}_t \leq c(\xi) \leq \sup_t \tilde{c}_t. \tag{11}$$

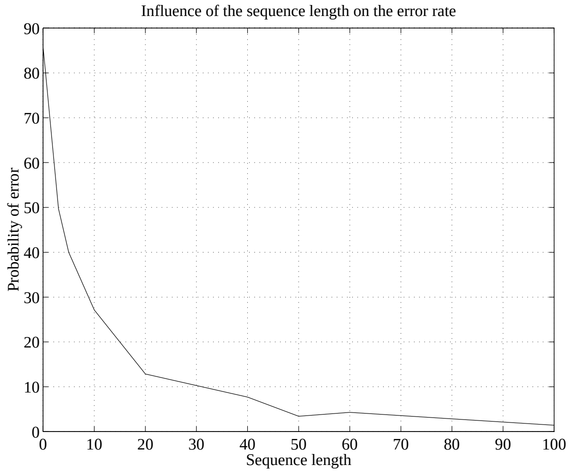

Figure 4: Evolution of the classification error rate when the sample length T increases.

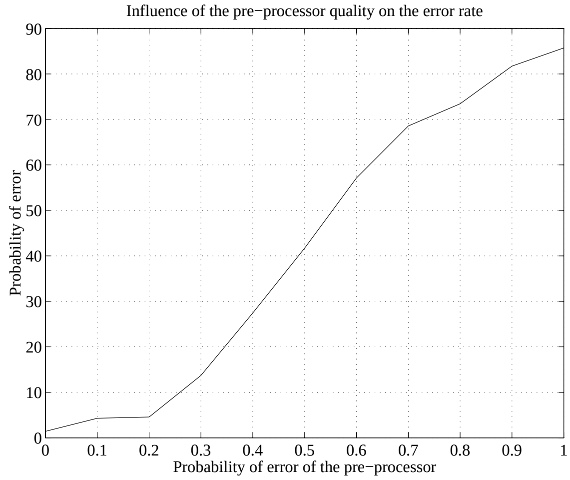

Figure 5: Evolution of the classification error rate when the performance of the pre-processor decrease.

4 PRELIMINARY RESULTS

In order to assess the validity of the concept of MDHMMS for the decomposition of mixtures of signals, several Monte-Carlo experiments on simple examples have been conducted. The goals of these experiments were to learn about the accuracy of the model for classification purpose, and to study the influence of the quality of the pre-processor on the recognition results. The DHMM dictionary contained three models $\Lambda = \{\lambda_1, \lambda_2, \lambda_3\}$, of respective order $N_1 = 1$, $N_2 = 2$, and $N_3 = 2$. The DHMMs output space $\mathcal{O}$ contained three elements ($M = 3$). The transition and emission matrices of the three DHMMs were

$$\mathbf{A_1} = \begin{pmatrix} 1 \end{pmatrix}, \qquad \mathbf{B_1} = \begin{pmatrix} 0.8 & 0.1 & 0.1 \end{pmatrix}$$

$$\mathbf{A_2} = \begin{pmatrix} 0.5 & 0.5 \\ 0.1 & 0.9 \end{pmatrix}, \qquad \mathbf{B_2} = \begin{pmatrix} 2/3 & 1/6 & 1/6 \\ 1/6 & 2/3 & 1/6 \end{pmatrix},$$

$$\mathbf{A_3} = \begin{pmatrix} 0.95 & 0.05 \\ 0.95 & 0.05 \end{pmatrix}, \qquad \mathbf{B_3} = \begin{pmatrix} 1/6 & 1/6 & 2/3 \\ 1/6 & 2/3 & 1/6 \end{pmatrix}.$$

In a first experiment, we supposed the stochastic matrix $\mathbf{F}_\Gamma$ to be binary. That is, we assumed that the pre-processor was perfect (no confusion between symbols), the only pre-processing effect accounted for by $\mathbf{F}_\Gamma$ was the "projection" of vector of symbols onto their subsets of distinct symbols with probability one, i.e., CDHMMs symbols $\bar{o}_t$ like $(\omega_1, \omega_2)'$, $(\omega_2, \omega_1, \omega_1)'$, or $(\omega_1, \omega_2, \omega_1)'$, are equally mapped onto $\tilde{o}_t = \{\omega_1, \omega_2\}$. Note that, in this case, the memoryless process $\tilde{o}_t = f(\bar{o}_t)$ is purely deterministic. The symbol sequences corresponding to the MDHMMs were generated by mixing sequences generated by the components DHMMs with f). At this stage of the experiments, our goal was to study the influence of the sequence length T on the decomposition/classification accuracy. If the set of HMM and their mixtures are identifiable, the error rate should tend to zero when T increases. This is indeed verified in Fig. 4.

In the next experiment, we supposed that the pre-processor had, in addition to its deterministic "projection" behavior described above, some confusion between "close" symbols. For example, if the input was $(\omega_1, \omega_2)'$ (or any equivalent combination of symbols), the output would be $\{\omega_1, \omega_2\}$ with probability $1 - \delta$, and any of $\{\omega_2\}$, $\{\omega_2\}$, or $\{\omega_1, \omega_2, \omega_3\}$ with probability $\delta/3$. That is, the pre-processor behaved perfectly with probability $1 - \delta$ and committed an error with probability δ. Figure 5 gives the classification/decomposition error rate of the system as a function of the probability of error of the pre-processor δ. The sample length T was set to 100. The results show that the performance of the classifier for decomposition of mixtures of DHMMs degrades smoothly with the performance of the classifier.

Other results will be shown at the conference.

REFERENCE

[1] Xianya Xie & R. J. Evans, "Multiple Target Tracking and Multiple Frequency Line Tracking Using Hidden Markov Models," *IEEE Trans. on Signal Processing*, Vol. 39, No. 12, pp. 2659–2676, Dec. 1991.

[2] P. D. Green, et al., "Auditory Scene Analysis and Hidden Markov Model Recognition of Speech in Noise," Proc. ICASSP'95, Detroit, USA, May 1995, pp. 401-404.

[3] L. B. White, "Cartesian Hidden Markov Models with Applications," *IEEE Trans. SP,* Vol. 40, No. 6, June 1992.

[4] C. Couvreur & Y. Bresler, "A Statistical Pattern Recognition Framework for Noise Recognition in an Intelligent Noise Monitoring System," *Proc. EURO-NOISE'95,* Lyon, France, March 1995, pp. 1007–1012.

[5] C. Couvreur & Y. Bresler, "Decomposition of Mixtures of Gaussian AR Processes," *Proc. ICASSP'95,* Detroit, USA, May 1995, pp. 1605–1608.

[6] C. Couvreur & Y. Bresler, "Dictionary-based Decomposition of Linear Mixtures of Gaussian Processes," *Proc. ICASSP'96,* Atlanta, USA, May 1996.

[7] C. Couvreur, "Hidden Markov Models and Their Mixtures," DEA Thesis, Department of Mathematics and Applied Mathematics, Catholic University of Louvain, Belgium, June 1996.

[8] B.-H. Juang & L. R. Rabiner, "A Probabilistic Distance Measure for Hidden Markov Models," AT&T Bell Lab. Tech. J., vol. 63, no. 7, pp. 1213–1243, Sep. 1984.

[9] P. R. Devijver & J. Kittler, *Pattern Recognition: A Statistical Approach,* Prentice-Hall, 1982.