K-Harmonic Means Clustering based Blind Equalization in Hostile Environments

Deepak Boppana   Sathyanarayana S Rao

Department of Electrical and Computer Engineering
Villanova University, PA 19085

Abstract- **A novel blind clustering based equalizer suitable for channels with inter-symbol interference (ISI), nonlinear distortions and co-channel interference (CCI) is proposed. Blind channel estimation is performed by partitioning the baseband data at the receiver into clusters that are identified using a new class of clustering algorithms known as K-Harmonic Means (KHM$_p$). The KHM$_p$ algorithms are insensitive to the initialization of the cluster centers owing to a built-in boosting function and provide reliable estimates of the cluster centers. The identified cluster representatives are then mapped to the corresponding combinations of input symbols using a discrete hidden Markov model formulation of the channel states and the mapping is used to compute the branch metrics in a cluster-based Viterbi detector. The performance of the proposed equalizer in hostile environments is illustrated with computer simulations.**

I. INTRODUCTION

High-speed digital communication suffers from inter-symbol interference (ISI), co-channel interference (CCI), noise and nonlinear distortions present in the channel. Under Gaussian white noise conditions, the maximum likelihood sequence detector (MLSD) outperforms both linear as well as nonlinear symbol decision equalizers. However, under the presence of nonwhite CCI and severe nonlinearities in the channel, MLSD is no longer the optimal solution. Approaches based on data clustering techniques have generated a lot of interest for their ability to avoid any explicit channel modeling [1]-[5]. These methods focus on the clusters formed by the data at the receiver, exploiting the fact that the channel distortions such as ISI, nonlinearities, and CCI are reflected in the shape and position of the clusters. As a result, using a *clustering* algorithm to identify the cluster representatives basically unravels the specific nature of the channel impairments. In the past, clustering techniques like the neural gas-clustering network [2] and the Iterative Self-Organizing Data Analysis Technique (ISODATA) [4] and have been used to perform this task.

In this paper, we propose a novel cluster-based MLSD that uses the K-Harmonic Means (KHM$_p$) class of clustering algorithms. The rest of the paper is organized as follows: In Section II, the concept of using clustering methods at the receiver is presented, with the KHM$_p$ algorithms described in Section III. Section IV, and V describe the channel mapping and sequence detection processes respectively. Simulation results for the nonlinear ISI and CCI impaired channels are presented in Section VI and the results are summarized in Section VII.

II. CLUSTERING–BASED EQUALIZATION

A. System Model

Let us consider the baseband model of a typical digital communication system corrupted by CCI as depicted in Fig.1. The channel, denoted by the FIR transfer function H(z), includes the modulation filter at the transmitter end, the multipath channel and filters at the front end of the receiver that are used to suppress the adjacent-channel interference (ACI) and perform low-pass filtering. The baseband received signal r(t), can be denoted as:

$$r(t) = s(t) + s_{co}(t) + w(t)$$
$$= f(I(t), I(t-1),......,I(t-L)) + s_{co}(t) + w(t) \quad (1)$$

where s(t) is the output of the desired channel, $s_{co}(t)$ is the co-channel interference component, w(t) is the zero mean white Gaussian noise with variance σ_n^2, and f (.) is the desired channel function representing channel characteristics like multipath effects and nonlinearities. s(t) and $s_{co}(t)$ are given by:

$$s(t) = \sum_{n=0}^{L-1} h(n) I(t-n) \quad (2)$$

$$s_{co}(t) = \sum_{i=1}^{k} \sum_{n=0}^{L_i-1} h_i(n) I_i(t-n) \quad (3)$$

where I(t) and I$_i$(t) are the desired and co-channel data symbols that are assumed to be equiprobable bipolar (±1) and mutually uncorrelated, h(n) and h$_i$(n) are the impulse response coefficients of the desired channel and the i-th co-channel having L and L$_i$ taps respectively and the total number of interfering co-channels is assumed to be k.
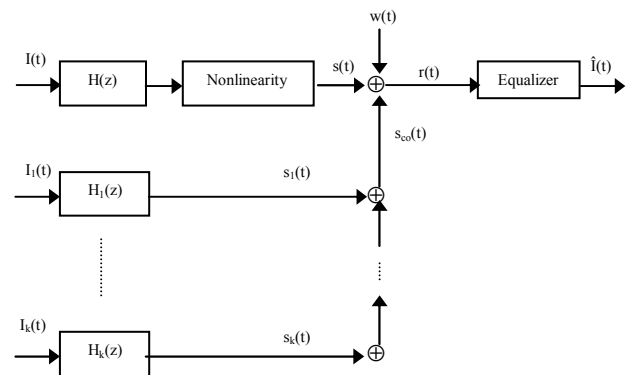

Fig. 1. Discrete-time model of a typical digital communication system

The signal-to-noise ratio (SNR) and the signal-to-interference ration (SIR) are calculated as follows:

$$SNR = \frac{\sigma_s^2}{\sigma_n^2} \text{ , and } SIR = \frac{\sigma_s^2}{\sigma_{co}^2}$$

where σ_s^2 and σ_{co}^2 are the signal power and the co-channel signal powers respectively.

The proposed equalizer assuming a single interfering co-channel is depicted in Fig.2.

B. Clustering Concepts

We observe that the finite length of the channel and nature of the input sequence I(t), result in the desired channel output sequence s(t) being a discrete values signal with 2^{L+1} different elements, called *desired channel states* m_k , $1 \le k \le 2^{L+1}$, corresponding to 2^{L+1} possible realizations of the sequence of transmitted bits (I(t), …., I(t-L) in the 1-dimensional space. If the channel has nonlinear characteristics, the desired channel states shift to new positions depending on the form of the nonlinearity. Similarly, assuming a single co-channel, the co-channel output sequence s_{co}(t) will be a discrete values signal with 2^{L_1+1} *co-channel states* m_{co}^j, $1 \le j \le 2^{L_1+1}$. The randomness of the additive white Gaussian noise results in the formation of clusters around the *observed states* that arise from the sum of the co-channel states plus each one of the desired states. The bipolar nature of the transmitted symbols results in the observed states being uniformly distributed around the corresponding desired states and we can assume the overall mean coincides with the desired states. We use a new class of clustering algorithms known as K-Harmonic Means (KHM$_p$) to identify the desired channel states. Once the desired channel states have been identified, the co-channel states are estimated by performing the clustering operation on the shifted observed data sequence, $\hat{r}(t) = r(t) - m_k$, where r(t) is the received data symbol related to the desired state m_k.

III. K-HARMONIC MEANS CLUSTERING

The frequently used clustering algorithms like the K-means and ISODATA have the intrinsic problem of depending heavily on the initialization of the cluster centers for achieving good performance. This is due to their winner-takes-all partitioning strategy, which results in a strong association between the data points and the nearest center and prevents the centers from moving out of a local density of data. This problem can be resolved by replacing the *minimum* distance from a data point to the centers, used in K-means, by the *Harmonic Averages (HA)* of the distances from the data point to all centers resulting in a continuous transition between the data points and centers. This new class of center-based iterative clustering algorithms known as the K-Harmonic Means clustering (KHM$_p$) [6], [7] is essentially insensitive to the initialization of the centers and significantly improves the quality of clustering results compared to the K-means and ISODATA under different initialization conditions.
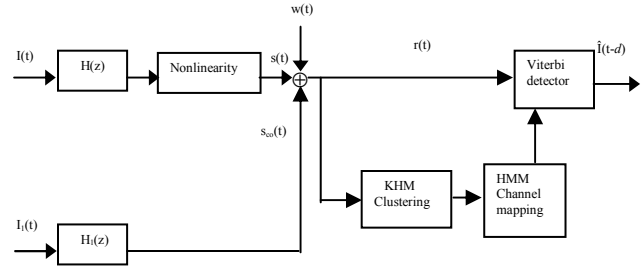

Fig. 2. Block diagram of the proposed cluster-based blind MLSD

We use the KHM$_p$ clustering algorithms to classify a block of received data r(t) of length N, into K clusters and determine the associated set M of cluster representatives. Since the number of clusters K into which the data must be classified is unknown, we obtain the clustering using the KHM$_p$ algorithms for increasing K until little reduction is obtained in the performance function J. The performance function J for the KHM$_p$ algorithms is defined as:

$$J = \sum_{i=1}^{N} HA\{\left\| r_i - m_k \right\|^p \mid m \in M, \, k = 1,....,K\}$$

$$= \sum_{i=1}^{N} \frac{K}{\displaystyle\sum_{k=1}^{K} \frac{1}{\left\| r_i - m_k \right\|^p}} \qquad (4)$$

The recursive formula for minimizing the performance function J, is obtained by taking the partial derivatives of J with respect to the center positions m_k, $k=1,…, K$ and setting them to zero as:

$$m_k = \sum_{i=1}^{N} \frac{1}{d_{i,k}^{p+2} (\displaystyle\sum_{l=1}^{K} \frac{1}{d_{i,l}^p})^2} r_i \Bigg/ \sum_{i=1}^{N} \frac{1}{d_{i,k}^{p+2} (\displaystyle\sum_{l=1}^{K} \frac{1}{d_{i,l}^p})^2} ,$$

where $d_{i,l} = \left\| r_i - m_l \right\|$. $\qquad (5)$

If we express m_k as $m_k = \dfrac{\displaystyle\sum_{i=1}^{N} p(\,m_k / r_i\,) * a(r_i) r_i}{\displaystyle\sum_{i=1}^{N} p(\,m_k / r_i\,) * a(r_i)}$,

with $a(r(t)) > 0$, and $\displaystyle\sum_{k=1}^{K} p(\,m_k / r_i\,) = 1$,

then $a(r(t))$, the weighting function of the data points, decides how much of each data point r(t) participates in the next iteration of calculating the new center locations whereas the term $p(\,m_k / r_i\,)$ decides the portion of $a(r_i) r_i$ that is associated with m_k and is known as the membership function.

We find that for $p > 2$, $a(r_i)$ has a smaller value for the data points that are closer to one of the centers. This property serves as a boosting function and boosts, in the next iteration, the participation of the data points that are not close to any centers. This dynamic weighting of the data points in effect flattens out a local density that has trapped more than one center and reduces the chance of multiple centers being trapped in a single local cluster of data, thus making the algorithm insensitive to the initialization of the centers [6]. The asymptotic computational complexity per iteration for computing the pair-wise distances from N data points to K centers is O(N*K) which is the same as that of K-means algorithm and the overall complexity is far less than that of the neural gas clustering network used in [2] that requires a large number of iterations to converge.

IV. HMM-BASED CHANNEL MAPPING

The estimated cluster representatives are related to one or more of the 2^{L+1} possible realizations of the sequence of input symbols, which can be named as their respective *labels*. We represent the unknown probabilities of specific clusters to correspond to specific labels by a probability matrix $b_n(m_k), 1 \le n, k \le 2^{L+1}$ and treat this matrix as an unknown parameter θ of discrete observations HMM. The problem then, can be viewed as a maximum likelihood parameter estimation problem and the EM algorithm can be applied to find the optimal parameters of the HMM which best match the given discrete observation sequence.

A. Discrete Hidden Markov Model Formulation

1. A block of received data r(t) of length T, $T \le N$ where N is the length of the block used for clustering, is quantized to the cluster representative closest to them, and is taken as the discrete observation sequence Y = (y(1), …, y(T)), $y(t) \in \{m_k; k = 1,...,K\}$ feeding the HMM.

2. There are $Z = 2^L$ states in the HMM and the states can be represented as S(t): (I(t-1)….I(t-L)) for a channel of memory L, where I(t) is the i.i.d. sequence of transmitted bipolar symbols.

3. We observe that each state transition from state i to state j results in an emission of a specific cluster representative with a certain *label* and that only specific transitions among the clusters are possible due to the finite length of the channel and the effect of ISI on successive received samples. Hence, the elements of the state transition probability matrix

$$a_{ij} = P[S(t+1) = j \mid S(t) = i], 1 \le i, j \le Z,$$

are assigned a value of 1/2 (2 possible symbols) if they correspond to a viable state transition and a value of zero if they do not.

4. The initial state distribution $\pi_i = P[S(1) = i]$, for $1 \le i \le Z$, is usually know at the receiver based on channel length measurements.

5. The desired channel mapping probability matrix $b_n(m_k), 1 \le n, k \le 2^{L+1}$, is considered as the unknown parameter θ to be estimated by applying the EM algorithm to the HMM.

B. Parameter Estimation Via The EM Algorithm

Unknown parameter estimation of an HMM using the EM algorithm is also known as the Baum-Welch algorithm or the forward-backward algorithm [8]. Following the maximum likelihood parameter estimation concept, the estimation is done iteratively, estimating the parameter θ that maximizes the probability $P(Y/\theta)$ in every iteration and the algorithm converges when $P(Y/\theta)$ exceeds a predetermined threshold.

1) Initialization: Let $\alpha_t(i)$ and $\beta_{t+1}(j)$ be the forward and backward parameters respectively.

Set $\alpha_1(i) = 1$ for the known initial state i and 0 otherwise, $\beta_{T+1}(j) = 1$ for $j = 1,…, Z$

Set $b_n(m_k) = 1/K$, for $n = 1,…, 2^{L+1}$, $k = 1,…, K$

2) Recursion: The following parameters are calculated in each iteration.

1. $\alpha_t(j)$ for t = 2, …, T+1, $j = 1, …, Z$

$$\alpha_t(j) = \sum_{i=1}^{Z} \alpha_{t-1}(i) P[S(t) = j \mid S(t-1) = i] b_n(y(t-1)) \quad (6)$$

2. $\beta_t(i)$ for t = T, T-1, …, 1, $i = 1,…, Z$

$$\beta_t(i) = \sum_{j=1}^{Z} \beta_{t+1}(j) P[S(t+1) = j \mid S(t) = i] b_n(y(t)) \quad (7)$$

3. $\xi_t(i, j)$, the probability of being in state i at time t and state j at time $t+1$

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_n(y(t)) \beta_{t+1}(j)}{\sum_{i=1}^{Z} \sum_{j=1}^{Z} \alpha_t(i) a_{ij} b_n(y(t)) \beta_{t+1}(j)} \quad (8)$$

4. Finally, $b_n(m_k)$, for $n = 1,…, 2^{L+1}$, $k = 1,…, K$ is

$$b_n(m_k) = \frac{\sum_{t=1, y(t)=m_k}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \xi_t(i, j)} \quad (9)$$

The recursion is repeated until convergence [4].

The different cluster representatives are then mapped to the respective labels for which they have the highest probability as revealed by the estimated matrix $b_n(m_k)$. Thus, the required information for maximum likelihood sequence detection is now available and the detector proceeds to perform signal detection.

V. CLUSTER-BASED VITERBI DETECTION

The channel mapping obtained using the procedure outlined in Section IV, is used to compute the distances D_k between the received symbol and the desired channel outputs m_k and a decision is made based on the path having the minimum cost. In the presence of co-channel interference, the distance metric is modified as the minimum distance between the received

data r(t) and each of the observed states formed due to the sum of the corresponding desired state m_k and the co-channel states m_{co}^j as described earlier in Section II.

$$D_k = \min_{j=1}^{L_1}\left\{(r - m_k - m_{co}^j)^T (r - m_k - m_{co}^j)\right\} \qquad (10)$$

Also, the number of possible searching paths in the Viterbi trellis can be reduced by noting the constraints placed on the allowable jumps among the clusters as described earlier. Like in the HMM, the number of states is given by $Z=2^L$ and there are only two possible state transitions at any time. Performing signal detection in the method outlined above is suitable for both linear and nonlinear channels and is especially found to be very efficient when the channel has nonlinear characteristics as shown in the computer simulations that follow.

VI. SIMULATIONS

In all the simulations that follow, we assume the transmitted symbols to be equiprobable bipolar symbols. The proposed scheme, as shown in Fig. 2, is investigated through bit error rate (BER) for different SNR and SIR values.

A. Linear channel with ISI

We initially consider a simple channel (L=1) without any nonlinearity and CCI to demonstrate the performance of the proposed KHM_p clustering and HMM based channel mapping methods. The channel is represented as:

r(t) = 0.5 I(t) + I(t-1) + w(t)

The signal to noise ratio (SNR) is taken as 12dB. The K-Harmonic Means clustering algorithm with $p =3$, N=200, and random data samples as the initial centers, identifies the cluster representatives as:

$m_1 = -1.5$, $m_2 = -0.5005$, $m_3 = 0.5095$ and $m_4 = 1.4969$

The K-Means algorithm on the other hand, when initialized randomly, generates varying cluster representatives based on the initialization. With the number of states of the HMM fixed at $Z = 2^L = 2$ and T=40, the channel mapping probability matrix $b_n(m_k), 1 \le n, k \le 4$ converges to the matrix shown in Table 1 in 8 iterations of the EM algorithm.

Each of the labels is then mapped to the cluster representative m_k for which it has the highest probability as revealed by the matrix $b_n(m_k)$. We can thus conclude from the table that the channel output m_1 corresponds to the sequence of input bits (-1 -1), m_3 to (-1 1) and so on.

TABLE 1: Probability matrix $b_n(m_k)$ for channel A

Label	Cluster representative			
I(t)I(t-1)	m_1	m_2	m_3	m_4
(-1 -1)	1	0	0	0
(-1 1)	0	0	1	0
(1 -1)	0	1	0	0
(1 1)	0	0	0	1

TABLE 2: Probability matrix $b_n(m_k)$ for channel B

Label:	Cluster representatives					
I(t)I(t-1)I(t-2)	m_1	m_2	m_3	m_4	m_5	m_6
(1 1 1)	1	0	0	0	0	0
(1 1 -1)	0	1	0	0	0	0
(1 -1 1)	0	0	1	0	0	0
(1 -1 -1)	0	0	0	0	1	0
(-1 1 1)	0	1	0	0	0	0
(-1 1 -1)	0	0	0	1	0	0
(-1 -1 1)	0	0	0	0	1	0
(-1 -1 -1)	0	0	0	0	0	1

Near optimal sequence detection can then be achieved by the Viterbi algorithm by using this mapping to compute the branch metrics associated with each of the legal state transitions in all the possible paths in the state trellis.

B. Channel with ISI and nonlinearity

We now consider an ISI impaired channel with nonlinearity but without any CCI as shown below:

r(t) = 0.5 I(t) + 0.7 I(t-1) + 0.5 I(t-2)
with nonlinearity: $r(t) + 0.05r(t)^2 - 0.1r(t)^3 + w(t)$

For a signal to noise ratio (SNR) of 10dB, the KHM_3 clustering algorithm identifies 6 clusters in the block of N=200 received data samples and the cluster centers are as follows:

$m_1 = 1.4154$, $m_2 = 0.6842$, $m_3 = 0.3183$,

$m_4 = -0.2641$, $m_5 = -0.6175$, $m_6 = -1.1532$,

We observe that the clustering algorithm has identified only 6 clusters whereas the maximum possible number of clusters is $K = 2^{L+1} = 2^3 = 8$. This is because of the high ISI and the symmetry in the channel that leads to overlapping clusters. This is revealed by the output of the HMM, which is the probability matrix $b_n(m_k)$, for $1 \le n \le 8, 1 \le k \le 6$ (converges in 9 iterations) as shown in Table 2. For example, it is revealed that both the labels (1 1 -1) and (-1 1 1) correspond to the cluster representative m_2 and the labels (1 -1 -1) and (-1 -1 1) correspond to m_5, etc. Thus the exact channel mapping is identified by the HMM process.

Fig. 3 compares the performance, in terms of the bit error rate (BER) for various SNR, of the proposed KHM_3 clustering based blind detector to that achieved using the K-Means clustering algorithm, and a classical MLSD with optimal linear channel estimator. The initial cluster centers for all the three methods were initialized randomly from the data available. We can observe from the figure that the proposed blind detector outperforms both the detector with the K-Means clustering,
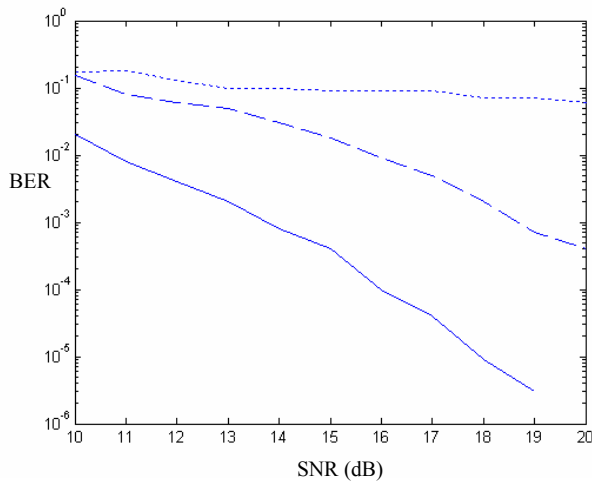
Fig. 3: ISI and nonlinearly impaired signal — Performance comparison: '—' proposed blind detector, '- - -' MLSD with linear channel estimator, '….' blind detector with K-Means clustering and random initialization of cluster centers.

Fig. 4: ISI, CCI, and nonlinearly impaired signal — Performance comparison: '—' proposed blind detector, '- - -' MLSD with linear channel estimator, '….' blind detector with K-Means clustering and random initialization of cluster centers.

which suffers severely from its inherent random initialization problem, as well as the non-blind classical MLSD with a linear channel estimator.

C. Channel with ISI, nonlinearity and CCI

Finally, we consider the entire structure presented in Fig 2. The ISI and nonlinearity are as previously described and we add a single co-channel interferer as given below with a SIR of 10dB.

$s_{co}(t) = \lambda(0.6I_1(t) + 0.8I_1(t-1))$ where λ determines the SIR. The co-channel states are computed as described in Section II and the distance metric given in equation (10) is used to compute the branch metrics in the Viterbi algorithm. Fig. 4 verifies the superior performance of the proposed scheme against the conventional MLSD as well as the detector with the K-Means clustering even in the presence of CCI.

VII CONCLUSIONS

A blind maximum likelihood sequence detector based on clustering concepts was proposed. The received data were grouped into clusters using the K-Harmonic Means unsupervised classification algorithm. The insensitivity of the KHM_p algorithm to the random initialization of the cluster centers resulted in providing good estimates of the possible noiseless discrete outputs of the channel. The EM algorithm was used to determine the channel mapping probability matrix that was modeled as the unknown parameter of a discrete HMM, and was subsequently used by the Viterbi detector to perform signal detection. The simulations confirm that the proposed method does not require any training sequences or explicit channel modeling and is well suited for ISI and CCI impaired linear as well as nonlinear channels.
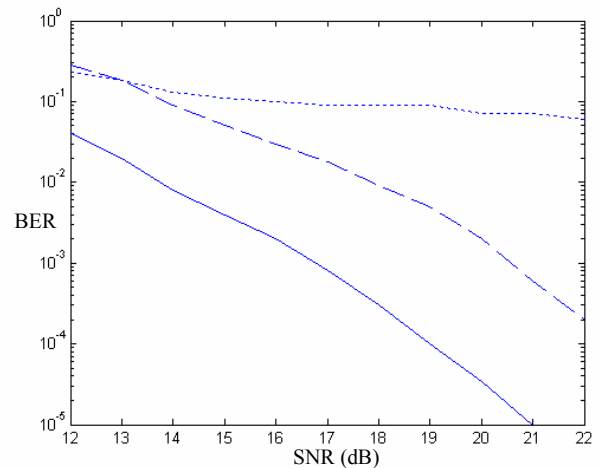
REFERENCES

[1] S. Theodoridis, C.F.N.Cowan, C.P.Callender, C.M.S.See "Schemes for equalization of communication channels with nonlinear impairments," IEE Proc. Communications, vol 142, no.3, pp.165-171, Jun. 1995.

[2] Y.J. Jeng and C.C. Yeh, "Cluster-based blind nonlinear-channel estimation," IEEE Trans. on Signal Processing, vol.45, pp.1161-1172, May 1997.

[3] G.K. Kaleh, R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," IEEE Trans. on Communications, vol.42, pp.2406-2413, July 1994.

[4] K. Georgoulakis, S. Theodoridis, "Blind and semi-blind equalization using hidden Markov models and clustering techniques," Signal Processing, 80, pp.1795-1805, Sep'00.

[5] Deepak Boppana, Sathyanarayana S Rao, "Clustering-based blind maximum likelihood sequence detection for GSM and TDMA systems", Proceedings of the 45th IEEE International Midwest Symposium on Circuits and Systems, August 2002.

[6] Bin Zhang, "Generalized K-harmonic means-boosting in unsupervised learning," Hewlett-Packard Laboratories Technical Report, Oct 2000.

[7] B. Zhang, M. Hsu, U. Dayal, "K-harmonic means," TSDM 2000, Lyon, France, Sep 2000.

[8] Jeff. A. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," ICSI, April 1998.