

Semi-blind Block Channel Estimation and Signal Detection Using Hidden Markov Models

Pei Chen and Hisashi Kobayashi
Department of Electrical Engineering
Princeton University
Princeton, New Jersey 08544, USA

Abstract - We propose two maximum likelihood based semi-blind block channel estimation and signal detection algorithms for multipath channels with additive Gaussian noise. The algorithms are based on the Baum-Welch algorithm and the segmental k-means algorithm for Hidden Markov Models (HMMs). By making use of a training signal, the algorithms are applied block-wise to sequential disjoint subintervals of the whole observation interval. We study the effects of block length in terms of the bit error rate (BER), the mean square error (MSE) of the estimated channel impulse response, and its Cramer-Rao lower bound. Our simulation results show that the BER performance does not suffer even for a short block length when a good initial estimate is available.

1 INTRODUCTION

In order to cope with an unknown and possibly time-varying radio channel, the problem of channel estimation is recognized as a key research problem, because a large class of interference cancellation algorithms and decoding algorithms require complete or partial knowledge of channel characteristics. The existing and proposed digital communication systems support the use of known training signals to estimate the channel. By incorporating the training signals into blind schemes, semi-blind channel estimation schemes are anticipated to improve the performance over pure blind or non-blind schemes.

Hidden Markov Models (HMMs), on the other hand, have been extensively studied and successfully applied to problems of sequential pattern recognition such as speech recognition and handwriting recognition [7]. The parameter estimation problem for HMMs can be solved efficiently by the Baum-Welch algorithm [1]. The reestimation formulas of the Baum-Welch algorithm can be interpreted as an application of the Expectation-Maximization (EM) [3] algorithm to HMMs, and the computational efficiency of the Baum-Welch algorithm comes from the efficient procedure of the forward-backward algorithm [7].

Recently the Baum-Welch algorithm for HMMs has been applied to maximum likelihood channel estimation [2,5]. In this paper, we propose a semi-blind block channel estimation and signal detection scheme which gives block-wise maximum likelihood channel estimate and MAP decoding by using the Baum-Welch algorithm in sequential disjoint subintervals of an observation. The advantage of block channel estimation lies in several aspects. Firstly, it can avoid a long delay to process the whole observation interval; secondly, it is suitable for a slow time varying channel; thirdly, it has lower computational complexity. We evaluate the performance in terms of the mean square error (MSE) and bit error rate (BER) for different block lengths.

The segmental k-means algorithm for HMMs uses the state-optimized likelihood as its optimization criterion [4]. We apply the two iterative steps of the segmental k-means algorithm to channel estimation and signal detection and obtain a very simple structure which gives the maximum likelihood sequence estimate and decision directed channel estimation. This algorithm can also be applied to the received signal, block by block. We evaluate the effects of block length by simulation, and compare the results with those obtained by the Baum-Welch algorithm.

The rest of the paper is organized as follows. Section 2 gives the channel model and its HMM description. Section 3 describes our semi-blind block channel estimation and signal detection scheme based on the Baum-Welch algorithm. Section 4 derives the two step iterative scheme based on the k-means algorithm. Section 5 evaluates the performance of the proposed schemes by BER, MSE and its Cramer-Rao lower bound for different block lengths. Section 6 presents our concluding remarks.

2 CHANNEL MODEL AND ITS HMM DESCRIPTION

We consider the following multipath channel model:

$$y_n = \sum_{l=1}^L h_l d_{n+1-l} + \nu_n, \quad (1)$$

where the channel has memory of length L ; $\underline{h} = (h_1, h_2, \dots, h_L)$ is the channel impulse response; ν_n 's are i.i.d. complex Gaussian random variables with zero mean and variance σ^2 ; $\underline{d} = (d_1, d_2, \dots, d_M)$ is the transmitted baseband signal including a training part and a data part.

For a given observation sequence $\underline{y} = (y_1, y_2, \dots, y_N)$, our goal is to find the maximum likelihood estimate of the channel parameter vector $\underline{\theta} = (h_1, h_2, \dots, h_L, \sigma^2)$ that maximizes $f(\underline{y}|\underline{\theta})$.

The above problem can be formulated as an HMM parameter estimation problem. An HMM models the case where the observation is a probabilistic function of the state of a Markov chain. An HMM is specified by three probability measures A , B and π , where A is the state transition probability matrix; B is a vector of conditional probability density functions (PDF) with each entry being the observation PDF in a state, and π is the initial state probability vector.

We define state $S_n = (d_{n+1-L}, \dots, d_{n-1}, d_n)$. If each element takes on K possible values, the total number of states is K^L . The parameters of the HMM can be found as:

- State transition probability $a_{ij} = P[S_{n+1} = j | S_n = i]$; data part:

$$a_{ij} = \begin{cases} 1/K, & \text{if } i \rightarrow j \text{ is permissible} \\ 0, & \text{otherwise} \end{cases}$$

training part:

$$a_{ij} = \begin{cases} 1, & \text{if } i \rightarrow j \text{ is permissible} \\ 0, & \text{otherwise} \end{cases}$$

- Initial state probability $\pi_i = P[S_1 = i]$:

$$\pi_i = \begin{cases} \frac{1}{K^L}, & \text{if starting from data part} \\ \delta(i - i_0), & \text{if starting from training part} \end{cases}$$

where i_0 is the starting state.

- Observation PDF in state i $b_i(y) = f(y|S_n = i, \underline{\theta})$:

$$b_i(y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|y - \sum_{l=1}^L h_l d_{i,L+1-l}|^2}{2\sigma^2}\right), \quad (2)$$

where $i = (d_{i,1}, d_{i,2}, \dots, d_{i,L})$.

3 BAUM-WELCH ALGORITHM

3.1 Baum-Welch Algorithm for Channel Estimation

The Baum-Welch algorithm is a computationally efficient reestimation procedure to solve the parameter estimation problem for HMMs. It finds the parameter $\underline{\theta}$ that maximizes the likelihood function $f(\underline{y}|\underline{\theta})$ by iteratively maximizing the Baum's auxiliary function

$$Q(\underline{\theta}|\underline{\theta}^{[k]}) = \sum_{\underline{S}} f(\underline{y}, \underline{S}|\underline{\theta}^{[k]}) \log f(\underline{y}, \underline{S}|\underline{\theta}) \quad (3)$$

over $\underline{\theta}$. Here $\underline{\theta}^{[k]}$ is the estimation of $\underline{\theta}$ in the k^{th} iteration; $\underline{S} = [S_1, S_2, \dots, S_N]$ are the unobserved states corresponding to the observations \underline{y} .

From Bayes' rule and the properties of the HMM, we obtain the following equations:

$$f(\underline{y}, \underline{S}|\underline{\theta}) = f(\underline{y}|\underline{S}, \underline{\theta})p(\underline{S}|\underline{\theta}), \quad (4)$$

$$p(\underline{S}|\underline{\theta}) = \pi_{S_1} \prod_{n=2}^N a_{S_{n-1}S_n}, \quad (5)$$

$$f(\underline{y}|\underline{S}, \underline{\theta}) = \prod_{n=1}^N f(y_n|S_n, \underline{\theta}). \quad (6)$$

Substituting Eqs.(4),(5) and (6) into Eq.(3), we have,

$$Q(\underline{\theta}|\underline{\theta}^{[k]}) = C + \sum_{n=1}^N \sum_{i=1}^{K^L} \lambda_n(i) \log f(y_n|S_n = i, \underline{\theta}), \quad (7)$$

where C is a constant independent of $\underline{\theta}$, and $\lambda_n(i)$ is defined as

$$\lambda_n(i) = f(\underline{y}, S_n = i|\underline{\theta}^{[k]}). \quad (8)$$

From Eqs.(2) and (7), the maximization problem is equivalent to minimizing

$$D(\underline{\theta}|\underline{\theta}^{[k]}) = \sum_{n=1}^N \sum_{i=1}^{K^L} \lambda_n(i) \left[\log \sigma^2 + \frac{|y_n - \sum_{l=1}^L h_l d_{i,L+1-l}|^2}{2\sigma^2} \right] \quad (9)$$

over $\underline{\theta}$. Therefore we can find $\underline{\theta}^{[k+1]}$ by solving

$$\nabla_{\underline{\theta}} D(\underline{\theta}|\underline{\theta}^{[k]})|_{\underline{\theta}=\underline{\theta}^{[k+1]}} = 0, \quad (10)$$

which readily leads to

$$\underline{h}^{[k+1]} = G^{-1}\underline{V}, \quad (11)$$

where $\underline{h}^{[k+1]} = (a_1^{[k+1]} b_1^{[k+1]} \dots a_L^{[k+1]} b_L^{[k+1]})^T$ with $a_l^{[k+1]}$ and $b_l^{[k+1]}$ being the real and imaginary parts of h_l in the $k+1^{\text{th}}$ estimation, respectively. G is a $2L \times 2L$ matrix with

$$\begin{bmatrix} \sum_{n=1}^N \sum_{i=1}^{K^L} \lambda_n(i) \text{Re}\{D_{i,lm}\} & -\sum_{n=1}^N \sum_{i=1}^{K^L} \lambda_n(i) \text{Im}\{D_{i,lm}\} \\ \sum_{n=1}^N \sum_{i=1}^{K^L} \lambda_n(i) \text{Im}\{D_{i,lm}\} & \sum_{n=1}^N \sum_{i=1}^{K^L} \lambda_n(i) \text{Re}\{D_{i,lm}\} \end{bmatrix}$$

as its $(l, m)^{\text{th}}$ subblock, where $D_{i,lm}$ is defined as

$$D_{i,lm} = d_{i,L+1-l}^* d_{i,L+1-m},$$

and \underline{V} is a $2L \times 1$ vector with

$$\begin{pmatrix} \sum_{n=1}^N \sum_{i=1}^{K^L} \lambda_n(i) \text{Re}\{d_{i,L+1-l}^* y_n\} \\ \sum_{n=1}^N \sum_{i=1}^{K^L} \lambda_n(i) \text{Im}\{d_{i,L+1-l}^* y_n\} \end{pmatrix},$$

as its l^{th} subvector.

When $\underline{h}^{2[k+1]}$ is available, $\sigma^{[k+1]}$ can be readily obtained as

$$\sigma^{2[k+1]} = \frac{\sum_{n=1}^N \sum_{i=1}^{K^L} \lambda_n(i) |y_n - \sum_{l=1}^L \hat{h}_l^{[k+1]} d_{i,L+1-l}|^2}{2 \sum_{n=1}^N \sum_{i=1}^{K^L} \lambda_n(i)}. \quad (12)$$

The forward-backward procedure provides an efficient way to compute $\lambda_n(i)$. Let $\alpha_n(i)$ and $\beta_n(i)$ be the forward and backward variables, respectively, i.e.,

$$\alpha_n(i) = f(y_n^r, S_n = i | \underline{\theta}^{[k]}), \quad (13)$$

$$\beta_n(i) = f(y_{n+1}^N | S_n = i, \underline{\theta}^{[k]}), \quad (14)$$

then

$$\lambda_n(i) = \alpha_n(i) \beta_n(i), \quad (15)$$

and $\alpha_n(i)$ and $\beta_n(i)$ can be computed recursively by

$$\alpha_n(i) = \sum_{i'=1}^{K^L} \alpha_{n-1}(i') a_{i'i} f(y_n | S_n = i, \underline{\theta}^{[k]}), \quad (16)$$

$$\beta_n(i) = \sum_{i'=1}^{K^L} \beta_{n+1}(i') a_{ii'} f(y_{n+1} | S_{n+1} = i', \underline{\theta}^{[k]}), \quad (17)$$

with the boundary conditions

$$\alpha_1(i) = f(y_1 | S_1 = i, \underline{\theta}^{[k]}) \pi_i, \quad (18)$$

$$\beta_N(i) = 1, \quad (19)$$

for $1 \leq i \leq K^L$.

3.2 Channel Estimation with Training Signal

For the training part, the transmitted sequence \underline{d} is known. The log-likelihood function is given by

$$\log f(\underline{y} | \underline{\theta}) = -N_t \log(2\pi\sigma^2) - \sum_{n=1}^{N_t} \frac{|y_n - \sum_{l=1}^L \hat{h}_l d_{n+1-l}|^2}{2\sigma^2}, \quad (20)$$

where N_t is the length of the training sequence. The maximum likelihood estimate of $\underline{\theta}$ can be simply found by

$$\nabla_{\underline{\theta}} \log f(\underline{y} | \underline{\theta}) |_{\underline{\theta}=\hat{\underline{\theta}}} = 0. \quad (21)$$

Hence we obtain

$$\hat{\underline{h}} = G_t^{-1} \underline{V}_t, \quad (22)$$

where G_t is a $2L \times 2L$ matrix with the $(l, m)^{th}$ subblock

$$\begin{bmatrix} \sum_{n=1}^{N_t} \text{Re}\{d_{n+1-l}^* d_{n+1-m}\} & -\sum_{n=1}^{N_t} \text{Im}\{d_{n+1-l}^* d_{n+1-m}\} \\ -\sum_{n=1}^{N_t} \text{Im}\{d_{n+1-l}^* d_{n+1-m}\} & \sum_{n=1}^{N_t} \text{Re}\{d_{n+1-l}^* d_{n+1-m}\} \end{bmatrix},$$

and \underline{V}_t is a $2L \times 1$ vector with the l^{th} subvector

$$\begin{pmatrix} \sum_{n=1}^{N_t} \text{Re}\{d_{n+1-l}^* y_n\} \\ \sum_{n=1}^{N_t} \text{Im}\{d_{n+1-l}^* y_n\} \end{pmatrix},$$

and

$$\hat{\sigma}^2 = \frac{1}{2N_t} \sum_{n=1}^{N_t} |y_n - \sum_{l=1}^L \hat{h}_l d_{n+1-l}|^2. \quad (23)$$

Comparing Eqs.(11),(12) and (22),(23), we can see that the channel estimation for the training part is a degenerated case of the Baum-Welch algorithm, since

$$\begin{aligned} \lambda_n(i) &= f(\underline{y}, S_n = i | \underline{\theta}^{[k]}) \\ &= \begin{cases} f(\underline{y} | \underline{\theta}^{[k]}), & \text{if } i = i_t \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \quad (24)$$

where i_t is a part of the training sequence.

3.3 Semi-blind Block Channel Estimation and Signal Detection

The Baum-Welch algorithm discussed in Section 3.1 uses the entire observation to iteratively update the estimation of the channel parameters. It may cause a large time delay in decoding and is not suitable to time varying channels. To deal with these problems, we divide the observation interval into sequential disjoint subintervals and apply the Baum-Welch algorithm to each subinterval. Generally speaking, if the observation interval is too short, the Baum-Welch algorithm suffers from the problem that it may converge to a local minimum. However, we can avoid this problem by using the training signal. We can obtain a good estimate of the channel for the subinterval corresponding to the training signal by Eqs.(22)-(23), and then use the estimates as initial values in adjacent subintervals. If the channel is slowly varying, a good initial estimate is made available by using the estimates from the adjacent subintervals, and thus the problem of converging to a local minimum can be avoided.

Notice that MAP decoding can be obtained as a by-product of the Baum-Welch algorithm. For each subinterval, after the final iteration, we decode the transmitted signal by

$$S_n = \text{argmax}_i \lambda_n(i). \quad (25)$$

Therefore, signal detection does not need any additional computation.

4 SEGMENTAL K-MEANS ALGORITHM

Instead of the maximum likelihood criterion

$$f(\underline{y} | \underline{\theta}) = \sum_{\underline{S}} f(\underline{y}, \underline{S} | \underline{\theta}) \quad (26)$$

used in the Baum-Welch algorithm, the segmental k-means algorithm uses

$$\max_{\underline{S}} f(\underline{y}, \underline{S}|\underline{\theta}) \quad (27)$$

as the optimization criterion for estimation the parameters of the HMM. It focuses on the most likely state sequence as opposed to summing over all possible state sequences. The algorithm involves iteration of two fundamental steps: (1) segmentation step, which finds the sequence that maximizes the joint likelihood function $f(\underline{y}, \underline{S}|\underline{\theta})$ for the current parameter estimates; (2) optimization step, which finds new estimates for $\underline{\theta}$ so as to maximize the above state-optimized likelihood. The convergence properties of the segmental k-means algorithm was proved in [4].

Applying the segmental k-means algorithm to our channel estimation and signal detection problem, we obtain a simple iterative structure as follows. At the k^{th} iteration, we have

- Segmentation step:

$$\begin{aligned} \underline{S}^{[k]} &= \underset{\underline{S}}{\operatorname{argmax}} f(\underline{y}, \underline{S}|\underline{\theta}^{[k-1]}) \\ &= \underset{\underline{S}}{\operatorname{argmax}} [\log f(\underline{y}|\underline{S}, \underline{\theta}^{[k-1]}) + \log P(\underline{S})] \\ &= \underset{\underline{S}}{\operatorname{argmax}} [\log f(\underline{y}|\underline{S}, \underline{\theta}^{[k-1]})] \end{aligned} \quad (28)$$

since all \underline{S} 's are equiprobable.

- Optimization step:

$$\begin{aligned} \underline{\theta}^{[k]} &= \underset{\underline{\theta}}{\operatorname{argmax}} f(\underline{y}, \underline{S}^{[k]}|\underline{\theta}) \\ &= \underset{\underline{\theta}}{\operatorname{argmax}} [\log f(\underline{y}|\underline{S}^{[k]}, \underline{\theta})]. \end{aligned} \quad (29)$$

The segmentation step gives the maximum likelihood sequence estimation for each iteration of estimated channel parameters and can be implemented efficiently by the Viterbi algorithm. The optimization step estimates the channel in a decision directed mode, i.e. it updates the estimates of the channel by treating the current estimate of the transmitted signal as a known signal.

As for the Baum-Welch algorithm, the segmental k-means algorithm can also be applied to the received signal block by block. We evaluate the performance of the algorithms for different block sizes in the next section.

5 PERFORMANCE EVALUATION

5.1 Cramer-Rao Lower Bound (CRLB)

The CRLB gives a lower bound for the variance of an unbiased estimate.

$$\operatorname{Var}[\hat{\theta}_i(\underline{Y})] \geq \frac{1}{I_{\theta_i}}, \quad (30)$$

where I_{θ_i} is the Fisher's information for θ_i

$$I_{\theta_i} = -E\left\{\frac{\partial^2}{\partial \theta_i^2} \log f(\underline{Y}|\underline{\theta})\right\}. \quad (31)$$

Using the identity

$$f(\underline{Y}|\underline{\theta}) = \sum_{\underline{S}} P(\underline{S}) f(\underline{Y}|\underline{\theta}, \underline{S}) = E_{\underline{S}}\{f(\underline{Y}|\underline{\theta}, \underline{S})\} \quad (32)$$

and Jensen's inequality

$$\log E_{\underline{S}}\{f(\underline{Y}|\underline{\theta}, \underline{S})\} \geq E_{\underline{S}}\{\log f(\underline{Y}|\underline{\theta}, \underline{S})\}, \quad (33)$$

we have

$$I_{\theta_i} \leq -E_{\underline{S}}E\left\{\frac{\partial^2}{\partial \theta_i^2} \log f(\underline{Y}|\underline{\theta}, \underline{S})\right\}. \quad (34)$$

By substituting Eq.(2) into $f(\underline{Y}|\underline{\theta}, \underline{S})$ of the above expression, we find

$$I_{a_i} \leq \frac{1}{\sigma^2} E_{\underline{S}}\left\{\sum_{n=1}^{N'} |d_{n+1-i}|^2\right\} = \frac{N'}{\sigma^2} E\{|d|^2\} \quad (35)$$

$$I_{b_i} \leq \frac{N'}{\sigma^2} E\{|d|^2\} \quad (36)$$

$$\begin{aligned} I_{\sigma^2} &\leq -E_{\underline{S}}\left\{\frac{N'}{\sigma^4} - \frac{\sum_{n=1}^{N'} E\{|Y_n - \sum_{l=1}^L h_l d_{n+1-l}|^2\}}{\sigma^6}\right\} \\ &= \frac{N'}{\sigma^4} \end{aligned} \quad (37)$$

where a_i and b_i are the real and imaginary parts of h_i , respectively, and N' is the length of subintervals. We see that the CRLB is inversely proportional to the block size.

5.2 Simulation Results

We evaluate the above semi-blind block channel estimation and signal detection schemes by simulation in terms of their BER performances and MSEs of the estimated channel impulse responses. We take an antipodal signal where $d_n \in \{+1, -1\}$ and a multipath channel with impulse response $\underline{h} = [0.408 \ 0.816 \ 0.408]$ (channel (b) in [6], chap.10). We choose two sets of parameters for initial estimates of \underline{h} :

$$\begin{aligned} \underline{h}_{\text{init1}} &= [0.577 \ 0.577 \ 0.577] \\ \underline{h}_{\text{init2}} &= [0.3 \ 0.9 \ 0.3] \end{aligned}$$

$\underline{h}_{\text{init1}}$ represents the case when we do not have any knowledge of the channel. In this case, we can only randomly choose an initial estimate, and the best we can do is to take an average value. $\underline{h}_{\text{init2}}$ represents the case when a good initial estimate is available from an adjacent subinterval. We test the performance for different

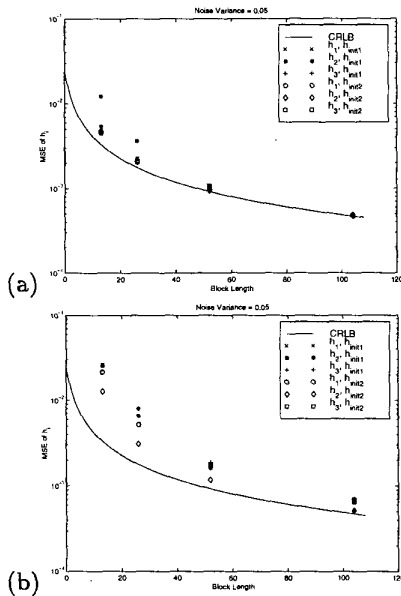


Figure 1: MSE of \hat{h} , SNR= 13dB (a) Baum-Welch algorithm; (b) segmental k-means algorithm

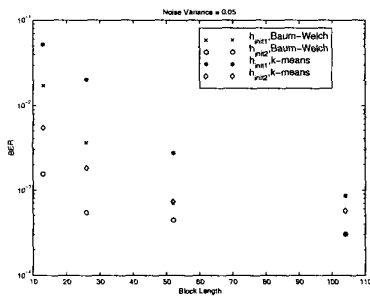


Figure 2: BER for different block lengths, SNR=13dB

block lengths $N' = 13, 26, 52, 104$ with SNR varies from 7dB to 13dB.

Fig. 1 and 2 illustrate the results for SNR= 13dB. The results for other SNR values exhibit similar patterns.

From Fig. 2 we can see that for both the Baum-Welch algorithm and the segmental k-means algorithm if the channel is completely unknown, there is a substantial performance degradation to be incurred by using short blocks. However, if a good initial estimate is available, the block length can be shortened to a certain extent with only a minor performance loss. Also notice that with a good initial estimate, the performance degradation of the segmental k-means algorithm compared with the Baum-Welch algorithm is reasonably small. The segmental k-means algorithm uses the Viterbi algorithm, whereas the Baum-Welch algorithm involves the forward-

backward algorithm. Thus, the latter is much more computationally involved. In view of complexity, the segmental k-means algorithm is more attractive.

6 CONCLUDING REMARKS

We have proposed two semi-blind block channel estimation and signal detection algorithms using the Baum-Welch algorithm and the segmental k-means algorithm for HMMs. The algorithms apply to the subintervals of the received signal instead of to the whole observation interval. This approach provides several advantages in terms of time delay, complexity and the ability to deal with a slowly fading channel. Our simulation results show that if a good initial estimate is made available by using a training signal, the algorithms with short block length suffer only a minor performance loss. Comparing with the Baum-Welch algorithm, the segmental k-means algorithm is more attractive when the complexity is a major issue.

ACKNOWLEDGMENTS

The present work has been supported by the National Science Foundation, the New Jersey Center for Wireless Telecommunications (NJCWT) and Mitsubishi Electric Research Laboratories (MERL).

REFERENCES

- [1] L. E. Baum, "An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes," *Inequalities*, Vol. 3, pp. 1-8, 1972.
- [2] H. A. Cirpan and M. K. Tsatsanis, "Stochastic maximum likelihood methods for semi-blind channel estimation," *IEEE Signal Processing Lett.*, Vol. 5, No. 1, pp. 21-24, 1998.
- [3] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.*, Vol. 39, No. 1, pp. 1-38, 1977.
- [4] B. H. Juang and L. R. Rabiner, "The segmental k-means algorithm for estimating parameters of Hidden Markov Models," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 38, No. 9, pp. 1639-1641, 1990.
- [5] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. Commun.*, Vol. 42, No. 7, pp. 2406-2413, 1994.
- [6] J. G. Proakis, *Digital Communications*, 3rd ed., McGraw-Hill, 1995.
- [7] L. R. Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition," *Proceedings of the IEEE*, Vol. 77, No. 2, pp. 257-285, 1989.