Speech Recognition Using Hidden Markov Model

Ankita Gupta*
Gajendra singh**
*-M.Tech(communication & Information).
Deptt. of Electronics & Communication, SSITM, Aligarh
Email:-ankita.gupta.ssitm.20@gmail.com
** H.O.D
Dept. of Electronics & communication , SSITM

Neha Sharma#,
#-Assistant Lecturer
Dept. of Electronics And Communication
ACET Aligarh
Email:-nsnehasharma14@gmail.com

Abstract-The aim of this paper on "Speech Recognition using Hidden Markov Model (HMM)" is to create and develop a limited vocabulary speech recognition system with the help of HMM to statically model the words in the dictionary. For this purpose we use a database that contains a samples of different speaker (for better accuracy, at last twenty speakers) are desirable. So, this speech recognition system is speaker dependent to test the theory presented in this paper and gets a practical experience of speech recognition using HMM the speech recognizer machine should be implemented on Matlab. This implementation is based on mathematical calculations and equation. Using these implementations, the different word samples spoken by different speakers undergoes through different sub process like segmentation of signal, clustering of signal. Then its training phase described in this paper to get a reliable training at least 20 samples of each word from 20 different speakers are required to record in an environment tree from external disturbance. With the help of these information's, a speech recognition model is ready to recognize the given speech.

KEY WORDS
HMMS – Hidden Markov Model

INTRODUCTION

The purpose of this paper on speech recognition using HMM is to get a deeper theoretical and practical understanding of speech recognizer. The works started by examine a currently existing state of speech recognizer. The work started by examine a currently existing state of speech recognizer called Hidden Markov Tool Kit (HTK) HTK is not easy to use, because it need to have its complete theoretical knowledge for its functioning with the experience from HTK, it was desired to make a system that would be easier to use. The web page is limited because it is not possible to try different settings. This is why we use a programming language Matlab for this purpose because it is easily extended with new features applying this knowledge in practical manner speech recognizer implemented in Matlab is ready to use.

What is speech recognition?

Speech recognition is the process of recognizing some specific characteristic features of an audiable signal.

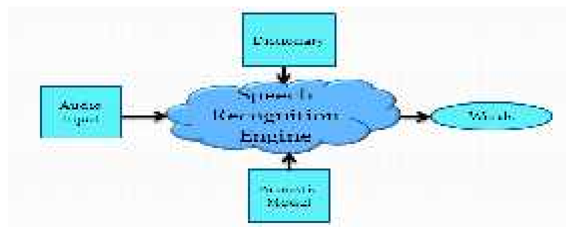
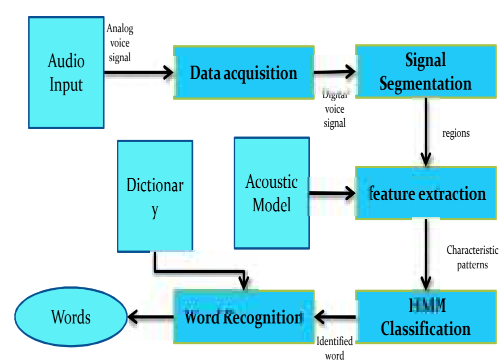


Figure (a): Basic Speech Recognition Process

Speech Recognition Engine



In speech recognition Engine, we have a caustic model which produce the cepstral coefficients & match that word with the trained HMM words dictionary to identify the test word a recent preponderance of recognizers based on statistical models, It was felt that a HMM clustering model would be more effective than a template based approach. A HMM based recognizer first processes the speech signal to convert it into digital signal, then it performs segmentation of this digital signal to derive its cepstral coefficients (characteristic feature of a signal) then after its training phase a dictionary of words is ready to recognize any test word.

What is HMM?

A Hidden Markov Model is a finite state machine having a fixed number of states in which Transition of states takes place from are state to another and also the next state only depends on present state. It is a finite set of states, with a probability distribution. The transition among the states is governed by transition probabilities. An outcome can be generated according to distribution. Only the outcome not the state is visible to an external observer. It is a statistical method of characterizing the spectral properties of frames of a pattern. The underlying assumption of HMM is that the speech can be well characterized as parametric random process and that the parameters of the stochastic process can be estimated in well defined manner. A HMM is characterized by three matrices i.e. A,B,π

A→State Transition Probability matrix (NòN)

B→Observation Probability Distribution Matrix (NòM)

π→Initial state Distribution Matrix (Nò1)

where, N = Number of states in HMM

M= Number of observation symbols.

Therefore HMM can be defined as.

N,M.λ =(A,B, π)

Where λ →Trained HMM

<u>Three Basic Issues For HMMs</u>

In order to apply HMMs for speech recognition, we need to address three problems stated below

1- <u>Evaluation Problem</u>

Given the λ observation sequence $O = \{O_1, O_2---O_T\}$

How do we efficiently compute P {O/λ}

<u>Solution :</u> Recursive procedures like forward and Backward Procedure exist which can compute p {O/λ}

- Define forward variable

  $\propto_t (i)=P\{O_1,O_2------O_t.q_t =i/\lambda\}$ (1)

  recursion

$\propto_{t+1} (i)=b_j(O_{t+1},) (\Sigma_I \propto t (i)\ a_{ij})$

- Similarly backward variable

  $\beta_t (i) = P\ \{O_{t+1},O_{t+1}----------O_t.q_t =i/\lambda\}$ (2)

  recursion

  $\beta_t (i) = (\Sigma_{j+}\beta_{t+1} (i)\ aijbj(O_{t+1}))$ (3)

  Initialsing :

  $\propto_1 (j)=\pi_i b_i(O_1)$ and $\beta_t (i) =1$ (4)

  Hence, $P\{O/\lambda\} = \Sigma_I \propto_t (i)\ \beta_t(i)$ (5)

2- <u>Decoding Problem</u>

Given the observation sequence $O = \{O_1,O_2-------O_t\}$ and a model λ, how do we choose a corresponding state sequence $Q=\{q_1, q_2--------q_t\}$ M which is optimal

in some meaningful sense.

<u>Solution-</u> Viterbi Algorithm finds the single best sequence or for given observation sequence O. The following equation are presented which is viterbi Algorithm.

a) Initialization:

$\delta_1 (i) = \pi_i b_i (O1),\ 1\le i \le N$ (6)

$\Psi_1 (i) = 0$ (7)

b) Recursion :

$\delta_t (i) = \max [\delta_{t-1} (i)a_{ij}]\ bj\ (o_t),\ 2 \le t \le t$ (8)

$1\le i \le N \qquad 1 \le j \le N$

$\psi_t (i) = \arg\max [\delta_{t-1} (i)\ aij],\ 2 \le t \le t$ (9)

$1\le i \le N \qquad 1 \le j \le N$

c) Termination :

$P * = \max [\delta_t (i)]$ (10)

$1\le i \le N$

$q*_t$ arg. Max $[\delta_t(i)]$ (11)

$1\le i \le N$

d) Path (State sequence) back tracking :

$q_t* \ \psi_{t+1}(q*_{t+1})$, t=t-1, t-2--------1 (12)

<u>3- Learning Problem</u>

This is the problem of parameter estimation. This is by far is the toughest problem of HMM. How do we adjust the model parameter.

λ = (A,B,π) to maximize. P(o/λ)?

Solution- We can, however, choose

λ = (A,B,π) such that probability p(O/λ) is locally maximized using iterative procedure such as Baum-welch method. But Baum Welch re-estimation procedure suffers from follow problems-

a) Numerical problem and hence hard to Implement.
b) Needs special scaling.
c) Needs multiple observation sequences. There exist one more method which can overcome these problems & can be used to train a HMM. This is segmental K-means method.

Segmental k-means uses the solution to problem 2 to modify the model parameters. This process undergoes following steps.

- Consider a starting HMM λ
- Calculate Transition Probability matrix A
- Calculate Emission matrix or observation symbol probability Distribution matrix B.
- Update HMM parameters according to Transition & Emission probability)
- Continue recursion & update the parameters of HMM.

<u>Applications</u>

Speech recognition using the technique Hidden markov Model is very useful in many applications and environments in our daily life. Generally speech recognizer is a machine which understands human and their spoken words in some way & can act thereafter.

It can be used in following fields-

1- In car environment to voice control non-critical operations, such as dialing any phone number.
2- On board Navigation, presenting the driving route to driver.
3- By applying voice control, traffic safety will be increased.
4- Facilitate a person with functional disability or other kinds of handicaps. To make their daily works easier with their voice they could operate the light switch, turn off/ On any domestic appliances. This leads to intelligent homes where these operations can be made available for common man as well as for handicapped.
5- Can be used in Han held digital recorder.
6- This technique can be useful in any field whether it is commercial, Industrial, Military, or forensic applications.

RESULT

To test the theory presented in this paper and get a practical experience of speech recognizer is to be implemented on Matlab. The implementation is based on mathematical calculations and equation that are explained as the three Basic problems of HMMs. (i.e. Evaluation problem, Decoding Problem, Learning Problem) Using these implementations HMMs for different words & speech can be designed or trained. In complete process to training a HMM different words spoke by several speaker undergoes through different sub processes like segmentation, derivation of cepstral coefficients then clustering and after that training of HMMs of different speech or word. For the above processes, we implement the whole process on Matlab because all these process and functions are predefined on Matlab. Therefore we need not to go in deep and large mathematical calculations An once the HMMs are trained speech recognizer machine is ready to recognize any test speech. To get a reliable training at least twenty samples of different words from at least twenty speakers are required to record un disturbance free environment. The training results in speech recognizer mode for different speech. The model represents a statically mode (Hidden Markov Model, HMM) for each word.

HMM Performance on Isolated Word Recognition

We conclude this section on isolated word recognition using HMMs by giving a set of performance results (in terms of average word error rate) on the task of recognizing isolated digits in a speaker independent manner. For this task, a training set consisting of 100 occurrences of each digit by 100 talkers (i.e., a single occurrence of each digit per talker) was used. Half the talkers were male; half female. For testing the algorithm, we used the initial training set, as well as three other independent test sets with the following characteristics:

TS2 : the same 100 talkers as were used in the training; 100 occurrences of each digit
TS3 : a new set of 100 talkers (50 male, 50 female); 100 occurrences of each digit
TS4 : another new set of 100 talkers (50 male, 50 female); 100 occurrences of each digit

The results of the recognition tests are given in Table1. The recognizers are the following:

LPC/DTW : Conventional template-based recognizer using dynamic time warping (DTW) alignment
LPC/DTW/VQ : Conventional recognizer with vector quantization of the feature vectors (*M* = 64)
HMM/VQ : HMM recognizer with *M* = 64codebook
HMM/CD : HMM recognizer using continuous density model with *M* — *5* mixtures per state
HMM/AR : HMM recognizer using auto regressive observation density

Table 5.2 Average Digit Error Rates for Several Recognizers and Evaluation Sets

Recognizer Type	Evaluation Set			
	Original Training	TS2	TS3	TS4
LPC/DTW	0.1	0.2	2.0	1.1
LPC/DTW/VQ	–	3.5	–	–
HMM/VQ	–	3.7	–	–
HMM/CD	0	0.2	1.3	1.8
HMM/AR	0.3	1.8	3.4	4.1

It can be seen that, when using a VQ, the performance of the isolated word recognizer degrades in both the conventional and HMM modes. It can also be seen that the performances of the conventional template-based recognizer, and the HMM recognizer with a continuous density model are comparable. Finally Table 1 shows that the autoregressive density HMM gives poorer performance than the standard mixture density model.

CONCLUSION

In this paper we have to present the theory of hidden Markov models in simple ways and concepts. Our main purpose is to focus on physical explanations of basic mathematics, therefore. We have avoided long drawn out proofs and derivations of the key results and to cussed primarily on trying to interpret the meaning of math and its practical implementation in real world systems. We have also illustrate some of the applications of theory of HMMs to simple problems in speech recognition, and point out how the technique could be applied to more advanced speech recognition problems.

For under standing the fundamentals of speech recognition using hidden Markov model, in real sense, a person needs hand on experience with the software, hardware and platform, Hence strongly encourage all serious readers of this paper to program the algorithms, implement the system, and literally build application. Without practical experience reader will not found the words written in this paper alive.

REFERENCES

[1] L.R. Rabiner, "A Tutorial on Hidden Markov Models and selected application in speech Recognition "Proceedings of IEEE trans.,Vol.77, No.2, FEBRUARY 1981

[2] L.A. Liporace, "Maximum likelihood estimation for multivariate observations of Markov sources, "IEEE Trans. Informat. Theory, Vol. IT 28, no.5, pp. 729-734, 1982.

[3] B.H. Juang, "Maximum likelihood estimation for mixture multivariate stochastic observations of Markov chains, " AT &T Tech. J., vol. 64,no. 6pp. 1235-1249, July- Aug. 1985.

[4] B.H.Juang, S.E.Levinson, and M.M. Sondhi, " Maximum likelihood estimation for multivariate mixture observations of Markov chains, "IEEE Trans. Informat. Theory, vol. IT-32, no.2, pp. 307-309, Mar, 1986.

[5] L.R. Rabiner, B.H. Juang, S.E.Levinson, and M.M.Sondhi, "Some properties of continuous hidden Markov model representations. "AT & T Tech. J., vol.64no.6, pp. 1251-1270,July-Aug. 1985.

[6] K.F.Lee and H.W. Hon, "Large-vocabulary speaker-independent continuous speech recognition, " in Conf. Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, pp. 123-126, Apr. 1988.

[7] L.R.Rabiner, S.E.Levinson, and M.M.Sondhi, "On the application of vector quantization and hidden Markov models to speaker-independent isolated word recognition, "Bell syst. Tech. J., vol. 62, no. 4, pp.1075-1105, Apr. 1983.

[8] R. Billi, " Vector quantization and Markov source models applied to speech recognition, "in Proc. ICASSP '82 (Paris, France), pp.574-577, May 1982.

[9] L.R. Rabiner, B.H.Juang, S.E.Levinson, and M.M. Sondhi, "Recogntionof isolated digits using hidden Markov models with continuous mixture densities, " AT & T Tech. J. vol .64, no.6, pp1211-1222, July-Aug. 1986.

[10] L.R. Rabiner, J.GWilpon, and B.H.Jaung, "A segmental k-means training procedure for connected word recognition, AT & T Tech.J., vol. 65, no. 3, pp. 21-31, May-June, 1986.

[11] L.R.Rabiner, J.G.Wilpon, and S.E.Levinson, " A speaker-independent, syntax-directed, connected word recognition system based on hidden Markov models and level building, "IEEE Trans. Acoust. Speech Signal Processing. Vol. ASSP-33, no. 3pp.561-573, June 1985.

[12] L.R.Rabiner, J.G.Wilpon, and B.H.Juang. "A model-based connected digit recognition system using either hidden Markov models or templates, "Computer, Speech, and Language, vol.1, no.2, 167-197, Dec. 1986.

[13] A.B.Portiz, "Linear predictive hidden Markov models and the speech signal, "in Proc. ICASSP 82(Paris, France), PP. 1291-1294, May 1982.

[14] B.H.Juang and L.R. Rabiner, "Mixture autoregressive hidden Markov models for Speech Signal Processing, Vol. ASSP-33, no.6, pp.1404-1433, Dec. 1985.

[15] M.J. Russeland R.K.Moore, "Explicit modeling of state occupancy in hidden Markov models for automatic speech recognition. "Proc. ICASSP 85 (Tempa, FL), pp. 5-8, Mar.1986.

[16] B.Lowerre an R.Reddy, "The HARPY speech understanding system, "In Trends in Speech Recognition, W.Lea, Editor. Englewood Cliffs, Nj: Prentice-Hall, 1980, pp. 340-346.

[17] L.R.Bahi, P.F. Brown, P.V. de Souza, and R.L. Mercer, "Maximum mutual information estimation of hidden Markov model parameters for speech recognition, " In Proc, ICASSSP 86 (Tokyo, Japan), pp. 49-52, Apr. 1986.

[18] Y.Ephram, A.Dembo, and L.R.Rabiner, "A minimum discrimination information approach for hidden Markov modeling, "in Proc. ICASSP 87 (Dallas, TX) Apr. 1987.

[19] B.H.Juang and L.R. Rabiner, " A probabilistic distance measure for hidden Markov models, "At & T Tech. J., vol.64, no.2, pp. 391-408, Feb. 1985.

[20] F.Jelinek and R.L. Mercer, " Interpolated estimation of Markov source parameters from sparse data, " in Pattern Recognition in Practice, E.S. Gelesma and L.N. Kanal, Eds. Amsterdam, The Netherlands: North Holland, 1980, pp. 381-397.

[21] R.Schwartz et al., "Context dependent modeling for acoustic phonetic recognition of continuous speech, " and Signal Processing, pp. 1205-1208., Apr. 1985.

[22] __, "On the use of hidden Markov models for speaker-independent recognition of isolated words from a medium size vocabulary. " AT & T Tech. J., vol. 63, no. 4, pp. 627-642, Apr. 1984.

[23] A.B. Portiz and A.G.Richter, "lsolated word recognition, "in Proc. ICASSP' 86 (Tokyo, Japan), pp. 705-708, Apr. 1986.

[24] R.P.Lippmann, E.A.Martin, and D.B.,Paul, " Multi style training for robust isolated word speech recognition, " in Proc. ICASSP 87 (Dallas, TX)pp. 705-708, Apr. 1987.

[25] D.B.Paul, " A Speaker Stree resistant HMM ISOLATED WORD RECOGNIZER, " in Proc. ICASSP 87 (Dallas, TX), pp. 713-716, Apr. 1987.

[26] V.N. Gupta M. Lennig and P. Mermelstein, "Integration of acoustic information in large vocabulary word recognizer, " in Conf. Proc. IEEE Int. Cont. on Acoustics, Speech, and Signal Processing. Pp. 697-700, Apr. 1987.

[27] S.E.Levinson, "Continuous speech recognition by means of acoustic-phonetic classification obtained from a hidden Markov model, "in Proc. ICASSP 87 (Dalls TX) Apr. 1987.

[28] J.G. Wilpon, L.R. Rabiner, and T.Martin, "An improved word detection algorithm for telephone quality speech incorporating both syntactic and semantic constraints, " AT & T Bell Labs Tech. J., Vol 63. No.3, pp. 479-498, Mar. 1984.

[29] J.G. Wilpon and L.R. Rabiner, " Application of hidden Markov models to automatic speech endpoint detection, " Computer Speech and Language, Vol 2, no. 3/4 , pp. 321-341, Sept. Dec. 1987.

[30] A. Averbuch et al., "Experiments with the TANGORA 20,000 WORD SPEECH RECOGNIZER, " IN Conf. Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, pp. 701-704, Apr. 1987.

[31] B.S. Atal and S.L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave, "J.Acoust. Soc. Am., vol. 50, pp. 637-655, 1971.

[32] F.L.Ltakura and S.Sati. " Analysis synthesis telephone based upon the maximum likelihood method, " in Proc. 6th Int. Congress on Acoustics (Tokyo, Japan) pp. C17-20, 1968.

[33] J. Makhoul, "Linear prediction: A tutorial review, " Proc. IEEE vol. 63, pp. 561, 580, 1975.

[34] J.D. Market and A.H. Gray Jr. , Linear Prediction of Speech. New York, NY: Sringer- Verlag, 1976.

[35] Y.Tokhura, " A weighted cepstral distance measure for speech recognition, " IEEE Trans. Acoust. Speech Signal Processing vol. ASSP-35 no. 10, pp. 1414-1422, Oct. 1987.

[36] B.H.Juang, L.R.Rabiner, and J.G. Wilpon, "On the use of bandpass littering in speech recognition," IEEE Trans. Acoust. Speech Signal Processing. Vol ASSP-35, no.7, pp. 947-954, July 1987.

[37] S. Furui, "Speaker independent isolated word recognition based on dynamics emphasized cepstrum," Trans. IECE of Japan, vol. 69, no. 12, pp.1310-1317, Dec. 1986.

[38] F.K.Soong and A.E.Rosenberg, " On the use instantaneous and transitional spectral information in speaker recognition," in Proc. ICASSP 86 (Tokyo, Japan). pp. 877-880, Apr. 1986.

[39] H.Bourlard, Y.Kamp, H.Ney, and C.J. Wilpon, and B.H. Juang, " A speaker independent connected speech recognition via dynamic programming and statistical methods," in Speech and Speaker Recognition, M.R. Schoreder, Ed. Basel, Switzerland: Karger, 1985, pp.115-148.

[40] C.J. Wellekens, "Global connected digit recognition using Baum-Welch algorithm," in Proc. ICASSP 86(Tokyo, Japan), pp. 1081-1084, Apr. 1986

[41] A.M. Derouault, "Context dependent phonetic Markov models for large vocabulary speech recognition," in Proc. ICASSP 87 (Dallas, TX), Paper 10,1.1, pp.360-363, Apr.1987.

[42] B.Merialdo, " Speech recognition with very large size dictionary," in Proc. ICASSP 87 (Dallas, TX), Paper 10.2.2, pp.364-367, Apr. 1987.

[43] Y.L.Chow et al., "BYBLOS: The BBN continuous speech recognition system,"in Proc. ICASSP 87 (Dallas, TX), Paper 3.7.1, pp. 89-92, Apr. 1987.

[44] L.E.Baum and T. Pertrie, " Statistical inference for probabilistic functions of finite state Markov chains," Ann Math. Stat., vol. 37, pp. 1554-1563. 1966.

[45] L.E.Baum and J.A. Egon, "An inequality with applications to statistical estimation for probabilistic functions of a Markov Process and to a model for ecology,"Bull Amer. Meteorol. Soc., Vol. 73, pp. 360-363, 1967.

[46] L.E.Baum and G.R. Sell, "Growth functions for transformations on manifolds, "Pac1 Math.,vol. 27, no. 2 pp. 211-227, 1968.

[47] L.E.Baum, T.Petrie, G. Soules, and N. Weiss. "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, Ann Math Stat., vol 41, No. 1, pp. 164-171, 1970

[48] L.E.Baum, "An inequality and associated maximization technique in statistical estimation for probabilistic function of Markov processes," Inequalities, vol.3, pp1-8 1972.

[49] J.K.Baker, "The dragon system- An overview, " IEEE Trans. Acoust. Speech Signal Processing, vol ASSP-23, no. 1, pp. 24-29, Feb. 1975.

[50] F. jelinek,"A fast sequential decoding algorithm using a stack, " IBM J. Res Develop., vol. 13, pp. 675-685, 1969.

[51] L.R. Bahi and F. jelinek, " Decoding for Channels with insertions, deletions, and substitutions with applications to speech recognition., "IEEE Trans. Informat. Theory. Vol IT-21, pp 404-411, 1975.

[52] F.Jelinek, L.R. Bahi, and R.L. Mercer, " Design of a linguistic statistical decoder for the recognition of continuous speech, " IEEE Trans. Informat. Theory, vol. IT-21, pp. 250-256, 1975.

[53] F.jelinek, " Continuous speech recognition by statistical methods, " Proc. IEEE, vol. 64, pp.532-536, Apr. 1976.

[54] R. Bakis, "Continuous speech word recognition via centi-second acoustic states,in Proc. ASA Meeting (Washington, DC), Apr. 1976.

[55] F. jelinek, L.R.Cahi, and R.L. Mercer, " Continuous speech recognition: Statistical methods, " in Handbook of Statistics, H.P.R.Krishaniad, Ed. Amsterdam, The Netherlands North Holland, 1982.

[56] L.R. Bahi, Jelinek, and R.L. Mecer, "A maximum likelihood approach to continuous speech recognition, " IEEE Trans. Pattern Anal. Machine Intell., vol PAMI-5 pp. 179-190, 1983.

[57] S.E. Levinson, L.R. Rabiner, and M.M. Sondhi, "An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition, "Bell Syst. Tech. J., vol. 62., no. 4, pp. 1035 1074, Apr. 1983.

[58] B.H. Juang. "On the hidden Markov model and dynamic time warping for speech recognition- A unfied view. " AT & T Tech. J., vol. 63, no.7, pp. 12131243, Sept. 1984.

[59] L.R.Rabiner and B.H. Juang, "An introduction to hidden Markov models, "IEEE ASSP Mag., vol. 3, no.1. pp. 4-6, 1986.

[60] J.S.Bridle, " Stochastic models and template matching: Some important relationships between two apparently different of Acoustics, Autum Conf., pp. 1-8, Nov 1984.

[61] J.Makhoul, 5. Roucos, and H.Gish, "Vector quantization in speech coading, "Proc. IEEE, vol. 73. No. 11. pp. 1551-1558, nov. 1985.

[62] S.E.Levinson, "Structural methods in automatic speech recognition. "Proc. IEEE, vol 73, no. 11, pp. 1625-1650, Nov. 1985.

[63] A.W.Darke, "Discrete-state Makov processes, "Chapter 5 in Fundamentals of Applied Probability Theory. New York, NY: McGraw-Hill, 1967.

[64] A.J.Viterbi, "Error bounds for convolutional codes and an asymptotically optimal decoding algorithm, "IEEE Trans. Informat. Theory, vol. IT-13, PP.260-269, Apr. 1967.

[65] G.D. Fornery, "The Viterbi algorithm, "Proc. IEEE, vol. 61, pp. 268-278, Mar. 1973.

[66] A.P.Dempster, N.M.Laird, and D.B.Rubin, "Maximum likelihood from incomplete date via the EM algorithm, "J.Royy.Stat.Soc. Vol. 39, no.1, pp.1-38, 1977.

