Feature Selection for Speech Recognition using Hidden Markov Model

[1]Ms. Priyanka Shinde, [2]Prof. P. M. Ghate

Department of Electronics and Telecommunication
Rajarshi Shahu College of Engineering, Pune

Abstract— **Present system focus on the challenging issue of selection of feature for HMM for speech recognition application. The features which do not contribute in distinguish between two states could be removed without affecting the usefulness of model. In this paper, Feature Saliency is introduced for selection of feature for Hidden Markov model. The feature saliency gives probability of relevance of feature that distinguish between state independent distributions. An expectation maximization algorithm is used for calculation of maximum a posteriori estimates. Exponential and beta priors are used to include cost in the process of selecting feature. The feature extraction process is implemented using MFCC (Mel Frequency cepstral Coefficients).Extracted MFCC features are given to pattern trainer and are trained by HMM to create HMM model for each word. EM algorithm is used to find out maximum likelihood. The speech recognition process depends on frequency analysis. This can be done because each person has some very unique characteristics to their voice that can be isolated in the frequency domain. This paper presents an approach to the recognition of speech signal using frequency spectral information with Mel frequency for the improvement of speech feature representation in a HMM based recognition approach. There are two strong reasons why Hidden Morkov Model is used. Very first reason is the models are very rich in mathematical structure and hence can form the basis for use in a wide range of applications. Second reason is the models, if applied properly, work very well in practice for several important applications.**

Index Terms: **Hidden Markov Model (HMM), Mel Frequency Cepstral Coefficients (MFCC), Expectation Maximization (EM).**

I. INTRODUCTION

Speech is natural vocalized and also a primary means of communication. Speech recognition is process in which sequence of spoken word is translated into text. This project is designed to simplify this communication barrier by helping the computer understand human speech through an speech recognition system. Communicating with computer through speech is more simple, fast and comfortable rather than using other medium like mouse keyboard for human being. Speech recognition also provides access for anybody who has a handicap that prevents use of a keyboard. The entire class of disabled people can use a computer. Speech recognition will provide potentially way to make their lives easier. Computers also need a way to be able to identify who is trying to use them. The most common method of user identification is through the use of passwords. Passwords protection is very effective using speech recognition for several reasons. As several characteristics to a person's voice that are unique to the individual. Because of these unique characteristics, a person's voice could be a very accurate way to authenticate a user. There are many algorithms and techniques are use.

There are two modes of speech recognition system: training mode and testing mode. In training mode all utterances of speaker are processed and feature vector s corresponding to utterances are found using feature extraction techniques like LPC, MFCC, and LPCC. The training vector has spectral features that distinguish different words. This training vector of spectral features is used for testing purpose. In testing mode test utterances are used for which system is trained to find. For each word test pattern is generated which is nothing but extracted features of that utterance used for testing. Thus test pattern is compared against training pattern using different classification techniques such as HMM, ANN, KNN etc. If the test pattern and training pattern matches then it means that particular pattern is recognized by training mode that corresponding pattern is displayed as output. This paper reports the findings of the speech recognition study using the MFCC and HMM techniques

II. LITERATURE SURVEY

The speech recognition idea was begun in 1940s , however basically the primary speech recognition was showed up in 1952 at the bell labs, that was about recognition of a digit in a commotion free condition .1950s was the establishment time frame for the discourse recognition innovation, in this period work was done on the foundational ideal models of the speech recognition that is automation and information theoretic models .In the 1960's we could perceive little vocabularies (request of 10-100words) of disconnected words, in light of basic acoustic-phonetic properties of speech sounds . The key innovations that were created amid this decade were channel banks and time normalization strategies. In 1970s the medium vocabularies (request of 100-1000 words) utilizing simple, template based, pattern recognition techniques were perceived. In 1980s expansive vocabularies (1000-boundless) were utilized and speech recognition issues based on statistical with large networks for language handling structures were addressed. The Hidden Markov Model (HMM) and the stochastic language model were developed is this time, which together were potential techniques for taking care of consistent speech recognition issue proficiently and with elite.

Hidden Markov models (HMMs) and their

augmentations Hidden semi-Markov models (HSMMs) are broadly utilized for modeling consecutive information. Speech recognition, video identification, and tool wear monitoring are only a couple of the fields in which HMMs and HSMMs have discovered far reaching application. These models are made out of time series of correlated hidden up and observed arbitrary variables, with the last appearing as a vector of features normally ascertained from raw information created by sensors or other observational channels. An imperative issue in the development of HMMs and HSMMs is the determination of which feature to use in the model. Techniques for highlight feature selection particularly intended for HMMs and HSMMs are the concentration of this paper.

Models developed utilizing a high-dimensional feature vector may contain highlights that don't add to distinguishing between the states. The guideline of miserliness would recommend evacuating these features if this should be possible without significantly influencing the usefulness of the model for estimating the fundamental states. One conceivable way to deal with reducing the dimensionality of the feature vector is to demonstrate each conceivable subset of features and select the model with the greatest likelihood for explaining the information. This approach becomes unfeasible as the quantity of features increases exponential growth in the quantity of feature of subsets. Further, this way to deal with selection of feature choice does exclude the cost of gathering each feature in the likelihood evaluation.

It is possible to consider cost as basis of feature selection. Cost paid for collection of feature is characterized as cost of test [1]. For including the cost in selection process several methods are studied which methods balance test cost with misclassification cost and focuses on decision systems [2] K-nearest neighbor [4], decision trees [3]. But these methods require labeled data because they use a mis-classification as a measure in the feature selection process. Ji and Carin [5] formulate the feature selection with cost problem as a partially observable Markov decision process (POMDP) in which the actions are the selection of features to sample, and the hidden states are mixture components. This POMDP formulation has several limitations, including significant computational requirements and the difficulty of handling continuous features.

In this paper, we propose feature saliency model that bridges subset-based and test-cost-based strategies by effectively optimizing model parameters and selecting relevant feature subsets given the cost of each feature. Our models, the feature saliency Hidden Markov Model (FSHMM) give maximum a posteriori (MAP) estimates and select important feature, under the suspicion that the number of states are known. Knowing the number of fundamental states makes it possible to utilize the expectation maximization (EM) algorithm. The EM results exact parameter selection, and decreases computational time and model complexity compared to different options that have been considered in the literature. In our approach, new parameters called feature saliencies are brought into the model and used to check how much a given feature distinguish the states.

III. METHODOLOGY:

An important issue in the construction of HMMs is the selection of which features to use in the model. Methods for feature selection specifically designed for HMM are the focus of this project. A high-dimensional feature vector model may not contain features that do not contribute to distinguishing between the states. The principle of parsimony would suggest remove these features if this significantly not affecting the usefulness of the model for estimating the underlying states. For reducing the dimensionality of the feature vector, one approach is to model every possible subset of features and selecting the model with the greatest likelihood for explaining the data. This will become impractical because of exponential growth in the number of feature subsets the number of features increases. This approach for selection of feature does not include the cost for collection of each feature in the likelihood evaluation. Another approach is test cost based methods in which test cost is balanced with misclassification cost and focus on decision systems, decision trees and k-nearest neighbor [9]. But these methods require labeled data as misclassification measure is used in feature selection process.

In this work, we propose Feature saliency model that bridge the literature on subset-based and test-cost-based approaches for efficiently optimizing model parameters and relevant feature selection subsets given the cost of each feature. Our models, the feature saliency hidden Markov model (FSHMM) and provide maximum a posteriori (MAP) estimates and select relevant features, by assuming that the number of states are known. It is possible to use the expectation maximization (EM) algorithm as number of states is known. The EM algorithm gives accurate parameter estimates, reduces computational time and also model complexity.

IV. SPEECH RECOGNITION BLOCK DIAGRAM:

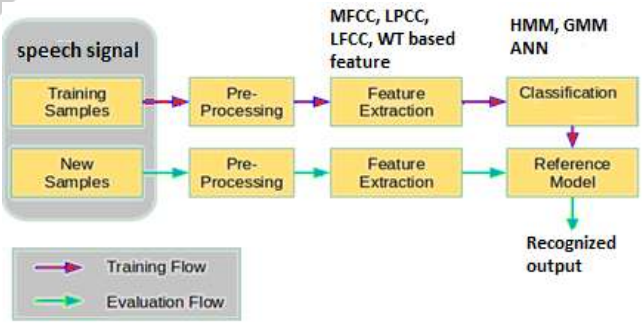

Fig 1: Block diagram speech recognition.

Speech recognition process consists of two main modules; one is feature extraction and other feature matching. The process that convert voice signal to feature vector is done by signal-processing. Fig.1 shows input training samples are given to pre – processing block and

output of it is noise free speech samples. These samples are used by feature extraction block and output of it is feature vector. The main purpose of feature extraction module is to convert speech waveform to some type of representation. Which is used for further analysis and processing, this extracted information is known as feature vector. Following are the few methods for implementing for extracting feature factor.

- MFCC (Mel-Frequency Cepstrum Coefficient)
- LPC (Linear Predictive Coding)

After obtaining the feature vector we build the acoustic model. The acoustic model is used compare the unknown voice sample. As shown in block diagram, Output of feature extraction block is given as input to classifier for the formation of acoustic model. Different types of acoustic models are:

- VQ-Code
- GMM-Gaussian Mixture Model

Feature vector of unknown sample is compared with the training database and depending upon the likelihood word is recognized. Here we are using context independent phoneme HMM.

V. FEATURE EXTRACTION:

MFCC coefficients are obtained from non-linear mel scale ie Mel Frequency Cepstrum which is short term power spectrum of speech signal. It is calculated as cosine transform of log power spectrum on mel scale. In MFC, on the mel scale the frequencies are similarly spaced. The human auditory system's response is approximated more closely in mel scale than the linearly spaced frequency bands used in cepstrum case. The mel scale allow better representation similar to human auditory system.

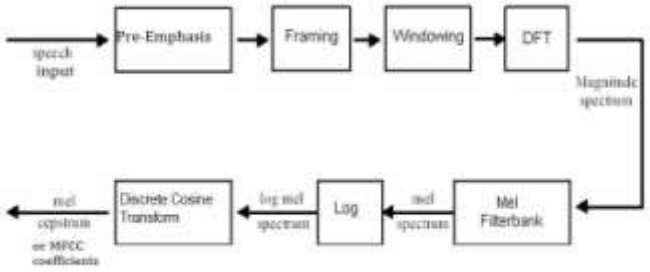

Fig 2:Block diagram of calculation of MFCC

Fig 2 shows training phase in which MFCC coefficients are extracted through the following steps.

1. Take FFT of windowed signal. Compute its squared magnitude. This gives power spectrum.
2. Pre-emphasis of spectrum is done to approximate unequal sensitivity of human perception at different frequencies.
3. Take integration of power spectrum within 50% overlapping critical band filter response. Triangular overlapping windows called mel filters are used for integration.
4. Integration of log power spectrum is done which compress the spectral amplitudes.
5. Take DCT, which will give cepstral coefficients.
6. Out of 20 or more coefficients typically 12 to 14 coefficients are used.

VI. FEATURE SALIENCY HIDDEN MARKOV MODEL:

ConsideraHMMwithcontinuousemissionsand*I*states.Let,

$y = \{y_{0,}y_{1,} \dots \dots \dots \dots y_T\}$-Sequenceofobserveddata,

where each $y_t \in R$

y_{lt}= observation for *l*-th feature at time *t* .

Let, unobserved states sequence

$$x = \{x_0, x_1, \dots \dots x_T\}$$

The transition matrix of the Markov chain associated with this sequence is denoted as A. The components of this transition matrix are denoted by

$$a_{ij} = p(x_t = j | x_{t-1} = i), \quad (1)$$

π = the initial state distribution.

In termsof these quantities, the complete data likelihood can be written as:

$$p(x, y, |\Delta) = \pi x_0 f_{x_0}(y_0) \prod a_{x_{t-1}, x_t} f_{x_t}(y_t) \quad (2)$$

Where Δ = set of model parameters

$f_{xt}(y_t)$ = emission distribution given state x_t.

For feature selection, feature saliency model for emission distributions is used [11]. If feature's distribution dependant on the underlying states then it is considered s relevant feature. Feature is irrelevant if its distribution is independent of the state. Let ,

$$Z = \{z_{0,}z_{1,} \dots \dots \dots \dots z_L\}$$

be the set of binary variables indicating the relevancy of each feature.

If z_l =1, *l*-th feature is relevant. Otherwise, if zl = 0, *l*-th feature is irrelevant.

The feature saliency ∂_l is defined as the probability that the *l*-th feature is relevant. Assuming the features are conditionally independent given the state allows the conditional distribution of *yt* given **z** and **x** to be written as:

$$p(y_t | z, x_t = i, \Delta) = \prod r(y_{lt} | u_{il}, \sigma_{il}^2)^{z_l} q(y_{lt} | \varepsilon_l, \tau_l^2)^{1-z_l} \quad (3)$$

Where $r(y_{lt} | \mu_{il}, \sigma_{il}^2)$ is Gaussian conditional feature distribution for *l*-th feature with state dependant mean μ_{il} and state dependant varience σ_{il}^2, and $q(y_{lt} | \varepsilon_l, \tau_l^2)$ is the state independent Gaussian feature distribution with mean ε_l and variance τ_l^2. For FSHMM the set of model parameters Δ is $\{\pi, A, \mu, \sigma, \tau, \varepsilon, \rho\}$.

The marginal Probability of z is:

$$P(z|\Delta) = \prod_{l-1}^{L} \rho l^{zl} (1 - \rho l)^{1-zl} \quad (4)$$

The joint probability of y_t and z given x is:

$$p(y_t, z | x_t = i, \Delta) = \prod\prod_{l-1}^{L} \rho_l r(y_{lt} | u_{il}, \sigma_{il}^2)^{zl} [(1 - \rho_l) q(y_{lt} | \varepsilon_l, \tau_l^2)^{1-zl}] \quad (5)$$

The marginal distribution for y given x can be found by summing (4) over z and is:

$$f_{xt}(y_t) = p(y_t, z / x_t = i, \Delta) = \prod\prod_{l-1}^{L} \{\rho_l r(y_{lt} / u_{il}, \sigma_{il}^2)^{zl} + [(1 - \rho_l) q(y_{lt} | \varepsilon_l, \tau_l^2)] \quad (6)$$

The complete data likelihood for the FSHMM is:

$$p(x, y, z|\Delta) = \pi x_0\, p(y_0, z \mid x_{0,}\Delta) \prod_{t-1}^{T} a_{x_{t-1},x_t}\, p(y_{t,}z|x_{t,}\Delta)(7)$$

VII. EM algorithm for HMM:

TheEMalgorithm,referredtoastheBaum-Welchalgorithm when used with HMMs ([3], [22]), is used to calculatemaximum likelihood (ML) estimates for the model parameters. The Baum-Welch algorithm has two steps: expectation step and maximization step, also abbreviated as E-step and M-step. The E-step findsthe expectedvalueofthecompletelog-likelihoodwithrespectto the state, given the data and the current model parameters. The M-step maximizes the expectation computed in the E-step to find next state model parameters. The Q function which is expectation of complete log likelihood given by:

$$Q(\Delta, \Delta') = \in [\log p(x, y|\Delta) \mid y, \Delta'](8)$$

In (8), Δ= set of model parameters for current Iteration

Δ′ = set of model parameters from preveous Iteration

Probabilities for E-step are calculated by using forward and backward algorithms. These probabilities are used in the M-step. These two steps are repeated until convergence expectation.

VIII. DATABASE

The recording is done for 15 male speakers at sampling rate of 8 KHz. The time given for recording speech samples is two seconds, because it was found that two seconds are enough for recording isolated words. If the time given for recording was more than two seconds that would result in having so much silence time in the recorded speech sample or the word's utterance. Recorded speech files were in .wave file. There are few isolated words as 'APPLE', 'LIME', 'PINEAPPLE', 'PEACH', 'ORANGE', 'KIWI', 'BANANA' spoken by different speakers. Hence there are total 15 samples in database. This database is used for training and for testing purpose. Hence there are total 105 samples of isolated words. The collected speech samples are then going to pass through the features extraction, features training and features testing stages.

IX.RESULT

Both training and testing sample features are classified by HMM classifier. There are 7 words. According to classes respective wave files were generated by HMM classifier. In this way, speech recognition system is made more robust and efficient and hence the performance of speech recognition system is improved. The result is satisfactory for isolated English words. Table I shows performance parameter of speech recognition.

$$\text{Recognition rate} = \frac{\text{successfully detected words}}{\text{Number of words in test set}}$$

Table I. Performance parameter

TP	88.57%
FP	11.43%
Accuracy	89.99 %

X. CONCLUSION:

The project work has been carried out for isolated word speech recognition in Hindi language. The input speech was sampled at 8 KHz and then processed with a Hamming window to obtain the feature vectors. Here at Training Phase, feature extraction methods standard MFCC is used. At Testing Phase, HMM classifier was used. The experiment was performed on a set of speech data consisting of words of English language recorded by male and speakers. MFCC and HMM have given us performance parameters of recognition accuracy.

REFERENCES:

[1] Pruthi, Tarun, Sameer Saksena, and Pradip K. Das. "Swaranjali: Isolated word recognition for Hindi language using VQ and HMM." International Conference on Multimedia Processing and Systems (ICMPS), IIT Madras. 2000.

[2] Kumar, Kuldeep, R. K. Aggarwal, and Ankita Jain. "A Hindi speech recognition system for connected words using HTK." International Journal of Computational Systems Engineering 1.1 (2012): 25-32.

[3] Sinha, S, Agrawal, S. S. and Jain, A. 2013. "Continuous density Hidden Morkov Model for context dependent Hindi speech recognition", Int. Conference on Advances in Computing, Communication and Informatics (ICACCI), pp. 1953-1958, IEEE.

[4] Aggarwal, R. K., and M. Dave. "Using Gaussian mixtures for Hindi speech recognition system." International Journal of Signal Processing, Image Processing and Pattern Recognition 4.4 (2011): 157-170.

[5] Gaikwad, Santosh K., Bharti W. Gawali, and Pravin Yannawar. "A review on speech recognition technique." International Journal of Computer Applications 10.3 (2010): 16-24.

[6] Rabiner, Lawrence R. "A tutorial on hidden Markov models and selected applications in speech recognition." Proceedings of the IEEE 77.2 (1989): 257-286.

[7] Saksamudre, Suman K., and R. R. Deshmukh. "Isolated Word Recognition System for Hindi Language." (2015).

[8] Ankit Kumar, Mohit Dua, Tripti Choudhary, "Continuous Hindi Speech Recognition Using Monophone based Acoustic Modeling", International Journal of Computer Applications 2014.

[9] Pratik K. Kurzekar, Ratnadeep R. Deshmukh , Vishal B. Waghmare, "A Comparative Study of Feature Extraction Techniques for Speech Recognition System",International Journal of

Innovative Research in Science,Engineering and Technology,Dec.2014.

[10] V.Vaidhehi1, Anusha J, Anand P, "Automatic Speech Recognition using Different techniques", International Journal of Science and Research, India Online ISSN: 2319-7064,May 2013

[11] Saksamudre, Suman K., and R. R. Deshmukh. "Comparative Study of Isolated Word Recognition System for Hindi Language." International Journal of Engineering Research and Technology. Vol. 4. No. 07, July-2015. ESRSA Publications, 2015.

[12] Thakur, Abhishek, and Rajesh Kumar. "Automatic Speech Recognition System for Hindi Utterances with Regional Indian Accents: A Review 1." (2013).

[13] Saini, Preeti, Parneet Kaur, and Mohit Dua. "Hindi Automatic Speech Recognition Using HTK." International Journal of Engineering Trends and Technology 4 (2013).

[14] Mishra, Neema, Urmila Shrawankar, and Vilas M. Thakare. "An Overview of Hindi Speech Recognition." arXiv: 1305.2847 (2013).

[15] Ms.Vrinda, Mr.ChanderShakhar 2013. "Speech Recognition System for English Language", International Journal of Advanced Research in Computer and Communication Engineering, vol.2, issue-1.

[16] Kumar, Mohit, Nitendra Rajput, and Ashish Verma. "A large-vocabulary continuous speech recognition system for Hindi." IBM journal of research and development 48.5.6 (2004): 703-715.

[17] Rabiner, Lawrence, and Biing-Hwang Juang. "Fundamental of Speech Recognition Prentice-hall International." (1993).

[18] J. A. Bilmes, ''A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models,'' Int. Comput. Sci. Inst., vol. 4, no. 510, p. 126, 1998.

[19] M. H. C. Law, M. A. T. Figueiredo, and A. K. Jain, ''Simultaneous feature selection and clustering using mixture models,'' IEEE Trans. Pattern Anal. Mach. Intell, vol. 26, no. 9, pp. 1154–1166, Sep. 2004.

[20] C. A. McGrory and D. M. Titterington, ''Variational Bayesian analysis for hidden Markov models,'' Austral. New Zealand J. Statist., vol. 51, no. 2, pp. 227–244, Jun. 2009.

[21] STEPHEN ADAMS, PETER A, BELING, RANDY COGILL "Feature selection for Hidden markov Models and Hidden Semi- Markov Models" IEEE Access, Volume 4,may 9,2016.

[22] Ankit Kuamr, Mohit Dua, Tripti Choudhary, " Continuous Hindi speech Recognition using Gaussian Mixture HMM" IEEE students conference on Electrical, electronics and Computer Science, volume 6, June 2014.

[23] Ashok shigli, Ibrahim patel, Dr. k. Shrinivas Rao "A spectral feature process for speech recognition using HMM with MFCC approach" National conference on computing and communication systems (NCCCS), Volume No.8, May 2012.