A Review Paper on MFCC based Hindi Speech Recognition System using HTK Toolkit

Miss. Nikita A. Dhanvijay[1] Prof. P.R.Badadapure[2]
[1]M.Tech Student [2]HOD
[1,2]Department of Electronic & Tele-Communication Engineering
[1,2]JSPM's ICOER Wagholi, Pune

Abstract— To implement the robot's capabilities, it is important for us to communicate with them efficiently. Thus, Human Robot interchange is attracting the observance of the researchers these days. A speech recognition technique has been developed using various characteristic extraction method like MFCC (mel frequency cepestral coefficient), LPC (linear predictive coding) and HMM (hidden markov model) is used as the classifier. Less work has been finished for Hindi language in this domain with a vocabulary size not very large. So, work finished for Hindi database, with a vocabulary size a bit protracted. HMM has been enforced using HTK Toolkit. Afterwards the presentation of both of the method used have been compared. The work has been finished using audacity for sound recordings and Cygwin to implement the HTK commands in Linux kind environment in windows platform. As well as, the tech improvement has been checked in the speaker dependent and speaker independent both kind of environments, whose present results, as well as, the comparison graph of the technique shows that MFCC done well as compared to LPC in each and every condition.

Key words: HTK Toolkit, MFCC

I. INTRODUCTION

To modify the robot capable to follow the commands through voice, Speech Recognition technique is emerging as a big need in the region of Robotics. As voice is the cut price and easily available biometric tool. Till now, minimum work has been finished in the region of Speech recognition with Hindi Language. Thus, Hindi vocabulary with rather maximum database having 35 words data has been used, so that the robot could follow the voice commands given in Hindi language word. There are many ways through which interchange is possible, whether the interaction is human-human interchange or it is human-robot interchange. Even if we analyze a human- human interchange, we can see that there are various ways (verbal or non-verbal) in which humans can interact with each other. So for human-robot interchange also, there are many ways for human to interact with the Robot, which change from visual to touch to voice.

A. Previous Work

Previously a speech recognition technique with a vocabulary size of 30 words had been performed by K. Kumar et al tried to execute the technique much more efficient as they maximize the vocabulary up to thirty Hindi language words. The technique also shows good presentation for speaker independent environments. He used Hidden Markov Toolkit to make the technique and trained the technique for 30 Hindi language words with the data collected from eight speakers. The technique overall presentation was 94.63%. Before that Tarun Pruthi et al. in 2000, described a Hindi language speech recognizer, for the Hindi digits data for 2 male speakers. He used LPC for property dehydration and HMM for recognition purposes. The technique was giving a good presentation but as it was speaker specific its presentation needed to be enhanced up to some extent. The vocabulary of only 10 digits data was quite minimize. Y. Lee et al. Tried to compare various speech property in terms of the class-relevant and class-irrelevant data based on Shannon Algorithm. The property compared by him where Mel scaled FFT, Cepstrum, mel-cepstrum, and wavelet property. The experiment was finished on the TIMIT corpus and the best property of speech to uniquely find it was given as Mel-Scaled FFT. M.A. Anusuya et al. improved a SR technique for Kannada speech recognition.

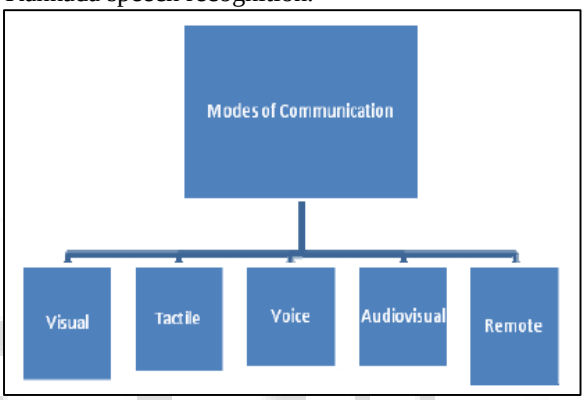

Fig. 1: Modes of Communication

II. LITERATURE SURVEY

This paper describes a technique to select a suitable property for speech recognition using data theoretic measure. Conventional speech recognition method heuristically choose a portion of frequency components data, cepstrum, mel-cepstrum, energy, and their time variation of speech waveforms as their speech property. However, these method never have good presentation if the selected property are not suitable for speech recognition. Since the recognition rate is the only presentation measure of speech recognition method, it is hard to judge how suitable the selected property is. To solve problem, it is important to analyze the property itself, and measure how well the property itself is. Method, we compare the Mel scaled FFT, cepstrum, mel-cepstrum, and wavelet property of the TIMIT speech data. The result shows that, among these property, the Mel-scaled FFT is the best property for speech recognition based on the measure. [1]

Pre-processing of speech signals is identify a crucial step in the development of a robust and efficient speech or speaker realization method. This paper deals with various speech processing method and the realization accuracy with respect to wavelet transforms method. It is shown that by applying wavelet transform method to the conventional technique the signal realization accuracy will

be maximized by using DWT and the wavelet packets for Clear and noisy speech signals data respectively. [4]

This paper shows the execution of *Swaranjali*, an experimental, speaker-dependent, real-time, isolated word recognizer for Hindi language. The results find with Swaranjali for tests conducted on a vocabulary of Hindi language digits for 2 male speakers are proposed in the end. The rest of the paper discusses the implementation of the method. The scheme present uses a standard execution, with some modifications to the noise elimination algorithm and the HMM training algorithm. [5]

The Speech is prominent mode of Communication among of human being. The communication among human computer interchange is called human computer interface. Speech has potential of being most important mode of interchange with computer .This paper find an overview of major method perspective of the fundamental progress of speech realization and also gives overview method developed in each stage of speech realization. This paper helps in choosing the system along with their relative merits & demerits. A comparative study of various method is finished as per stages. This paper is concludes with the decision on property direction for developing system in human computer interface technique using Marathi Language. [10]

III. PROPOSED WORK

A speech realization system has been developed is shown in Fig. 2, for which following steps are implemented:

1) Firstly a data collection with a vocabulary size of 35 Hindi words, have been processed. The voice samples to process the collection are taken from 2 male speakers and 3 female speakers. Audacity is used for Data collection processed. After getting a fully processed data collection, the collection is then used to train the method as well as to test it.
2) Finally, using the HTK Toolkit and the data collection processed, the speech realization method is improved by applying MFCC or LPC as a property dehydration method and HMM as a classifier. For parameterization presented from wave files to MFCC or LPC as well as to execute HMM efficiently HTK Toolkit has been used.

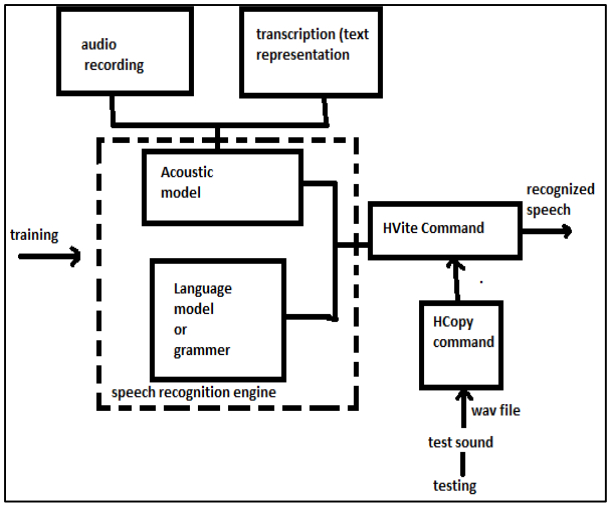

Fig. 2: The general framework of proposed model.

IV. SYSTEM SPECIFICATION

MFCC: The property used here is Mel- Frequency Cepstrum Coefficients for which we dehydrate the cepestral envelop in which the formants shown in Fig.3 shows the MFCC coefficients. The formants of a voice signal show the unique features of a voice, using which the speaker can be recognized. For this reason, the speaker recognition as well as speech realization method here uses the concept of formants to find the speakers, as well as realization the speech.

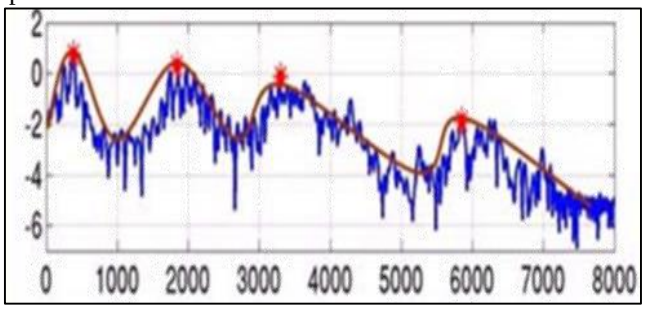

Fig. 3: Example of Spectral envelop

The smooth curve connecting the formants is the spectral envelope and our first task in the process of dehydrating MFCC coefficients is to dehydrate the spectral envelope.

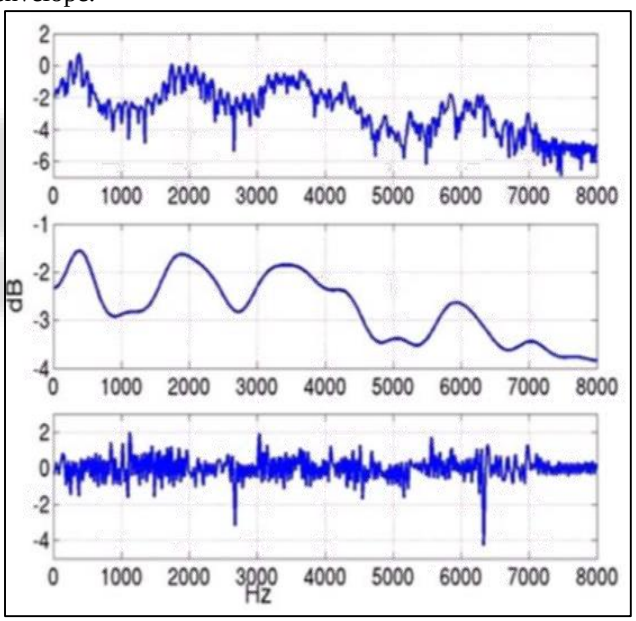

Fig. 4: Extraction of Spectral envelop from Spectral Details

To achieve this separation we first find the IFFT of the signal, then after IFFT of the log spectrum would find the signal in pseudo frequency axis. The technique of separating spectral envelope from original speech signal is shown in Fig. 4, 5 and 6.

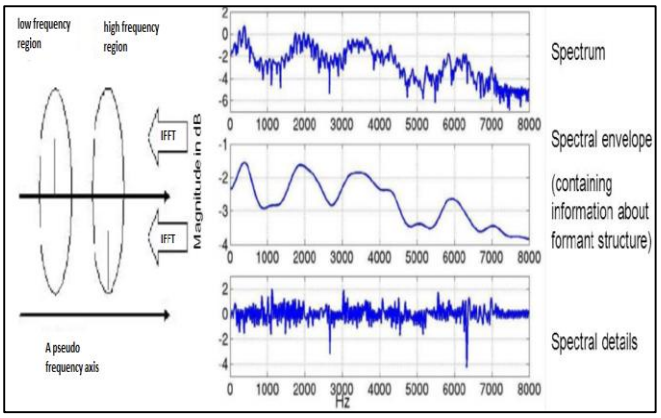

Fig. 5: Process representing separation of spectral envelop from spectral details.

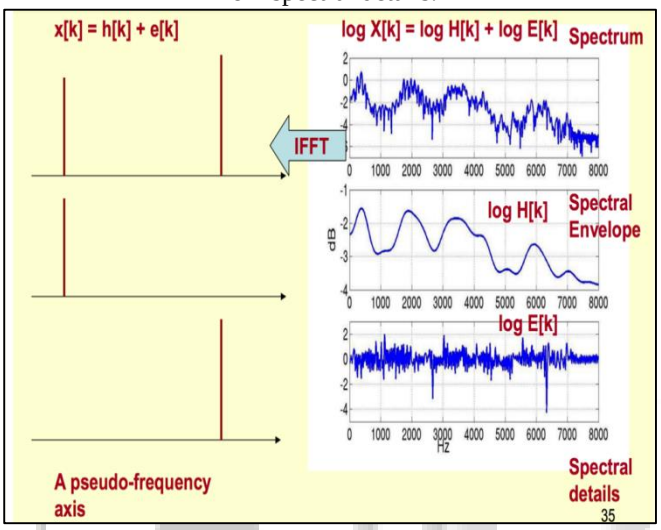

Fig. 6: Process of obtaining spectral envelop

A. LPC

As the name suggests Linear Predictive Coding find coefficients decreasing the difference between the actual speech samples of signal and the linearly predicted ones. It is a very reliable technique. Mostly Auto Regression technique is used for speech signal.

B. Hidden Markov Model

Numbers of terms are convergence while studying Hidden Markov Model. First of all, we should be aware of what Markov features is some of the systems exhibit the features in which the future states of the technique is dependent on the current state of the method. This features is known as Markov features and the technique which exhibit such a feature are called Markov method. When the Markov method with hidden states are converted into statistical Markov Model, it is known as Hidden Markov Model. Viterbi Algorithm is used for finding the most likely future states based on the current state of the method.

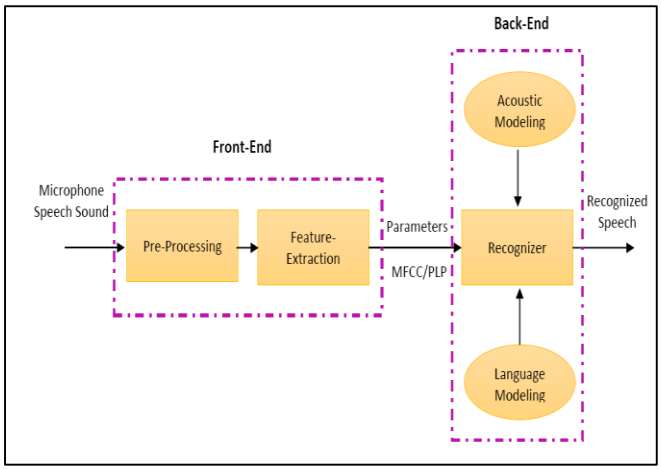

Fig. 7: Block diagram of system

V. CONCLUSION

Speech realization technique in all of the cases the accuracy is minimized with the maximized number of speakers in the train database. As well as the technique shows less accuracy as the degree of independence of the technique is maximized. The technique performs is not good for Speaker independent environment. LPC yield poor results in all the cases. So these drawbacks and limitations could be reduced. The technique could be made more accurate in the case of more number of speakers as well. The various new technologies could be merged to make the method more accurate with MFCC. The vocabulary in this thesis work is of 35 Hindi language words, it could also be extended to make the method more capable to realize more Hindi language words. As well as the method could be improved for other Indian Languages as well.

REFERENCES

[1] Y. Lee and K.-W. Hwang, "Selecting Good Speech Features for Recognition," ETRI, vol. 18, Apr. 1996
[2] http://www.ghacks.net/2009/04/02/web-browser-opera-face-gestures/.
[3] A. N. Kandpal and M. Rao, "Implementation of PCA and ICA for Voice Recognition and Seperation of Speech," in proc. of IEEE International Conference on Advanced Management Science (ICAMS), vol. 3, pp. 536-538, 2010.
[4] M. A. Anusuya and S. K. Katti, "Mel Frequency Discrete Wavelet Coefficients for Kannada Speech Recognition using PCA," in Proc. of Int. Conf. on Advances in Computer Science, 2010. [5] Tarun Pruthi, Sameer Saksena and Pradip K Das, "Swaranjali: Isolated word recognition for hindi language using VQ And HMM," in proc. Of International conference on multimedia processing and systems, Aug. 2000.
[5] D. Spiliotopoulos, I. Androutsopoulos, and C. D. Spyropoulos, "Human- Robot Interaction based on Spoken Natural Language Dialogue", in proc. Of the European workshop on service and humanoid robots.
[6] A. A M Abushariah, T. S. Gunawan, O. O. Khalifa, and M. A. M. Abushariah, "English Digits Speech Recognition System based on Hidden Markov Models," in Proc. Of IEEE International conference on computer and communication engineering, pp.1-5, may 2010.

[7] K. Kumar and R. K. Agarawal, "Hindi speech recognition using HTK," in International Journal of Computing and Business Research, vol. 2, May, 2011.
[8] S. Young, et al., the HTK Book. December, 1995.
[9] Gaikwad, S.K. and Gawali, B.A. A review on speech recognition technique. In International Journal of Computer Applications, volume 10, November, 2010.
[10] H.-J. Bohme, T. Wilhelm, and J. Key. An approach to multi-modal human-machine interaction for intelligent service robots. Robotics and Autonomous Systems, elsevier science, 44:83–96, December 2004.
[11] S. Mallat, "A wavelet tour of signal processing", Academic Press, 1998.
[12] Y.T.Chan "Wavelet Basics",Kulwer Academic Publications,©1995.
[13] J.S.Walker, "Wavelets and their Scientific Applications", Chamman and Hall/CRC, © 1999.
[14] Daubechies, "Ten lectures on wavelets," society for industrial and Applied mathematics, 1992.
[15] Nikhil Rao,"Speech compression using wavelets", ELEC 4801 THESIS PROGEC, School of Information Technology and Electrical Engineering, The university Of Queensland, October 2001.
[16] H.Hermansky,"Perceputal Linear Predictive (PLP) analysis of speech", J. Acoust. Soc. Am., 87(4):1738-1752, 1990.
[17] H. Hermansky, N. Morgan, "Rasta Processing of Speech", IEEE Trans. on Speech and Audio Proc., Vol.2, No.4, 1994.
[18] M.A.Anusuya and S.K.Katti, "Kannada speech recognition using Discrete Wavelet Transform-PCA", International conference on computer applications-2010, Dec.24-27,Pondicherry,India
[19] M.A.Anusuya and S.K.Katti, " Mel-frequency discrete wavelet coefficients for kannada speech recognition using PCA", International conference on Advances in computer science, Dec.21-22,2010, Trivandrum, Kerala, India.

