



# **PERFORMANCE COMPARISON OF ROBUST SPEECH RECOGNITION USING DIFFERENT FEATURE EXTRACTION TECHNIQUES**

**Divya Gupta; Poonam Bansal; Kavita Choudhary**

Jaagan Nath University, Jaipur

GGSIU University, Delhi

Jagan Nath University, Jaipur

[fromdivya81@gmail.com](mailto:fromdivya81@gmail.com); [Pbansal89@gmail.com](mailto:Pbansal89@gmail.com); [kavita@jagannathuniversity.org](mailto:kavita@jagannathuniversity.org)

*Abstract. The principal target of talk affirmation zone is to make techniques and structures for talk commitment to machine. Talk is the basic techniques for correspondence between individuals. For reasons going from inventive enthusiasm about the segments for mechanical affirmation of human talk abilities to longing to robotize fundamental errands which require human machine associations and research in modified talk affirmation by machines has pulled in a ton of thought for quite a while. In light of genuine advances in authentic exhibiting of talk, customized talk affirmation structures today find expansive application in assignments that require human machine interface, for instance, modified call taking care of in telephone frameworks, and request based information systems that give invigorated travel information, stock esteem references, atmosphere reports, Data section, voice correspondence, access to information: travel, keeping cash, Commands, Avoinics, Automobile passage, talk elucidation, Handicapped people (amaze people) general store, railroad reservations.*

*Keywords: Speech recognition, noise, robustness, distortion modeling, compensation, uncertainty processing, joint model training.*

## **1. INTRODUCTION**

Human dependably recognize speaker while they are conversing with each other .The speaker may exhibit in a similar place or in better places. Along these lines a visually impaired individual can recognize a speaker construct solely with respect to his/her vocal attributes. Creatures likewise utilize these attributes to recognize their natural one .The improvement of productive Speaker Identification framework has been a subject of dynamic research amid most recent two decades since they have a substantial number of potential applications in many fields that require precise client distinguishing proof, for example, shopping by phone, bank exchange, gets to control and voice message and so forth. Speaker acknowledgment is a non specific term utilized for two related issues: Speaker ID and confirmation. In the distinguishing proof assignment the objective is to perceive the obscure speaker from an arrangement of N known speakers. In confirmation, a character assert (e.g., a username) is given to the recognizer and the objective is to acknowledge or dismiss the given personality claim.

In this paper we focus on the ID undertaking. Speaker distinguishing proof can be additionally partitioned into two branches:"

1. Open-set speaker distinguishing proof (Speaker from outside the preparation set might be inspected).
2. Closed set speaker distinguishing proof (The speaker is constantly one of a shut set utilized for preparing).

Contingent upon the calculation utilized for the recognizable proof, this assignment can likewise be isolated into ward (The speaker must express one of a shut arrangement of words) and autonomous ID (The speaker may articulate any sort of words).

## 2. METHODOLOGY

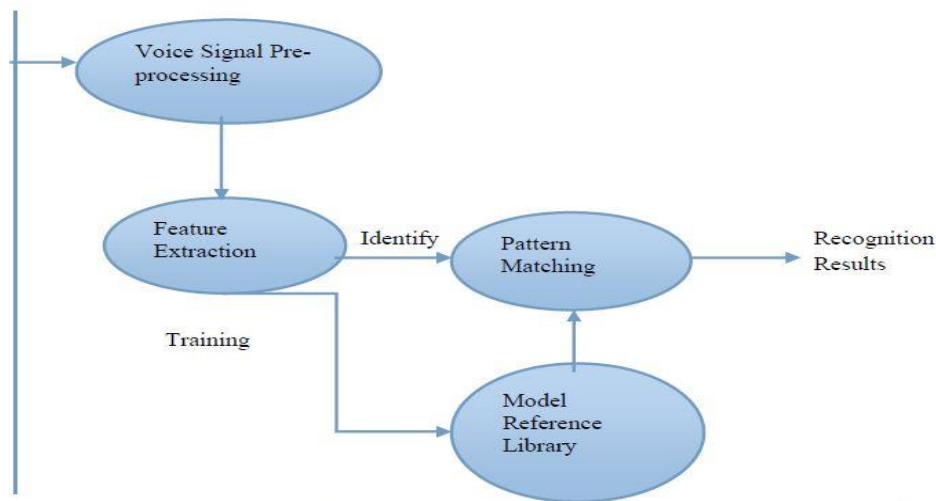


Figure 1: Basic of Speech Recognition System

### 2.1 Feature Extraction

In simple words, all the important characteristics of the input voice that define it are extracted and a voice template is generated. Feature extraction is mainly used to compute a feature vector sequence to represent the input signal.

There are three stages of feature extraction:

1. *Speech Analysis or Acoustic Front End*: In this stage, spectrotemporal extraction of the input signal are performed and original data specifying the envelop of short speech interval are generated.
2. *Compilation*: In this stage, a feature vector is combined having static and dynamic features.
3. *Transformation*: This stage, although not always present, deals with transformation of feature vector into more compact vectors supplied to the recognizer.

The most agreed upon features in these stages are:

1. A programmed framework ought to help in recognizing distinctive, however comparable sounding discourse.
2. Without the assistance of preparing information, a programmed making of acoustic models ought to be made for these sounds.
3. Invariant insights crosswise over speakers and talking situations ought to be appeared.

*Pattern Matching*: Computers are then used in the recognition process for matching. The stored voice template is matched with the readings of the input data provided.

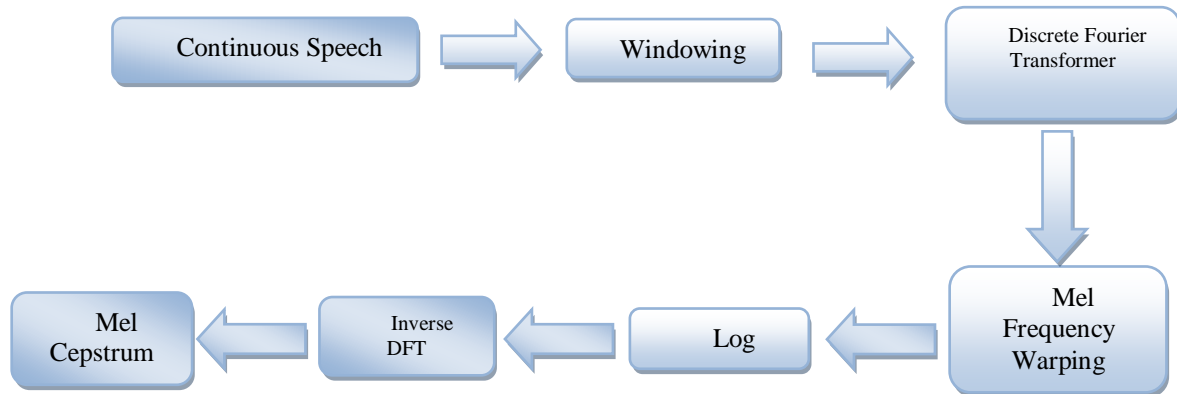
*Model Reference Library*: The reference library comprises of all the discourse flag layouts which are contrasted and the information motion amid Pattern Matching, to distinguish and perceive the discourse flag input. [6]

Speaker distinguishing proof is the way toward recognizing a speaker that implies who is talking. The procedure of speaker distinguishing proof incorporates two primary strides:

Feature extraction

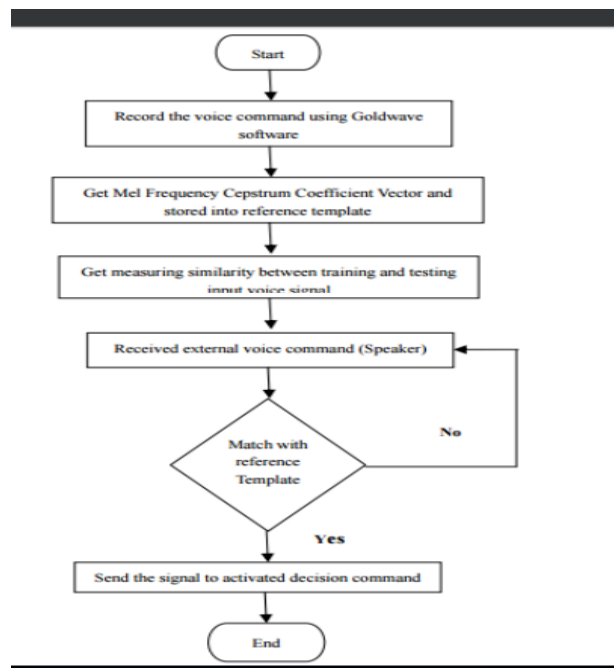
- Feature matching

Feature extraction techniques reduce the inputs vector’s dimensionality. It does so without affecting the power of signal. The features to be extracted from a signal are the pitch, the formant frequencies and the loudness of the speaker.



**Fig. 2.** MFCC Feature Extraction

The above figure speaks to the element extraction graph. From one side we input the consistent discourse signals for the procedure of windowing. During the time spent windowing the interruptions which are available toward the begin and in addition toward the finish of the casing are limited. After this procedure, the nonstop discourse flag is changed over into windowed outlines. These windowed casings are passed into the discrete Fourier transformer which changes over the windowed outlines into greatness range. Presently in the following stride, ghostly investigation is finished with a settled determination along a subjective recurrence scale that is the Mel-recurrence scale which delivers a Mel-range. This range is then passed to Log and afterward to reverse of discrete Fourier change which delivers the last outcome as Mel-Cepstrum. The Mel-Cepstrum comprises of the elements that are required for speaker ID



**Fig 3:** Cepstrum coefficient evaluation

**Table 1.** Comparison of feature extraction techniques

Technique	Features	Advantages	Disadvantages
<b>Linear Predictive Coding</b>	Compresses the given speech, efficient in identifying process	Good analysis techniques for extracting features and hence encoding the speech at low bit rate	Scaling down of outcome due to noise
<b>Cepstral Coefficients</b>	Fourier frequency transform based	Encoding at high bit rate	Inconsistent with human hearing due to linearly spaced filters
<b>Linear Predictive Cepstral Coefficients Coding</b>	straight prescient coefficient is spoken to in cepstrum space then the acquired coefficients are direct prescient cepstral coefficients	more robust and reliable than LPC.	Due to linearly spaced filters
<b>Mel-Frequency Cepstral Coefficients</b>	Uses Filter Bank coefficients	utilizes Mel recurrence scaling which is extremely surmised to the human sound-related framework.  more vigorous, dependable list of capabilities for discourse acknowledgment than the LPC coefficients	

**2.2 Vector Quantization**

VQ is a procedure of taking an expansive arrangement of highlight vectors and delivering a littler arrangement of measure vectors that speaks to the centroids of the dispersion.

The system of VQ comprises of separating few agent include vectors as a productive methods for describing the speaker particular elements. By methods for VQ, putting away each and every vector that we create from the preparation is incomprehensible. By utilizing these preparation information elements are grouped to frame a codebook for every speaker. In the acknowledgment organize, the information from the tried speaker is contrasted with the codebook of every speaker and measure the distinction. These distinctions are then use to settle on the acknowledgment choice.

**2.3Timit Database**

After performing MFCC on a single input wave, the next step is to add the signal waves from timit database. Our aim is to add 100 samples, 50 male and 50 female, from Timit database to our training and testing folder. MFCC and Pattern Matching will then be performed on all the Samples from the database.

**3. CONCLUSION**

Speech Recognition is one of the most useful procedure nowadays. Various feature extraction techniques defined are vital and by introducing each subsequent one the flaw of the preceding technique has been removed. MFCC feature extraction technique is the most efficient as it provides information about lower frequencies and noises rather than the higher ones. With the help of mel spaced filterbank it helps in realtime analysis. It

removes all inconsistencies involved due to noise or unregular filters involved in LPC or LPCC. Timit database acts a storage backend for the training and testing speech data and hence in recognition process.

## REFERENCES

- [1] M.A.Anusuya and S.K.Katti, "Speech Recognition by Machine: A Review", (IJCSIS) International Journal of Computer Science and Information Security, vol. 6, no. 3, pp.181-205, 2009.
- [2] Mohit Dua, R.K.Aggarwal, Virender Kadyan and Shelza Dua, "Punjabi Automatic Speech Recognition Using HTK", IJCSI International Journal of Computer Science Issues, vol. 9, issue 4, no. 1, July 2012.
- [3] Rajesh Kumar Aggarwal and M. Dave, "Acoustic modeling problem for automatic speech recognition system: advances and refinements Part (Part II)", Int J Speech Technol, pp. 309– 320, 2011.
- [4] Kuldeep Kumar, Ankita Jain and R.K. Aggarwal, "A Hindi speech recognition system for connected words using HTK", Int. J. Computational Systems Engineering, vol. 1, no. 1, pp. 25-32, 2012.
- [5] Kuldeep Kumar R. K. Aggarwal, "Hindi speech recognition system using HTK", International Journal of Computing and Business Research, vol. 2, issue 2, May 2011.
- [6] R.K. Aggarwal and M. Dave, "Performance evaluation of sequentially combined heterogeneous feature streams for Hindi speech recognition system", 01 September 2011.
- [7] Anusuya, M. A., & Katti, S. K.. Front end analysis of speech recognition: A review. International Journal of Speech Technology, Springer, vol.14, pp. 99–145, 2011.
- [8] Jacob Benesty, M. Mohan Sondhi, and Yiteng Huang, Handbook of Speech Processing, Springer, 2008.
- [9] Wiqas Ghai and Navdeep Singh, "Literature Review on Automatic Speech Recognition", International Journal of Computer Applications vol. 41–no.8, pp. 42-50, March 2012.
- [10] R K Aggarwal and M. Dave, "Markov Modeling in Hindi Speech Recognition System: A Review", CSI Journal of Computing, vol. 1, no.1, pp. 38-47, 2012.
- [11] Dev, A. (2009) 'Effect of retroflex sounds on the recognition of hindi voiced and unvoiced stops', Journal of AI and Soc., Springer, vol. 23, pp. 603-612.
- [12] Paul A.K., Das D., Kamal M.M., 2009. Bangla Speech Recognition System Using LPC and ANN, Seventh International Conference on Advances in Pattern Recognition, IEEE Xplore, (Kolkata, Feb. 4-6 2009), 171–174.
- [13] Thiang, Suryo Wijoyo, 2011. Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot , Proc. of. Int.Conf. on Information and Electronics Engineering, IPCSIT vol.6, (IACSIT Press, Singapore).
- [14] Ooi Chia Ai, M. Hariharan,, Sazali Yaacob, Lim Sin Chee, 2012. Classification of speech dysfluencies with MFCC and LPCC features, Expert Systems with Applications, Vol.39 (2), 2157– 2165.
- [15] Engin Avci , Zuhtu Hakan Akpolat, 2006. , Speech recognition using a wavelet packet adaptive network based fuzzy inference system, Expert Systems with Applications, Volume 31, Issue 3, pp. 495–503.
- [16] Vimal Krishnan V.R, Babu Anto P, 2009. Features of Wavelet Packet Decomposition and Discrete Wavelet Transform for Malayalam Speech Recognition, International Journal of Recent Trends in Engineering, Vol. 1(2), 93-96.
- [17] Yang Jie, 2009. Noise robust speech recognition by combining speech enhancement in the wavelet domain and Lin-log RASTA, ISECS International Colloquium on Computing, Communication, Control, and Management, IEEE Xplore, (Aug. 8-9, 2009), Vol. 2, 415-418.
- [18] Shivesh Ranjan, 2010. Exploring the Discrete Wavelet Transform as a Tool for Hindi Speech Recognition, International Journal of Computer Theory and Engineering, Vol. 2, No. 4, 642-646.
- [19] Sonia Sunny, David Peter S., K. Poulouse Jacob. 2011, Wavelet Packet Decomposition and Artificial Neural Networks based Recognition of Spoken Digits, International journal of machine intelligence, Vol.3, issue 4, 318-321.
- [20] M.A.Anusuya, 2011. Comparison of Different Speech Feature Extraction Techniques with and without Wavelet Transform to Kannada Speech Recognition, International Journal of Computer Applications, Vol. 26, No.4, 19-24.
- [21] Picone J.W., 1993. Signal Modelling Technique in Speech Recognition, Proc. of the IEEE, Vol. 81, No.9, 1215-1247.
- [22] Rabiner L., Juang B. H., 1993. Fundamentals of Speech Recognition, Prentice-Hall, Englewood Cliffs, NJ.
- [23] Jeremy Bradbury, 2000. Linear Predictive Coding.
- [24] S. Mallat, 1999. A wavelet Tour of Signal Processing, Academic Press, San Diego.
- [25] K. P Soman, K.I Ramachandran, N.G Resmi, 2010. Insight into Wavelets From Theory to Practice, PHI Learning Private Ltd, New Delhi.
- [26] Elif Derya Ubeyil, 2009. Combined Neural Network Model Employing Wavelet Coefficients for ECG Signals Classification, Digital signal Processing, Vol 19, 297-308.

- [27] S. Chan Woo, C.Peng Lin, R. Osman, 2001. Development of a Speaker Recognition System using Wavelets and Artificial Neural networks, Proc. of 2001 Int. Symposium on Intelligent Multimedia, Video and Speech processing, (Hong Kong, May 2-4, 2001), 413-416.
- [28] S. Kadambe, P. Srinivasan, 1994. Application of Adaptive Wavelets for Speech, Optical Engineering , Vol 33(7), 2204- 2211.
- [29] S .G. Mallat 1989. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation, IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol.11, 674-693.
- [30] [http://en.wikipedia.org/wiki/Discrete\\_wavelet\\_transform](http://en.wikipedia.org/wiki/Discrete_wavelet_transform)
- [31] Fecit Science and Technology Production Research Center, 2003. Wavelet Analysis and Application by MATLAB6.5 [M], Electronics Industrial Press, Beijing.