

IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

VOLUME : 1 ISSUE : 8 Print / Issue Publication Date: 09-Oct-2016

ISSN : 2455-2143

Indexed In

Google SCHOLAR OPEN ACCESS doi Crossref

WWW.IJEAST.COM

editor@ijeast.com

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 8, ISSN No. 2455-2143, Pages 108-112
Published Online June - July 2016 in IJEAST (http://www.ijeast.com)

NEURAL NETWORKS FOR HINDI SPEECH RECOGNITION

Poonam Sharma
Department of CSE & IT
The NorthCap University,
Gurgaon, Haryana, India

Anjali Garg
Department of CSE
The NorthCap University,
Gurgaon, Haryana, India

***Abstract*— Automatic Speech Recognition System has been a challenging and interesting area of research in last decades. But very few researchers have worked on Hindi and other Indian languages. In this paper a detailed study of using various neural networks for Hindi speech recognition with their detailed comparison is shown. In the first phase various MFCC, LPC and PLP features are calculated. In the second phase these features are fed to various neural networks to see their performance. Results show that the probabilistic neural networks give better performance as compared to the other methods.**

Keywords— Mel Frequency Cepstral Coefficients (MFCC); Linear Predictive Code (LPC); Perceptual Linear Prediction (PLP); Probabilistic Neural Network (PNN)

I. INTRODUCTION

Automatic Speech Recognition (ASR) has gained significant progress in technology as well as in application. There exist vast performance gap between human speech recognition (HSR) and ASR which has restrained its full acceptance in real life situation. Over more than 50 years of research and advancement, speech recognition has gained huge success. But still performance is the major bottleneck for its practicality especially when it comes to Hindi language.. As a lot of research experiments and results are achieved in English language throughout the world but a limited success is achieved for Hindi Speech recognition. Moreover, Hindi is fourth-most spoken language in the whole world Therefore; there is huge scope to develop such systems in Hindi language.

Features are extracted from input speech sample and in addition to training vector, it is sent for training using Neural Networks in supervised learning. Outputs are adjusted in accordance with targets. In modern ASR system, researchers use combination of basic technique in order to enhance recognition rate. The recognition rate is determined in terms of accuracy. In this work, the database of 150 samples (75 by male speaker, 75 by female speaker) is created and features are extracted using a combination of three feature extraction techniques, Mel frequency Cepstral Coefficient, Predictive Linear Coding and Perceptual Linear Prediction (MFCC-LPC-PLP). Neural Network is trained by these samples and samples are tested against various neural networks.

In Fig 1 the basic architecture of probabilistic network is shown. As speech model can also be compared to any Bayesian network because given the language model and acoustic model we have to find the probability of a particular word being spoken, therefore probabilistic neural networks are well suited for speech recognition.

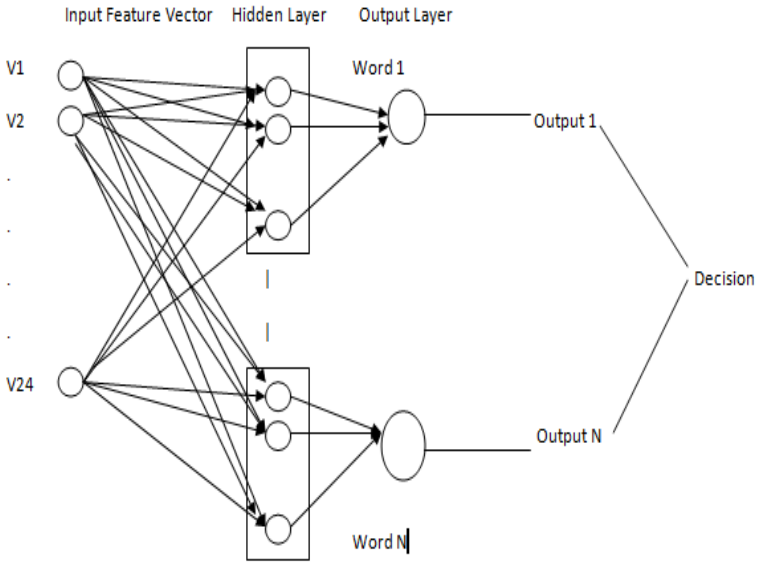


Fig 1. Architecture of Probabilistic Neural Networks

Back propagation networks work on the philosophy that It is better to learn from the errors. In this network the errors coming between output and target vectors are sent back again while training the network to increase the accuracy of the system. Basic architecture is shown in Fig 2.

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 8, ISSN No. 2455-2143, Pages 108-112
Published Online June - July 2016 in IJEAST (http://www.ijeast.com)

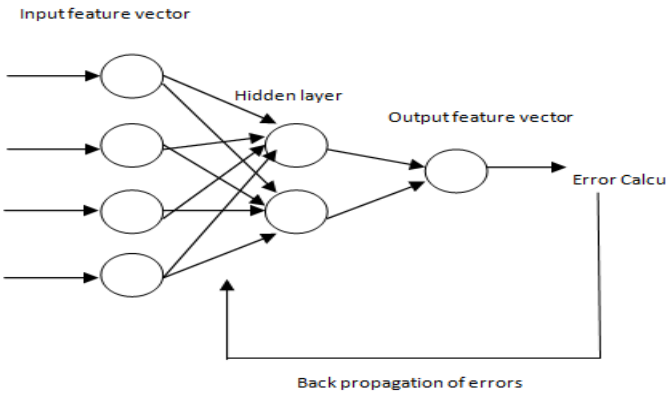


Fig 2. Architecture of Back Propagation Neural Networks

Linear Neural networks and perceptron are the simplest neural networks. These work better when the database is less and not much complexity is involved within the system. The basic architectures are shown in Fig 3 and Fig 4.

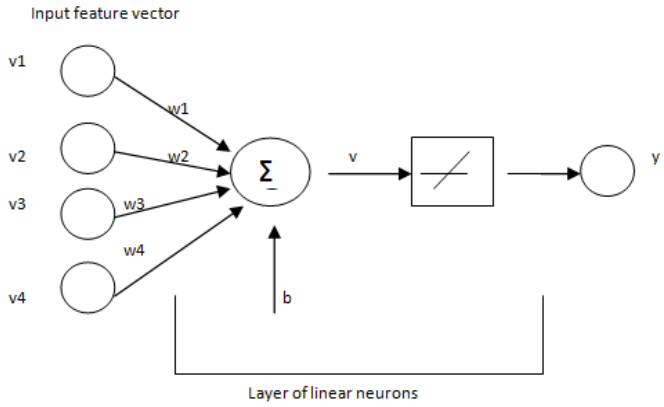


Fig 3. Architecture of Linear Neural Networks

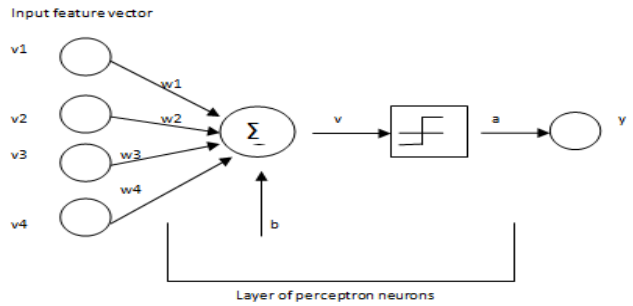


Fig 4. Architecture of Perceptron Neural Networks

II. PROPOSED METHOS

In this paper, the algorithm is designed for Hindi speech recognition based on the results obtained from different feature extraction techniques namely MFCC, PLP and LPC.

A. Algorithm for Recognition

The proposed algorithm is implemented on Matlab 2012a. The following steps are followed for recognizing speech:
Step 1: At input, speech signal p_i is given.

Step 2: Perform windowing using hamming window of 25ms and do Discrete Fourier Transformation.

$$w(n) = 0.54 - 0.46 \cos(2\pi n / N), \ 0 \le n \le N \quad (1)$$

$$X(k) = \sum_{n=0}^{N} x(n) e^{-j2\pi kn/N}, \ 0 \le k \le N\text{-}1 \quad (2)$$

Step 3: Compute features using Mel frequency cepstral coefficient (MFCC) and Mel frequency is given as below:

$$Mel(f) = 2595 \log_{10}(1 + f/700) \quad (3)$$

Step 4: Compute features using Linear Predictive Coding(LPC).

Step 5: Compute features using Perceptual Linear Prediction (PLP).

Step 6: The final input vector obtained after merging features from step 3 and step 4 is:

$$F = [M_L \ C_L \ P_L]$$

Step 7: Create final target vector for training.

Step 8: Import final input vector and target vector and create different neural networks. For probabilistic neural network we can use the output criteria :

$$O_i = \left[\frac{1}{(\sqrt{(2\pi\sigma^2}^P}\right]\left(\frac{1}{Q}\right)\sum_{q=1}^{Q} exp\left\{-\frac{||V-X^q||^2}{2\sigma^2}\right\} \quad (4)$$

Step 9: Train the network and simulate results.

Step 10: Test the selected sample against Neural Network.

Step 11: If word spoken correctly, then Speech Recognized and display CORRECT;
else, Speech Not Recognized and display INCORRECT.

Step 12: Plot performance plot.

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 8, ISSN No. 2455-2143, Pages 108-112
Published Online June - July 2016 in IJEAST (http://www.ijeast.com)

2.2 *Flowchart*

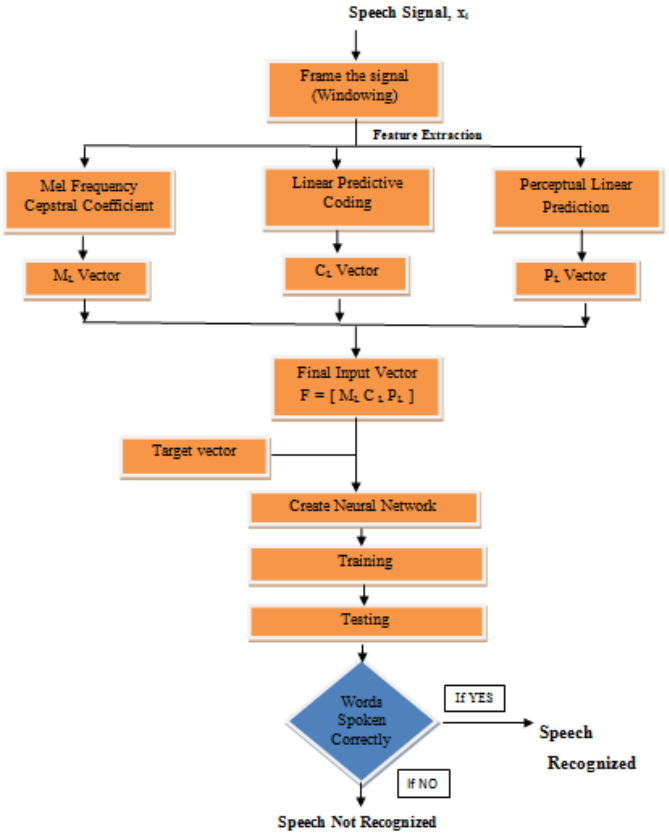


Fig 5: Flowchart

III. . EXPERIMENT AND RESULT

The below are the simulated results. In this paper, we have compared several Neural Networks and Feature Extraction techniques

3.1. *Output of Algorithm*

82.66% of average accuracy is achieved for male speaker and 80% is obtained for female speaker. So, at the output 81.33% of average accuracy is achieved.

Table 1: Output of algorithm

	Words Spoken	**Words Recognized Correctly**	**Average accuracy**
Male Speaker	75	62	82.66%
Female Speaker	75	60	80%
Output of Algorithm			81.33%

For male speaker, the output gives better results. The best performance is achieved at 0.03 at epoch 20. The Performance plot is shown below:

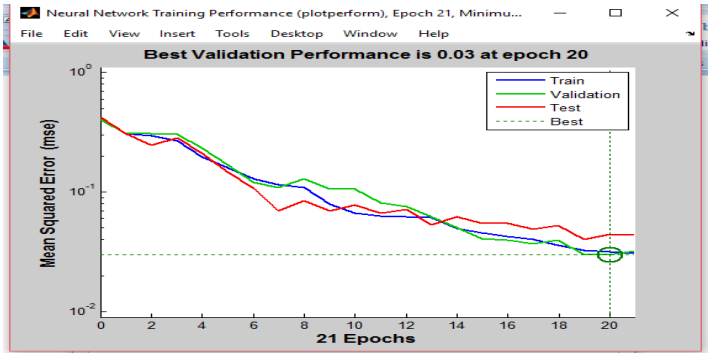


Fig 6. Probabilistic Neural Network Training Performance Plot of "male speaker"

3.2 Simulated Results of various Feature Extraction Techniques

1. *Results when Only Mel Frequency Cepstral Coefficient (MFCC):*

Table 2: Accuracy using only MFCC

	Words Spoken	**Words Recognized Correctly**	**Average accuracy**
Male Speaker	75	60	80%
Female Speaker	75	57	76%
Average Accuracy using only MFCC			78%

2. *Results of Only using Linear Predictive Coding (LPC):*

Table 3: Accuracy using only LPC

	Words Spoken	**Words Recognized Correctly**	**Average accuracy**
Male Speaker	75	51	68%
Female Speaker	75	53	70.66%
Average Accuracy using only LPC			69.33%

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 8, ISSN No. 2455-2143, Pages 108-112
Published Online June - July 2016 in IJEAST (http://www.ijeast.com)



3. Results of using *Only Perceptual Linear Predictive (PLP):*

Table 4: Accuracy using only PLP

	Words Spoken	**Words Recognized Correctly**	**Average accuracy**
Male Speaker	75	48	64%
Female Speaker	75	45	60%
Average Accuracy using only PLP			62%

4.Results when *Combination of MFCC, LPC and PLP are used.*

Table 5: Accuracy using MFCC-LPC-PLP

	Words Spoken	**Words Recognized Correctly**	**Average accuracy**
Male Speaker	75	62	82.66%
Female Speaker	75	60	80%
Average Accuracy using MFCC, LPC and PLP			81.33%

COMPARISON OF VARIOUS FEATURE EXTRACTION TECHNIQUES

Table 6: Comparison of feature extraction technique

SNo.	**Technique**	**Average Accuracy**
1.	Mel-Frequency Cepstral Coefficient (MFCC)	78%
2.	Linear Predictive Coding (LPC)	69.33%
3.	Perceptual Linear Predictive (PLP)	62%
4.	Mel-Frequency Cepstral Coefficient – Linear Predictive Coding - Perceptual Linear Predictive (MFCC-LPC-PLP)	81.33%

3.3 Simulated Results of various Neural Networks

1. *Probabilistic Neural Network*

Table 7: Probabilistic Neural Network

	Words Spoken	**Words Recognized Correctly**	**Average accuracy**
Male Speaker	75	62	82.66%
Female Speaker	75	60	80%
Probabilistic Neural Network			81.33%

2. *Feed Forward Back Propagation Network*

Table 8: Feed Forward Back Propagation Network

	Words Spoken	**Words Recognized Correctly**	**Average accuracy**
Male Speaker	75	59	78.66%
Female Speaker	75	60	80%
Feed Forward Back Propagation Network			79%

3. *Perceptron Neural Network*

Table 9: Perceptron Neural Network

	Words Spoken	**Words Recognized Correctly**	**Average accuracy**
Male Speaker	75	54	72%
Female Speaker	75	56	74.66%
Perceptron Neural Network			73.33%

4. *Linear Neural Network*

Table 10: Linear Neural Network

	Words Spoken	**Words Recognized Correctly**	**Average accuracy**
Male Speaker	75	53	70.66%

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 8, ISSN No. 2455-2143, Pages 108-112
Published Online June - July 2016 in IJEAST (http://www.ijeast.com)



Female Speaker	75	51	68%
Linear Neural Network			69.33%

COMPARISON OF DIFFERENT NEURAL NETWORKS

Table 11. Comparison of different Neural Networks

SNo.	Technique	Average Accuracy
1.	Probabilistic Neural Network	81.33%
2.	Feed-Forward BPN	79%
3.	Perceptron Neural Network	73.33%
4.	Linear Neural Network	69.33%

IV. CONCLUSION

In this paper we have compared various neural network techniques for Hindi Speech recognition and it was observed that probabilistic neural networks work better as compared to other state of the art networks. This work can be extended further by incorporating some other features for increasing the recognition accuracy. Also the comparison is done only for isolated words so it can be extended to continuous speech also. Same techniques can be applied to other Indian languages also for designing their recognition systems.

V. REFERENCES

[1] J. M. Baker ; L. Deng ; J. Glass S. Khudanpur ; C. Lee, N. Morgan,; D. O'Shaugnessy. (2009) : Research developments and directions in speech recognition and understanding, part 2, IEEE Signal Process. Mag., vol. **26**, no. 4, pp. 78–85.

[2] L.Dengand ; X. Li. (2013). Machine learning paradigms for speech recognition: An overview, IEEE Trans.Audio, Speech, Lang. Process., vol. **21**, no. 5, pp. 1060–1089.

[3] S.Sinha ; S.S. Aggarwal ; Aruna Jain. (2013). Continuous Density Hidden Markov Models for Context Dependent Hind Speech Recognition, ICACCI,pp. 1953-1958.

[4] A.Mohamed ; T.Sainath ; G.Dahl ; B.Ramabhadran ; G.Hinton ; M. Picheny. (2011). Deep belief networks using discriminative features for phone recognition, in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), pp. 5060–5063.

[5] T.Yoshioka ; M.J.F.Gales. (2014). Environmentally robust ASR for deep neural network acoustic models, Computer Speech and Language, pp. 65-86.

[6] O.A.Hamid ; A.R. Mohamed ; H.Jiang. (2014). Convolution Neural Networks for speech recognition. IEEE/ACM transactions on Audio Speech and Language Processing, Vol **22**, No 10, pp. 1533-1545.

[7] L.R.Rabiner ; B.H.Juang. (1986). An introduction to Hidden Markov Models, IEEE Signal Process. Mag., pp. 4–16.

[8] Sanghmitra V. Arora. (2013). Effect of time Derivatives of MFCC Features on HMM Based Speech Recognition System, ACEEE Internation Journal on Signal and Image Processing, Vol **4**, No 3, pp. 50-55..

IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE AND TECHNOLOGY

AUTHORS

Poonam Sharma

Anjali Garg

ABOUT IJEAST

International journal of Engineering Applied Science and Technology (IJEAST) is a peer-reviewed, open access journal that publishes high - quality research paper in the field of Engineering, Applied Science and Technology.

IJEAST aims to provide a platform for researchers, academicians, and professionals to share their innovative ideas, research findings, and practical experiences with the global scientific community.

JOURNAL METRICS

H-INDEX	i10-INDEX	IJEAST CITATIONS	PEER-REVIEWED & OPEN ACCESS
33	154	9,647 (IJEAST CITATIONS)	PUBLISHED MONTHLY

PEER REVIEWED

All submission are rigorously peer reviewed to ensure quality.

AUTHORS ARE OUR PRIORITY

We value our authors and recognize their contribution to the advancement of research and knowledge.

OPEN ACCESS

Free and unrestricted access to research for all.

GLOBAL REACH

Connecting researchers and professionals worldwide.

TIMELY PUBLICATION

We ensure a swift and efficient publication process.

For more information, visit our website
www.ijeast.com

editor@ijeast.com

www.ijeast.com

India

2455-2143