Continuous Bangla Speech and Speaker Detection By Deep Scaly Neural Network.

Yasin Ali Khan
Department of Computer Science and Engineering,
Chittagong University of Engineering and Technology,
Chittagong-4349, Bangladesh
shihabyasin@gmail.com

Mohammed Moshiul Hoque
Department of Computer Science and Engineering,
Chittagong University of Engineering and Technology,
Chittagong-4349, Bangladesh
Email:

Abstract*— Speech recognition is a complex cognitive task .To build interactive intelligent system in Bengali , we need Bangla Speech Recognition System with high accuracy .Being motivated by human ear construction , we develop a Deep Scaly Neural Network in this respect .We use Fuzzy Equivalence Relations for pattern classification .This technique allows ,construction of clusters in a natural adaptive way (contrasts with Fuzzy C Means).This biologically inspired recognition system is driven by well-formed mathematical formulation simulating human ear .Statistics shows that, this proposed architecture performs well (above 92% accuracy)in continuous Bangla speech recognition and speaker detection.

Keywords—Fuzzy Equivalence Relations,Scaly Neural Network, Bangla speech recognition.

I. INTRODUCTION

Automated speech recognition research is progressing continuously since 1930.Different technologies have been used in this respect like HMM[1],ANN-HMM[2][3],HMM-RNN[4] etc. Bangla is an international language spoken by around 8% of world population[5]. In this paper we've built a scaly neural network[6][7]for recognizing continuous Bangla speech and speaker. Krause and Hackbarth's[8]scaly architecture helps reducing communication overhead. We also tries to emulate human ear cochlear processing[9] in some degree using MFCC feature. Fuzzy Equivalence Relations are used for pre-processing feature clustering in a natural and adaptive way[10][11].

II. RELATED WORKS

Some preliminary work reported in literature for Bangla Speech Recognition based on phonemes[12],letters [5], words [13][14][15], small [16] or medium vocabulary speech system [17].Recent works : Rahman et al., 2003[18], S.A. Hossain, 2004[19], Abul et al., 2007[5]. Though some handful works we need continuous research for industry standard Bangla Recognition System.

• MOTIVATION

As technology demands robust Bangla speech recognition system different researchers relentlessly working in this arena. Motivation comes from previous many researchers work like K. Roy et. al [20], Wouter Gevaert et.al[21],
Akram M. Othman, and May H. Riadh[6], Md. Farukuzzaman Khan[22][15], Md. Saidur Rahman et. al[17], Abul Hasanat et.al[5], Nusrat Jahan Lisa et.al.[23], A H M. Rezaul Karim[24], K. J. Rahman et.al.[21], and many more.

• CHALLENGES

For different cognitive task different neural network architecture performs best and still there is no method to select the best one[25].So, we repeatedly searching best network topology in trial and error fashion[26].

• PROBLEMS

Taking speech signal in a perfect silent environment was tough and electronic noise inherent in microphone disturbs the process slightly.

III. SYSTEM ARCHITECTURE

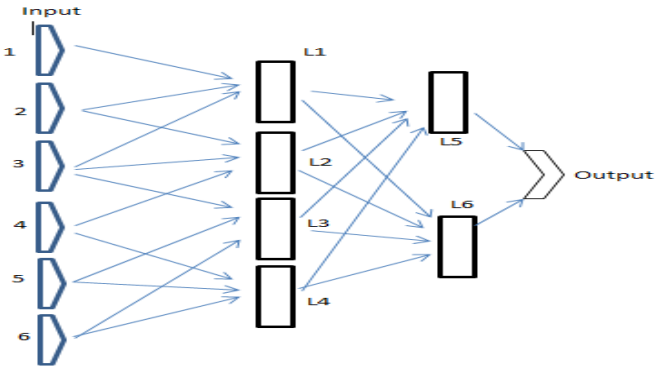

Fig : Scaly Neural Network for speaker and speech Detection.

In every hidden layer we use 50 neurons. 102 sized feature vector is used. Every input node takes 17 features. Overlapped of 34 frames in each hidden layer.

IV. NEURAL NETWORKS FOR SPEECH RECOGNITION

Speech recognition is inherently very complex dynamic task and we can say it a hyper-computation or super-Turing computation task[27].Neural Network, a machine learning tool that can perform some complex cognitive task like pattern recognition, memorizing, prediction etc. in a fairly good manner[28][29][30][31]. Researchers claims this tool will perform better in coming decades and industry standard

speech recognition system is building around this technology[32][33].

- *Reason For Scaly*

There is no good way for determining size of a neural network for a given task. Increasing neural or connections between them not necessarily increase performance[34]. So, we choose localized structure for good scaling a fully connected neural network.

- *Limitation*

Though reduced connectivity results savings in computational cost, slight decrease in robustness of the neural network may occur.

V. EXPERIMENTS

1. ENVIRONMENT:

Silent room and speaker with a microphone.

2. Acoustic Signal Capturing:

A close talk microphone is used.10 Bangla sentences have taken from 10 different speakers, each 3 times.

3. Sampling and Quantification:

We have used a sampling rate of 11025KHz .After that quantification is done.

4. Silence removal and End point Detection

We use a novel approach described in[35]. After that processed signals are stored as .wav files.

5. Normalization:

We use Matlab **mapminmax** function for [-1,+1] mapping.

6. Pre-emphasis:

We use a first-order high-pass filter by Matlab filter.

7. Windowing:

To extract invariance properties from captured speech we use function **enframe**[36]and Hamming windowing is used.

8. Feature Extraction (MFCC):

We use function **melcepst** (**Voicebox).**Some researchers reports for Bangla Speech Recognition MFCC39 is good[27].

9. Feature Clustering

Feature forms in a natural way in every human voice differently depending on their vocal cord frequency[37]. So, we use Fuzzy Equivalence Relations for feature clustering in a natural way[38].

10. Training :

70% samples for training chosen in random order.Standard Back-propagation algorithm is used. We train in batch mode.

11. Testing:

15% for validating and 15% for testing speaker dependent and similarly for speaker independent mode testing.

12. SENTENCES USED

1.	ঢাকা বাংলাদেশের রাজধানী
2.	আমি তোমাকে ভালবাসি
3.	পাখিটি আকাশে উড়ে
4.	তুমি কোথায় যাও
5.	তুমি কি করো

13. RESULTS

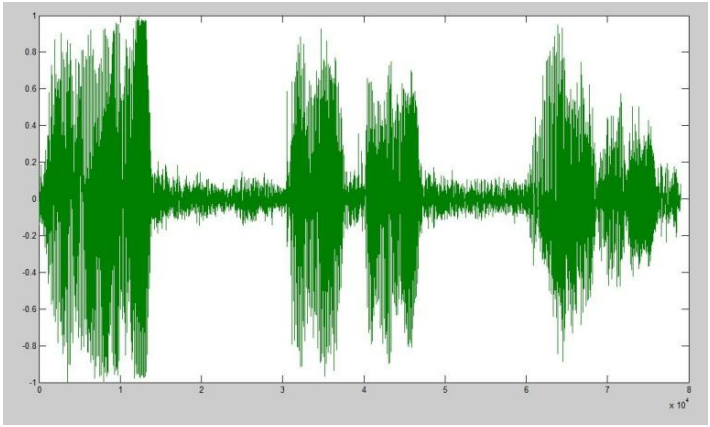

Fig: Sentence 1 Time Domain Representation

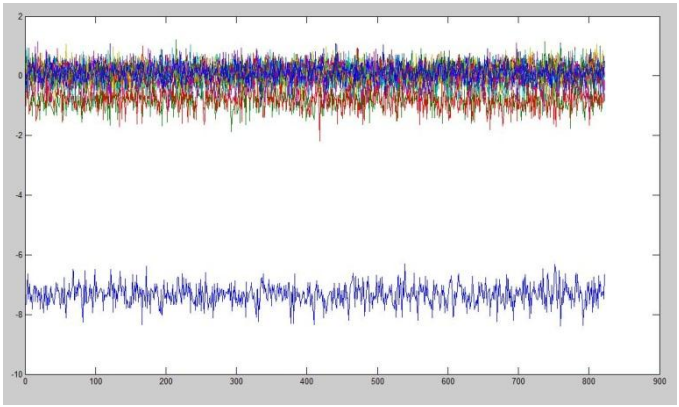

Fig : After MFCC Extracted

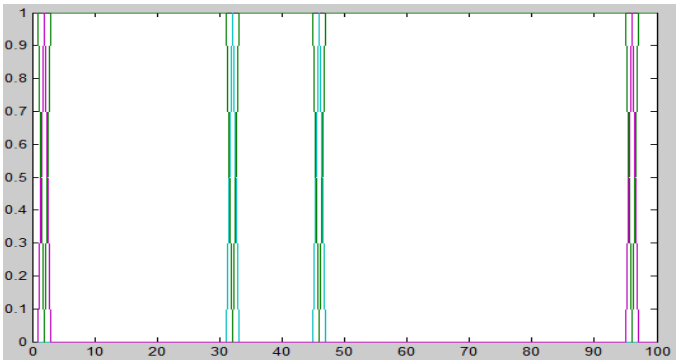

Fig : After Clustering By Recurrence Relations

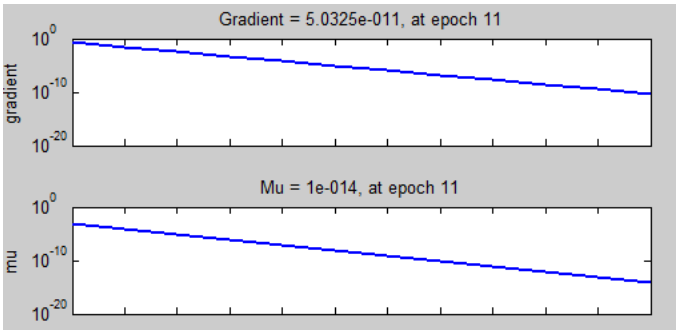

Fig : One of different stages of training status

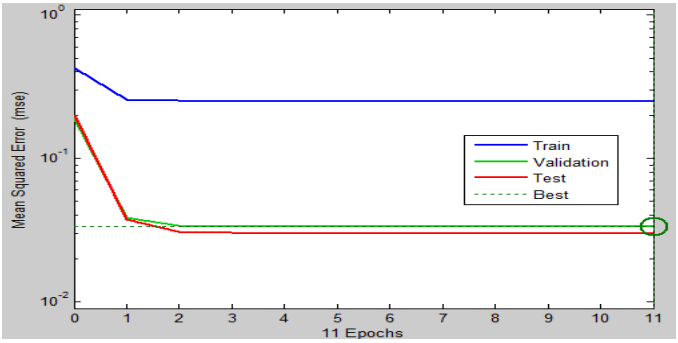

Fig : One of many performance evaluation graphs

Based on classification accuracy we see the following results:

Speaker Dependent Mode	Speaker Independent Mode
93.11%	91.55%

VI. CONCLUSION

Perfect silence is important for good recognition. Again using high quality noise-cancelling microphone is recommended. Considering all these technicalities we hope performance rate will increase slightly.

REFERENCES

[1]B. H. Juang and L. R. Rabiner, "Automatic Speech Recognition-A Brief History of the Technology", Elsevier Encyclopedia of Language and Linguistics, Second Edition, 2005.

[2] H.A. Bourlard and N. Morgan, Connnectionist SpeechRecognition: A Hybrid Approach, Kluwer Academic Publishers, 1994.

[3] Qifeng Zhu, Barry Chen, Nelson Morgan, and AndreasStolcke, "Tandem connectionist feature extraction forconversational speech recognition," in InternationalConference on Machine Learning for Multimodal Interaction,Berlin, Heidelberg, 2005, MLMI'04, pp. 223–231, Springer-Verlag.

[4] A. J. Robinson, "An Application of Recurrent Nets toPhone Probability Estimation," IEEE Transactions onNeural Networks, vol. 5, no. 2, pp. 298–305, 1994.

[5] Abul Hasanat, Md. Rezaul Karim, Md. ShahidurRahman and Md. Zafar Iqbal, "Recognition ofSpoken letters in Bangla", 5th ICCIT 2002, EastWest University, Dhaka, Bangladesh, 27-28 December 2002.

[6] Akram M. Othman, May H. Riadh, "Speech Recognition Using Scaly Neural Networks", World Academy of Science, Engineering and Technology 38 2008.

[7] Smith, A. D. "Isolated Word Recognition Using Gradient Back-Propagation Neural Networks", MEng. Dissertation, University Of Newcastle Upon Tyne, 1990.

[8] Harrison, T. D., and Fallside, F., ""A Connectionist Model For Phoneme Recognition In Continuous Speech", Proc. Of The IEEE Int. Conf. On Acoustics, Speech and Signal Processing, ICASSP '89, pp. 417- 420, Glasgow 1989.

[9] Krause, A., and Hackbarth, H., "Scaly Artificial Neural Networks For Speaker-Independent Recognition Of Isolated Words", Proc. Of The IEEE Int. Conf. On Acoustics, Speech and Signale Processing, ICASSP '89, pp.21-24, 1989.

[10] Klir, G. and Bo Yuan, Fuzzy Sets and Fuzzy Logic (1995) ISBN 978-0-13-101171-7.

[11] Demirci M., Recasens J., Fuzzy groups, fuzzy functions and fuzzy equivalence relations, Fuzzy Sets and Systems, 144 (2004), 441-458.

[12]S. M. Jahangir Alam, an M.Sc. Thesis on "SystemDevelopment for Bangla Phoneme Recognition",Dept. of Computer Science & Engineering, Islamic University, Kushtia-7003, July-2004.

[13] Kaushik Roy, Dipankar Das and M. Ganjer Ali,"Development of the Speech Recognition Systemusing Artificial Neural Network", 5th ICCIT 2002,East West University, Dhaka, Bangladesh, 27-28December 2002.

[14] Md. Farukuzzaman Khan, "Computer Recognitionof Bangla Speech", M.Phill. Thesis, Dept. ofComputer Science and Engineering, IslamicUniversity, Kushtia, September, 2002.

[15] Md. Farukuzzaman Khan, Md. Mijanur Rahman,Md. Mostafizur Rahman, "Development of BanglaVoice Command Driven DOS Utility System",Journal of Applied Science & Technology, IslamicUniversity, Kushtia-7003, Bangladesh, Vol. 03,No. 02, p93-98, December-2003.

[16] Md. Saidur Rahman, "Small Vocabulary SpeechRecognition in Bangla Language", M.Sc. Thesis,Dept. of Computer Science & Engineering,Islamic University, Kushtia-7003, July-2004.

[17] Md. Rabiul Huq, "A medium vocabulary speech totext system", M. Sc. Thesis, Dept. of ComputerScience & Engineering, Islamic University, Kushtia-7003, February-2005.

[18] K. J. Rahman, M. A. Hossain, D. Das, T. Islam, and M. G. Ali , "Continuous bangla speech recognition system" in Proc.6th International Conference on Computer and Information Technology (ICCIT03), Dhaka, Bangladesh, 2003.

[19] S. A. Hossain, M. L. Rahman, F. Ahmed, and M. Dewan, "Bangla speech synthesis, analysis, and recognition: an Overview" , in Proc. NCCPB, Dhaka, 2004.

[20] Md. Ali Hossain, Md. Mijanur Rahman, Uzzal Kumar Prodhan, Md. Farukuzzaman Khan, "Implementation Of Back-Propagation NeuralNetwork For Isolated Bangla Speech Recognition", International Journal of Information Sciences and Techniques (IJIST) Vol.3, No.4, July 2013.

[21] Wouter Gevaert, Georgi Tsenov, Valeri Mladenov," Neural Networks used for Speech Recognition",Journal of Automatic Control, University of Belgrade, VOL. 20:1-7, 2010.

[22]Md. Farukuzzaman Khan and Dr. Ramesh Chandra Debnath, "Comparative Study of Feature Extraction Methods for Bangla Phoneme Recognition", 5thICCIT 2002, East West University, Dhaka, Bangladesh, PP 2728, December 2002.

[23]Nusrat Jahan Lisa et.al, "Performance Evaluation of Bangla Word Recognition Using Different Acoustic Features", IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.9, September 2010.

[24]A H M. Rezaul Karim, Md. S. Rahman, Md. Zafar Iqbal, "Recognition of Spoken Letters in Bangla", Proc. of 6th ICCIT, Dhaka, 2002.

[25]Hush, D. R., and Horne, B. G., "Progress In Supervised Neural Networks: What's New Since Lippmann?", IEEE Signal Processing Magazine , pp. 8-39, January 1993.

[26] Burr, D. J., "Experiments On Neural Net Recognition On Spoken And Written Text", IEEE trans. on Acoustics, Speech and Signal Processing, vol. 36, no. 7, pp. 1162-1168, July 1988.

[27]Hava Siegelmann (April 1995). "Computation Beyond the Turing Limit". Science**268** (5210): 545–548.

[28]Haykins Simon. "Neural Networks, A comprehensive Foundation". Macmillan College Publishing Company, New York 1994.

[29]Mohammad H.Hassoun,"Fundamental of Artificial Neural Networks",Prentice Hall,2008.

[30]S.Rajasekaran,G.A.Vijayalakshmi Pai,"Neural Networks, Fuzzy Logic, and Genetic Algorithms",Prentice Hall,2007.

[31]N.K.Bose,P.Liang,"Neural Networks with Graphs,Algorithms and Applications",McGrawHill,2008.

[32]Christopher M. Bishop,"Neural Networks for Pattern Recognition ", CLARENDON PRESS, OXFORD,1995.

[33]Satu Virtanen, Kosti Rytknen "Neural Network", Helsinki Universityof Technology, web site http://www.askcom/tlark.neural networks.html

[34]Hush, D. R., and Horne, B. G., "Progress In Supervised Neural Networks: What's New Since Lippmann?", IEEE Signal Processing Magazine , pp. 8-39, January 1993.

[35] G. Saha, Sandipan Chakroborty, Suman Senapati ,"A New Silence Removal and Endpoint Detection Algorithm for Speech and Speaker Recognition Applications".

[36]Voicebox: Speech processing toolbox for matlab. "http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html".

[37]J.C.Simon "Spoken Language Generation and Understanding", 1980.

[38]George J.Klir/Bo Yuan,"FUZZY SETS AND FUZZY LOGIC",Prentice Hall,2005.