Composite Feature Selection Method based on Spoken Word and Speaker Recognition

Dipen Nath
Asst. Professor, Dept. of Computer Science and IT
Mangaldai College
Mangaldai, Assam, India

Sanjib Kr. Kalita, PhD
Asst. Professor, Dept. of Computer Science
Gauhati University
Guwahti-14, Assam, India

ABSTRACT
The aim of this paper is to measure the recognition capability of composite features extracted from speech signal and compare the result with other individually considered features for both spoken word and speaker based recognitions. Standard features like formants (F1, F2, F3), Linear Predictive Coefficients (LPC) and Mel Frequency Cepstral Coefficients (MFCC) along with various combinations among them are considered for the task to arrive at the conclusion. Six different speakers and six different strings (words) are considered in the present study. The threshold is set through an iterative approach for both spoken word and speaker recognition experiments. The mixing of LPC and MFCC is found to be the most promising combination among all others. Another interesting conclusion that we can draw from the study that the composite feature approach gives accuracy very near to 100% in case of speaker recognition task as compared to spoken word recognition task..

General Terms
Digital Signal Processing, Pattern Recognition, Artificial Neural Network, Automatic Speech and Speaker Recognition

Keywords
Feature Extraction and Selection, Feed Forward Neural Network, Speech and Speaker Recognition

1. INTRODUCTION
Spoken word and speaker recognition are the two domains which are of immense interest among speech researchers of the signal processing industry around the globe. The uniqueness of the anatomical structure of human vocal tract is a key factor in identifying spoken words and speakers through their voice signal. Spoken word recognition is an important recognition task whereby people have the option of comparing pre-stored words with new comers with new words. Various keyboard and mouse based data entry interfaces can be replaced by voice based applications which minimizes the overhead of typing and mouse handling activities. On the other hand, speaker recognition is another process of recognizing a person by speaker's voice. Voice comes under the category of biometric identity due to the differences in anatomical structure of the speakers. Identifying a person by his voice has several advantages. Remote persons can easily be authenticated using their voice patterns [25].

Like any other pattern recognition systems, speech and speaker recognition systems involve two phases namely, 'training' and 'testing' in the supervised approach [24]. Training is the process of familiarizing the system with the voice characteristics of the speakers enrolled in the system. Testing is the actual recognition task. Feature vectors that represent the voice characteristics of the speakers are extracted from the training utterances and are used in the formation of the reference model through the use of neural network training module. During testing, similar feature vectors are extracted from the test utterances (unknown) with the same process, and the degree of their match with the reference is obtained after using some similarity measurement technique. The level of their match is used to arrive at the decision whether the test utterance is acceptable or to be rejected for further processing using some of the static as well as dynamic threshold determination techniques.

2. LITERATURE REVIEW
Speech researchers across the world are focusing in the automatic speech recognition (ASR) problems treating it as a major challenge [1]. The interdisciplinary nature of speech technology constitutes another intricacy for speech researchers. Speech analysis and recognition tasks have been explored using different techniques and features. Some of the well known techniques are stated below.

- Artificial Neural Network (ANN)
- Fuzzy Logic in ASR
- Vector Quantization (VQ)
- Hidden Markov Models (HMM)
- Dynamic Time Warping (DTW)
- Principal Component Analysis (PCA)
- Independent Component Analysis (ICA)

Bishnu Prasad Das et. al [2] worked on a system to recognize the digits 'zero' to 'nine' spoken by a set of 28 English speakers. Words are classified using a combination of features based on LPC, MFCC, STE and ZCR. The recognition accuracy was 85% which they claimed better as compared to others using these features individually. The same task was practiced by others and is reported in the extant literature. The overall accuracy was supposed to be enhanced by enhanced by combining more features of the speech samples according to them.

Praveen N et.al [3] reported a speaker recognition technique based on spectral characteristics and mel-frequency cepstral coefficients. They studied the recognition using the two classifiers i.e. multilayer perceptron and Euclidean distance classifier and found their recognition accuracy as 83.38% and 96.18% respectively.

Adjoudj Reda et. al [1] has got a speaker recognition accuracy of 97% for his own data set in an attempt taking into account various date sets with MFCC and ANN as the supporting tools.

Kshamamayee Dash et.al [4] reported another interesting MFCC and ANN based speaker recognition project and applied it to a speech of some unknown speaker. After investigating the extracted features of the unknown speech and then comparing them to the stored features for each different speaker, the results they found were having efficiency 85%. They gave emphasis on collecting 100 such speech instances in future and to calculate the MFCC features for NN training to get more accurate figures for identification.

Lajish V. L et.al [5] has modelled the speaker identity based on the non-linear properties of the speech samples. The speaker identification experiments are conducted based on Phase Space Point Distribution (PSPD). The PSPD features obtained from five vowels are used for speaker identification purpose using the feed forward multi layer perceptron. The experiment is repeated by taking different combination of PSPD, MFCC, pitch and first formant frequency. The experimental results indicate that the proposed phase space approach by itself is still below (31.60%) than that of MFCC features (46.21%). The results further show that the combined approach of mixing PSPD features, MFCC, pitch and first formant frequency offers enormous improvement in speaker identification (on an average of 83.40%) accuracy, which instigates us to go for the proposed work.

Dipen Nath et.al [21] has tested the efficiency of a speaker recognizer with different combinations of features. They found the feature '*Formant+LPC'* as the optimal one amongst other three mentioned high accuracy feature sets. The recognition rate reaches up to 100% for one of the two tested data sets. Two different experiments with different data sets with different sampling rates gave them a strong evidence for concluding the study to support '*Formant+LPC'* feature for speaker recognition. Also the strings with more phonemic contents are found to be a better choice for higher speaker recognition rate. They proposed to extend the work to distinguish among male and female speakers and were insisting on repeating the experiment using some standard data sets and compare the end results.

E. M. Mohammed et. al [23] tried evaluating spoken language identification using LPC and MFCC with artificial neural network as classifier. But the experiments concluded with MFCC and ANN as the better combination to reach even up to 100% recognition, as compared to considering LPC for feature extraction part.

3. THEORETICAL BACKGROUND

3.1 Formant and LPC

Formants are extracted and removed from the speech signal in LPC analysis and the intensity & frequency of the remaining buzz [6, 7, 8, 9] is then estimated. The method of removing the formants is termed as inverse filtering and the remaining signal part is called as the residue. In LPC system, each sample of the signal is expressed as the linear combination of the previous samples and the respective equation is called the linear predictor. Hence it is called linear predictive coding (LPC). The coefficients of the difference equation (the prediction coefficients) characterize the Formants.

A predictor polynomial, defined as the Fourier transform of the corresponding second order predictor is given by [4, 7, 13]

$$A_k(e^{jw}) = 1 - \alpha_k e^{jw} - \beta_k e^{-j2w} \tag{1}$$

where α_k and β_k are the real valued prediction coefficients. From equation (1), we get

$$|A_k(e^{jw})|^2 = 1 + \alpha_k^2 + \beta_k^2 - 2\alpha_k(1-\beta_k)\cos\omega - 2\beta_k\cos(2\omega)$$

$$= (1+\beta_k)^2 + \alpha_k^2 + \frac{\alpha_k^2(1-\beta_k)^2}{4\beta_k} - 4\beta_k\cos\omega + \left[\frac{\alpha_k^2(1-\beta_k)^2}{4\beta_k}\right]^2 \tag{2}$$

The parameter β_k corresponds to the bandwidth of the resonator and defined as negative logarithm of $(-\beta_k).[A_k(e^{jw})]^2$. The formant frequency is given by

$$F_1 = \arccos\left[\frac{-\alpha_k(1-\beta_k)}{4\beta_k}\right] \tag{3}$$

Using equation (1), the corresponding predictor error can be written as

$$\begin{aligned} &E(\omega_{k-1}, \omega_k \mid \alpha_k, \beta_k) \\ &= (1+\alpha_k^2+\beta_k^2)r_k(0) - 2\alpha_k(1-\beta_k)r_k(1) - 2\beta_k r_k(2) \end{aligned} \tag{4}$$

where, $r_k(\gamma)$ are the autocorrelation coefficients where γ =0, 1, 2. It is thus found that $|\cos\omega| < 1$, and thus the values of α_k and β_k are taken as-

$$\left.\begin{aligned} &\alpha_k < 2 \\ &-1 < \beta_k < \frac{-\alpha_k^2}{4} \end{aligned}\right\} \tag{5}$$

3.2 MFCC

LPC and Formant frequency estimation [7, 10, 11] methods were used for feature extraction purpose earlier. But recently the Mel Frequency Cepstral Coefficient (MFCC) has been widely used in speech processing applications.

MFCC [22] is based on the human peripheral auditory system. As the human perception of the frequency content's of speech signal does not follow linear scales, thus for each tone with an actual frequency, a subjective pitch is measured on a scale called the 'Mel Scale'. The continuous speech signal is blocked into some finite number (N) of samples with adjacent frames being separated by *M* (*M* < *N*). The first frame consists of the starting *N* samples and the second frame begins *M* samples after the first frame, and overlaps it by *N* - *M* samples. This process continues until the entire speech is accounted for within few frames. The next step in the processing is to window each individual frame so that the signal discontinuities get minimized at the beginning and end of each of the frames. The concept here is to minimize the spectral distortion by using the window to taper the signal to zero at the beginning and end of each frame. We define the window as $w(n),\ 0 \le n \le N-1$, where *N* is the number of samples in each frame. The result of the windowing is the signal

$$y_l(n) = x_l(n)w(n), \quad 0 \le n \le N-1 \quad (6)$$

Fast Fourier Transform deals with converting each frame of *N* samples from time domain into frequency domain. The FFT is a high-speed algorithm to implement the Discrete Fourier Transform (DFT), which is defined on the set of *N* samples $\{x_n\}$, as

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi kn/N}, \qquad k = 0,1,2,...,N-1 \quad (7)$$

The result after this step is often referred to as *spectrum* or *periodogram*. In the present study the number of mel spectrum coefficients, *K*, is chosen as 12. *Cepstrum calculation* is derived from the Fourier Transform of the recorded speech signal, where the frequency bands are positioned logarithmically, whereas the same were not positioned logarithmically in the Fourier Transform. As the frequency bands are positioned logarithmically in MFCC processor, it approximates the human system response more closely than in any other system. In the mel frequency cepstral coefficients, the calculation of mel cepstrum is same as the real one, except the mel cepstrum's frequency scale is warped to keep up a correspondence with the mel scale.

3.3 Feed Forward ANN

As compared to few normally available non-linear methods of discrimination, feed forward artificial neural networks [1, 11, 13, 19, 20] are more widely used in solving classification problems because of its straight forward approach. Neural networks emerge as one of the class of flexible non parametric classification methods which is used frequently for classification. Feed forward neural networks provide a flexible way to generalize linear regression functions. We start with the simplest but most common form i.e. MLP (Multi Layer Perceptron) with one hidden layer only. This work can be further tested under multi layer perceptron with multiple hidden layer purviews.

3.4 Levenberg Marquardt Algorithm

The Levenberg-Marquardt Algorithm (LMA) is a curve-fitting algorithm used in the present study. Least squares problems arise when fitting a parameterized function to a set of measured data points by minimizing the sum of the squares of the errors between the data points and the function [14, 15, 16]. Non linearity in the parameters is the reason for the least square problems. Nonlinear least squares methods involve an iterative improvement to parameter values to reduce the sum of the squares of errors between the function and the measured points. The Levenberg-Marquardt curve-fitting method is a combination of two minimization methods, namely, the Gradient Descent method and the Gauss-Newton method. In case of Gradient Descent method, the sum of the squared errors is minimized by updating the parameters in the direction of the greatest reduction of the least squares objective. But, in the Gauss-Newton method, the sum of the squared errors is reduced by assuming that the least square function is locally quadratic, and it finds the minimum of the quadratic. The Levenberg-Marquardt method acts like a Gradient Descent method, when the parameters are far apart from their optimal value and it acts like the Gauss-Newton method when the parameters are close to their optimal value. MATLAB is used to train the proposed network by implementing LMA as back propagation algorithm. Validation vectors are used to stop training early if the network performance on the validation vectors fails to improve or remains the same. Test vectors are used for further generalizing the network well, but it does not affect the training process.

3.5 Spoken Word and Speaker Recognition

Spoken word recognition concentrates on recognizing particular strings (words) spoken by someone, but speaker recognition concentrates on identifying the one who speaks the certain words [17, 18]. The aim in spoken word recognition is to recognize the unknown word from a set of known words (closed set spoken word recognition). On the other hand, speaker identification (SI) is to recognize the unknown speaker from a set of known speakers (closed set speaker identification). Both spoken word and speaker recognition systems are composed of the following modules [11, 13]:

- ***Front-end processing*** – It converts the sampled speech signal into set of feature vectors characterizing the properties of spoken words or speakers that can separate different words or speakers. Frontend processing is performed both in training and testing phases.
- ***Speaker modeling*** - This part performs a reduction of feature data by modeling the distributions of the feature vectors.
- ***Speaker database*** - The speaker models are stored here.
- ***Decision logic*** – It makes the final decision about the identity of the word or the speaker by comparing unknown feature vectors to all models in the database and selecting the best matching model.

4. EXPRIMENTAL SETUP

4.1 Speech Database

The data set is recorded in 16KHz sampling frequency. Words considered are 'Green', 'Indigo', 'Red', 'Logoff', 'Restart' and 'Shutdown'. A total of six speakers are involved in preparation of the data set. *Male-age-34*, *Male-age-15*, *Female-age-14, Male-age-34, Male-age-24* & *Female-age-28* are the corresponding contributors for the Data Set.

The data set is composed of four males and two females (with six different words) who have contributed to prepare the whole dataset containing six words. So we have [(50x6) x3] = 900 utterances in the data set. The experiment is carried out with Intel (R) core (TM) i-5 2430M CPU @2.40 GHz 2.40 GHz processor and 3.00 GB RAM. Windows 7 Ultimate (32-bit o/s) and MATLAB version 7.11.0 (R2010b) is used for the experiment part and Goldwave Version 5.58 is used for recording of the sound samples.

4.2 Network Architecture

A three layer feed forward neural network has been finally selected for the recognizers in the present study which is shown in Figure-1. The network consists of 32 input nodes with 10 numbers of hidden nodes in the single hidden layer and 6 output nodes for both the cases of spoken word and speaker recognition.

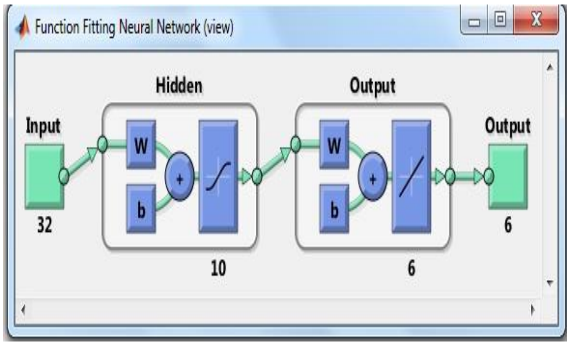

Figure- 1: ANN architecture with 'LPC+MFCC' as input feature for both Speaker and Spoken Word Recognition

To train the network, LMA [14, 15, 16] has been used. The feature vector which is the output of the feature extraction block has been normalized and used as input to the feed forward based recognizer. A total of 630(70%), 135(15%) and 135(15%) utterances of the dataset are used for the network training, validation and testing procedures respectively. Finally after getting satisfactory regression results as well as mean square results, we select certain network for future recognition activities. A total of 90 utterances are taken for each of the two kinds of recognitions, i.e. spoken word and speaker recognitions, considering 0.45 as the threshold value, which is determined by an iterative approach of threshold determination as stated below.

Step-1: A high value of thresh hold is set against the chosen feature

Step-2: Recognition rate is calculated for the first time

Step-3: The threshold value is decreased by a step of 0.05

Step-4: Recognition rate is again calculated

Step-5: Repeat the steps 3 and 4 until two consecutive iterations yield same result.

Step-6: Select the thresh hold as the final one for the said feature

Step-7: Store the threshold against the feature chosen

Step-8: Repeat the step-1 through step-7 to choose a high recognition final feature with the final threshold.

The above steps can be realised clearly from the Table-1, Figure-2, Table-2 and Figure-3 respectively.

Table- 1: Iterative threshold determination table for Speaker recognition

Threshold	LPC	MFCC	LPC+ MFCC
0.80	82.22	93.33	93.33
0.75	85.56	95.00	96.67
0.70	90.56	95.56	97.22
0.65	93.89	96.11	97.22
0.60	95.56	96.11	97.78
0.55	97.22	96.11	98.33
0.50	97.78	96.67	98.89
0.45	98.33	96.67	99.44
0.40	98.33	96.67	99.44

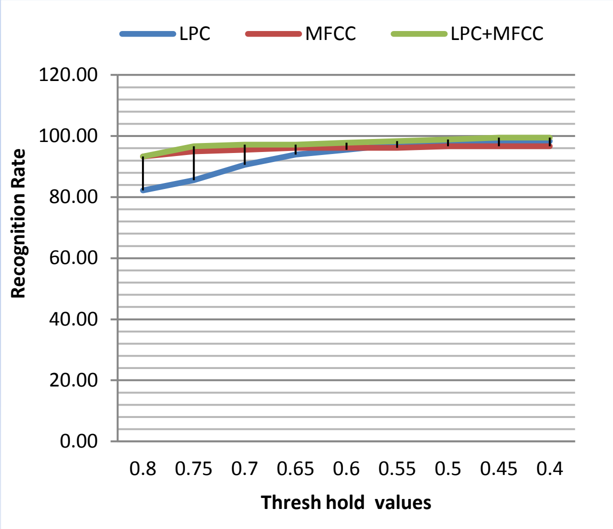

Figure-2: Threshold determination in Speaker recognition

Table 2: Iterative threshold determination table for Spoken word recognition

Threshold	LPC	MFCC	LPC+ MFCC
0.80	68.33	79.44	79.44
0.75	72.78	84.44	85.00
0.70	79.44	87.78	88.89
0.65	83.33	90.00	88.89
0.60	88.33	92.22	94.44
0.55	90.00	92.78	95.00
0.50	92.22	93.89	96.67
0.45	93.89	95.56	97.78
0.40	95.00	95.56	97.78

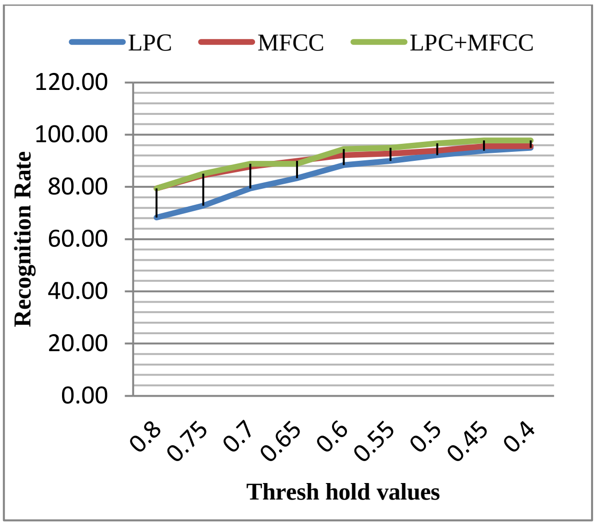

Figure-3: Threshold determination in spoken word recognition

5. RESULT AND DISCUSSION

5.1 Feature Selection

The features i.e. formants (F1, F2, F3), LPC, MFCC, formant + LPC, formant + MFCC, LPC + MFCC, formant+LPC+MFCC are studied here. The data set is prepared taking into account the words 'Red', 'In-di-go', 'Green' by three speakers and 'Log-off', 'Re-start' and 'Shut-down' by another three speakers with a total of six speakers. But amongst them the 'LPC+MFCC' combination is found to be close to 100% recognition as compared to other combinations or individual considerations. This feature is found to be a better choice for both spoken word and speaker recognition domains as shown in Figure-4 and Figure-5 below.

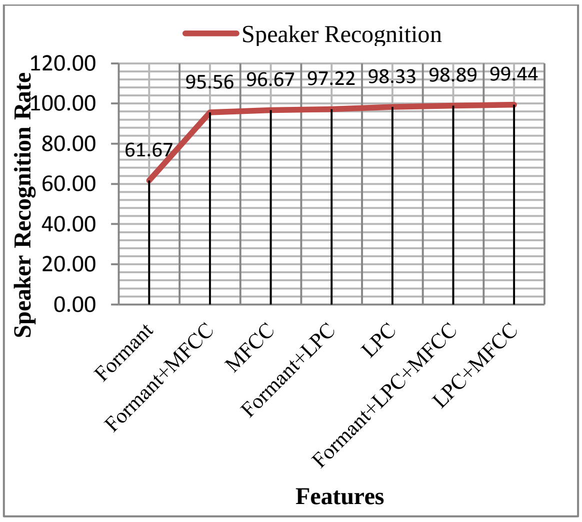

Figure-4: Speaker recognition rate among various features

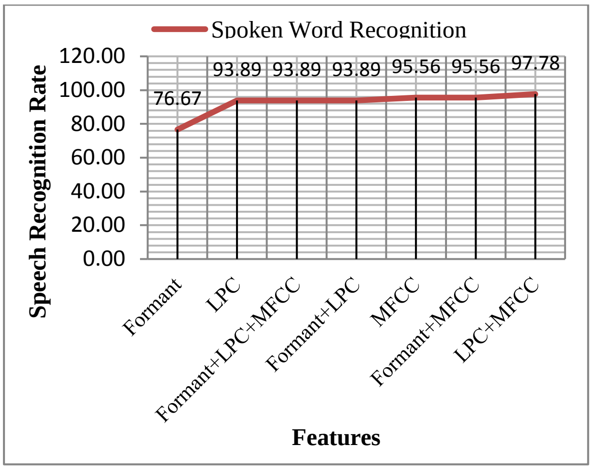

Figure-5: Spoken word recognition rate among various features

6. CONCLUSION AND FUTURE WORK

The efficiency of both the recognizers i.e. spoken word based recognizer and speaker based recognizer is examined with different feature combinations. The feature '*LPC+MFCC'* is found to be the most promising feature amongst other (i.e. LPC, MFCC when considered individually) for higher recognition of the spoken words as well as and the speakers. This experiment is one of its first kind to investigate the promising feature taking into account both spoken word and speaker based recognitions maintaining accuracy up to 100%. Both experiments for spoken word and speaker support the common feature '*LPC+MFCC'*. Also the strings with more phonemic contents are considered better choice for higher recognition rate in case of speaker recognition.

The proposed work can be extended for distinguishing genders to minimize the work space enhancing both space and time complexities. Few interesting results are expected after comparing the present result with some standard data based results.

7. ACKNOWLEDGMENTS

We are very much thankful to Deepjyoti, Deep, Dulumani, Abhinav, Kaushik and Ankita for their 900 careful utterances (spoken words) present in the data set. It is because of the data for which it becomes possible to accomplish the entire task bringing into light the novel inferences.

8. REFERENCES

[1] Adjoudj Reda, Boukelif Aoued, "Artificial Neural Network & Mel-Frequency Cepstrum Coefficients-Based Speaker Recognition", 3rd International Conference: **S**ciences of **E**lectronic, **T**echnologies of **I**nformation and **T**elecommunications--TUNISIA, March 27-31, 2005

[2] Das, B. P., Parekh, Ranjan. "Recognition of Isolated Words using Features based on LPC, MFCC, ZCR and STE with Neural Network Classifiers". International Journal of Modern Engineering Research (IJMER) ,Vol.2, Issue.3, May-June 2012 pp-854-858 [ISSN: 2249-6645]

[3] Praveen N, Tessamma Thomas, "Text dependent speaker recognition using MFCC features and BPANN", International Journal of Computer Applications (0975 – 8887), Volume 74– No.5, July 2013.

[4] Kshamamayee Dash, Debananda Padhi, Bhoomika Panda, Prof. Sanghamitra Mohanty, “Speaker Identification using Mel Frequency Cepstral Coefficient and BPNN”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 4, April 2012 ISSN: 2277 128X

[5] Lajish V.L , Sunil Kumar R.K and Vivek P, “Speaker identification using a nonlinear speech model and ANN”, International Journal of Advanced Information Technology (IJAIT) Vol. 2, No.5, October 2012

[6] Thiang, Suryo Wijoyo. “Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot”. 2011 International Conference on Information and Electronics Engineering IPCSIT vol.6 (2011) © (2011) IACSIT Press, Singapore

[7] Talukdar, P. H., Bhattacharjee, U., Goswami, C. K. , Barman, J., “Cepstral Measure of Boro Vowels through LPC-Analysis”, Journal of the CSI, Vol. 34 No 1, Jan – Mar, 2004.

[8] Kalita S.K., Dutta R., and Talukdar P. H., “A spectral analysis of Bodo and Assamese vowels”, Abstracts 3rd International Conference on “Computers and Devices for Communication”. CODEC – 06, Kolkata, India, pp. 41, 2006.

[9] Braman, J., Kalita, S., Talukdar, P. H., “Features extraction of bodo vowels through lpc-analysis”, Proceedings of Frontiers of Research on Speech and Music (FRMS-2004), 2004

[10] Hasan Rashidul, Jamil Mustafa, Rabbani Golam, Rahman Saifur, “Speaker identification using mel frequency cepstral coefficients”, 3rd International Conference on Electrical & Computer Engineering, Dhaka, Bangladesh, ICECE 2004, 28-30 December 2004

[11] Rabiner L., Juang B. H. and Yegnanarayana B. – “Fundamentals of Speech Processing”, Pearson Education, ISBN 978-81-775-8560-5 (2011)

[12] D.Ripley, “Neural Networks and Related Methods for Classification”, Journal of the Royal Statistical Society. Series B (Methodological), Vol. 56, No. 3(1994), pp. 409-456, 1994.

[13] Rabiner L. and Juang B. H. – “Fundamental of Speech Processing”, Prentice-Hall, 1993

[14] K. Levenberg. “A Method for the Solution of Certain Non-Linear Problems in Least Squares”. The Quarterly of Applied Mathematics, 2: 164-168 (1944).

[15] M.I.A. Lourakis., “A brief description of the Levenberg-Marquardt algorithm” implemented by levmar, Technical Report, Institute of Computer Science, Foundation for Research and Technology, - Hellas, 2005.

[16] Mark K. Transtrum and James P. Sethna “Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization,” Preprint submitted to Journal of Computational Physics, January 30, 2012.

[17] Vibha Tiwari, “MFCC and its applications in speaker recognition”, International Journal on Emerging Technologies 1(1): 19-22(2010) ISSN: 0975-8364, 2010.

[18] S.Khan, Mohd Rafibul lslam, M. Faizul, D. Doll, “Speaker recognition using MFCC”, presented in IJCSES ,International Journal of Computer Science and Engineering System, **2**(1**):** 2008.

[19] Bishop, C., “Neural Networks for Pattern Recognition”, Oxford University Press, Oxford, 1995.

[20] Haykin, S., “Neural Networks - A Comprehensive Foundation”, 2nd ed. Prentice-Hall, Englewood Cliffs, 1998.

[21] Dipen Nath, Sanjib Kr Kalita, “Feature Selection Method for Speaker Recognition using Neural Network”, International Journal of Computer Applications(0975-8887), Volume 101- No.3, September, 2014[ISBN: 973-93-80883-64-3, ISSN: 0975-8887]

[22] Sanjib Kr. Kalita, Dipen Nath, “SPEECH UTTERANCE CLASSIFICATION USING MEL-FREQUENCY CEPSTRAL COEFFICIENT (MFCC) AND ROC ANALYSIS”, a chapter in the book titled “Advanced Computer Science & Applications” pp. (54-64), 2014 [ISBN: 978-93-83252-08-4]

[23] E. M. Mohammed, M. S. Sayed, A. M. Moselhy, A. A. Abdelnaiem, “LPC and MFCC Performance Evaluation with Artificial Neural Network for Spoken Language Identification”, International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 6, No. 3, June, 2013

[24] Soyuj Kumar Sahoo, Tarun Choubisa & S. R. Mahadeva Prasanna (2012) Multimodal Biometric Person Authentication : A Review, IETE Technical Review, 29:1, 54-75

[25] Rajkumar Palaniappan, Kenneth Sundaraj, Nizam Uddin Ahamed, Agilan Arjunan & Sebastian Sundaraj (2013) Computer-based Respiratory Sound Analysis: A Systematic Review, IETE Technical Review, 30:3, 248-256

9. AUTHOR PROFILE

Dipen Nath is pursuing his PhD at Gauhati University, Assam, India. He has completed his MSc in Computer Science from Gauhati University in 2004 and B.Sc (honours in Mathematics) in the year 2002 in the same university. Currently he is working as an Assistant Professor at Mangaldai College under Gauhati University. He has publications in International Journal and has authored few peer-reviewed articles and presented the same in international/national conferences and seminars.

Dr. Sanjib Kr. Kalita is an Assistant Professor, Dept. of Computer Science, Gauhati University, Assam, India. Dr. Kalita has previously worked at various government Colleges and Universities. He was the head in the dept. of Computer Science, Handique Girls’ College, Guwahati (2002-09). He headed the dept. of Computer Science, Gauhati University, Kokrajhar campus (presently Bodoland University) (2009-12). He has done his PhD from Gauhati University. He has authored over thirty peer-reviewed articles, and one book.; email: sanjib959@gauhati.ac.in