Journal of Theoretical and Applied Information Technology
15th November 2022. Vol.100. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A NOVEL APPROACH FOR AUTOMATIC SPEAKER IDENTIFICATION OF ASSAMESE LANGUAGE USING COSINE SIMILARITY AND ABSOLUTE MFCC FEATURE MATRIX

ANKUMON SARMAH[1], RIZWAN REHMAN[2], PRIYAKSHI MAHANTA[3], KANKANA DUTTA[4], KAUSTUVMONI BORDOLOI[5], KIMASHA BORAH[6], HARJINDER SINGH[7]

[1,2,3,4,5,6] Assistant Professor, DIBRUGARH UNIVERSITY, Centre for Computer Science and Applications, India
[7]Assistant Professor, D. H. S. K. COLLEGE, Department of BCA and Computer Science, India
e-mail: [1]ankumonsarmah2009@gmail.com, [2]rizwan@dibru.ac.in, [3]priyakshimahanta@dibru.ac.in, [4]kankanadutta@dibru.ac.in, [5]kaustuvmoni@dibru.ac.in, [6]kimashaborah@dibru.ac.in, [7]singhanjum5@gmail.com

ABSTRACT

Automatic speaker Identification (ASI) is always challenging work for researchers. ASI is a process where a speaker is recognized automatically from his/her voice sample by comparing it with their previously recorded voices. The machine learning approach has been gaining popularity in recent years for ASI. Different machine learning approaches used in ASI in recent years are Convolutional Neural Network (CNN) [14,15,16], Deep Neural Network (DNN) [10,11,12,13], Artificial Neural Network (ANN) [17,18]. This research aims to build an automatic speaker identification system for the Assamese language, which is spoken in the North-Eastern part of India and is one of the low-resource languages. So far, cosine similarity and parallel processing methods have not been used for speaker identification in the Assamese Language, which is the novelty of the current work. The model developed in this work uses Mel-frequency cepstral coefficient (MFCC) to extract important features of speakers' voices to create a training sample set in the first phase. In the present approach, we have used the Speaker's absolute feature vectors (MFCC) directly, without any averaging, in order to retain and exploit the Speaker's unique characteristics. In the second phase, the features in the training sample set of the first phase are compared with the real-time test voice samples using the cosine similarity method to identify the Speaker automatically. Parallel processing is used to compare all the coefficients in the test voice sample with the training voice sample to make the system faster. The effectiveness of the proposed method has been established in terms of precision, recall, f1 score, and accuracy value. The model demonstrated an accuracy of 91% for speaker identification in the Assamese language.
Keywords: *Mel- Frequency Cepstral Coefficient (MFCC), Speaker Identification, Cosine Similarity, Automatic Speaker Identification (ASI*), *Assamese*

1. INTRODUCTION

Automatic Speaker Recognition is a process where a speaker of a particular language is identified from his voice samples. The automatic speaker recognition can be done by comparing the real-time voice sample of the Speaker with his/her previously recorded voice samples. A voice sample is initially in Analog form, which is difficult to process due to its infinite points. Therefore, it is converted to a digital form [1]. The recorded voice samples are converted to digital form before being separated into training and test data sets. Test data sets samples are compared by the model developed from the training data sets to recognize the Speaker. Speaker recognition can be broadly categorized as Speaker Identification and Speaker verification. Speaker identification is recognizing a speaker from previously-stored voice samples of speakers [2], and Speaker verification is the process where a speaker is verified as an authenticated user of the system or not. In speaker verification, there are two categories, text-dependent and text-independent [2]. In a text-dependent system, the process requires speech from a fixed word or sentence, while in the case of a text-independent system, the speech samples may be completely different from what was spoken in the training phase [8].

Despite all the progress in the ASI field, it is still a challenging area, specifically for low-resource languages. Developing systems for extensive vocabulary databases, particularly for low-resource languages with a dearth of standard speech corpus, is one of the difficulties. Nowadays, it is imperative and valuable to develop an ASI system for low-resource languages since, at present, only about 100 languages have Automatic Speech Recognition (ASR) capability out of the roughly 7000 languages spoken worldwide [30]. In this paper, a low-resource language of northeast India, i.e., Assamese, has been considered for the study since a minimal amount of work has been done for the Assamese language. Assamese is an Indo-Aryan language spoken mainly in the northeast Indian state of Assam. [19]. The total population of Assamese speakers is nearly 15.09 million, making up 48.38% of the state's population, according to the Language census of 2011[20]. Feature extraction is the next important part of the Speaker identification process, for which we have used the MFCC technique for the current work. There are different methods for feature extraction from voice data, namely Mel Frequency Cepstral Coefficients (MFCC), Linear Prediction Coefficients (LPC), and perceptual linear prediction coefficients (PLPs) [2] [7]. Much research work has been done for ASI using MFCC for different languages like Berber, English, Many Indian languages, Chinese, Arabic, Italian, French, German, and Spanish [6][2][1][25][8].

After feature extraction, the next phase in Speaker identification is identifying the user with the help of a similarity measure. Different similarity measures are Cosine distance, Manhattan distance, Euclidean distance, Minkowski distance, and Jaccard similarity [8][22]. The advantage of cosine similarity over Euclidean distance is that even if the two similar data objects are far apart by the Euclidean distance because of the size, they could still have a smaller angle between them [22] which can suggest their similarity. Apart from this, Cosine similarity has a disadvantage: the Magnitude of vectors dealing with term frequency does not play any role in this similarity measurement [21]. Much research has been done on Speaker recognition using different approaches for many languages. Soufiane Hourri and Jamal Kharroubi, in their work on text-independent speaker identification, showed that Cosine similarity gave a better result than Euclidean and Manhattan distance [8]. In this paper, we have used Cosine similarity for similarity measure.

Cosine similarity is a measure of similarity between two vectors of equal length. The formula to find the cosine similarity between two vectors is given below in equation 1.

$$\cos(A,B) = \left(\frac{A.B}{||A||.||B||}\right) = \left(\frac{\sum_{i=1}^{n}(A_i B_i)}{\left(\sum_{i=1}^{n}\sqrt{A_i^2}\right)\left(\sum_{i=1}^{n}\sqrt{B_i^2}\right)}\right) \quad (1)$$

In the above equation, Ai and Bi are the components of the two vectors. The cosine similarity between two vectors is measured in terms of the angle formed by the two vectors [27]. If the value is 1, it means the angle formed by them is 0^o which shows that the two vectors are similar, whereas -1 means the two vectors are opposite, or we can say they are dissimilar.

The Mel-Frequency Cepstrum (MFC) represents the short-period power spectrum of the sound wave. The collection of coefficients of MFC is known as MFCC (Mel frequency cepstral coefficient), which is based on the acoustic characteristics of humans [3]. MFCC is a widely used feature extraction method in automatic speaker identification because its coefficients are based on human hearing perceptions [5]. For generating the MFCC, the first step is Frame blocking, followed by Windowing [9]. In the third step, FFT is computed where each frame is converted from the time domain to the frequency domain. According to psychophysical studies, the human perception of frequencies does not follow a linear scale; therefore, transformation is required from the linear scale of frequencies in the Mel scale [6]. The approximate formula to compute the mels for a given frequency f in Hz is given below:

$$Mel(f) = 2595 * \log10(1 + \frac{f}{700}) \quad (2)$$

In their work [23][32] have performed Automatic Speaker Recognition in the Assamese language, where the authors have used neural network with MFCC but no other literature where ASR has been done in the Assamese language using Cosine Similarity method. Also, to preserve the Speaker's distinctive qualities, we have used the Speaker's feature vectors (MFCC) directly in the current method without any averaging. So, in

our work, we have proposed a simple but high accuracy-oriented ASI method using cosine similarity for Assamese language speakers. The statistical validation of the proposed method is evaluated in terms of accuracy, precision and recall value.

2. LITERATURE REVIEW

Much research has been done in the field of Speaker recognition. It is challenging for the researchers because of various factors such as recording environments, variation of voice with time, health conditions, [6], etc. Table 1 provides an overview of various related work done in the field of speaker recognition.

Table 1: Related work

Sl. No.	Title	Dataset Used	Technique used	Validation method	Reference Number
1	Speaker Recognition for Hindi Speech Signal using MFCC-GMM Approach	Hindi speech sample of 15 speakers (10 male, 5 female)	Combination of MFCC and Gaussian mixture model	Accuracy	[1]
2	Automatic Speaker Recognition using MFCC and Artificial Neural Network	IITG Multivariability Speaker Recognition Database	Multilayer perceptron (MLP) feedforward neural network	Accuracy	[2]
3	Real Time Speaker Recognition System using MFCC and Vector Quantization Technique	TIDIGIT database	MFCC and Vector Quantization Technique	Accuracy	[3]
4	Speaker Recognition using MFCC, shifted MFCC with Vector Quantization and Fuzzy	Dataset consist 1760 training utterances and 680 testing utterances.	MFCC and Shifted MFCC with Vector Quantization and fuzzy	Accuracy	[4]
5	SVM based Emotional Speaker Recognition using MFCC-SDC Features	IEMOCAP database	multiclass Support Vector Machine (SVM) classifier	Accuracy	[5]
6	A Novel Scoring Method Based on Distance Calculation for Similarity Measurement in Text-Independent Speaker Verification	FSCSR corpus	a novel scoring method based on distance calculation for similarity measurement [8]	EER	[8]
7	Text dependant speaker recognition using MFCC, LPC and DWT	Vowel-emphatic Algerian Berber dataset	MFCC feature, their first and second derivatives and discrete wavelet transform (DWT) followed by linear predictive coding (LPC)	Recognition Rate	[6]
8	Speaker identification based on normalized pitch frequency and Mel Frequency Cepstral Coefficients	8 speakers repeating each sentence 10 times for getting 80 speech samples (Primary Dataset)	cepstral features and the Normalized Pitch Frequency (NPF)	recognition rate	[9]

9	An MFCC-based text-independent speaker identification system for access control	Various Primary dataset	MFCC and GMM	Accuracy	[25]
10	Assamese Speaker Recognition Using Artificial Neural Network	Primary dataset of 10 Speakers	MFCC, LPC and Artificial Neural Network	Accuracy	[23]
11	Speaker identification model for Assamese language using a neural framework.	Primary dataset of 5 speakers from four different dialects of Assamese language	The model presented here employs a unique Self Organizing Map (SOM) embedded in a probabilistic neural network (PNN) and learning vector quantization (LVQ) neural framework.	Accuracy	[32]

3. METHODOLOGY

In the proposed method, the MFCC matrices obtained from the speech signals of the vowels spoken by each Speaker are used to create the model by separating the matrices into test and training sets. Each ***mi*** (each matrix in Test Data matrices) is compared with training samples with the help of Cosine Similarity method. Each row of a ***mi*** is compared with the corresponding row of a training sample with the help of cosine similarity, and the highest total sum is used to identify the Speaker. The block diagram of the proposed method and the algorithm is shown in *Figure 1* and in section *3.1*, respectively.

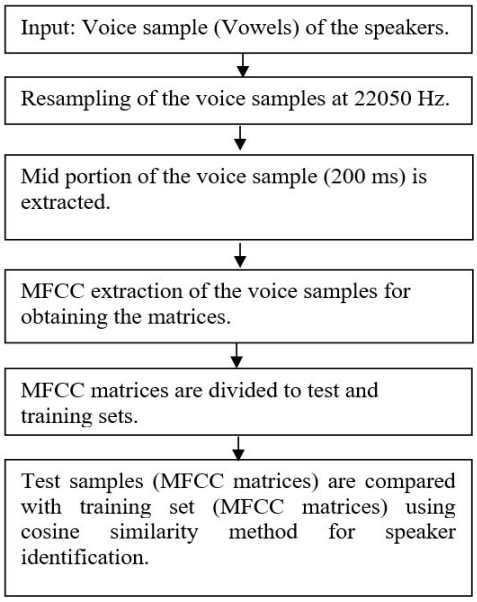
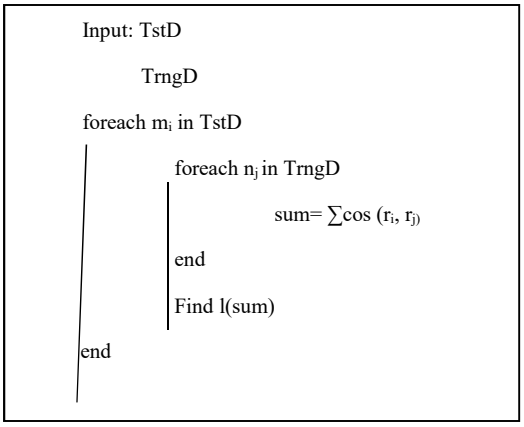


Figure 1: The process of the proposed method for Speaker Identification from Assamese vowel

3.1 Algorithm of The Proposed Method:



Where, ***TstD***= Test data matrices, m_i= Each matrix in Test Data matrices, r_i= Each row of m_i, ***TrngD***= Training data matrices, n_j= Each matrix in Training data matrices, r_j= Each row of n_i, $\sum cos\ (r_i, r_j)$= sum of cosine similarity of each r_i of m_i with corresponding row r_j from n_j, ***l(sum)***= Largest sum

3.2 Algorithm Complexity

Since our algorithm mainly focuses on matrix comparison using Cosine similarity, the time required is comparatively high. Therefore, we have applied parallel processing using python programming language to test our code, which reduces the time required to identify the Speaker. The complexity of the algorithm from Section 3.1 is shown below:

Computing ***l(sum)*** between each m_i with all n_j is $O(n^2)$ where ***n*** is the dimensionality of each row.

4. DATASET

Dataset plays a leading role in the Speaker identification system [1]. Assamese vowel consists of total 11 characters (অ, আ, ই, ঈ, উ, ঊ, ঋ, এ,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ঐ, ঔ, ঈ) [24].

The present work's dataset comprises 26 speakers (12 male and 14 female). The speakers recorded each vowel 40 times, thus creating 280 minutes of the voice sample. Many works have been done using very few speakers ranging from 06 to 30 for speaker identification [1][5][6][25][9][29]. Each speech sample collected at a 44100-sample rate is then re-sampled with 22050 Hz frequency since the components above the Nyquist Frequency are not represented [31]. As the middle portion of every sample contains the highest peak frequency, so the middle 200 milliseconds of each sample is selected to find the features using the Mel frequency cepstral coefficient. Since most of the signal information is represented by the first few coefficients between 12 and 20 [9], so we have used 12 coefficients of MFCC; therefore, each matrix formed from each sample has 12 columns and 39 rows. Thirty-nine rows in each matrix are due to the same length of 200 milliseconds of each sample. Speaker information of the dataset used is shown in *Table 3*.

Table 3: Speaker information of the datasets used

Sl. No.	Speakers	Gender	Age	Profession
1	Speaker 1	Male	25	Student
2	Speaker 2	Male	24	Student
3	Speaker 3	Female	23	Student
4	Speaker 4	Male	24	Student
5	Speaker 5	Male	36	Banking Professional
6	Speaker 6	Female	24	Student
7	Speaker 7	Female	35	Teacher
8	Speaker 8	Female	21	Student
9	Speaker 9	Male	22	Student
10	Speaker 10	Male	24	Student
11	Speaker 11	Female	23	Student
12	Speaker 12	Male	24	Student
13	Speaker 13	Female	24	Student
14	Speaker 14	Female	28	Teacher
15	Speaker 15	Male	24	Student
16	Speaker 16	Male	33	Teacher
17	Speaker 17	Female	23	Student
18	Speaker 18	Male	24	Student
19	Speaker 19	Female	23	Student
20	Speaker 20	Male	22	Student
21	Speaker 21	Female	24	Student
22	Speaker 22	Female	25	Research Scholar
23	Speaker 23	Female	24	Student
24	Speaker 24	Female	23	Student
25	Speaker 25	Male	24	Student
26	Speaker 26	Female	23	Student

5. EXPERIMENTAL RESULTS

The Experimental result of the proposed work is given in terms of Speaker Identification

Accuracy. The accuracy of the speaker identification as provided by the model developed is shown in *Table 4*. Precision, recall, and f1-score are shown in *Table 5*.

Table 4: Accuracy of Speaker Identification of the proposed method

Sl. No	Speaker	Speaker identification Accuracy (%)
1	Speaker 1	97.24
2	Speaker 2	100
3	Speaker 3	89.28
4	Speaker 4	82
5	Speaker 5	98
6	Speaker 6	99
7	Speaker 7	100
8	Speaker 8	93
9	Speaker 9	92
10	Speaker 10	90
11	Speaker 11	96
12	Speaker 12	70
13	Speaker 13	100
14	Speaker 14	100
15	Speaker 15	56
16	Speaker 16	95
17	Speaker 17	91
18	Speaker 18	93
19	Speaker 19	98
20	Speaker 20	100
21	Speaker 21	95
22	Speaker 22	100
23	Speaker 23	40
24	Speaker 24	98
25	Speaker 25	72
26	Speaker 26	95
Average	89.98	

e accuracy of recognition is equated by using the following equation:

Accuracy= (Number of correct recognition/Total number of sample) x 100 (3) [1]

Table 4 contains the accuracy of each Speaker identified by the system and the average accuracy. The average accuracy is calculated by summing the individual accuracy and then dividing it by the total number of speakers [26]. Other metrics like precision, recall, and F1 score was also evaluated to assure the consistency of the result. The results of the same are shown in Table 6.

Precision was used to determine the number of times our model was correct in predicting the proper Speaker. The recall is for how many positive labels the model successfully identified out of all the possibilities. F1 score is calculated as a weighted average of recall and precision [26]. From Figure 2, it can be deduced that for speaker 13, we get a perfect precision and recall value, while for the other Speaker's precision range is between [0.67 – 0.99], and the recall range is between [0.40 – 1], which is quite satisfactory. In the present approach, we have used the Speaker's feature vectors (MFCC) directly, without any averaging, in order to retain and exploit the Speaker's unique characteristics.

Table 5: Precision, recall and f1-score value of each Speaker

Speaker	Precision	Recall	F1-Score	Support
Speaker 1	0.99	0.97	0.98	109
Speaker 2	0.97	1.00	0.98	60
Speaker 3	0.93	0.89	0.91	28
Speaker 4	0.94	0.82	0.87	88
Speaker 5	1.00	0.98	0.99	110
Speaker 6	0.95	0.99	0.97	73
Speaker 7	0.89	1.00	0.94	110
Speaker 8	0.90	0.93	0.91	81
Speaker 9	0.94	0.92	0.93	108
Speaker 10	0.92	0.90	0.91	63
Speaker 11	0.87	0.96	0.92	56
Speaker 12	0.67	0.70	0.68	91
Speaker 13	1.00	1.00	1.00	110

Speaker 14	0.98	1.00	0.99	113
Speaker 15	0.60	0.56	0.58	72
Speaker 16	1.00	0.95	0.98	110
Speaker 17	1.00	0.91	0.95	86
Speaker 18	0.94	0.93	0.94	88
Speaker 19	0.93	0.98	0.95	81
Speaker 20	0.98	1.00	0.99	109
Speaker 21	0.91	0.95	0.93	109
Speaker 22	0.91	1.00	0.95	110
Speaker 23	0.80	0.40	0.53	50
Speaker 24	0.75	0.98	0.85	110
Speaker 25	0.89	0.72	0.79	110
Speaker 26	0.98	0.95	0.96	110
Accuracy			0.91	2345
macro avg.	0.91	0.90	0.90	2345
weighted avg.	0.92	0.91	0.91	2345

Figure 2 shows the data as represented in Table 5, in the form of a bar chart for comparison between precision, recall and f1-score of 26 Speakers. It can be seen in the figure that for speakers 12, 15 and 23, the parameters precision, recall and f1-score are low, suggesting a low recognition rate, while for the other speakers, it can be seen that the same parameters are high, suggesting a high recognition rate. Sarma M. and Kandarpa K. S., in their work, Speaker identification model for Assamese language using a neural framework [32], where they applied an ANN based prototype model for vowel-based speaker identification got an average accuracy of 88.9% for 20 speakers. In our work, the model we obtained from the dataset can provide an average recognition rate of about 89% for 26 speakers speaking Assamese vowels.

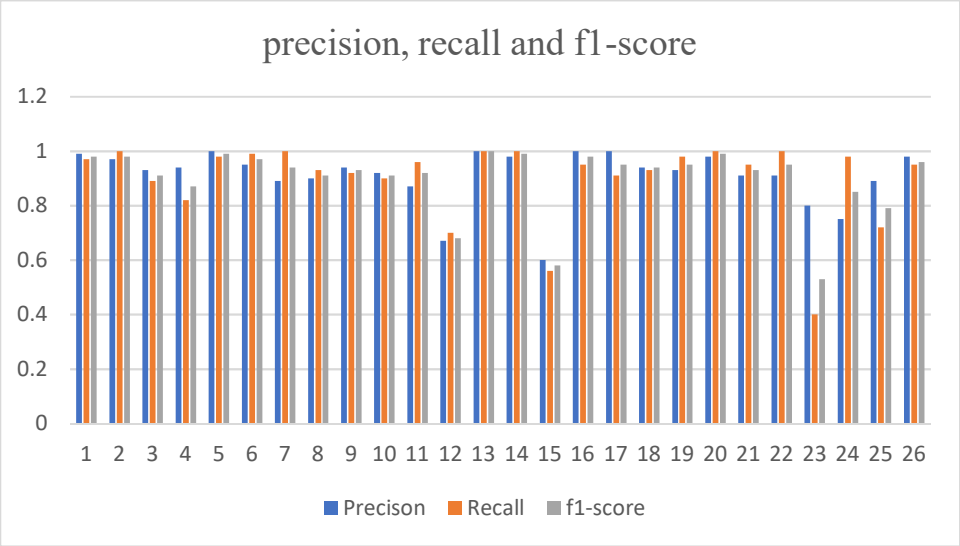


Figure 2: Graph of precision, recall and f1-score of 26 Speakers

6. DISCUSSION AND FUTURE WORKS

Given that MFCC has a higher amount of low-frequency filtering than LFCC and that this spectral area contains more speaker information, MFCC performed best when comparing the outcomes of the two techniques. Therefore, the system was tested with the help of the MFCC and Cosine Similarity index. The model showed a satisfactory accuracy rate of 91% for the speaker identification system in the Assamese language. One of the reasons that the system is fast is that it can identify a speaker directly from its feature vector (MFCC), so instead of having a training phase, the samples in the test set are compared directly with the samples in the training set for identifying the Speaker. This method can be applied for creating models for different other languages. The same

approach can be used for Speaker recognition for voice-enabled operating machines. The accuracy of the system can be enhanced with a dataset of more number of speakers. Various work has been done on Speaker identification and Speaker verification using cosine similarity [27][28]. In the current work it has been observed that in few cases due to the acoustic similarity of speakers' vowel pronunciation some speakers are falsely recognized. From the present dataset, we found that most of the testing samples of speaker-15 are identified as speaker-12 due to their acoustic similarity of pronunciation of Assamese vowels. Apart from that the model in most of the cases is accurate in identifying the exact Speaker as well as the male and female speakers. These observations can be helpful in designing a speaker identification system for the Assamese language as well as other North-Eastern languages, which are similar in terms of the phone. The noise in the voice samples has to be addressed for the system to be more robust, which leads to our future work in this field.

The novelty of the current work is that cosine similarity and parallel processing approaches have not yet been applied to speaker identification in the Assamese language. There are a few works where the authors have employed neural networks with MFCC to do automatic speaker recognition in the Assamese language, but there is no other literature where ASR has been done in the Assamese language using the cosine similarity method. Additionally, we have employed the Speaker's feature vectors (MFCC) directly in the current technique without any averaging in order to preserve the Speaker's distinguishing characteristics. The accuracy of the suggested strategy was 97%, 94%, and 91% for 10, 20, and 26 speakers, respectively.

ACKNOWLEDGEMENT

The authors would like to acknowledge all the Speakers who gave their valuable time to record all the Assamese vowels.

REFERENCES

[1] Maurya, Ankur, Divya Kumar, and R. K. Agarwal. "Speaker recognition for Hindi speech signal using MFCC-GMM approach." *Procedia Computer Science* 125 (2018): 880-887.

[2] Devi, Kharibam Jilenkumari, Ayekpam Alice Devi, and Khelchandra Thongam. "Automatic Speaker Recognition using MFCC and Artificial Neural Network." *International Journal of Innovative Technology and Exploring Engineering* 9 (2019): 39-42.

[3] Bharti, Roma, and Priyanka Bansal. "Real time speaker recognition system using MFCC and vector quantization technique." *International Journal of Computer Applications* 117.1 (2015).

[4] Bansal, Priyanka, Syed Akhtar Imam, and Roma Bharti. "Speaker recognition using MFCC, shifted MFCC with vector quantization and fuzzy." *2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI)*. IEEE, 2015.

[5] Mansour, Asma, and Zied Lachiri. "SVM based emotional speaker recognition using MFCC-SDC features." *International Journal of Advanced Computer Science and Applications* 8.4 (2017).

[6] Chelali, Fatma Zohra, and Amar Djeradi. "Text dependant speaker recognition using MFCC, LPC and DWT." *International Journal of Speech Technology* 20.3 (2017): 725-740.

[7] BHATT, SHOBHA, ANURAG JAIN, and Amita Dev. "Monophone-based connected word Hindi speech recognition improvement." *Sādhanā* 46.2 (2021): 1-17.

[8] Hourri, Soufiane, and Jamal Kharroubi. "A novel scoring method based on distance calculation for similarity measurement in text-independent speaker verification." *Procedia computer science* 148 (2019): 256-265.

[9] Nasr, Marwa A., et al. "Speaker identification based on normalized pitch frequency and Mel Frequency Cepstral Coefficients." *International Journal of Speech Technology* 21.4 (2018): 941-951.

[10] Lei, Yun, et al. "A novel scheme for speaker recognition using a phonetically-aware deep neural network." *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2014.

[11] L. Ferrer, Y. Lei, M. McLaren and N. Scheffer, "Study of Senone-Based Deep Neural Network Approaches for Spoken Language Recognition," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 1, pp. 105-116, Jan. 2016, doi: 10.1109/TASLP.2015.2496226.

[12] Garcia-Romero, Daniel, et al. "Improving speaker recognition performance in the

domain adaptation challenge using deep neural networks." *2014 IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 2014.

[13] P. Matějka et al., "Analysis of DNN approaches to speaker identification," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 5100-5104, doi: 10.1109/ICASSP.2016.7472649.

[14] Dimitri Palaz, Mathew Magimai.-Doss, and Ronan Collobert, "Analysis of CNN-based speech recognition system using raw speech as input," *in Proc. of Interspeech*, 2015

[15] H. Muckenhirn, M. Magimai.-Doss and S. Marcell, "Towards Directly Modeling Raw Speech Signal for Speaker Verification Using CNNS,",*IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 4884-4888, doi: 10.1109/ICASSP.2018.8462165.

[16] A. M. Jalil, F. S. Hasan and H. A. Alabbasi, "Speaker identification using convolutional neural network for clean and noisy speech samples," *First International Conference of Computer and Applied Sciences (CAS)*, 2019, pp. 57-62, doi: 10.1109/CAS47993.2019.9075461.

[17] M. M. Hossain, B. Ahmed and M. Asrafi, "A real time speaker identification using artificial neural network," *10th international conference on computer and information technology*, 2007, pp. 1-5, doi: 10.1109/ICCITECHN.2007.4579414.

[18] Pawar, R. V., P. P. Kajave, and S. N. Mali. "Speaker Identification using Neural Networks." *Iec (prague)*. 2005.

[19] Wikipedia contributors. "Assamese language." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 24 Jan. 2022. Web. 16 Feb. 2022

[20] Wikipedia contributors. "Assamese people." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 12 Feb. 2022. Web. 16 Feb. 2022.

[21] A. Heidarian and M. J. Dinneen, "A Hybrid Geometric Approach for Measuring Similarity Level Among Documents and Document Clustering," *2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService)*, 2016, pp. 142-151, doi: 10.1109/BigDataService.2016.14.

[22] Pandit, Shraddha, and Suchita Gupta. "A comparative study on distance measuring approaches for clustering." *International journal of research in computer science* 2.1 (2011): 29-31.

[23] Medhi, Dr-Bhargab & Talukdar, Prof.. "Assamese Speaker Recognition Using Artificial Neural Network", *IJARCCE.* 321-324.10.17148/IJARCCE.2015.4377.

[24] Barua, Hemchandra, Hema Kosha, January 1, 1900

[25] Liu, Jung-Chun, et al. "An MFCC-based text-independent speaker identification system for access control." *Concurrency and Computation: Practice and Experience* 30.2 (2018): e4255.

[26] Nammous, Mohammad K., Khalid Saeed, and Paweł Kobojek. "Using a small amount of text-independent speech data for a BiLSTM large-scale speaker identification approach." *Journal of King Saud University-Computer and Information Sciences* (2020).

[27] Ahmad, Waquar, Harish Karnick, and Rajesh M. Hegde. "Cosine distance metric learning for speaker verification using large margin nearest neighbor method." *Pacific Rim Conference on Multimedia*. Springer, Cham, 2014.

[28] George, Kuruvachan K., et al. "Analysis of cosine distance features for speaker verification." *Pattern Recognition Letters* 112 (2018): 285-289.

[29] Dutta, Munmi, et al. "Closed-set text-independent speaker identification system using multiple ANN classifiers." *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014*. Springer, Cham, 2015.

[30] Chuangsuwanich, Ekapol. "Multilingual techniques for low resource automatic speech recognition". Massachusetts Institute of Technology Cambridge United States, 2016.

[31] https://support.ircam.fr/docs/audiosculpt/3.0/co/sampling.html, Sunday, 25 September 2022, Greenwich Mean Time (GMT), 25/09/2022

[32] Sarma, Mousmita, and Kandarpa Kumar Sarma. "Speaker identification model for Assamese language using a neural framework." *The 2013 international joint conference on neural networks (IJCNN)*. IEEE, 2013