International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-10, August 2019

Automatic Vowel Recognition from Assamese Spoken words

P Sarma, M Mitra, M P Bhuyan, V Deka, S Sarmah, S K Sarma

Abstract*: Vowel plays the most important role in any speech processing work. In this research work, recognition of Assamese vowel from spoken Assamese words is explored. Assamese is a language which is spoken by major people in Brahmaputra Valley of Assam, Assam is a state which is situated in the North-East part of India. This automatic vowel recognition system is implemented by using three efficient techniques Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Random Forest (RF) classifier. The database used in the experiments is specially designed for this purpose. A list of phonetically vowel rich Assamese words is prepared for the experiment. As an initial effort, twenty different (20) words uttered by fifty-five (55) speakers are taken. Utterances from both male and female speakers are collected. Each utterance was repeated two times by every speaker. A database of the total of 2200 samples is prepared. After experimenting on different samples it is seen that Random Forest (RF) is giving the best performance compared to the other two classifiers. The performance of the system is shown with testing dataset and comparison is done. Outcome of this research work will enhance the Machine Translation from Assamese to any other language.

Keywords: LPC, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest (RF).

I. INTRODUCTION

A. Assamese Language

The root of Assamese language is Sanskrit and it is an Indo-Aryan language. Assamese is spoken by millions of people in the world, basically, all the eight states of North East India. There are eight oral vowel phonemes, two semivowels and twenty-one consonants in Assamese alphabet. There are three broad clusters of Assamese language; they are Indic, Dardic, and Iranian. Indic is generally known as Indo-Aryan. There are several different dialects in the Assamese language.

This variation is seen mostly in different geographic areas of Assam. *Table -I* shows the Assamese vowels with their way of pronunciations and International Phonetic Alphabet *(IPA)* symbols [1]. In this work, speech samples are collected which are uttered in standard Assamese dialect and these samples are adopted for official work. According to eminent linguistics of Assam, Dr. B.K. Kakati, Assamese is a composite language, consisting of many words from Indo-Aryan and Indo Chinese origin [2]. A few Assamese vowels have both primary and secondary appearance. The secondary appearance of a vowel is a symbol; it comes into sight with a consonant in a word. It may possible to get some other vowels which are different in the script but pronunciation is same as the other. For these types of vowels, 100% correct detection becomes crucial. In the *IPA* chart, these vowels are not included as they pronounced as the mother vowel.

B. Vowels and Consonants

Vowels and consonants together form a word. A sentence is formed by the meaningful and correct chunk of more than one word. In most of the languages in the world, the word is formed with consonants and vowels. Depending on how different articulatory parts react, speech is classified into voiced, unvoiced and silence. A spoken language becomes transparent only because of voiced signals. Voiced sounds are produced by periodic vibration of vocal cords. At some specific frequencies, resonances are produced at the vocal tract. Vowel sounds are periodic in nature and they carry more energy than consonant utterance. The vocal cord does not vibrate when consonants are pronounced, so they are unvoiced. When there is no speech, it is called silence. As shown in *Table-I*, according to the position of tongue hump at the time of pronunciation, vowels are classified into the front, central and back. On the other hand, consonants are stated in terms of stroking point of the tongue within the mouth, so named as labial, veler, glottal, etc.

Manuscript published on 30 August 2019.
*Correspondence Author(s)
P Sarma*, Department of Information Technology, Gauhati University Guwahati, India.
S Mitra, Department of Information Technology, Gauhati University Guwahati, India.
M P Bhuyan*, Department of Information Technology, Gauhati University Guwahati, India.
V Deka, Department of Information Technology, Gauhati University Guwahati, India.
S Sarmah, Department of Information Technology, Gauhati University Guwahati, India.
S K Sarma, Department of Information Technology, Gauhati University Guwahati, India.

Table- I: Assamese vowel phoneme' pronunciation with IPA symbols

Tongue Position		Front		Central		Back	
Shape of Lips		Unrounded		Neutral		Rounded	
Height of the Tongue ↓	Space in the oral cavity ↓	IPA	Assamese Vowel Phoneme	IPA	Assamese Vowel Phoneme	IPA	Assamese Vowel Phoneme
High	Close	/i/	ই			/u/	উ
High-Mid	Half Close	/e/	এ'			/o/	ও
Low - Mid	Half Open	/ɛ/	এ			/ɔ/	অ'
Low	Open			/a/	আ	/ɒ/	অ

C. Format Frequencies

Speech features are some frequency components of the speech signal. They are represented in three-dimensional coordinates. Air produced at lungs comes out of the mouth through the vocal tract. The vocal tract acts as a filter (resonator) and curbs the insignificant signals. It also amplifies some other major signals. These amplified signals are called formants or resonance frequency. Formant values of speech signal carry most of the acoustic characteristics of the utterance. Formally they can be named as the peak of the spectrum of speech signal [3]. First (*F1*), second (*F2*) and third (*F3*) formant frequencies are considered most important for any phonemic and acoustic analysis of human speech. The reason for taking only the first three formant values is because they can symbolize the geometry of the human vocal tract easily. The three-dimensional representation of a speech signal is called a spectrogram. Time and frequency are plotted on X-axis and Y-axis respectively. The relatively dark portions as shown in the lower part of *Fig. 1*, represents intensity in a different frequency range. These are the formats of the speech signal. Amplitude noticed on the positive (+) side of X-axis represents air pressure compression, on the X-axis means normal pressure and amplitude on (-) side of X-axis means rarefaction pressure [4]. *Fig. 1* shows the spectrogram hence the formants of Assamese word /ai/ **(আই)**.

D. Objectives

In this research work we are performing below mentioned steps to extract vowels form Assamese utterance using *KNN*, *SVM* and *RF* then compared the findings. The research work is divided as following:

i) Recording of the selected words by a sophisticated system at .wav format with the help of Audacity software.
ii) Using LPC method, estimation of first two formants of those recorded voice samples is done.
iii) Build up the automatic Vowel Recognizers system using the three techniques *SVM*, *RF* and *KNN*.
iv) The classification will be tested with test dataset, result and performance will be analysed.

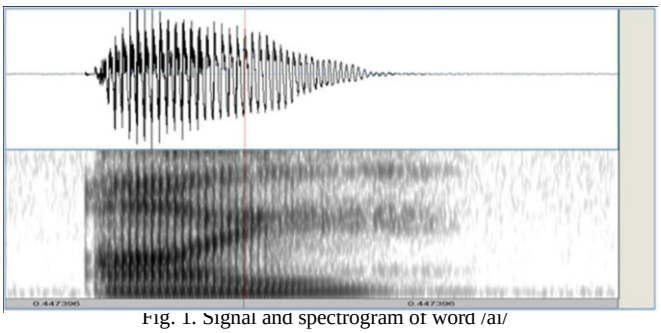

Fig. 1. Signal and spectrogram of word /ai/

Rest of the paper is organized as follows: *Section II* explains the related works in this research, *Section-III* describes the proposed methodology, *Section-IV* describes the Working and Implementation, *Section-V* describes and analyzes the results, and *Section-VI* concludes the present work and proposes the future work.

II. RELATED WORKS

To build an automatic vowel recognition system from spoken words, for the background study, lots of research papers are surveyed and a few of them are mentioned which are used for the present proposed model building.

In 2012, V. Anantha Natarajan and S. Jothilakshmi were able to show segmented speech from a nonstop chain of speech. Then using *SVM* classifier and formant frequency they were able to classify vowel and consonants. They used speech samples from the television broadcast [5].

Mousmita Sarma and Kandarpa Kumar Sarma in 2013 used a new approach for speaker identification using vowel which was segmented from isolated words. The approached used by them were *SOM* (Self-Organizing Map) and *PNN* (Probabilistic Neural Network) [6].

In the year 2013, Bhaskarjyoti sarma and his team showed a way to analyze eight standard Assamese vowels using formant recognition and normalization. They have found and mentioned in the article that *F1* and *F2* can distinguish two vowels using an average approach. This system's recognition rate was 84%-96% [7].

In 2015, Kandarpa Kumar Sarma and Mridusmita Sharma talked about *RNN* (Recurrent Neural Network) and *KNN* (K-Nearest Neighbors) based algorithms to detect vowel sounds using feature vector.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

They also tried vowel identification from different dialects of Assamese language using the same Acoustic Phonetic Features as a feature vector. Their approach shows that *KNN* based algorithm gives better recognition rate than *ANN*-based method [8]. In the year 2017, Ahmed H and Esraa H. Abdul Ameer carried out a broad survey on the five categorization algorithms such as Decision tree, *SVM*, *KNN*, *Naive Bayes* and *HMM*. They have also described the improvements in the algorithms and the features found by different researchers from time to time [9]. In February 2019, D. Dutta, R.D.Choudhury and S.Gogoi wrote a paper about the recent improvement on speech unit detection from already available speech classifier and speech features (from speech corpora). They compared the speech parameters found in different speech corpus. The parameters they used for comparisons are the number of speakers, age and gender, recording environment, etc. Techniques for signal gaining as well as pre-processing were also addressed in their work [10].

III. METHODOLOGY

The proposed model is designed according to the plan and collected the initial speech samples which were recorded in a sophisticated speech recording studio. The samples of recoded voice are collected from people which are in the age group of 18 to 30 years; since this group has proper dimension and shape of their vocal tract. The formant frequency of every vowel phoneme is different. Only first and second formant frequencies of the samples are considered because it can carry enough information for this research work. There are 20 phonetically rich (vowels) of Assamese words are selected which are used to build the database. These twenty segments were uttered by fifty-five male and female speakers. Each word was uttered for two times by every speaker. As a whole, we were able to get a total of 2200 good samples for the experiment in the proposed model. *Fig. 2* shows the block diagram of the proposed model. With the help of audacity tool, all the samples are recorded in a noise-free environment in 44.1 Kz with 16-bit rate mono channel. During this sample acquiring process, the error occurs due to injected noise and silences. These unwanted signals degrade the performance of the recognition process. Preprocessing is done to remove those elements. Cool edit pro, audacity is used for filtering, coding and voice editing. Audacity is mainly used to remove all unwanted signals (noise and silence) from the samples.

A. LPC for Formant Extraction

This method is most effective for low bit rate speech encoding and analysis. Linear Predictive Coding (*LPC*) is the most accurate formant estimation method when it operates on a raw speech signal initially pre-emphasizing is done on the signal. It is a 1st order high pass filter and able to eliminate consequences of any abrupt changes of a continuous signal.

In the next phase, framing is done on the speech signal. The duration of the units is generally (10-20) ms. In this work, we are taking 20ms as frame length and 10 ms as overlapping length. Windowing is performed on the frames to flat the input signal as it makes computation easy. A hamming window function is multiplied with the framed signal to get a better spectral analysis. The basic idea behind *LPC* is that every speech signal can be represented as linear combinations of the preceding ones. The following equation is called as a linear predictor and so-called as linear predictive coding.

$$\hat{s}(n) = \sum_{k=1}^{p} (\alpha_k s(n-k)) + Gu(n) \quad (1)$$

The α_k part for $k= 1,2,....p$ are constants (co-efficient) throughout the whole analysis frame. $u(n)$ filters excitation, G is the gain. So using the above equation $\hat{s}(n)$ is guessed. *LPC* is an efficient way of speech feature extraction. One advantage is that it reduces bit rates of the signal. So less bandwidth works properly, several users can be increased significantly. We have built *SVM*, *KNN*, and *RF* classifier models and trained every one of them with 1800 formant (*F1*, *F2*) feature values taken from the training dataset. Based on this training all three models were able to predict 200 unknown formant

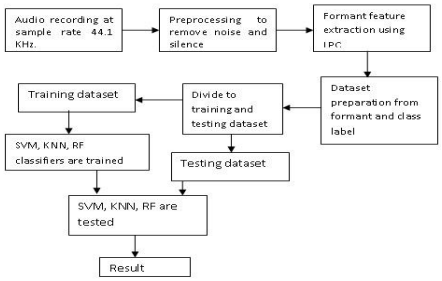

Fig. 2. Block diagram of the proposed system

some unknown samples are tested on these three classifier models to find out the recognition rate of Assamese vowels from word utterances.

B. Classifier

After formant features were extracted, the *SVM*, *RF* and *KNN* classifiers are trained. The training was done to recognize vowels depending on formant estimation values. Short descriptions of the three classifiers are given below.

1) Random Forest (RF)

Random forest is a tree-structured algorithm which builds several decision trees. This algorithm works by choosing *n* features from *m* numbers of total available features [11]. It gives a majority class constructed from several building trees. The pseudo-codes of *RF* are as follows:

- Random selection of *n* features out of *m* feature and $n< m$.
- Use the best split to select the root node and construct different decision trees based on *n_estimator* parameter value.
- Ending in each case is predicted for every decision tree.
- For each target of every tree, the vote is planned.
- The maximum vote for any target is measured as the final prediction.

In the present experiment 10 is taken as *n_estimators*.

2) K Nearest Neighbor (KNN) Classifier

Initially, this algorithm was used for pattern classification. This is a supervised algorithm and basically used for classification and regression purpose [12].

The algorithm is good and not complex to implement but suffers from some drawbacks.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The algorithm exploits the whole training data individually and thus predicts the class of test data set. The pseudo-code of *KNN* classifier is:

- *K* is determined; it represents the total numbers of nearest neighbors.
- Distance is measured between the training features and chosen test features.
- Sorting is done on a distance matrix. Depending on k^{th} minimal distance, the nearest neighbor is calculated.
- Y category nearest neighbors are collected.
- The maximum number of targets, available in the nearest selected neighbors is taken as an objective of test features.

In the implementation, the value of *K* is selected as 5 and Euclidian distance is taken as distance type.

3) Support Vector Machine (SVM)

Support vector machine is a supervised algorithm which is used as a binary classifier. It can be used for regression also. Application of this algorithm is basically seen at classification problems. It works under the principle of creation of a hyper-plane that separates data into different classes [13]. The algorithm takes a set of data as input and outputs a line separating those data into different classes.

At first attempt, *SVM* tries to find out a line or hyper-plane between the two classes of data. According to the algorithm, points from both the classes closest to the separator line are found out. They are named as support vectors. Next distance between the support vectors and the line are calculated. This is the distance called margin. Aim of the *SVM* mechanism is to maximize this margin. For the nonlinear arrangement of the given dataset *SVM* uses nonlinear kernel function. Steps are given below: 1. Start

2. Dataset input
3. Dataset classification
4. Four kernel functions Linear, Polynomials, Sigmoid and Radial Based are applied to the dataset.
5. Specify the hyper plane
6. The accuracy obtained, if not satisfactory go to step 4
7. End

In this research work, Radial based function Kernel is used.

IV. WORKING ON THE FRAMEWORK/IMPLEMENTATION

A. Sample Collection

For training and testing total twenty (20) words including eight (8), Assamese vowels were recorded from the equal number of male and female speakers respectively. For recording the speech signal, a PC headset and sound recording software audacity are used. The sampling frequency is 44100 Hz and the channel is Mono. The recorded voice samples for male and female speakers were saved as *.wav* file with the help of audacity software.

B. Formant feature extraction

After pre-processing formant features were extracted the by using the *LPC* method. *Fig. 3* shows a screenshot giving the formant feature for vowel 'অ'

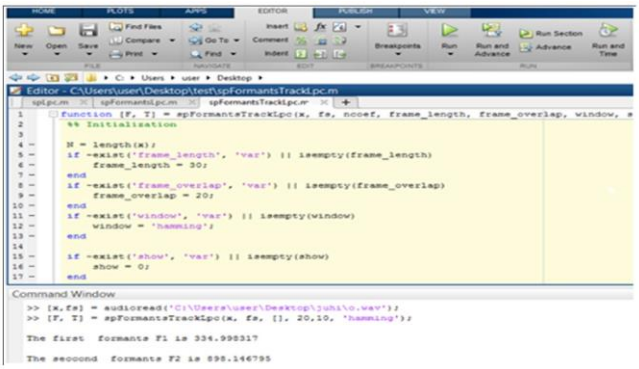

Fig.3. Screenshot of formant feature of 'অ' voice sample

C. Training and testing the dataset

After determining the formant features of all selected *.wav* files, train and test database are prepared. *Fig. 4 (a and b)* train data set, *Fig. 5 (a and b)* test data sets are shown.

	A	B	C
1	First Formant	Second Formant	Class
2	455	917.1	অ
3	544.26	1043.2	অ
4	514.78	1036.54	অ
5	457.35	507.25	অ
6	555.33	1042.24	অ
7	534.98	1096	অ
8	571.86	944.03	অ
9	480.2	859.66	অ
10	433.06	903.37	অ
11	456.42	849.83	অ
12	380.01	708.04	অ
13	412.36	622.43	অ
14	553.96	1102.98	অ
15	564.72	1031.61	অ
16	507.86	922.82	অ
17	548.56	976.04	অ
18	498.42	795.43	অ
19	494.95	865.48	অ
20	505.26	802.65	অ
21	593.7	768.99	অ
22	701.88	1102.08	অ
23	656.42	938.02	অ

Fig. 4. (a) Screenshot of train dataset

	A	B	C
1979	234	1105.39	ৰাঙ - অ,ঙ
1980	245.27	1075.44	ৰাঙ - অ,ঙ
1981	247.2	1155.34	ৰাঙ - অ,ঙ
1982	254.9	666.37	ৰাঙ - অ,ঙ
1983	245.79	1175.77	ৰাঙ - অ,ঙ
1984	235.25	1404.96	ৰাঙ - অ,ঙ
1985	282.95	1153.22	ৰাঙ - অ,ঙ
1986	269.2	1152.2	ৰাঙ - অ,ঙ
1987	288.67	1563.03	ৰাঙ - অ,ঙ
1988	251.59	1054.42	ৰাঙ - অ,ঙ
1989	259.72	1049.46	ৰাঙ - অ,ঙ
1990	271.11	1070.81	ৰাঙ - অ,ঙ
1991	280.7	1143.45	ৰাঙ - অ,ঙ
1992	283.15	578.21	ৰাঙ - অ,ঙ
1993	272.11	820.98	ৰাঙ - অ,ঙ
1994	304.08	748.06	ৰাঙ - অ,ঙ
1995	272.11	820.98	ৰাঙ - অ,ঙ
1996	304.08	748.06	ৰাঙ - অ,ঙ
1997	267.81	1130.75	ৰাঙ - অ,ঙ
1998	257.69	701.05	ৰাঙ - অ,ঙ
1999	247.31	1030.9	ৰাঙ - অ,ঙ
2000	269.82	812.97	ৰাঙ - অ,ঙ
2001	269.64	771.69	ৰাঙ - অ,ঙ

Fig. 4. (b) Screenshot of train dataset

Retrieval Number: J13010881019/19©BEIESP
DOI: 10.35940/ijitee.J1301.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

	A	B	C
1	First Formant	Second Formant	Class
2	427.14	1017.1	অ
3	335.91	952.59	অ
4	557.35	907.25	অ
5	565.33	942.24	অ
6	560.33	1062.03	অ
7	472.87	588.86	অ
8	703.75	875	অ
9	485.48	806.28	অ
10	786.94	1016.57	অ
11	399.36	838.91	অ
12	551.2	995.35	অ'
13	494.75	1073.54	অ'
14	508.88	913.01	অ'
15	526.18	975.86	অ'
16	558.82	1051.04	অ'
17	543	733.15	অ'
18	443.27	932.93	অ'
19	552.19	915.36	অ'
20	630.08	997.48	অ'
21	526.87	1036.62	অ'
22	336.42	1368.16	আ
23	823.76	1431.03	আ
24	813.59	1552.49	আ

Fig. 5. (a) Screenshot of test dataset

	A	B	C
179	265.65	1180.51	মা - আ
180	297.47	1054.73	মা - আ
181	286.68	1008.05	মা - আ
182	236.84	1107.86	গা -আ
183	206.13	1198.2	গা -আ
184	250.65	1841.29	গা -আ
185	231.24	1032.34	গা -আ
186	245.95	1485.76	গা -আ
187	287.7	1711.82	গা -আ
188	220.08	1217.71	গা -আ
189	309.36	1500.35	গা -আ
190	368.77	881.76	গা -আ
191	479.87	1220.79	গা -আ
192	270.99	1309.49	নাও - আ,ও
193	150.25	666.05	নাও - আ,ও
194	195.32	1135.21	নাও - আ,ও
195	243.44	1058.48	নাও - আ,ও
196	266.16	1529.36	নাও - আ,ও
197	274.77	1117.27	নাও - আ,ও
198	262.22	641.22	নাও - আ,ও
199	289.09	1025.07	নাও - আ,ও
200	255.23	1418.17	নাও - আ,ও
201	254.43	1427.69	নাও - আ,ও

Fig. 5. (b) Screenshot of test dataset

D. Results of three classifier

KNN and *SVM* generate better result in case of image and text classification. This research work is based on speech signal and we observed better performance in *RF* classifier for all the selected samples.

The model is tested with test dataset to find out the classifications. *Fig. 6 (a)* shows the result of the *KNN* classifier, from *Fig.6 (a)*, we can see the various values of *precision*, *recall*, *F1-score* and *support vector* for all the twenty samples, *Fig. 6(b)* is the confusion matrix for the above said *KNN* classifier. Similarly, the result of *SVM* and *RF* classifiers are shown in *Fig. 7 (a and b)* and *Fig. 8 (a and b)* respectively. *Fig. 9* shows the screenshot of a few test results of the three classifiers.

```
>>> predictions = KNN.predict(X_test)
>>> accuracy = accuracy_score(Y_test,predictions)
>>> print("The accuracy of KNN is:",accuracy*100)
The accuracy of KNN is: 66.0
>>> print(classification_report(Y_test, predictions))
              precision    recall  f1-score   support

           অ       0.53      0.80      0.64        10
          অ'       1.00      0.60      0.75        10
    অলপ - অ       0.89      0.80      0.84        10
          আ       0.75      0.90      0.82        10
    আঠ - আ       0.88      0.70      0.78        10
    আম - আ       1.00      0.60      0.75        10
          ই       0.47      0.90      0.62        10
          উ       0.47      0.70      0.56        10
    উট - উ       0.62      0.80      0.70        10
          এ       0.46      0.60      0.52        10
         এ'       0.67      0.40      0.50        10
    এক - এ       0.88      0.70      0.78        10
          ও       0.83      0.50      0.62        10
     গা -আ       0.50      0.80      0.62        10
    জাম - আ       0.60      0.90      0.72        10
  নাও - আ,ও       0.44      0.40      0.42        10
    মই - ই       1.00      0.60      0.75        10
    মা - আ       0.75      0.60      0.67        10
ল'বা - অ',আ       0.83      0.50      0.62        10
    সোপ - ও       1.00      0.40      0.57        10

   micro avg       0.66      0.66      0.66       200
   macro avg       0.73      0.66      0.66       200
weighted avg       0.73      0.66      0.66       200
```

Fig. 6. (a) Screenshot of the KNN classifier

```
>>> print(confusion_matrix(Y_test, predictions))
[[8 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0]
 [3 6 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0]
 [2 0 8 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 9 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 2 7 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0]
 [1 0 0 1 0 6 0 1 0 1 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 9 0 0 1 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 1 7 0 1 0 0 0 1 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 1 1 8 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 1 1 0 6 0 0 0 0 0 1 0 0 1 0]
 [0 0 0 0 0 0 0 1 1 1 3 4 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 1 0 0 0 2 7 0 0 0 0 0 0 0 0]
 [0 0 0 0 1 0 0 2 2 0 0 0 5 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 1 0 0 0 0 0 8 0 1 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 9 0 0 1 0 0]
 [0 0 0 0 0 0 2 0 0 0 0 0 0 0 4 4 0 0 0 0]
 [0 0 0 0 0 0 0 0 1 0 0 0 0 1 1 1 6 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 3 1 0 0 6 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 3 0 1 0 1 5 0]
 [1 0 1 0 0 0 1 1 1 0 0 1 0 0 0 0 0 0 0 4]]
```

Fig. 6. (b) Screenshot of the KNN classifier confusion matrix

```
>>> predictions = SVM.predict(X_test)
>>> accuracy = accuracy_score(Y_test,predictions)
>>> print("The accuracy of SVM is:",accuracy*100)
The accuracy of SVM is: 82.0
>>> print(classification_report(Y_test, predictions))
              precision    recall  f1-score   support

           অ       1.00      0.80      0.89        10
          অ'       0.90      0.90      0.90        10
    অলপ - অ       0.89      0.80      0.84        10
          আ       0.90      0.90      0.90        10
    আঠ - আ       1.00      0.80      0.89        10
    আম - আ       0.89      0.80      0.84        10
          ই       0.73      0.80      0.76        10
          উ       0.73      0.80      0.76        10
    উট - উ       0.89      0.80      0.84        10
          এ       0.78      0.70      0.74        10
         এ'       0.69      0.90      0.78        10
    এক - এ       0.82      0.90      0.86        10
          ও       0.89      0.80      0.84        10
     গা -আ       0.80      0.80      0.80        10
    জাম - আ       0.67      0.80      0.73        10
  নাও - আ,ও       0.75      0.90      0.82        10
    মই - ই       0.80      0.80      0.80        10
    মা - আ       0.89      0.80      0.84        10
ল'বা - অ',আ       0.64      0.70      0.67        10
    সোপ - ও       1.00      0.90      0.95        10

   micro avg       0.82      0.82      0.82       200
   macro avg       0.83      0.82      0.82       200
weighted avg       0.83      0.82      0.82       200
```

Fig. 7. (a) Screenshot of SVM classifier

```
>>> print(confusion_matrix(Y_test, predictions))
[[8 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0]
 [0 9 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
 [0 1 8 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0]
 [0 0 0 9 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
 [0 0 0 1 8 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0]
 [0 0 0 0 0 8 0 1 0 0 1 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 8 0 0 0 1 0 0 0 0 1 0 0 0 0]
 [0 0 1 0 0 0 0 8 0 0 0 0 0 0 1 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 8 0 2 0 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 7 0 0 0 0 0 0 1 0 2 0]
 [0 0 0 0 0 0 0 0 0 0 9 1 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 1 0 9 0 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 1 1 0 0 0 0 8 0 0 0 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 8 1 1 0 0 0 0]
 [0 0 0 0 0 0 1 0 0 0 0 0 0 0 8 0 0 1 0 0]
 [0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 9 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 8 0 0 0 0]
 [0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 8 0 0 0]
 [0 0 0 0 0 0 0 1 0 0 0 0 0 2 0 0 0 0 7 0]
 [0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 9]]
```

Fig. 7. (b) Screenshot of SVM classifier confusion matrix

```
>>> predictions = RF.predict(X_test)
>>> accuracy = accuracy_score(Y_test,predictions)
>>> print("The accuracy of RF is:",accuracy*100)
The accuracy of RF is: 86.5
>>> print(classification_report(Y_test, predictions))
              precision    recall  f1-score   support

           অ       1.00      0.80      0.89        10
          অ'       1.00      0.90      0.95        10
    অলপ - অ       0.91      1.00      0.95        10
          আ       0.90      0.90      0.90        10
    আঠ - আ       0.89      0.80      0.84        10
    আম - আ       0.89      0.80      0.84        10
          ই       0.64      0.90      0.75        10
          উ       0.73      0.80      0.76        10
    উট - উ       0.89      0.80      0.84        10
          এ       0.75      0.90      0.82        10
         এ'       0.89      0.80      0.84        10
    এক - এ       0.91      1.00      0.95        10
          ও       0.90      0.90      0.90        10
     গা -আ       0.80      0.80      0.80        10
    জাম - আ       0.82      0.90      0.86        10
  নাও - আ,ও       0.82      0.90      0.86        10
    মই - ই       0.82      0.90      0.86        10
    মা - আ       1.00      0.80      0.89        10
ল'বা - অ',আ       1.00      0.80      0.89        10
    সোপ - ও       1.00      0.90      0.95        10

   micro avg       0.86      0.86      0.86       200
   macro avg       0.88      0.86      0.87       200
weighted avg       0.88      0.86      0.87       200
```

Fig. 8. (a) Screenshot of RF classifier

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

```
>>> print(confusion_matrix(Y_test, predictions))
[[ 8  0  0  0  0  0  1  0  0  0  0  0  1  0  0  0  0  0  0  0]
 [ 0  9  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0 10  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  9  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  1  8  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  1  8  0  0  0  1  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  9  0  0  0  0  0  0  1  0  0  0  0  0  0]
 [ 0  0  1  0  0  0  0  8  0  0  0  0  0  0  1  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  2  8  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  1  0  0  9  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  1  8  1  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0 10  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  1  0  0  0  9  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  8  0  1  1  0  0]
 [ 0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  9  0  0  0  0]
 [ 0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  9  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  9  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  1  8  0  0]
 [ 0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  1  0  0  0  8]
 [ 0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  9]]
```

Fig. 8 (b) Screenshot of RF classifier confusion matrix

V. RESULT AND ANALYSIS

After training the models are tested by 200 formant feature values of selected samples where each sample has 10 formant features. Then, the recognition rate per sample and Mean recognition rate of three classifiers are obtained as depicted in *Table -II*.

```
File  Edit  Format  Run  Options  Window  Help
array(['আ'], dtype=object)
>>> predictions = KNN.predict([ [1015.95,1497.19]])
>>> predictions
array(['আম - আ'], dtype=object)
>>> predictions = KNN.predict([ [818.67,1322.06]])
>>> predictions
array(['আই - আ'], dtype=object)
>>>
>>> predictions = KNN.predict([ [ 454.37,1877.5]])
>>> predictions
array(['এক - এ'], dtype=object)
>>> predictions = KNN.predict([ [ 341.02,1164.63]])
>>> predictions
array(['ই'], dtype=object)
>>> predictions = KNN.predict([ [ 366.71,653.86]])
>>> predictions
array(["অ'"], dtype=object)
>>> predictions = SVM.predict([ [ 237.82,1596.34]])
>>> PREDICTIONS

>>> predictions
array(['জাম - আ'], dtype=object)
>>> predictions = SVM.predict([ [ 412.88,1808.32]])
>>> predictions
array(["অ'ৰা - অ',আ"], dtype=object)
>>> predictions = SVM.predict([ [ 810.87,1072.58]])
>>> predictions
array(['সোণ - ও'], dtype=object)
>>> predictions = SVM.predict([ [ 248.46,1160.89]])
>>> predictions
array(['মই - ই'], dtype=object)
>>> predictions = RF.predict([ [247.62,1526.49]])
>>> predictions
array(['মা - আ'], dtype=object)
>>> predictions = RF.predict([ [ 360.29,1425.55]])
>>> predictions
array(['গা -আ'], dtype=object)
>>> predictions = RF.predict([ [ 241.01,1604.42]])
>>> predictions
array(['নাও - আ,ও'], dtype=object)
```

Fig. 9. Screenshot of testing result of three classifiers

From *Table-II*, it is observed that the recognition rate of *KNN* classifier is less or equal to *RF* classifier in every sample. The recognition rate of *KNN* classifier is also less or equal to *SVM* classifier in every sample except the 'ই'(/i/) and 'জাম' sample shown in *Sl. No. 4.* and *15.* /i/ is a vowel and at the time of its utterance tongue height becomes high, oral cavity space changes to close, lip shape becomes unrounded and tongue goes to the front. a/ is a vowel and at the time of its utterance tongue height becomes low, oral cavity space changes to open, lip shape becomes neutral and tongue goes to the central. The above mentioned unusual case may be due to pronunciation uniqueness of the vowels /i/ and /a/.From *Table-II* it is observed that *SVM* classifier gives equal or less vowel recognition rate than *RF* classifier in most of the cases. But exceptions are noticed at *Sl. no. 6* and *12*. For এ' (/e/) and এ(/ɛ/) samples, *SVM* recognition rates are more than *RF* classifier. /e/ and /ɛ/ are vowels with utterance specification of high-mid tongue height, half-closed oral cavity space, unrounded lip shape and tongue goes to the front. *Fig. 10* shows a bar graph of recognition rate for three classifiers *KNN*, *SVM*, and *RF*. The X-axis represents the classifiers and Y-axis represents the accuracy percentage. This *Fig. 10* is created from classifier's mean values from *Table-II*. It is noticed that *KNN* classifier provides a low recognition rate than the other two classifiers in most of the cases. It may happen for the following possibilities:

- The Euclidean distance which is chosen at tuning time is not so appropriate for the dataset that is why classification is not 100% correct.
- There may have low variance or high correlation feature spaces in the dataset itself.
- The value of *K* which is taken at tuning time probably is not perfect for this experimental dataset.

SVM classifiers provide low mean recognition rate than *RF* classifiers because whatever the kernel function is chosen at the time of tuning for this dataset, it might not transfer all the low dimensional data instances into higher dimensional data instances. This is the observation that the specified hyper- plane could not separate all the classes.

Table-II: Recognition rates of three classifiers

SL No	Input voice Sample	Predicted Vowels (Manually)	IPA label	Recognition Rate Of KNN	Recognition Rate Of SVM	Recognition Rate Of RF
1	অ	অ	/ɒ/	80	75	80
2	অ'	অ'	/ɔ/	60	90	90
3	আ	আ	/a/	90	90	90
4	ই	ই	/i/	90	80	90
5	এ	এ	/Ɛ/	60	90	90
6	এ '	এ'	/e/	40	90	80
7	উ	উ	/u/	70	80	80
8	ও	ও	/o/	50	80	90
9	অলপ	অ	/ɒ/	80	80	100
10	আম	আ	/a/	60	80	80
11	আঠ	আ	/a/	70	80	80
12	এক	এ	/Ɛ/	70	80	90
13	উট	উ	/u/	80	80	80
14	ল'ৰা	অ', আ	/ɔ/,/a/	50	70	80
15	জাম	আ	/a/	90	80	90
16	সোন	ও	/o/	40	90	90
17	মই	ই	/i/	60	80	90
18	মা	আ	/a/	60	80	80
19	গা	আ	/a/	80	80	80
20	নাও	আ,ও	/a/,/o/	40	85	90
Over all Mean				66	82	86.5

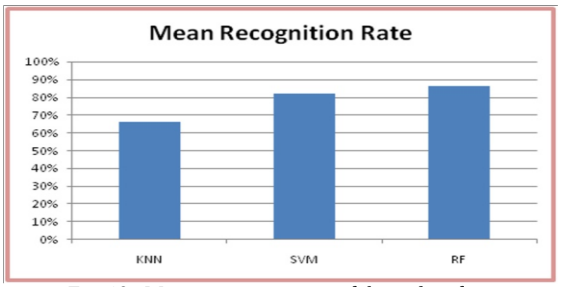

Fig. 10. Mean recognition rate of three classifiers

At last, it can be concluded that Random Forest (*RF*) provides better recognition rate than other two classifiers in the experiment which is shown in *Fig. 10*. But, it could provide a much better result than the present result, if all the decision trees are highly uncorrelated.

VI. CONCLUSION AND FUTURE WORK

This approach of identifying vowels from spoken Assamese word is an original research work of the authors. In this research work, three classifier techniques are used to recognize vowels correctly. It is seen that Random Forest (*RF*) was able to show a good performance with 86.5% accuracy. In most of the cases, *SVM* was also able to show good recognition rate but as a whole *RF* shows very good performance almost for all the samples. This is because RF uses multiple decision trees and it can give the most correct result using the voting outcome of the multiple trees. As seen from *Table-II* vowel recognition rate for all the three techniques are not very competent which is below 90% of the mean value. One obvious reason is due to overlapped *F1* and *F2* values for some of the samples. This partly covered values of *F1* and *F2* pilots to a confusion situation of the recognizer. Unwanted injection of noise is also one reason for not getting excellent result. Pronunciation inclination is also observed among the phonemes. Some high energetic and raising pitch influences the pronunciation of its neighbor phonemes. The way how a speaker pronounces a word also influence recognition of phonemes. For example, in *Table-II* sample number 20 has two consecutive vowel utterances with a different way of pronunciation, /a/ and /o/. There are some scopes to improve these drawbacks. Hence, in the future, we shall try to collect more standard data and to find out the exact boundary for every selected sample to improve the recognition rate of the system.

REFERENCES

1. M.Devi and P.H.Talukdar (2015) "A Study on Estimation of Formants of Assamese Vowel and Words using LPC Method", *Journal of Harmonized Reserch*, 3(3) : 116 – 121
2. Banikanta Kakati "Assamese its formation and development" 5th ed (2007) *Guwahati, India, LBS publication*
3. G.K.Vallabha and B. Tuller "Systematic errors in the formant analysis of steady- state vowels*."* (2002) *Speech Communication*, 38(1-2) : 141- 160
4. J. N. Holmes, W. J. Holmes, and P. N. Garner "Using formant frequencies in speech recognition" (1997) *Proc. EUROSPEECH 4* :2083-2086
5. V.A. Natarajan and S. Jothilakshmi (2012) "Segmentation of continuous speech into consonant and vowel units using formant frequencies**",** *International Journal of Computer Applications*, 56(15) : 0974 – 8887
6. M.Sarma and K.K. Sarma (2013) "Vowel phoneme segmentation for speaker identification using an ANN based framework**",** *Journal of Intelligent System*, 22(2):111 – 130
7. B.Sarma, (2013) "Normalization and automatic recognition of Assamese vowels," *International Journal of Computer Applications*, 75(11) : 0975 – 8887
8. K.K.Sarma and M.Sharma, (2015) "Dialectal Assamese vowel speech detection using acoustic phonetic features, KNN and RNN*",2nd International Conference on Signal Processing and Integrated Networks , Noida, India*, IEEE : 674-678
9. A.H. Aliw and E. H. A. Ameer," (2017) Comparative study of five text classification algorithms with their improvements ", *International Journal of Applied Engineering Research* 12 (14) : 4309-4319.
10. D. Dutta, R.D.Choudhury and S.Gogoi,(2019) "Speech Corpora, Feature Extraction Techniques and Classifiers with Special Reference to Automatic Speech Recognition", *International Journal of Computer Sciences and Engineering*, 7(2).

11. W.Koehrsen,(2017) "Random Forest Simple Explanation",Medium, Dec. 27, 2017 [Online].Available:https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d
12. M.Bazmara, K Nearest Neighbor Algorithm for Finding Soccer Talent, Researchgate, April 2013 .[online]. Available:https://www.researchgate.net/publication/237080861_K_Nearest_Neighbor_Algorithm_for_Finding_Soccer_Talent retrieved on 13th May, 2019
13. "Support vector machine", *Wikipedia.* Accessed May 18, 2019.[Online].Available: https://en.wikipedia.org/wiki/Support-vector_machine.

AUTHORS PROFILE

P Sarma is currently working as an Assistant professor in the department of Information Technology, Gauhati University, Guwahati, India. She did BE in Computer Science & Engineering, MTech in Information Technology, PhD in Speech Processing. Her research areas are Speech Processing, Machine Learning, Natural Language Processing, Image Processing. Email: parismita.sarma@gmail.com

M Mitra is a MTech student in the department of Information Technology, Gauhati University. Her research area is Speech Processing.

M P Bhuyan is a research scholar in the department of Information Technology, Gauhati University, Guwahati, India. He did BE in Computer Science & Engineering and MTech in Information Technology. His research areas are Artificial Intelligence, Natural Language Processing, Parallel Computing and Machine Learning. Email: mpratim250@gmail.com

V Deka is currently working as an Assistant professor in the department of Information Technology, Gauhati University, Guwahati, India. He is the HoD of the department of Information Technology. He did MSc in Computer Science, MTech in Information Technology and PhD in IoT. His research areas are Machine Learning, IoT, Information Theory, Natural Language Processing.

S Sarmah is currently working as an Assistant professor in the department of Information Technology, Gauhati University, Guwahati, India. He did MSc in Computer Science, MTech in Information Technology, and PhD in Computer Networks. His research areas are Computer Networks, Cryptography, Machine Learning.

S K Sarma is currently working as a professor in the department of Information Technology, Gauhati University, Guwahati, India. He is a computer scientist and published various research papers in renowned conferences and journals. His research areas are Natural Language Processing, Artificial Intelligence, Machine Learning, Computer Networks, etc. He has guided many PhD students.