Speech Recognition of Isolated Words using a New Speech Database in Sylheti

Gautam Chakraborty, Navajit Saikia

Abstract*: With the advancements in the field of artificial intelligence, speech recognition based applications are becoming more and more popular in the recent years. Researchers working in many areas including linguistics, engineering, psychology, etc. have been trying to address various aspects relating to speech recognition in different natural languages around the globe. Although many interactive speech applications in "well-resourced" major languages are being developed, uses of these applications are still limited due to language barrier. Hence, researchers have also been concentrating to design speech recognition system in various under-resourced languages. Sylheti is one of such under-resourced languages primarily spoken in the Sylhet division of Bangladesh and also spoken in the southern part of Assam, India. This paper has two contributions: i) it presents a new speech database of isolated words for the Sylheti language, and ii) it presents speech recognition systems for the Sylheti language to recognize isolated Sylheti words by applying two variants of neural network classifiers. The performances of these recognition systems are evaluated with the proposed database and the observations are presented.

Keywords*: Automatic Speech Recognition, Mel Frequency Cepstral Coefficient, Sylheti, Under-resourced Language, Feed-forward neural network, Recurrent Neural Network.

I. INTRODUCTION

Speech is a primary mode of communication among humans. Each uttered word in a language contains linguistic contents (vowel and consonant speech segments) specific to the language. With the advances in machine or artificial intelligence, it has become more pertinent to use voice for man-machine interaction. Even with a small vocabulary of isolated words, speech recognition is used in mobile telephony, interactive television, support systems for differently abled people, robotics, etc. Automatic Speech Recognition (ASR) involves conversion of a given speech signal into a machine readable format and then transforming it into desired outputs which can be used in applications for practical purposes [1]. As a pattern recognition problem, a speech recognition system compares a given test pattern with the training pattern of each speech classes for classification. ASR systems, starting from isolated digit recognition to continuous speech recognition, have evolved significantly in various languages.

Speech recognition deals with isolated words, connected words or continuous speech depending on the requirements of applications using the speech databases varying from small vocabulary to large vocabulary [1],[2],[3],[17],[30],[38]. An ASR system may also be speaker dependent or speaker independent[2],[56]. Technologies like document preparation or retrieval, command and control, automated customer service, etc. use speaker independent speech recognizers. Speaker dependent systems are used in the applications like interactive voice response system, computer game control, etc. Factors that affect the performance of an ASR system include type of speech, dependency of speaker, vocabulary size, age variation, etc.[1]. A generic ASR system ideally consists of three stages [5]:

i) *Signal Pre-processing* involves the extraction of voiced parts from a speech signal through a series of signal analysis. It derives the voiced parts in digital form by exercising an input speech signal through analog-to-digital conversion, pre-emphasis filtering followed by windowing.

ii) *Feature Extraction* computes features from each voiced part in the pre-processed signal. Some popular features used in ASR systems are Linear Predictive Coding (LPC) coefficients [5],[6], Mel Frequency Cepstral Coefficients (MFCC) [4],[6],[7],[8],[10], short-time energy [6], i-vector [11], etc. The mel-frequency scale in MFCC coefficients being proportional to the logarithm of the linear frequency below 1 kHz, it closely reflects the human perception; and hence, MFCC features are mostly used in ASR systems.

iii) *Classification* is the process of mapping the feature vector of an input word into 1 out of N word classes of the considered vocabulary during testing. Some popularly used classifiers in ASR are Artificial Neural Network (ANN) [5],[10],[12],[13], Hidden Markov model (HMM) [14],[15], Dynamic Time Warping (DTW) [16],[17], Deep Neural Network [9],[47],[51], etc. The application of ANN in designing ASR system is still being used by researchers [5],[6],[19],[20],[21],[22],[23],[36],[40],[42] despite the developments in the field of deep neural network (DNN) in recent times. ANN has been a popular choice because of its following characteristics [24]:

Manuscript published on 30 September 2019
* Correspondence Author

Gautam Chakraborty*, Department of Electronics & Telecommunication Engineering, Assam Engineering College, Guwahati, India. Email: gauchak2012@gmail.com.

Navajit Saikia, Department of Electronics & Telecommunication Engineering, Assam Engineering College, Guwahati, India. Email: navajit.ete@aec.ac.in.

- Non linearity: Neural network has the ability to learn and model the non-linear and complex relationships between the inputs and outputs to perform tasks like pattern recognition and classification,

multi-dimensional data processing, etc.

- Robustness: Because of the inherent parallel structure, neural network can continue to work even if any element of the network fails.
- Adaptability: Neural network has the self learning capability, and hence does not require reprogramming in a dynamic environment.

Many speech interactive applications are already being used which facilitates major languages or "well-resourced languages" like English, Chinese, French, German, Russian, Hindi, etc. However, language barrier seen in human interaction demands for ASR systems in "under-resourced languages" [25],[29]. An under-resourced language is one which has some shortfalls such as lack of writing system, lack of linguistic study, limited or unavailability of electronic speech resources, etc.[25],[26]. A comprehensive list of 7,111 languages in the world is available in [27] which includes both the well- and under- resourced languages. Researchers in recent years have reported speech recognition solutions for some of the under-resourced languages [28],[29],[58],[59].

The language Sylheti is an under-resourced language with limited linguistic study [31],[32],[33],[41], few printed and electronic literatures [34], limited linguistic expertise, etc. There is also no record of any speech and language technology on it. Sylheti belongs to the Indo-Aryan language family [32] and more than ten million people speak Sylheti globally. It is spoken largely in the Sylhet Division of Bangladesh and also spoken in the northern part of Tripura and the Barak Valley region of Assam, India. Sylheti is written in its unique Sylheti Nagari script and the alphabets are presented in [34]. Sylheti has a total of 32 alphabets comprising of 5 vowels and 27 consonants. A phonetic study in the Sylheti language has been carried out only recently in [32] which introduces the phonemes present in Sylheti first time. Some characteristics such as distinctive way of pronunciation, de-aspiration and deaffrication, etc. of this language are also worth mentioning here [32].

From the perspective of this work, a brief introduction to ANN is presented here. ANN as a machine learning model replicates the human brain which uses a large collection of nonlinear information processing elements (called artificial neurons or nodes or units) and it is organized in three layers: one input layer, one or more hidden layers and one output layer [24]. There may be one or many hidden layers depending on the network architecture (single layer or multi-layer). The net input to a neuron is equal to the weighted sum of the outputs of the neurons feeding it. For example, consider that the neuron Y receives inputs from three neurons X1, X2 and X3 through communication links having weights w_1, w_2 and w_3 respectively as shown in Figure 1. If x_1, x_2 and x_3 are the respective outputs of X1, X2 and X3, the net input to Y is:

$$y_{in} = w_1x_1 + w_2x_2 + w_3x_3 \qquad (1)$$

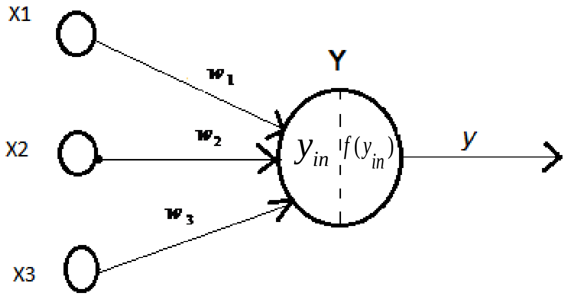


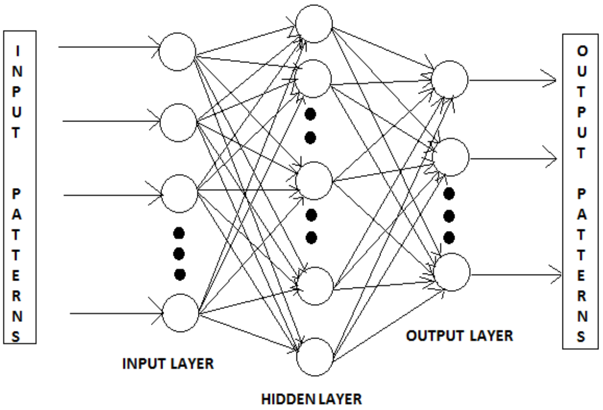
Figure 1. Model of an artificial neuron

The weight associated with a communication link measures the quantity of knowledge of the local network formed by the two neurons. The output y of neuron Y depends on an activation function $f(.)$ according to:

$$y = f(y_{in}) \qquad (2)$$

The activation function used in neural network is generally non-linear which can be chosen from sigmoid function, hyperbolic tangent function, etc.[5]. This output y may be connected to one or more neurons in next layer. During training, weights associated with the communication link between two neurons are adjusted to resolve the differences between the actual and predicted outcomes for subsequent forward passes in the network. It is to be noted that in ANN classifier, the number of nodes in input and output layers match the number of input features and output classes respectively.

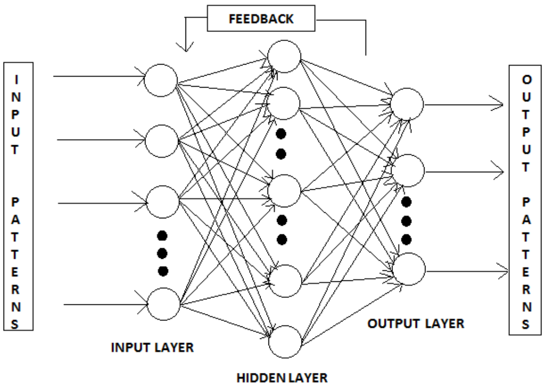
Based on the interconnections of neurons, ANN may be broadly categorised into two types: feed-forward and feedback or recurrent [24] neural network which are shown in Figure 2 (with one input, one hidden and one output layers). While a feed-forward neural network (FFNN) propagates data from input to output through the hidden nodes, a recurrent neural network (RNN) uses feedback by using an internal state memory as shown in Figure 2(b). Due to the use of feedback, RNN has the ability to deal with time-varying dynamic inputs. To be noted that the stability of an ANN depends on the number of neurons in the hidden layer(s).



(a) Architecture of FFNN

Retrieval Number: C5874098319/2019©BEIESP
DOI:10.35940/ijrte.C5874.098319
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication



(b) Architecture of RNN
Figure 2. Types of ANN

The remaining part of the paper is organized as follows: Section 2 presents a review of literatures on isolated word recognition in different languages. A new speech database for the Sylheti language is introduced in Section 3. Section 4 presents the proposed ASR system for isolated word in Sylheti. Experimental results are presented in Section 5 followed by discussions. Section 6 concludes this work.

II. LITERATURE REVIEW

Speech recognition has been a subject of interest for researchers for decades leading to the developments of speech interfaces for desktop and handheld devices. The first ASR system designed by researchers at Bell Laboratories in 1952 [35] could recognize ten digits (0-9) spoken by one speaker. In subsequent times, ASR systems for isolated words, connected words and continuous speeches in various languages are reported by employing different acoustic features and classification models [6],[18],[36],[37],[39],[44]. Although people used hidden Markov model in ASR during 1980s and 1990s, technology transition has been witnessed towards ANN significantly due to certain advantages as discussed in Section 1 and also towards DNN in recent times. As the present work confines to the recognition of isolated words in the language Sylheti, the literature review in the following concentrates on available ASR systems for isolated words in major and under-resourced languages.

A neural network based approach with MFCC, LPC and short-time energy features to recognize isolated English digits is presented by B. P. Das and R. Parekh [6]. The authors report an overall recognition accuracy of 85% with their own database. In [8], N. Seman et al. present a model to recognize isolated Malay words in live video and audio recorded speeches. The authors have used MFCC features and the multi-layer neural network to derive an average classification rate of 84.73%. A recognition system for Chinese digits is proposed in [49] by using MFCC features and neural network classifier where an average recognition rate of 97.4% is reported. I. Kipyatkova and A. Karpov introduce an RNN based ASR system in Russian language with word error rate (WER) of 14% [50]. The speech recognition models proposed in [48] for Arabic digits and words use two variants of neural network- multi-layer perceptron and Long Short-Term Memory(LSTM). For both the models, the authors have reported recognition rates of around 95% in case of digits as well as words. To recognize isolated Turkish digits, three ASR systems are proposed in [4] by applying different types of neural networks. Taking MFCC features of each digit, these systems present recognition accuracies in the range from 98.12% to 100%. In [5], authors present three ASR systems to recognise isolated digits in Assamese by using LPC features. These systems use FFNN, RNN and Cooperative Heterogenous ANN (CHANN). An ASR framework for recognizing isolated Bangla words is reported in [7] which employs MFCC features. The authors use a semantic time delay neural network in the work and report a recognition accuracy of 82.66%. J.T. Geiger et al. [39] have applied RNN in a hybrid NN-HMM system architecture considering the medium-vocabulary 2nd CHiMEM speech corpus in experiments. P. Sarma and A. Garg propose an ASR system in [40] to recognize Hindi words with a neural network classifier. MFCC and PLP features used here and an average recognition accuracy of 79% is reported. In [13], Marathi isolated words are considered to build an ASR system using neural network classifier which also presents an overall classification rate. Another ASR model for Malayalam isolated words is presented in [42] which uses ANN classifier and a combined feature set comprising of MFCC, energy and zero crossing. The authors have reported a recognition accuracy of 96.4%. M. Oprea and D. Schiopu [12] propose an ASR system to recognize Romanian isolated words by using neural network classifier. S. Furui presents a speaker-independent ASR system which is used to recognise names of Japanese cities [44]. With a vocabulary of 100 city names, the system has a recognition rate of 97.6%. M. K. Luka et al. [10] use neural network classifier to design an ASR system for Hausa language by using MFCC features. In [59], the authors introduce a neural network based ASR system for Gujarati isolated words by using two acoustic features MFCC and real cepstral coefficients [59] and report the comparison results in terms of average classification rates. In recent years, researchers are employing DNN in speech recognition. Authors in [45] use a DNN classifier with MFCC features for English digit recognition. When tested with the database of English digits constructed by Texus Instrument, an average recognition accuracy of 86.06% is reported in this work. By considering German speech data, [43] presents an ASR system by employing convolutional neural network which derives a WER of 58.36% and a letter error rate of 22.78%. Authors in [46] use LSTM RNN to construct an ASR system which is tested on two speech databases- the Augmented Multi-party Interaction (AMI) corpus and Corpus of Spontaneous Japanese Interaction (CJI). Another ASR system proposed in [47] considers three African languages Swahili, Amharic and Dinka which employs DNN classifier.

From the above discussion, it is observed that researchers are using ANN and its variants in recent times also to design ASR systems in some major languages [6],[48],[49],[50] due to their attractive characteristics as discussed in Section 1. It is to be also noted that ANN is being popularly used for digit and isolated word recognition in under-resourced languages [4],[5],[7],[8],[10].

These ANN based ASR systems are reported to deliver good recognition rates. As discussed in Section 1, Sylheti is an unexplored and under-resourced language from both linguistic as well as technological points of view. Speech recognition for Sylheti has not been considered yet except an ASR system reported for isolated digits pronounced in Sylheti from 0 to 9 [36] as a part of our initiative. Further to state here that there is no speech database available in the Sylheti language in electronic form for applications in speech or speaker recognition. Considering the above observations, we concentrate here on two aspects as follows:

- To construct a new speech database of small vocabulary for isolated Sylheti words. In doing so, the possible future use of the database in ASR environment can also to be considered. The Sylheti words (except the digits) are to be chosen based on phonetic studies made in [31],[32] such that the words are phonetically rich.
- To design ASR systems for isolated Sylheti words by using FFNN and RNN types of neural networks.

The following section presents the proposed Sylheti speech database for isolated words.

III. CONSTRUCTION OF NEW SYLHETI SPEECH DATABASE

A speech database (or corpus) is a collection of utterances for a particular language and it is an important resource for building a speech recognizer. The samples in such database are used for training and testing of an ASR system. In constructing a speech database, phonetic/linguistic level discussion in the language is found to be relevant. As an under-resourced language, a study in Sylheti is also carried out here based on its phonetic/linguistic characteristics and accordingly a brief comparison on phonemic structure of Sylheti language with two major languages English and Standard Colloquial Bangla (SCB) is presented below. Thereafter, the work on constructing a new database in Sylheti is discussed. Although the ASR systems which are to be proposed here do not involve phoneme recognition, it is aimed in the following to construct a standard speech database in Sylheti considering the phonetically rich words.

Each speech utterance is represented by some finite set of symbols known as phonemes which describe the basic units of speech sound [52]. Phonemic status of a sound is not same across languages. Moreover, the number of phonemes in one language varies from another language. The phoneme inventory of Sylheti presented in [32] shows that Sylheti has some specific phonemes which are not present in SCB or in major language like English. This phonetic study also presents a significant reduction and reconstruction compared to that of SCB. Also, Sylheti language has the nature of obstruent weakening which employs de-aspiration, spirantization and deaffrication. Altogether Sylheti has 22 phonemes as shown in Table 1, out of which 5 are vowel and 17 are consonant [32]. On the other hand, SCB has 37 phonemes, aggregated from 7 vowel and 30 consonant phonemes [52]. Five vowel phonemes /i/, /e/, /a/, /u/ and /ə/ in Sylheti are common to Bangla. The two other vowel phonemes of Bangla, /o/ and /æ/, are merged with the vowel phonemes /u/ and /e/ respectively in Sylheti due to restructuring in articulation [32]. Again, out of the 17 consonant phonemes in Sylheti, 13 phonemes are also present in SCB [52] and they are /b/, /t/, /g/, /m/, /n/, /ʃ/, /s/, /h/, /r/, /l/, /ŋ/, /t̪/, /d̪/. The other 4 consonant phonemes /z/, /x/, /ɖ/ and /Φ/ are specific in Sylheti language. The 17 consonant phonemes /p/, /p^h/, /b^h/, /t^h/, /d/, /d^h/, / t̪^{h}/, / d̪^{h}/, /c/, /c^h/, /k/, /k^h/, /g^h/, /w/, /j/, /ɟ/, / ɟ^h/ available in SCB are not present in Sylheti.

Table 1 : Sylheti Phonemes

Vowel phonemes	**Consonant phonemes**
/i/, /e/, /a/, /u/, /**ɔ**/	/b/, /t/, /g/**,** /m/, /n/, /ʃ/, /s/, /h/, /r/, /l/, /ŋ/, /t̪/, /d̪/, /z/, /x/, /ɖ/, /Φ/

When Sylheti is compared with English, it is observed that the English language has a total of 12 vowel phonemes [52]. All the 5 vowel phonemes in Sylheti are also present in English. Therefore, the remaining 7 vowel phonemes in English (/ə/, /æ/, /ɪ/, /ɒ/, /ɜ/, /ʌ/, /ʊ/) are specific to the language. On the other hand, 12 consonant phonemes /b/, /t/, /g/, /m/, /n/, /ʃ/, /s/, /h/, /r/, /l/, /ŋ/, /z/ in Sylheti [32] are also present in English [52]. Hence, Sylheti has 5 language specific consonant phonemes /t̪/, /d̪/, /x/, /ɖ/ and /Φ/, when compared with English. The English language has 12 specific consonant phonemes which are /p/, /d/, /ɵ/, /k/, /f/, /v/, /Ʒ/, /tʃ/, /dƷ/, /w/, /j/, /ð/. Therefore, there is enough scope to study the Sylheti language from the linguistic point of view and [32],[41] may be consulted for detail in this regard. This also entails that Sylheti can be considered to be studied from the technological viewpoints. This phonetic study will facilitate a database in transcribed form which may be used in studying the phoneme-based speech recognition and speaker recognition problems in Sylheti. The construction of a Sylheti speech database of small vocabulary considering of isolated words is presented in the following.

In constructing this new Sylheti speech database of isolated words, 30 most frequently-used mono syllabic words are considered which are phonetically rich. Out of these words, 10 are the utterances of the digits 0-9 in Sylheti and 20 are other Sylheti words among which few are taken from [32]. Table 2 lists the isolated words in Sylheti for the proposed database and it shows the meaning in English of each word and the phonemes present (in bold letters).

Table 2: Isolated Sylheti words in the proposed database. The phonemes present in the words are shown in bold letters.

Sylheti word	*Meaning*	*Sylheti word*	*Meaning*
[s**ui**njɔ]	Zero	[**e**x]	One
[d̪**ui**]	Two	[t̪i**n**]	Three
[s**a**ir]	Four	[**Φ**as**]**	Five
[s**ɔ**y]	Six	[**s**at̪]	Seven
[a**ʈ**]	Eight	[nɔy]	Nine
[**d̪**an]	Donate	[**d̪**an]	Paddy
[**p**ua]	Boy	[**p**uri]	Girl
[**d̪**ud̪]	Milk	[**b**ari]	Home
[**p**ul]	Flower	[**b**ari]	Heavy
[**b**ala]	Good	[**b**ala]	Bracelet
[j**a**mai]	Husband	[**b**ɔu]	Wife
[ba**t̪**]	Boiled Rice	[ba**t̪**]	Arthritis
[ma**t̪**a]	Head	[mu**x**]	Face
[**g**a]	Body	[**g**a]	Wound
[**g**ai]	Stroke	[**g**ai]	Cow

The construction of speech database primarily involves speech acquisition and labeling [53].

The acquisition of speech utterances may be from read out speech or from spontaneous speech [53],[54]. Both of these approaches have their intrinsic advantages and limitations [53]. As the first attempt to construct a Sylheti speech database, the case of read out speech is chosen here. Therefore, speakers are asked to read out the Sylheti words shown in Table 2 and the utterances are captured. In recording the speech utterances, the following hardware and software are used:

- Microphone: iBall Rocky unidirectional microphone (frequency range from 20Hz to 20KHz)
- Laptop: Intel Core i3 processor and 2 GB RAM, manufactured by ASUS
- Operating system: Windows 7
- Voice recording software: PRAAT (version praat5367_win64)

Further the following parameters are set up during recording:

- Sampling rate: 16 KHz
- Channel mode: Mono
- Environment: Noise-free closed room environment
- Encoding format: WAV with 16 bit PCM
- Distance of microphone from speaker's mouth: 10-12 cm

This speech database consists of data recorded from 10 native speakers including 8 male speakers and 2 female speakers who are willing to participate to contribute during the construction of this database. These speakers do not have any history of speech disorders. As there is no specific rule about the male-female proportion in construction of speech database, literatures [51],[56],[58] have considered various proportions like 60%-40%, 70%-30%, 65%-35%, etc. The speakers in this work are chosen from Sylheti speaking areas in the Karimganj district of the state of Assam and the Kailasahar and Kumarghat districts of the state of Tripura, India where they have been living since their childhood. The ages of the participating speakers are in the range from 25 to 70 years. Six speakers are in the age group 25 to 45 years and have a graduate degree. The other 4 speakers are in the age group 46 to 70 years and are undergraduates. By choosing the speakers in different age group, the variations in speech characteristics with age are taken care of [55]. Apart from Sylheti, speakers can also speak English and Hindi. All the speakers are asked to utter (read out) each of the 30 Sylheti words in Table 2 for 10 times. The samples are recorded and stored according to: *speakernumber_age_gender_utteredword_utterancenumber .wav*. Thereby, a total of 300 utterances are recorded for each speaker. This exercise derives a speech database containing 3000 speech samples of isolated Sylheti words. Duration of recording for this new Sylheti speech database is approximately 5 hours. In the labeling process [53], the recorded utterances are verified by carefully listening the target words and presence of any irregular noise or quiet segments in the recorded samples are examined. For each recorded utterance, the voiced parts are manually extracted by selecting their beginning and end points and the unwanted silent parts are removed. This is done by using PRAAT software. The labeling exercise is rechecked by another verifier to confirm that only the voiced parts are retained from the recorded utterances in the final database.

The following section presents two ASR systems for recognizing isolated Sylheti words which are taken from the above-presented Sylheti speech database and also for analyzing the performance of the proposed systems.

IV. PROPOSED SPEECH RECOGNITION SYSTEM FOR SYLHETI LANGUAGE

It is observed in Section 2 that the many ASR systems [4],[6],[7],[8],[42],[49] developed in recent times for "well-resourced" as well as "under-resourced" languages use MFCC features and neural network classifiers due to their distinct characteristics as presented in Section 1. In view of the above, an architecture for ASR system which employs MFCC features and ANN classifier as shown in Figure 3 is proposed in this work. We consider to use two different types of ANN classifiers to derive two ASR systems for recognizing isolated words in Sylheti.

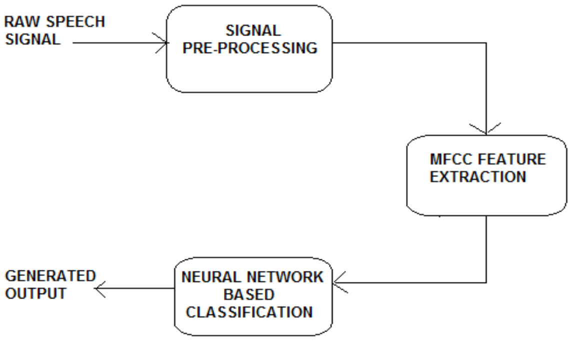


Figure 3. Architecture for ASR system employing MFCC features and ANN classifier

The functions of the each block in Figure 3 are described in the following.

A. Signal Pre-processing

The signal pre-processing involves analog-to-digital (A/D) conversion, end point detection, pre-emphasis filtering and windowing. In A/D conversion, the input speech signal is sampled at 8 KHz and quantized with 16 bits/sample to derive a digital signal. The voiced part of the speech signal is extracted from the digital signal by locating the beginning and end points in the utterance (end point detection). One popular method to extract the voiced part is to compute the zero-crossing rate. Here, the rate at which the speech signal crosses the zero amplitude line by transition from a positive to negative value or vice versa is measured. Voiced part exhibits a low zero-crossing rate. Another method to extract voiced part is short-time signal energy. After extraction of the voiced part, a pre-emphasis filter is used to emphasize the high-frequency components in the voiced part. It helps either to separate the signal from noise or to restore distorted signal. Here, a first-order high-pass finite impulse response (FIR) filter is applied to spectrally flatten the signal. Authors consider the following FIR filter for pre-emphasis [5]:

$$x_p[n] = x_v[n] - 0.95\, x_v[n-1] \qquad (3)$$

where, $x_v[n]$ is the input signal (voiced part) to the pre-emphasis filter and $x_p[n]$ is the output.

Due to time varying nature, speech signal is divided into short segments (of duration ranging from 5 to 100 ms) called *frames* [5]. Frames are assumed to be stationary and speech analysis is carried out on the frames.

In ASR systems, generally overlapping frames are considered with a frame duration in the range from 20ms to 40ms and with an overlapping of 5ms to 15ms [40]. In this proposed system, a frame duration of 32ms with an overlap of 10ms is considered.

The objective of the windowing stage is to minimize the spectral discontinuities at the boundaries of the frame. The windowing operation can be expressed with the following equation [5]:

$$F_w[n] = w[n]F[n]\ ;\ 0 \le n \le N-1 \qquad (4)$$

where, N is the number of frames in a speech sample, $F[n]$ is a frame, $w[n]$ is the window function and $F_w[n]$ is the windowed version of $F[n]$. Researchers usually apply Hamming window in speech analysis [4],[5],[36]. The coefficients of the Hamming window are computed according to:

$$w[n] = 0.54 - 0.46\cos\left[\frac{2\prod n}{N-1}\right]; 0 \le n \le N-1 \qquad (5)$$

This work also uses the Hamming window.

B. MFCC feature extraction

Feature is a set of representative values extracted from an input speech sample that uniquely characterize the sample. Here, windowed version of each frame in a speech signal is considered independently to compute a feature set for the frame. The feature sets of all the frames are then concatenated to derive the features for the input speech signal. In the following, the computation of the feature set for the frame $F[n]$ from its windowed version $F_w[n]$ is presented.

As discussed in Section 1, ASR systems for different languages use MFCC coefficients as features due to their high resemblance with human hearing system [4],[5],[6],[8]. The mel frequency scale is approximately linear up to about 1000Hz in the frequency and well approximates the sensitivity of the human ear. Therefore, the proposed ASR systems for Sylheti language also use a set of MFCC coefficients as the features for a frame. The block diagram for computing the MFCC coefficients at frame level is presented in Figure 4.

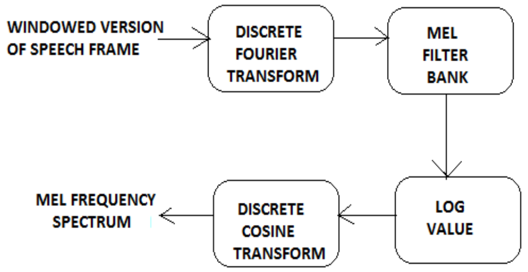

WINDOWED VERSION OF SPEECH FRAME
DISCRETE FOURIER TRANSFORM
MEL FILTER BANK
LOG VALUE
DISCRETE COSINE TRANSFORM
MEL FREQUENCY SPECTRUM

Figure 4. Computation of MFCC coefficients

The first block in MFCC computation finds the discrete Fourier transform (DFT) coefficients from the windowed version of an input speech frame deriving the amplitude spectrum. The DFT coefficients are usually obtained by employing the fast Fourier transform (FFT). The mel filter bank converts the frequency scale to the mel-scale, which is performed according to:

$$f_{mel} = 2595\log_{10}[1 + \frac{f_{linear}}{700}] \qquad (6)$$

where f_{mel} is the mel frequency corresponding to the linear frequency f_{linear}. Finally, log is taken from the output f_{mel} and discrete cosine transform (DCT) is applied to it to obtain the magnitudes of the resulting spectrum [4]. The methodology described above to extract MFCC features from the frames of a speech signal was proposed by Davis and Mermelstein in 1980.

As the first 12 to 13 MFCC coefficients contain maximum information present in a speech frame [40], we here consider the first 13 MFCC coefficients of a frame as features to represent the frame. Let $\{c_{n,i} \mid i = 1,2,\cdots,13\}$ represent the first 13 MFCC coefficients corresponding to the mel frequencies $\{f^{i}_{mel} \mid i = 1,2,\cdots,13\}$ for the n^{th} frame of an utterance. For a mel frequency f^{i}_{mel}, the mean of all the MFCC coefficients derived from the frames of an utterance is computed according to:

$$m_i = (\sum_{n=1}^{N} c_{n,i})/N, \quad i = 1,2,\cdots,13 \qquad (7)$$

The set of mean values $\{m_i \mid i = 1,2,\cdots,13\}$ acts as the features for the utterance.

C. Neural network based classification

As mentioned in Section 2, the present work proposes to employ two variants of ANN as classifiers in the ASR systems. The role of the ANN classifier is to classify an input speech by measuring its similarity with a reference pattern derived through training phase. The proposed ASR systems for isolated Sylheti words use FFNN and RNN separately for classification. Each of the neural networks is designed with the following parameters:

Input, output and hidden layers: Both FFNN and RNN networks are structured with one input layer, one hidden layer and one output layer. The number of neurons in input layer is taken as 13 as the feature set $\{m_i \mid i = 1,2,\cdots,13\}$ is used to represent an utterance. The output layer contains 30 neurons corresponding to the 30 different words to represent 30 different classes. The selection of an appropriate number of neurons in the hidden layer is challenging in the design of neural network. Using too few neurons in hidden layer results in underfitting whereas a large number of hidden neurons may cause overfitting [24],[57]. The number of hidden neurons may be chosen according to the three rule-of-thumb approaches [57]:

- It is between the input and output layer sizes.
- It is smaller than double of the input layer size.
- It is the sum of the output layer size and 2/3 of the input layer size.

However, these rules may not result an optimum hidden layer. Therefore, trial and error approach with backward or forward selections is generally adopted to find the optimum network architecture [57].

In the present work, the number of neurons in hidden layer is decided empirically as discussed above. The observed performances for both the FFNN and RNN networks suggest 46 neurons in the hidden layer. A detail description of these performances is presented in the next section.

Activation function: The non-linear activation (transfer) functions logsigmoid and tansigmoid are used respectively in this study for the output and hidden layers. The basic reasons of using sigmoid function are its smoothness, continuity and positive derivation. The logsigmoid function in the output layer produces the network outputs in the interval [0,1] i.e. output of one class is closer to be 1 once the word is detected and 0 otherwise. Again in tansigmoid function, it's output is zero centered in between -1 to 1 and hence optimization is easier.

Training algorithm: In both the ASR systems, the scaled conjugate gradient back-propagation method is used to train the networks due to its better learning speed [4],[5]. Many other authors have also used this training algorithm due to the above said advantage [6],[12]. As a supervised algorithm, this back-propagation method optimizes the weights of the neurons by using a loss/cost function [5] and produces faster convergence than other methods.

The following section presents the experimental setup and observations of the proposed ASR systems.

V. EXPERIMENTAL RESULT AND ANALYSIS

This work performs two sets of experimentations relating to the above-said two ASR systems for isolated Sylheti words. The first set deals with the FFNN based ASR system and that of second set deals with the RNN based ASR system. The following parameters are considered during experimentations:

1. Features: The set of 13 MFCC-based features $\left\{m_i \middle| i=1,2,\cdots,13\right\}$ for each utterance as presented in Section 4(B).
2. Classifiers: FFNN and RNN types as presented in Section 4(C).
3. Activation functions: tansigmoid for hidden layer and logsigmoid for output layer.
4. Training and testing datasets: The database for Sylheti language presented in Section 3 has a total of 3000 utterances of 30 words, where each word is uttered 10 times by each speaker. Out of the 3000 utterances, 1500 utterances comprising of 50 utterances of each word are considered for training the networks. The other 1500 utterances are used for testing.
5. Convergence: Targeted mean-squared error (MSE) of 0.001 during training.
6. Performance measure: The performances of ASR systems are studied in terms of Percentage recognition rate (%RR), which is computed according to:

$$\%RR = \frac{\text{Number of correct word recognitions}}{\text{Total number of word utterances used in testing}} \times 100\% \qquad (8)$$

As discussed in the previous section, the performances of the proposed ASR systems change when the number of neurons in the hidden layer is varied. In order to decide the optimum number of neurons in hidden layer for the proposed ASR systems, we conducted training and testing of the networks by varying numbers of neurons in the hidden layer according to the three rules-of-thumb mentioned in Section 4(C). Better performances are observed when the size of hidden layer is set at 38 as per the third rule (i.e., the number of neurons is equal to the sum of the output layer size and 2/3 of the input layer size) out of the three rules. However, to achieve superior performances, the trial and error approach is adopted in backward and forward directions by taking hidden layer neurons in the range 36 to 50. Figure 5 presents plots of the observed performances of the systems using the FFNN and RNN networks. It can be observed in the plots that the maximum performance of 84.5% is obtained for the proposed FFNN based ASR system when the hidden layer contains 46 neurons. Similarly, for RNN based system, the hidden layer with 46 neurons derives the best performance of 86.6%. We, therefore, consider 46 neurons in the hidden layers of the proposed systems.

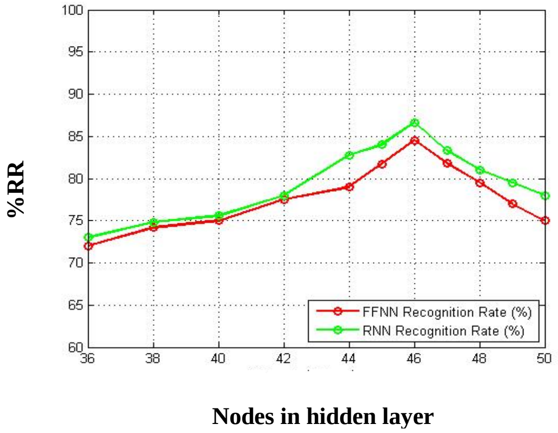


Figure 5. Observed Performance plots with different number of neurons in the hidden layer

A neural network model stops its training when any one of two conditions are met: a) the maximum number of epochs is reached, or b) performance is converged to the goal. In the presented work, the first condition is satisfied. A convergence plot is often generated in the training phase to show the closeness of the network outputs to the target values. It presents the MSE values between the corresponding network outputs and targeted values [5],[36]. Figure 6 presents convergence plots for both the proposed ASR systems in terms of MSE values. It is observed that the convergence of the RNN based ASR system is better than that of the FFNN based system. This is due to the inherent nature of feedback looping in RNN, which tries to adjust the errors of outputs of the neurons during training.

To further examine the robustness and performances of the proposed systems in terms of variations in training and testing samples, different combinations from the available 3000 word utterances of the proposed database are considered for training and testing. The total 3000 utterances are divided into four non-overlapping groups G1, G2, G3, and G4.

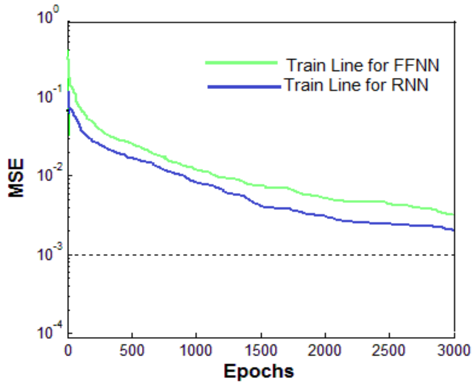


Figure 6. Convergence plots for the proposed ASR systems

In each group, 750 utterances (25 utterances of each of the 30 words) are considered. Out of these four groups, two groups are considered for training and that of other two groups are used for testing. Thereby, a total of ${}^4C_2 = 6$ different training and testing datasets are used. The corresponding observed recognition rates for both the proposed systems are presented in Table 3.

Table 3. Performances of both the ASR systems

ASR system using	Training Dataset	Testing Dataset	%RR	Average %RR
FFNN	G1,G2	G3,G4	83.9	84.55
	G1,G3	G2,G4	84.5	
	G1,G4	G2,G3	85	
	G2,G3	G1,G4	85.8	
	G2,G4	G1,G3	84.6	
	G3,G4	G1,G2	83.5	
RNN	G1,G2	G3,G4	85	86.38
	G1,G3	G2,G4	88.3	
	G1,G4	G2,G3	87	
	G2,G3	G1,G4	86.6	
	G2,G4	G1,G3	86	
	G3,G4	G1,G2	85.4	

It may be observed from the above experimentations that both the proposed systems perform more or less consistently when different training and testing Sylheti datasets are used. This implies good robustness of both the systems to variations in datasets. However, the RNN based ASR system derives better recognition accuracy (average %RR of 86.38) than that of the ASR system using FFNN (average %RR of 84.55). The better performance with the RNN classifier may be due to its inherent feedback characteristics as discussed in Section 1. Due to speech variability in age variation (which affect the performance of any ASR system) in the constructed Sylheti speech database, it is also noticeable the minor deterioration of recognition results in the presented ASR systems. Thus, from the generated results it can be concluded that the observed performances of the ASR systems for Sylheti presented above are comparable to the performances of similar systems available for other languages [6],[7],[8],[40],[50] and hence are considered to be satisfactory.

VI. CONCLUSION

Speech Recognition using neural network has been an area of research interest for long, and many ASR systems have been proposed for different languages around the globe. This paper has considered the "under-resourced" Sylheti language. As no speech database for Sylheti in electronic form is available, a new speech database of isolated Sylheti words has been proposed which can be used by researchers working in the domains of speech processing in Sylheti. This paper has also presented two ASR systems for the Sylheti language to recognize isolated Sylheti words by applying two variants of neural network classifiers, FFNN and RNN. It has been observed that the overall performance of ASR system using the RNN network (recognition rate:86.38%) is better than that of the FFNN based ASR system (84.55%) which is due to the feedback of RNN. One of our future works will concentrate on updating this constructed Sylheti database to include connected words and also to design ASR system for recognizing connected words in Sylheti. Another future work will be to employ DNN in ASR system for Sylheti. Also, the problem of speaker identification will be taken up for the Sylheti language.

REFERENCES

1. C. Kurian, "A Survey on Speech Recognition in Indian Languages", International Journal of Computer Science and Information Technologies, vol. 5, no. 5, 2014, pp. 6169-6175.
2. R. Matarneh, S. Maksymova, V. V. Lyashenko and N. V. Belova, "Speech recognition systems: A comparative Review", IOSR Journal of Computer Engineering, vol. 19, no. 5, 2017, pp. 71-79.
3. S. K.Gaikwad, B.W.Gawali and P. Yannawar, "A Review on Speech Recognition Technique", International Journal of Computer Applications, vol. 10, no.3, Nov. 2010, pp. 16-24.
4. G. Dede and M. H. Sazli, "Speech recognition with artificial neural networks", Elsevier journal of Digital Signal Processing, vol.20, no. 3, May, 2010, pp.763-768.
5. M. Sarma, K. Dutta, and K. K. Sarma, "Assamese Numeral Corpus for speech recognition using Cooperative ANN architecture", International Journal of Computer, Electrical, Automation, Control and Information Engineering vol.3, no.4,2009.
6. B. P. Das and R. Parekh, "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, no. 3, May-June 2012, pp. 854-858.
7. Y.A. Khan, S. M. Mostaq Hossain and M. M. Hoque, "Isolated Bangla word recognition and Speaker detection by Semantic Modular Time Delay Neural Network (MTDNN)", 18th International conference on Computer and Information Technology, Dhaka, Bangladesh , 21-23 Dec. 2015.
8. N. Seman, Z. A. Bakar and N. A. Bakar, "Measuring the performance of Isolated Spoken Malay Speech Recognition using Multi-layer Neural Network", International Conference on Science and social Research(CSSR 2010), Kualalumpur, Malaysia, December, 2010.
9. A. Mohammed, G. E. Dahl, and G. Hinton, "Acoustic Modeling using Deep Belief Networks", IEEE transactions on Audio, Speech and Language Processing, vol.20, no.1, January 2012, pp. 14-22.
10. M. K. Luka, I. A. Frank and G. Onwodi, "Neural Network Based Hausa Language Speech Recognition", International Journal of Advanced Research in Artificial Intelligence, vol. 1, no. 2, 2012, pp. 39-44..
11. A. Kanagasundaram, "Speaker Verification using I-vector Features", a PhD thesis of Queensland University of Technology, 2014.

12. M. Oprea and D. Schiopu, "An artificial neural network-based isolated word speech recognition system for the Romanian language", 16th International conference on System Theory, Control and Computing (ICSTCC), 12-14 Oct., Sinaia, Romania, 2012.
13. K. R. Ghule and R. R. Deshmukh, "Automatic Speech Recognition of Marathi isolated words using Neural Network", International Journal of Computer Science and Information Technologies, vol.6(5), 2015, pp. 4296-4298.
14. M. K. Sarma, A. Gajurel, A. Pokhrel and B. Joshi, "HMM based isolated word Nepali speech recognition", Proceedings of International conference on Machine learning and Cybernetics, Ningbo, China,2017.
15. S. S. Bharali and S. K. Kalita, "A comparative study of different features for isolated spoken word recognition using HMM with reference to Assamese language", International Journal of Speech Technology, Springer ,vol. 18, no. 4, 2015, pp. 673–684.
16. S. Xihao and Y. Miyanaga, "Dynamic time warping for speech recognition with training part to reduce the computation", International Symposium on Signals, Circuits and Systems ISSCS2013, 11-12 July, 2013.
17. B.W.Gawali, S. Gaikwad, P. Yannawar and S.C. Mehrotra, "Marathi isolated word recognition system using MFCC and DTW features", ACEEE International Journal of Information Technology, vol. 01, no. 01, Mar 2011, pp. 21-24.
18. C. Madhu, A. George and L. Mary, "Automatic language identification for seven Indian languages using higher level features", IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), Kollam, India, 2017, pp. 1-6.
19. P. Swietojanski, "Learning Representations for Speech Recognition using Artificial Neural Network", a doctoral thesis, 2016.
20. M. Borsky, "Robust recognition of strongly distorted speech", a doctoral thesis, 2016.
21. S. G. Surampudi and Ritu Pal, "Speech Signal processing using Neural Networks", IEEE International Advance Computing Conference (IACC 2015), Bangalore, India, 12-13 June 2015.
22. A. Zaatri, N. Azzizi and F. L. Rahmani, "Voice Recognition Technology using Neural Networks", Journal of New Technology and Materials, vol. 5, no. 1, 2015, pp. 26-30.
23. O.I. Abiodun, A Jantan, A.E.Omolara, K.V. Dada, N.A. Mohamed and H. Arshad, "State-of-the-art in artificial neural network applications: A survey", Heliyon, an Elsevier Journal, vol. 4, no. 11, 2018.
24. L. Fausett, "Fundamentals of Neural Networks: Architecture, Algorithms and Applications", Prentice-Hall, Inc., New Jersey 1994.
25. L. Besacier, E. Barnard, A. Karpov and T. Schultz, "Automatic Speech Recognition for Under-Resourced Languages: A Survey", Speech Communication, vol. 56, January, 2014, pp. 85-100.
26. V. Berment, "Methods to computerise "little equipped" languages and group of languages", PhD. Thesis, J. Fourier University-Grenoble I, May 2004.
27. "Ethnologue Languages of the World" https://www.ethnologue.com/statistics/status.
28. H.B.Sailor, M.V.S. Krishna, D. Chhabra, A. T. Patil, M.R. Kamble and H.A. Patil, "DA-IICT/IIITV system for low resource speech recognition challenge 2018", Interspeech 2018, 2-6 September 2018, Hyderabad, pp. 3187-3191.
29. M.A. Hasegawa-Johnson, P. Jyothi, D. McCloy, M. Mirbagheri, G. M. di Liberto, Amit Das, B. Ekin, C. Liu, V. Manohar, H. Tang, E. C. Lalor, N. F. Chen, P. Hager, T. Kekona, R. Sloan and A. K. C. Lee, "ASR for Under-Resourced Languages From Probabilistic Transcription", IEEE/ACM transactions on Audio, Speech, and Language processing, vol. 25, no. 1, January 2017.
30. K. Kumar, R.K. Aggarwal and A. Jain, "A Hindi speech recognition system for connected words using HTK", International Journal of Computational Systems Engineering, vol. 1, no. 1, 2012, pp. 25-32.
31. A. Gope and S. Mahanta, "Lexical Tones in Sylheti", 4th International Symposium on Tonal Aspects of Languages,Nijmegen,Netherlands,May 13-16,2014
32. A.Gope and S. Mahanta, "An Acoustic Analysis of Sylheti Phonemes", Proceedings of the 18th International Congress of Phonetic Sciences. Glassgow, UK, 2015.
33. A.Gope and S. Mahanta, "Perception of Lexical Tones in Sylheti", Tonal Aspects of Languages 2016, 24-27 May 2016, Newyork.
34. D. M. Kane, "Puthi-Pora:'Melodic Reading' and its use in the Islamisation of Bengal", Doctoral Dissertation, University of London, 2008.
35. K.H.Davis, R. Biddulph and S. Balashek, "Automatic Recognition of Spoken Digits", Journal of the Acoustic Soc. of America, vol. 24, no. 6, 1952, pp. 627-642.
36. G.Chakraborty, M.Sharma, N. Saikia and K. K.Sarma, "Recurrent Neural Network Based Approach To Recognise Isolated Digits In Sylheti Language Using MFCC Features", Proceedings of International conference on telecommunication, power analysis and computing techniques(ICTPACT-2017),Chennai, India, 6-8 April,2017.
37. H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for Spoken word recognition", IEEE Trans. Acoustic, Speech Signal Processing, vol. 26 , no.1, Feb 1978 , pp. 43-49.
38. X. Lei, A. W. Senior, A. Gruenstein and J. Sorensen, "Accurate and Compact Large vocabulary speech recognition on mobile devices", INTERSPEECH 2013, Lyon, France, 25-29 August 2013, pp. 662-665.
39. J.T.Geiger, Z.Zhang, F.Weninger, B. Schuller and G. Rigoli, "Robust speech recognition using long short term memory recurrent neural networks for hybrid acoustic modeling", Conference of the International Speech Communication Association, 14-18 September 2014, Singapore INTERSPEECH 2014
40. P. Sharma and A. Garg, "Feature Extraction and Recognition of Hindi Spoken Words using Neural Networks", International Journal of Computer Applications, vol. 142, no.7, May 2016., pp. 12-17.
41. A. Goswami, "Simplification of CC sequence of Loan words in Sylheti Bangla", Language in India, vol 13, no. 6 June,2013.
42. M. MoneyKumar, E. Sherly, and W. M. Varghese, "Isolated Word Recognition system for Malayalam Using Machine Learning", Proc. of the 12th Intl. Conference on Natural Language Processing, Trivandrum, India. December 2015, pp. 158–165.
43. J. Kunze, L. Kirsch, I. Kurenkov, A. Krug, J. Johannsmeier, and S. Stober, "Transfer Learning for Speech Recognition on a Budget", Proceeding of the 2nd workshop on Representation learning for NLP, Vancouver, Canada, August 3, 2017, pp. 168-177.
44. S. Furui, "Speaker-Independent Isolated Word Recognition Using Dynamic Features of Speech Spectrum", IEEE transactions on Acoustic,Speech and Signal processing, vol. 34 , no.1, 1986, pp. 52-59.
45. D. Dhanashri and S.B. Dhonde, "Isolated word speech recognition system using Deep Neural Networks", Proceedings of the International Conference on Data Engineering and Communication Technology, vol. 1, 2017, pp. 9-17.
46. T. Hori, C. Hori, S. Watanabe and J. R. Hershey, "Minimum word error Training of Long short-term memory Recurrent Neural Network Language models for Speech recognition", 41st IEEE International conference on Acoustic, Speech and Signal processing, Shanghai, China, vol. 2016-May, 2016, pp. 5990-5994.
47. A. Das, P. Jyothi, and M. H. Johnson, "Automatic Speech Recognition using Probabilistic transcriptions in Swahili, Amharic and Dinka", Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH 2016, San Francisco, USA, 8-12 September, 2016, pp. 3524-3528.
48. N. Zerari, S. Abdelhamid, H. Bouzgou, and C. Raymond, "Bidirectional deep architecture for Arabic speech recognition", Open Comput. Sci., DE GRUYTER, 2019, pp. 92–102.
49. C. Xu, X. Wang and S. Wang, "Research on Chinese Digit Speech Recognition Based on Multi-weighted Neural Network", IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application, 2008, pp. 400-403.
50. I. Kipyatkova and A. Karpov, "Recurrent Neural Network- based Language modeling for an Automatic Russian Speech Recognition System", Proceeding of the Artificial Intelligence and Natural Language and Information Extraction, Social Media and Web Search FRUCT Conference (AINL-ISMW FRUCT), 9-14 Nov. 2015, St. Petersburg, Russia.
51. D. T. Toledano, M. P. Fernandez-Gallego and A. Lozano-Diez, "Multi-resolution speech analysis for automatic speech recognition using deep neural networks: Experiments on TIMIT", PLoS ONE, vol. 13, no. 10, October 10, 2018.
52. B. Barman, "A contrastive analysis of English and Bangla phonemics", The Dhaka University Journal of Linguistics, vol. 2, no.4, August 2009.
53. W. Hong and P. Jin'gui, "An undergraduate Mandarin Speech Database for Speaker Recognition Research, Oriental COCOSDA International Conference on Speech Database and Assessments", Urumqi, China, 10-12 August, 2009.

Retrieval Number: C5874098319/2019©BEIESP
DOI:10.35940/ijrte.C5874.098319
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

54. C. Kurian, "A Review on Speech Corpus Development for Automatic Speech Recognition in Indian Languages", International Journal of Advanced Networking and Applications, vol.6, no.6, 2015, pp. 2556-2558.
55. B. Das, S. Mandal, P. Mitra and A. Basu, "Effect of aging on speech features and phoneme recognition: a study on Bengali voicing vowels", International Journal of Speech Technology, Springer, vol. 16, no. 1, March 2013, pp. 19-31.
56. M. Dua, R. K. Aggarwal and M. Biswas, "Performance evaluation of Hindi speech recognition system using optimized filterbanks", Engineering Science and Technology, an International Journal, Vol. 21, no. 3, June 2018, pp. 389-398.
57. F. S. Panchal and M. Panchal, "Review on methods of selecting number of Hidden nodes in Artificial Neural Network", International Journal of Computer Science and Mobile computing, vol. 3, no. 11, Nov.2014, pp. 455 – 464.
58. B. Deka, J. Chakraborty, A. Dey, S. Nath, P. Sarmah, S. R. Nirmala, and S. Vijaya, "Speech Corpora of Under Resourced Languages of North-East India", Oriental COCOSDA 2018, 7-8 May 2018, Miyazaki, Japan.
59. Desai V. A and Dr. V. K. Thakar, "Neural Network based Gujarati Speech Recognition for Dataset Collected by in-ear microphone", 6th International Conference On Advances In Computing & Communications, ICAAC 2016, 6-8 September,2016, Cochin, India, pp. 668-675.

AUTHORS PROFILE

Gautam Chakraborty, a research scholar in the department of Electronics & Telecommunication Engineering, Assam Engineering College, Guwahati, India, under Gauhati University, is currently working as an Assistant Professor at NERIM, Guwahati, Assam since 2010. His research interests include speech processing, cloud computing, etc. He has authored many research papers in national and international conference proceedings.

Dr. Navajit Saikia, currently Associate Professor in the Department of Electronics and Telecommunication Engineering, Assam Engineering College, Guwahati, India, has over 23. years of professional experience. His research interests include image processing, speech processing, reversible logic, information security, etc. He has co-authored several research papers in journals and conference proceedings. He is reviewer of many international journals and also has served as reviewer/ TPC member in many national/ international conferences.