RESEARCH ARTICLE OPEN ACCESS

Automated Speech Recognition System – A Literature Review

Manjutha M [1], Gracy J [2], Dr P Subashini [3], Dr M Krishnaveni [4]

Research Scholar [1] & [2], Professor [3], Assistant Professor [4]

Department of Computer Science

Avinashilingam Institute for Home Science and Higher Education for Women

Coimbatore, Tamil Nadu, India

ABSTRACT

Most natural form of human communication depends on speech. In order to understand human speech by enabling machines, computers can act as an intermediate for human expert, so that it can respond accurately and reliably to human voices. This can be achieved by speech recognition system, which permits a data processor to identify the words a person speaks in a microphone or telephone, and converts them into written text. Speech Recognition (SR) and its application is the thrust area in research during the past three decades which is carried out on various aspects, especially in the field of Information and Communication Technology (ICT) for speeding up scientific advancements. Also, people interact with system to utilize the technology in greater amount without the acquaintance of operating computer keyboard. At present, Automatic Speech Recognition (ASR) is effectively utilized for communication between human and machines. This paper analysis the accuracy of feature extraction based on modeling which is implemented using MFCC and HMM for two different type connected and continuous speech. The recognition result shows that the overall system accuracy for connected word is 69.22 % and continuous word is 50%.

Keywords — Speech Recognition, Analysis, Feature Extraction, Training, Testing, Modeling

I. INTRODUCTION

Speech processing is a unique discipline of signal processing. Study of speech signal and its processing method are the principles of speech processing. The speech processing application plays a major part in day-to-day life of commercial applications like Bank, Travel, Telecommunications and Voice Dialling. Some of the major growing applications are Language Identification, Speech Enhancement, Spoken Dialog System, Speaker Recognition and Verification, Speech Coding, Emotion and Attitude Recognition, Speech Segmentation and Labelling, Speech Recognition, Prosody, Text-to-Speech Synthesis, and Audio Visual Signal Processing. Input speech is given to the machine which accepts the command and translates into text format known as Speech Recognition System or Automatic Speech Recognition or Computer Speech Recognition or Speech to Text. Speech recognition systems analyze and train an individual speech that exploit to tune the recognition of specific voice which produces a more accurate result. Speech recognition consist of Vector Quantization, Feature Extraction, Dynamic Time Warping, Hidden Markov Models, Gaussian Mixture Model, Decision-tree based Clustering, training with Expectation Maximization (EM), Language Models, Speaker Adaptation and Finite-State Formulation.

In this paper, continuous and connected words are considered and MFCC features were extracted from the speech corpus. The extracted speech signal is trained by HMM model. Finally, the output result is compared with connected and continuous speech.

A. History of Speech Recognition

In 1952, Audrey system designed at Bell Laboratories was the first speech recognition system which recognized only digits spoken by a single person. Ten years later, IBM produced in 1962 which recognized 16 English words. In collaboration, Soviet Union, United States, England and Japan developed a hardware which recognized 4 vowels and 9 consonants. Carnegie Mellon's "Harpy" speech-understanding system recognized 1011 words between 1971 and 1976. Threshold Technology and Bell Laboratories are the first commercial speech recognition companies that interpret multiple person voice. A new statistical method called *Hidden Markov Model* (HMM) was introduced in 1980 which expanded to recognize hundred words to several thousand words and to recognize an unlimited number of words. Children could train to respond their voice in the form of Worlds of Wonder's Julie doll in 1987. In 1985, Kurzweil text-to-speech recognizes 5000 word vocabulary which is established by IBM. Dragon launched the first consumer speech recognition product called Dragon Dictate which recognizes 100 words per minute and also the system took 45 minutes to train the program. In 1996, Voice Activated Link (VAL) from Bell South launched a dial-in interactive voice recognition system which gave the information based on what the speaker said through phone. In 2001, speech recognition system attained 80% accuracy [14]. Ten years later, Google's English Voice Search system integrated 230 billion words from actual user. In 2015, Google's speech recognition experimented with Connectionist Temporal Classification (CTC) trained Long Short-Term Memory (LSTM) approaches which is implemented in Google Voice [12]. Various techniques suggested by many researchers for developing different applications in speech recognition are elaborated in this paper. This article is organized as follows. Section 2 presents classification of the speech recognition system. Section 3 analyses the related work carried out in the area of speech recognition system. Section 4 explains the methodologies used in speech recognition. Section 5 presents the conclusion and future extension of the research work.

II. CLASSIFICATION OF SPEECH RECOGNITION

Speech depends on speaker, bandwidth of speech and size of vocabulary, the different aspects where the intelligent have worked on. Speech recognition can be classified into various types based on utterance, speaker mode and vocabulary size.

1) *Classification of Speech Recognition based on the Utterances*

Speech recognition utterance can be classified into four categories namely, Isolated Word, Connected Word, Continuous Word, Spontaneous Word.

Isolated Word

Isolated word accepts and recognizes single word or single utterance at a time, and also, pronunciation needs pause between each utterance which is known as listen or non-listen state. This system has annoyances when selecting a different limit that alters the entire results. Example for Isolated word is 'start', 'stop', 'read', etc. Table I shows isolated words speech recognition system for different languages.

Connected Word

Connected word recognition system is identical to the isolated word. However, it allows speakers to pronounce the word together in minimum interval of time. The utterance can be a single word, or a collection of a few words, a single sentence, or even multiple sentences. Table II shows the list of language based connected words recognition system using different techniques.

Continuous Speech

Continuous speech allows user to speak typically, while the computer takes up the task of determining the content. Recognizers with continuous speech capabilities are most difficult to determine the utterance of speech boundary. Some of the authors experimented in continuous speech recognition system in different languages as shown in Table III.

TABLE I
LIST OF ISOLATED WORDS SPEECH RECOGNITION SYSTEM FOR DIFFERENT LANGUAGES

TABLE II
LIST OF CONNECTED WORDS RECOGNITION SYSTEM FOR DIFFERENT LANGUAGES

TABLE III
LIST OF CONTINUOUS SPEECH RECOGNITION SYSTEM FOR DIFFERENT LANGUAGES

Spontaneous Speech

Authors	Title
I.Mohamed Kalith et al. [15]	Isolated to Connected Tamil Digit Speech Recognition System Based on Hidden Markov Model (2016)
Bacha Rehmam et al. [5]	Artificial Neural Network based Speech Recognition using DWT Analysis Applied on Isolated Words from Oriental Languages (2015) Urdu
K.Venkataramana et al. [19]	Developing Telugu Speech Recognition System using Sphinx (2015)
Preeti Saini, et al.[31]	Hindi Automatic Speech Recognition using HTK (2013)
Ms.Vimala.C et al. [29]	Speaker Independent Isolated Speech Recognition System for Tamil Language using HMM (2012)

Recognition system deals with wide range of spontaneous speech phenomena, such as wrong starts, ungrammatical constructions, coughing, filled pauses, hesitations, laughter and other natural behaviors not found in regular speech.

Authors	Title
Khalil Ahammad et al. [20]	Connected Bangla Speech Recognition using Artificial Neural Network(c)
Annu Choudhary, et al. [3]	Automatic Speech Recognition System for Isolated and Connected Words for Hindi Language by using HTK(2013)
Kuldeep Kumar and R.K. Aggarwal [22]	A Hindi Speech Recognition System for Connected Words using HTK(2012)

Spontaneous speech recognition is complex in nature because of the large vocabulary size. Only a few experts worked in spontaneous speech recognition system as shown in Table IV.

Authors	Title
C.Sivaranjani et al. [6]	Syllable based Continuous Speech Recognition for Tamil Language(2016)
Parwinder Kaur et al. [30]	Detection of Syllables in Continuous Punjabi Speech Signal and Extraction of Formant Frequencies of Vowels(2015)
Maya Moneykumar et al. [24]	Malayalam Word Identification for Speech Recognition System(2014)
Suma Swamy et al. [35]	An Efficient Speech Recognition System (2013)
Lakshmi A et al. [23]	A Syllable based Continuous Speech Recognizer for Tamil(2006)

2) *Classification of Speech Recognition based on the Speaker Mode*

Speech recognition can be classified into speaker dependent and speaker independent systems.

Speaker Dependent System

Speaker dependent system is developed to work for a specific user whose voice has been trained already. The speaker dependent system is generally easier to develop, less expensive and produces more accuracy. Table V shows the connected digits for speaker dependent system.

TABLE IV

SPONTANEOUS SPEECH RECOGNITION SYSTEM FOR LARGE VOCABULARY

Authors	Title
Dufour, R et al. [9]	Spontaneous Speech Characterization and Detection in Large Audio Database(2009)
T.Nagarajan and et al. [36	Sub-band based Group Delay Segmentation of Spontaneous Speech into Syllable-like Units(2004)
Zhirong Wang and et al. [45]	Non-Native Spontaneous Speech Recognition through Polyphone Decision Tree Specialization(2003)
T.Sloboda, A.Waibel	Dictionary Learning for Spontaneous Speech Recognition(1996)
John Butzberger et al. [17]	Spontaneous Speech Effects in Large Vocabulary Speech Recognition Applications(1992)

TABLE V
CONNECTED DIGITS FOR SPEAKER DEPENDENT SYSTEM

Authors	Dataset	Methodology	Accuracy
K.P. Unnikrishnan et al. (1991) [18]	36 digit strings recorded by single male speaker Sampling rate 12kHz Digitized with 14bit A/D converter 288 digits used for segmentation	Analog Perceptron Learning Algorithm HMM model and DTW	Test set of 144 utterance is 99.3%

Speaker Independent System

Speaker independent system is developed to operate for any kind of speaker. In order to implement speaker independent system, large speech database need to be created using speech samples from different speakers. This systems are the most difficult to develop and most expensive and accuracy is lower than speaker dependent systems. Table VI shows the continuous speech for speaker independent system.

TABLE VI
CONTINUOUS SPEECH FOR SPEAKER INDEPENDENT SYSTEM

Authors	Dataset	Methodology	Accuracy
Mohamed S., and Ahmed H. Kandil (2016) [27]	276 utterances segmented into 2544 constituent syllable	Frame length 30ms Overlapping 60% Hamming Window Pre-processing-MIR© software	91.5%

Speaker Adaptive System

Speaker adaptive system has the ability to adapt the characteristics of a new speaker. Due to this adaptable characteristic, the recognition accuracy of speaker adaptive system can be improved gradually.

3) Classification of Speech Recognition based on the Vocabulary Size

The size of vocabulary is another important factor that makes speech recognition system different from another. The vocabulary size is divided into five categories namely, Small Vocabulary has tens of words, Medium Vocabulary has hundreds of words, Large Vocabulary has thousands of words, Very Large Vocabulary has ten thousands of words and Out-of-Vocabulary recognize words which are not trained.

III. LITERATURE REVIEW

Thiang, et al. (2011) presented speech recognition using Linear Predictive Coding (LPC) and Artificial Neural Network (ANN) for controlling movement of mobile robot. Input signals were sampled directly from the microphone and then the extraction was done by LPC and ANN [39]. Ms.Vimala.C and Dr.V.Radha (2012) proposed speaker independent isolated speech recognition system for Tamil language. Feature extraction, acoustic model, pronunciation dictionary and language model were implemented using HMM which produced 88% of accuracy in 2500 words [29]. Cini Kurian and Kannan Balakrishnan (2012) found development and evaluation of different acoustic models for Malayalam continuous speech recognition. In this paper HMM is used to compare and evaluate the Context Dependent (CD), Context Independent (CI) models and Context Dependent tied (CD tied) models from this CI model 21%. The database consists of 21 speakers including 10 males and 11 females [7]. Suma Swamy et al. (2013) introduced an efficient speech recognition system which was experimented with Mel Frequency Cepstrum Coefficients (MFCC), Vector Quantization (VQ), HMM which recognize the speech by 98% accuracy. The database consists of five words spoken by 4 speakers at ten times [35]. Annu Choudhary et al. (2013) proposed an automatic speech recognition system for isolated and connected words of Hindi language by using Hidden Markov Model Toolkit (HTK). Hindi words are used for dataset extracted by MFCC and the recognition system achieved 95% accuracy in isolated words and 90% in connected words [3]. Preeti Saini et al. (2013) proposed Hindi automatic speech recognition using HTK. Isolated words are used to recognize the speech with 10 states in HMM topology which produced 96.61% [31]. Md. Akkas Ali et al. (2013) presented automatic speech recognition technique for Bangla words. Feature extraction was done by, Linear Predictive Coding (LPC) and Gaussian Mixture Model (GMM). Totally 100 words recorded in 1000 times which gave 84% accuracy [25]. Maya Moneykumar, et al. (2014) developed Malayalam word identification for speech recognition system. The proposed work was done with syllable based segmentation using HMM on MFCC for feature extraction [24]. Jitendra Singh Pokhariya and Dr. Sanjay Mathur (2014) introduced Sanskrit speech recognition using HTK. MFCC and two state of HMM were used for extraction which produces 95.2% to 97.2% accuracy respectively [16]. In 2014, Geeta Nijhawan et al. developed real time speaker recognition system for Hindi words. Feature extraction done with MFCC using Quantization Linde, Buzo and Gray (VQLBG) algorithm. Voice Activity Detector (VAC) was proposed to remove the silence [10].

IV. SPEECH RECOGNITION METHODOLOGIES

Speech recognition is initiated with speaker producing an utterance which consists of audio waves. The audio waves are captured by a microphone and converted into electric signal which is again converted into digital signal in order to be understood by the speech system. The relevant information about the given utterance is extracted for accurate recognition. Finally, speech recognition system finds the best match by tuning the utterance. The architecture of speech recognition is shown in Fig. 1.

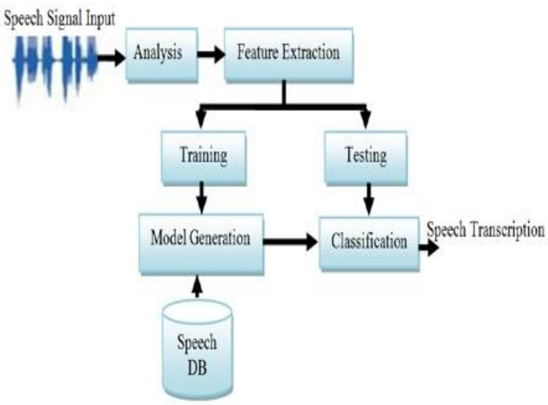

Figure 1. Automatic speech recognition system architecture

1) Analysis Techniques

Analysis is the initial stage of speech recognition system which involves sampling, windowing, framing and noise cancelation. Analysis deals with framing size for segmenting speech signal and contains various types of information about speaker due to vocal tract, behavior feature and excitation source which is explained in the analysis types such as segmentation analysis, sub-segmental analysis and supra-segmental analysis.

Segmentation Analysis:

Speech is analyzed using frame size and shift in the range of 10-30 ms to extract vocal tract information of speaker.

Sub-segmental Analysis:

The size of the frame and shift range around 3-5 ms extract the characteristic of the speaker from excitation state.

Supra-segmental Analysis:

Speech is analyzed, using frame size and behavior characteristics of the speaker. Different techniques used in analysis phase are shown in Table VII.

2) Feature Extraction

According to speech recognition theory, it should be possible to recognize speech directly from the digitized waveform. Since speech signals are unstable in nature, statistical representations should be generated for compressing the speech signal variability which is achieved by performing feature extraction. In order to transform the time domain signal into an effective parametric representation feature extraction is used. Most widely used extraction technique is MFCC. The block diagram of MFCC techniques is shown in Fig. 2. Different techniques proposed by many researchers for feature extraction are listed in Table VIII.

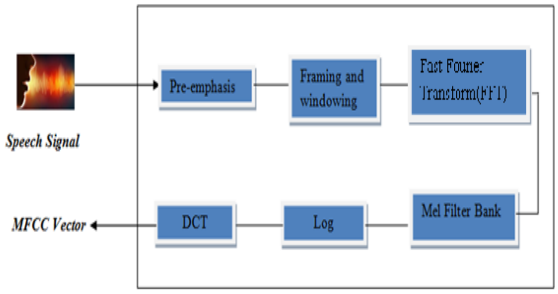

Figure 2. Block Diagram for MFCC Feature Extraction

The MFCC processor involves the following steps:

Pre-emphasis:

In the pre processing, the speech signal increases the amplitude of high frequency bands and decrease the amplitudes of lower bands which is implemented by FIR filter.

Framing and windowing:

The speech signal is split into number of frames. The frame size considered as 25 ms, hamming windowing is applied in order to minimize the signal discontinuities at the starting each edge of the frames.

Fast Fourier Transformer (FFT):

Each frame of N samples is converted in to time domain into frequency domain.

Mel Filter Bank:

The scale of frequency is converted from linear to mel scale which is called mel filter bank.

Logarithm:

Logarithm is taken for the mel filter bank which is known as log mel spectrum.

Discrete cosine Transform (DCT):

The log mel scale is again converted in to frequency domain to time domain which produces the feature of MFCC.

3) Recognition

Recognition is broadly classified into three, namely acoustic-phonetic approach, pattern-recognition approach and artificial intelligence approach. In the training phase of recognition system, parameters of the classification model are estimated using a large number of training classes. During the testing phase the features of a test speech are matched with the trained speech model of each and every class. Various techniques for recognition system are listed in Table IX.

4) Modeling Technique

The aim of modeling technique is to create speaker models using speaker specific feature vector. The speaker modeling techniques are separated into two categorization namely, speaker identification and speaker recognition. The speaker identification technique automatically identifies who is speaking on the basis of individual information integrated in speech signal. Speaker recognition technique is the identification of a person from the characteristics of specific speaker voice. Existing modeling techniques for speech recognition system is given in the Table X.

TABLE XI
CONTINUOUS AND CONNECTED SPEECH ACCURACY

Words	Spoken Words	Trained Words	Accuracy
Continuous	14	9 words	50%
Connected	39	20 words	69.22%

TABLE VIII
DIFFERENT TECHNIQUES IN SPEECH RECOGNITION FOR FEATURE EXTRACTION

Authors	Dataset	Methodology	Accuracy
Desai Vijayendra et al. (2016) [8]	In-ear microphone recorded up to 2.3kHz-4kHz Noise hum -90dB around 200 Hz	MFCC and Real Cepstral Coefficient (RC) Conjugate Gradient (CG)Algorithm and Levenberg-Merquardt (LM) Algorithm	Conjugate Gradient (CG) Algorithm
Vimala. C, V. Radha (2016) [43]	10 Tamil spoken digits (0-9) and 5 spoken names from 30 different speakers Sampling rate 16 kHz	Five pass Pre-processing Gammatone Filtering and Cochleagram Coefficients (GFCC) Multi Taper Windowing for Yule Walker (MTYWGFCC) MTYW-GFCC with Formant Frequencies (MTYW-GFCC-FF) Frequency Warping and Feature Normalization using LPC and CMN (FWCMN MTYW-GFCC-FF)	FWCMN-MTYW-GFCC-FF increased the WRR up to 99.06% for the HMM Techniques
Shaik Shafee et al. (2015) [34]	8 Isolated Telugu word by 6 male and female of the age group15-40	MFCC and Gamma tone coefficients (GFCC) K-means Algorithm and ANN	GFCC features giving better results
Purnima Pandit, et al.(2014) [32]	Ten different speakers are used to record Gujarati digits	MFCC Feature Matching Using Dynamic Time Warping (DTW)	95.56%
Geeta Nijhawan, et al. (2014)[10]	Five speakers used to record Hindi words	MFCC Linde, Buzo and Gray (LBG) algorithm VQ Model VQLBG Algorithm	80%
Md. Akkas Ali et al. (2013) [25]	Bengali speech is recorded in different time	LPC,GMM and MFCC	84% rate for 100 Bangla word
V. Anantha Natarajan and et al. (2012) [41]	Television broadcasts containing both male and female speaker recorded	Frame length-20ms shift- 5ms sampling rate - 8khz Hamming Window	83.46% by Gaussian Kernel
Satyanand Singh and Dr. E.G. Rajan (2011) [33]	70 speakers (61 males + 9 females) 6650 utterance Sampling rate 8.0khz	MFCC , Inverted MFCC using Triangular Filters and Gaussian Filters Modulo 2 Addition (EXOR) Used VQ-Linde Buzo Gray (LBG) Algorithm as Speaker Modeling Paradigm	98.57% in 2 seconds

TABLE IX

Authors	Dataset	Techniques
V.Naveen Kumar and et al. (2015) [42]	Stuttered speech is recorded	Frame length-20ms Pre-Emphasis- FIR filter Hamming Window Mel Frequency Wrapping Discrete Cosine Transform (DCT) VQ DTW
Hazrat Ali et al. (2015) [13]	Urdu Digits recorded by 10 speakers of both male and female	MFCC, Support Vector Machines (SVM) Random Forests Linear Discriminant Analysis
Kishori R. Ghule et al. (2015) [21]	Medicinal plants names recorded by 100 speaker of different age group from 20-50	Extraction by Discrete Wavelet Transforms(DWT), ANN
Thiang and et al. (2011) [39]	Robot control words are recorded	LPC , ANN Back propagation method is used for training
Mousmita Sarma et al. (2009) [28]	Three group of Assamese Language Dardic (Pisacha), the Indic (Indo-Aryan) and Iranian	Recurrent Neural Networks (RNNs) Multi-Layer Perceptron (MLP) Levenberg-Marquardth Back Propagation Algorithms
Azam Beg and et al.(2008) [4]	0-9 Urdu digits recorded by 15 speakers	Multi-Layer Feed-Forward Networks Fast Fourier Transformer (FFT) Discrete Fourier Transformer (DFT)

DIFFERENT TECHNIQUES FOR RECOGNITION SYSTEM

TABLE X EXISTING MODELING TECHNIQUES FOR SPEECH RECOGNITION

Authors	Dataset	Methodology
Wunna Soel and et al. (2015) [44]	12 vowels , 33 consonants and 4 medial in Myanmar language	MFCC for extraction DFT Acoustic model computed with HMM based GMM, SVM, Conditional Random Fields (CRFs) speech decoded by Viterbi Algorithm
Harpreet Kaur and et al. (2015) [11]	10 Punjabi alphabets recorded by 9 speaker	MFCC and HMM Testing by Viterbi Algorithm
Jitendra S.P et al. (2014) [16]	50 utterances by 10 speaker	MFCC and HMM 15 Grammar Model used for training and testing
AN. Sigappi and et al. (2012) [2]	100 Railway Station Names in Tamilnadu by both male and female in the age 21-60	HMM with 5 states and 4 Mixtures Auto associative Neural Networks (AANN) structure
Cini Kurian et al. (2012) [7]	Malayalam speech is recorded by 21 speakers (11 female and 10 males)	CI,CD and CD-TIED type of acoustic model Baum-Welch Re-estimation, HMM for testing and trained model in Viberti Algorithm
A.Srinivasan et al. (2009) [1]	letter ‘zha’ in Tamil language recorded by 3 males and 3 females Sampling rate 16kHz	LPC with Wavesurfer tool implemented in first order Three-State HMM

V. RESULT AND DISCUSSION

The dataset consists of 13 connected words and 14 continuous words which are recorded by 7 male speakers and 7 female speakers and the system is trained with 29 words totally. The recognition accuracy is calculated for connected and continuous words, also overall recognition rate for connected and continuous speech using MFCC with HMM model is shown in Table XI.

VI. CONCLUSION AND FUTURE WORK

Speech Recognition has been in development of more than 60 years. The various speech recognition methodologies and approach is available to enhance the recognition system. The fundamentals of SR system, various approaches existing for developing an ASR system are explained and compared in this paper. In recent years large vocabulary independent continuous speech has highly enhanced. From the review, it is concluded that HMM based MFCC feature is more suitable for speech recognition requirements and produces more good results than other models. In this paper, MFCC feature is extracted and the speech is trained by HMM model which is implemented for both connected and continuous speech. In order to improve the accuracy, other modeling techniques will be implemented in future.

REFERENCES

[1] A.Srinivasan, K.Srinivasa Rao, K.Kannan and D.Narasimhan, "Speech Recognition of the letter 'zha'(H) in Tamil Language using HMM", *International Journal of Engineering Science and Technology*, Vol.1(2),2009, pp.67-72.

[2] AN. Sigappi and S. Palanivel, "Spoken Word Recognition Strategy for Tamil Language", *International Journal of Computer Science Issues(IJCSI)*, Vol. 9, Issue 1, No 3, ISSN (Online):1694-0814 , January 2012, pp.227-233.

[3] Annu Choudhary, Mr. R.S. Chauhan, Mr. Gautam Gupta, "Automatic Speech Recognition System for Isolated & Connected Words of Hindi Language By Using Hidden Markov Model Toolkit (HTK)", in *Proceedings of International Conference on Emerging Trends in Engineering and Technology*, DOI: 03.AETS.2013.3.234, *22*-24th February 2012, pp.244– 252.

[4] Azam Beg and S. K. Hasnain, "A Speech Recognition System for Urdu Language", *in Lecture Notes in Computer Science (LNCS) Wireless Networks, Information Processing and Systems: Springer*, 2008, pp. 118-126.

[5] Bacha Rehmam, Zahid Halim, Ghulam Abbas, Tufail Muhammad "Artificial Neural Network-Based Speech Recognition Using DWT Analysis Applied on Isolated Words from Oriental Languages", *Malaysian Journal of Computer Science*, Vol. 28(3),2015, pp.242-262.

[6] C.Sivaranjani, B. Bharathi, "Syllable Based Continuous Speech Recognition for Tamil Language", *International Journal of Advanced Engineering Technology*, Vol. 7, Issue 1, E-ISSN 0976-3945, March 2016, pp. 01-04.

[7] Cini Kuriana, Kannan Balakrishnan, "Development & evaluation of different acoustic models for Malayalam continuous speech recognition", in *Proceedings of International Conference on Communication Technology and System Design 2011 Published by Elsevier Ltd*, December 2011, pp.1081-1088.

[8] Desai Vijayendra and Dr.Vishvjit K. Thakar, "Neural Network based Gujarati Speech Recognition for Dataset Collected by in-ear Microphone", *in Proceedings of 6th International Conference on Advances in Computing & Communications ICACC 2016*, Cochin, ISSN: *1877-0509*, 6-8th September 2016, pp.668-675.

[9] Dufour, R., Jousse, V.,Estève, Y., Béchet, F.,Linarès, G., "Spontaneous Speech Characterization and Detection in Large Audio Database", *in Proceedings of 13th International Conference on Speech and Computer (SPECOM 2009)*, St Petersburg(Russia), 21-25th June 2009.

[10] Geeta Nijhawan and Dr. M.K Soni, "Real Time Speaker Recognition System for Hindi Words", *International Journal of Information Engineering and Electronic Business,* Vol. 6, DOI: 10.5815/ijieeb.2014.02.04, April 2014, pp. 35-40.

[11] Harpreet Kaur and Rekha Bhatia, "Speech Recognition System for Punjabi Language", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 5, Issue 8, ISSN: 2277 128X, August 2015, pp. 566-573.

[12] Haşim Sak, Andrew Senior, Kanishka Rao, Françoise Beaufays and Johan Schalkwyk (September 2015): Google voice search: faster and more accurate.

[13] Hazrat Ali, An Jianwei and Khalid Iqbal, "Automatic Speech Recognition of Urdu Digits with Optimal Classification Approach", *International Journal of Computer Applications,* Vol.118, No.9, May 2015, pp. 1-5.

[14] http://www.pcworld.com/article/243060/speech_recognition_through_the_decades_how_we_ended_up_with_siri.html

[15] I. Mohamed Kalith, David Ashirvatham, Samantha Thelijjagoda, "Isolated to Connected Tamil Digit Speech Recognition System Based on Hidden Markov

Model", *International Journal of New Technologies in Science and Engineering*, Vol.3, Issue 4, ISSN:2231-5381, April 2016, pp.51-60.

[16] Jitendra Singh Pokhariya and Dr. Sanjay Mathur, "Sanskrit Speech Recognition using Hidden Markov Model Toolkit", *International Journal of Engineering Research & Technology (IJERT)*,Vol.3, Issue 10, ISSN: 2278-0181, October-2014, pp.93-98.

[17] John Butzberger, Hy Murveit, Elizabeth Shriberg, Patti Price, "Spontaneous Speech Effects In Large Vocabulary Speech Recognition Applications", *in Proceedings of DARPA Speech and Natural Language Workshop*, Morgan Kaufmann, New York, 1992, pp.339-343.

[18] K.P.Unnikrishnan, John J. Hopfield and David W.Tank, "Connected-Digit Speaker-Dependent Speech Recognition Using a Neural Network with Time-Delayed Connections", *IEEE Transaction on Signal Processing*,Vol.39, No.3,DOI:10.1109/78.80888, March 1991, pp.698-713.

[19] K.Venkataramana, P.Jayaprakash, P.S. Nagendra Babu "Developing Telugu Speech Recognition System using Sphinx-4", in *Proceedings of National Conference on Emerging Trends in Computing, Communication & Control Engineering*, Chennai, ISSN (Online):2319–8753, 10 September 2015, pp.41-48.

[20] Khalil Ahammad, Md. Mahfuzur Rahman, "Connected Bangla Speech Recognition using Artificial Neural Network", *International Journal of Computer Applications*, Vol.149, No.9, ISSN 0975–8887, September 2016, pp.38- 41.

[21] Kishori R. Ghule and Ratnadeep R. Deshmukh, "Automatic Speech Recognition of Marathi isolated words using Neural Network", *International Journal of Computer Science and Information Technologies (IJCSIT)*, Vol. 6(5), ISSN: 0975-9646 2015, pp.4296-4298.

[22] Kuldeep Kumar and R.K. Aggarwal, "A Hindi speech recognition system for connected words using HTK", *International Journal Computational Systems Engineering*, Vol.1, No.1, Online *ISSN:* 2046-3405, 2012, pp.25-32.

[23] Lakshmi A, Hema A Murthy, "A Syllable Based Continuous Speech Recognizer For Tamil", in *Proceedings of the Ninth International Conference on Spoken Language Processing, INTERSPEECH 2006 (ICSLP)*, Pittsburgh, Pennsylvania, USA,17-21st September 2006, pp.1878-1881.

[24] Maya Moneykumar, Elizabeth Sherly, Win Sam Varghese, "Malayalam Word Identification for Speech Recognition System" *An International Journal of Engineering Sciences*, Special Issue iDravadian , Vol. 15 ISSN: 2229-6913 (Print), December 2014, pp. 22-26.

[25] Md. Akkas Ali, Manwar Hossain, Mohammad Nuruzzaman Bhuiyan, "Automatic Speech Recognition Technique for Bangla Words", *International Journal of Advanced Science and Technology*, Vol. 50, January, 2013, pp.51-60.

[26] Megha Agrawal, Tina Raikwar, "Speech Recognition Using Signal Processing Techniques", *International Journal of Engineering and Innovative Technology (IJEIT)*, Vol.5, Issue8, ISSN: 2277-3754, February 2016, pp. 65-68.

[27] Mohamed S. Abdo and Ahmed H. Kandil, "Semi-Automatic Segmentation System for Syllables Extraction from Continuous Arabic Audio Signal", *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol.7, No.1, 2016, pp.535-540.

[28] Mousmita Sarma, Krishna Dutta and Kandarpa Kumar Sarma, "Assamese Numeral Corpus for Speech Recognition using Cooperative ANN Architect", *International Journal of Electrical and Electronics Engineering*,Vol.3,Issue8,Nov 2009,pp.456-465.

[29] Ms.Vimala.C and Dr.V.Radha, "Speaker Independent Isolated Speech Recognition System for Tamil Language using HMM", in *Proceedings International Conference on Communication Technology and System Design 2011*, Procedia Engineering 30 ISSN: *1877-7058*, 13March2012, pp.1097 – 1102.

[30] Parwinder Kaur, Ranpreet Kaur, Amanpreet Kaur , "Detection of Syllables in Continuous Punjabi Speech Signal and Extraction of Formant Frequencies of Vowels", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 5, Issue 4, ISSN: 2277 128X, 2015,pp.1091-1095.

[31] Preeti Saini, Parneet Kaur, Mohit Dua, "Hindi Automatic Speech Recognition Using HTK", *International Journal of Engineering Trends and Technology (IJETT)"*, Vol.4, Issue 6, ISSN:2231-5381, June 2013, pp.2223-2229.

[32] Purnima Pandit, Shardav Bhatt "Automatic Speech Recognition of Gujarati digits using Dynamic Time Warping", *International Journal of Engineering and Innovative Technology (IJEIT)*, Vol.3, Issue 12, ISSN: 2277-3754, June 2014, pp.69-73.

[33] Satyanand Singh and Dr. E.G. Rajan, "Vector Quantization Approach for Speaker Recognition using MFCC and Inverted MFCC", *International Journal of Computer Applications*, Vol.17, No.1, ISSN: *0975 – 8887*, March 2011, pp.1-7.

[34] Shaik Shafee, B.Anuradha, "Isolated Telugu Speech Recognition using MFCC and Gamma tone features by Radial Basis Networks in Noisy Environment",

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)", Vol.3, Issue 3, ISSN (Print): 2320-9798, March 2015, pp.1481-1488.

[35] Suma Swamy,K.V Ramakrishnan, "An Efficient Speech Recognition System", *Computer Science & Engineering: An InternationalJournal(CSEIJ)*,Vol.3,No.4,DOI:10.5121/cseij.2013.3403 August 2013, pp.21-27.

[36] T. Nagarajan and H. A. Murthy, "Sub-band based group delay segmentation of spontaneous speech into syllable-like units", *EURASIP Journal on Advances in Signal Processing*, ISSN: 1687-6180 (Online), December 2004, pp. 2614–2625.

[37] T.Nagarajan, Hema A. Murthy and Rajesh M. Hegde, "Segmentation of speech into syllable-like units", in *Proceedings* 8th European Conference on Speech Communication And Technology, Geneva,2003, pp.2893-2896.

[38] T.Sloboda, A.Waibel, "Dictionary Learning for spontaneous speech recognition", in *Proceedings of* International Conference on Spoken Language Processing *96*, Vol.4, 1996, pp.2328-2331.

[39] Thiang and Suryo Wijoyo, "Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot", in *Proceedings of International Conference on Information and Electronics Engineering (IPCSIT)*, Singapore, IACSIT Press, Vol.6, 2011, pp.179-183.

[40] V. Anantha Natarajan and S. Jothilakshmi, "Segmentation of Continuous Tamil Speech into Syllable like Units", *Indian Journal of Science and Technology*, Vol. 8(17), ISSN (Online): 0974-5645, August 2015, pp.1-5.

[41] V. Anantha Natarajan S. Jothilakshmi, "Segmentation of Continuous Speech into Consonant and Vowel Units using Formant Frequencies", *International Journal of Computer Applications*, Vol. 56, No.15, ISSN: *0975 – 8887*, October 2012, pp.24-27.

[42] V.Naveen Kumar, Y Padma Sai and C Om Prakash, "Design and Implementation of Silent Pause Stuttered Speech Recognition System", *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, Vol.4, Issue 3, ISSN (Online): 2278 – 8875, March 2015 , pp.1253-1260.

[43] Vimala. C, V. Radha, "Efficient Acoustic Front-End Processing for Tamil Speech Recognition using Modified GFCC Features", *International Journal Image, Graphics and Signal Processing*, DOI: 10.5815/ijigsp.2016.07.03, 2016, pp.22-31.

[44] Wunna Soe and Dr. Yadana Thein, "Syllable-Based Speech Recognition System for Myanmar", *International Journal of Computer Science, Engineering and Information Technology (IJCSEIT)*, Vol.5, No.2, DOI: 10.5121/ijcseit.2015.5201 1, April 2015, pp.1-13.

[45] Zhirong Wang, Tanja Schultz, "Non-Native Spontaneous Speech Recognition through Polyphone Decision Tree Specialization", in *Proceedings* 8th European Conference on Speech Communication and Technology, *Geneva*, Vol.2003, No.9, ISSN 1018-4074, pp.1449-1452.