JCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Volume-3, Issue-2 E-ISSN: 2347-2693

An Overview of Speech Recognition Techniques and the Progress in Speech Technology for Indian Scripts

Bharat C. Patel[1], Jagin M. Patel[2], Manish M. Kayasth[3*]

[1]Smt. Tanuben and Dr.Manubhai Trivedi college of Information science, Surat, Gujarat, India
[2]M. K. Institute of Computer studies, Bharuch, Gujarat, India
[3]Assistant Professor and HOD, UCCC & SPBCBA & SDHG College of BCA and IT, Surat, India

Available online at: www.ijcseonline.org

Received: Jan /14/2015 Revised: Feb/05/2015 Accepted: Feb/21/2015 Published: Feb/28/ 2015

Abstract— Speech is widely regarded as the most natural form of communication between human beings. The field of speech recognition has facilitated the development of man-machine conversations, leading to the creation of automatic speech recognition systems. These systems enable interactions between humans and machines, and their applications are diverse and extensive. This paper aims to provide an overview of speech recognition techniques, explore the various application areas of speech recognition systems, and review the progress made in speech technology, particularly for Indian scripts. Additionally, the paper focuses on feature extraction techniques and classification methods used in speech recognition systems. The primary objective of this paper is to facilitate newcomers in understanding the flow of speech recognition and guide them in their pursuit of further research in this field.

Keywords— Human-Machine Interaction, Speech Recognition System, Feature Extraction Methods, Classification Techniques

I. Introduction

Automatic speech recognition is the field of pattern recognition. Speech recognition indeed has a rich history, starting back in the 1950s [1]. It has evolved significantly over the years and has become a crucial technology in natural language processing (NLP). Speech recognition systems play a vital role in various applications and industries. Voice assistants rely on speech recognition to understand and respond to user commands. Its significance spans across a multitude of domains, including voice assistants, transcription services and more [2]. By understanding and interpreting spoken words, machines can facilitate seamless human-machine communication, transforming the way we interact with technology.

Transcription services also heavily rely on speech recognition to convert spoken language into written text accurately. The development of accurate and robust speech recognition systems has greatly improved human-machine communication. It allows users to interact with technology more naturally by speaking commands or queries instead of relying solely on text-based input. This advancement has had a transformative impact on various domains, including smart homes, customer service, healthcare, and more. As research in speech recognition continues to progress, we can expect further advancements in the accuracy, speed, and versatility of these systems. The ability of machines to understand and interpret spoken language is a fundamental aspect of enhancing human-machine interaction and enabling a wide range of innovative applications.

Voice-activated search is one of the key features offered by voice assistants. Instead of manually typing queries into search engines, users can simply speak their requests, making the process faster, more convenient, and hands-free. Speech recognition technology accurately converts spoken language into text, allowing the voice assistants to understand user commands and retrieve relevant information from the web [3].

Voice assistants go beyond just search capabilities. They can perform tasks like setting reminders, sending text messages, making phone calls, playing music, providing weather updates, and controlling smart home devices. This hands-free interaction with technology brings added convenience, especially in situations where manual input may be cumbersome or impossible, such as when driving, cooking, or multitasking.

Moreover, speech recognition technology has proven to be beneficial for individuals with disabilities. People with mobility impairments or those with conditions that affect their ability to use traditional input devices can use speech recognition to interact with computers and mobile devices, thus enabling greater accessibility and inclusivity.

The ongoing advancements in speech recognition technology continue to enhance the capabilities and accuracy of voice assistants, making them even more valuable in our day-to-day lives. As research and development in this field progress, we can expect further improvements in speech recognition accuracy, language coverage, and natural language understanding, enabling even more seamless and intuitive human-machine interaction.

Given the broad range of applications and the increasing demand for accurate and reliable speech recognition, a comprehensive review of the techniques employed in this field becomes crucial. Understanding the advancements, challenges, and limitations of existing approaches is essential for driving further progress and innovation.

This research paper aims to provide a comprehensive review of speech recognition techniques. It will explore methodologies such as acoustic modeling, language modeling, and decoding algorithms, which are integral to the functioning of speech recognition systems. Additionally, the impact of deep learning and neural network models on the evolution of speech recognition will be examined. By analyzing and synthesizing key research papers and studies, this review seeks to present a comprehensive overview of the current state-of-the-art techniques and identify potential avenues for future research.

II. LITERATURE REVIEW

Many research work done on speech recognition at national as well as international level. This section concentration on research work carried out on Indian script. Many researchers have contributed their work and still the work is going on. In this section, we are focusing on research work done in eight Indian scripts: Gujarati, Hindi, Marathi, Punjabi, Bengali, Malayalam, Kannada and Tamil.

Review of Speech Recognition System in Gujarati language:

Patel and Desai [4] proposed a model to recognize spoken Gujarati numerals. The authors collected sample data from 300 speakers and created dataset for Gujarati numerals. The proposed system utilizes the Mel Frequency Cepstral Coefficient (MFCC) method for feature extraction from spoken Gujarati numerals and the Dynamic Time Warping (DTW) technique for comparing test patterns with trained patterns to recognize spoken numerals. They achieved an average of 71.17% accuracy rate for all Gujarati numerals from 0 to 9.

Patel and Virparia [5] presented a small vocabulary speech recognition system for Gujarati word. The speech input is collected from 31 speakers out which 12 females and 19 males speaker respectively. Speech is inputted and recognized words are displayed as Gujarati text on screen. The accuracy rate 90.88% and 85.28% is obtained with male and female data respectively. The average recognition accuracy rate of proposed system is 88.71%.

Patel and Desai [6] presented a model that accepts isolated spoken Gujarati numeral as input and translate them into editable textual form. The proposed model extract features using MFCC technique and classification is carried out using K-Nearest Neighbor (K-NN). They collected sample data of Gujarati numerals from 350 speakers and achieved average success rate is about 78.13%.

Pipalia and Dave [7] designed a hybrid method (HMM/ANN) for isolated Gujarati word recognition. The proposed system is designed in MATLAB. The authors collected data from 3 female and 2 male speakers for proposed work and recorded them using Audacity software. The data consists of 30 Gujarati word including Gujarati numbers and routine words. Using experiment, the proposed model obtained average accuracy of 70.57% and using Genetic algorithm the accuracy is increases to 79.14%.

Pandit and Bhatt [8] addressed automatic speech recognition system for Gujarati digit. The authors used MFCC technique to extract features from Gujarati digit and Dynamic Time Warping algorithm is employed to recognize Gujarati digit. In the experiment, they obtained success rate 84.44% and the success rate of the proposed system was increased to 95.56% when extra care should be taken during recording of digit.

Review of Speech Recognition System in Hindi language:

Panwar et. al. [9] proposed speech recognition system to recognize Hindi digits 0 to 9. They used two feature extraction techniques such as Wavelet Based Features

(WBF) and MFCC. The experimental result shows that the success rate is increased by 4% compared to MFCC.

Kumar et. al. [10] presented a connected-words speech recognition system for Hindi language. The recognition system used Mel frequency cepstral coefficients to extract features from speech and system is trained to recognized word using hidden Markov models (HMMs). The training data is collected from 12 speakers. The experimental results show that the presented system provides the overall word-accuracy of 87.01%, word-error rate of 12.99%, and word-correction rate of 90.93% respectively.

Kumar and Aggarwal [11] proposed Hindi speech recognition system which convert acoustic wave into textual form. They trained 30 isolated Hindi words and Hidden Markov model is used to recognize Hindi spoken word. The overall accuracy of proposed model is 94.63%.

Verma and Mishra [12] studied various alternative representations of the speech signal that have the potential to contribute to the improvement of the recognition performance. They collected speech sample of Hindi numerals from 24-speakers and extract the features using LPC(Linear Prediction Coding), PLP(Perceptual Linear Predictive), wavelet based feature extraction techniques and HMM for recognition purpose. They achieved accuracy rate is 66%, 77.3% and 89.85% for different feature extraction techniques LPC, PLP, and db10 respectively for speaker independent system.

Saini et. al. [13] proposed automatic speech recognition system for Hindi language. The proposed system extracted feature from Hindi word using MFCC technique and training and testing is perform using HMM model. The authors collected 113 Hindi words from nine speakers and train them using proposed system. The experimental results show that the overall accuracy of the proposed system with 10 states in HMM topology is 96.61 and 95.49%.

Review of Speech Recognition System in Marathi language:

Kayte in [14] proposed speaker dependent, real-time speech recognition system which recognized an isolated word in Marathi language. The proposed system used Linear Prediction Coding and Vector Quantization (VQ) front-end for processing speech signals and Hidden Markov Models (HMMs) for recognition purpose. The overall accuracy obtained of the proposed system was 94.63%.

Patil and Pardeshi [15] presented connected word speech recognition system for Marathi language. The proposed system is implemented using MFCC feature extraction technique and Continuous Density Hidden Markov Model (CDHMM) for classification of Marathi word. The authors used speech segmentation methods such as Short Time Energy (STE) and Spectral Centroide (SC) to separate voice and unvoiced speech data. The proposed system is trained with 10 Marathi sentences; each sentence is uttered 10 times for training and 20 times for testing by four different users. The overall accuracy rate is obtained more than 80%.

Khetri et. al. in [16] described isolated Marathi word recognition system. The authors used three different databases of IWAMSR. The MFCC and VQ algorithm is used to extract feature from Marathi word and classification of word respectively. The three databases database1, database2 and database3 contain unknown speech samples 63, 90 and 36 were tested on the proposed system and obtained an accuracy rate is near about 85.71%, 31.11% and 88.88% respectively.

Nimbhore et. al in [17] addressed the proposed system which calculate the performance of pitch frequency estimation of Marathi spoken numbers. They utilized autocorrelation and cepstral methods to estimate speech frequency of Marathi numbers. The pitch frequencies are normalized using PRAAT tool. The experimental result show that the average mean of pitch frequency of numbers varies from 1.48 to 2.03 and standard deviation varies from 0.84 to 1.38 in Hz.

Gaikwad et. al. in [18] proposed Polly-clinic inquiry system(PCIS) using IVR in Marathi Language. The proposed PCIS model will be advantages to farmer as it is provide quick solution of query related to doctor inquiry on 24/7 basis. The proposed system trained with 1000 sentences and 50 isolated words. The overall accuracy of the proposed system is 88%.

Review of Speech Recognition System in Punjabi language:

Katyal et. al. [19] proposed speech recognition system in Punjabi isolated word. The authors utilized Ensemble Empirical mode decomposition (EEMD) which is adaptive and appears to be suitable for non-linear and non-stationary emotional speech signal analysis. They used neural network for classification purpose and found that the average accuracy of a word lies between 85% to 94%.

Ghai and Singh [20] created a continuous speech recognition system for Punjabi language using HTK toolkit on the Linux platform. The authors created in-house speech corpus of Punjabi language. They collected 100 sentences of continuous speech in Punjabi language from the speaker and stored them in .wav format. Average number of words per sentence is 7. Overall recognition accuracy has been found to be 82.18% at sentence level and 94.32% at word level with the help of proposed system.

Kumar [21] developed a small vocabulary speaker independent isolated word recognition system for Punjabi language. The author has created vocabulary of 500 isolated words and employed the Hidden Markov Model and Dynamic Time Warp technique to recognized spoken word. The accuracy achieved with proposed system is 91.3% and 94.0% using HMM and DTW respectively.

Ghai and Singh [22] described a phone based approach has applied for isolated and connected word speech Punjabi language speech recognition system. In the first experiment of isolated word speech recognition, they collected test speech samples for 70 words from 5 speakers including 3 males and 2 females. In this experiment, the proposed system obtained accuracy 92.05% for acoustic word model and 97.14% for acoustic triphone model respectively. In the second experiment of connected word speech recognition, authors collected test speech samples for 50 sequences of words from 6 speakers including 3 males and 3 females. Each sequence of speech sample contained 6 words. In this particular experiment, the accuracy rate 87.75% for acoustic word model and 91.62% for acoustic triphone model respectively obtained with proposed model.

Dua et. al. [23] proposed an isolated word Automatic Speech Recognition system for Punjabi language. The system is implemented using Hidden Markov Model Toolkit (HTK) on Linux platform. The proposed system is trained with 115 isolated Punjabi words and each word is uttered 3 times by 8 speakers so total words are file is 2760. The MFCC technique has been used for feature extraction and performance is evaluated using HMM. In the first experiment, overall performance in a class room environment is 95.63% and 94.08% is obtained in a class room environment and open space environment respectively.

Review of Speech Recognition System in Bengali language:

Kumar et. al. in [24] described spoken Bengali numerals recognition system in noisy environment. The system extracted MFCC features from spoken numerals and the classification is based on the dynamic time warping. The Recognition accuracy rate of the system is obtained more than 90% for has speaker dependent spoken numeral.

Mandal et. al. in [25] introduced Bengali Automatic Speech Recognition(ASR) system Shruti-II made for the visually challenged people and such people can send an E-mail using this system. The proposed system converts standard Bengali continuous speech to Bengali Unicode.

Das et. al. in [26] presented Bengali speech corpus development for speaker independent continuous speech recognition. They collected speech samples from two age groups: younger group of belong to 20 to 40 years age and older group of belong to 60 to 80 years age and created Bengali speech corpus. Authors used MFCC method to extract the feature; Hidden Markov Toolkit (HTK) has been used for aligning speech data and check the quality of speech corpus.

Hossain et. al. in [27] proposed isolated Bengali speech recognition system. They collected ten digits from 10 speakers and extract the features of digits using MFCC technique and recognized them using Back-Propagation Neural Network. They achieved recognition rate about 96.332% for known speakers and 92% for unknown speakers.

Review of Speech Recognition System in Malayalam language:

Kurian and Balakrishnan [28] addressed speaker independent speech recognition system for Malayalam digits. The author collected strings/sentences of Malayalam digit from 21 speakers. Out of it, there are 10 male and 11 female speakers of age group between 20 and 40. The created database consist of 420 sentences, each sentence contains 7 Malayalam digits. The phonetically important characteristics of speech are extracted as MFCC and Malayalam digit is recognized using Hidden Markov Models. The proposed system achieved 98.5% word recognition accuracy and 94.8% sentence recognition accuracy on a test set of continuous digit recognition task.

Mohamed and Nair [29] proposed small vocabulary continuous Malayalam speech recognition system. The whole corpus consisted of 108 sentences with 540 words and a total of 3060 phonemes. They applied 3 feature sets such as MFCC, delta (first-order) MFCC and delta-delta (second-order) MFCC on proposed system and Malayalam words recognized using HMM-based acoustic model. The authors achieved 94.67% word accuracy and 93.33% sentence correct using proposed model.

Anand et. al. [30] described large vocabulary continuous speech recognition (LVCSR) system for the Malayalam language with an application for visually challenged. The speech data was collected from 80 speakers (40 male and 40 female speakers) of age group 20-30 years, 31-40 years and 41-50 years. They used MFCC method to extract acoustic feature from input signal and HMM based model is used for recognition purpose. The system achieved word accuracy of 75% in average and further increased up to 80% in average, by implementing speaker adaptation technique.

Kurian and Balakrishnan [31] addressed connected digit speech recognition system for Malayalam language. The authors utilized Perceptual Linear Predictive cepstral coefficient for speech parameterization and continuous density Hidden Markov Model in the recognition process. Viterbi algorithm is used for decoding. The training database contains utterances of 21 speakers which is collected in office environment. The system achieved accuracy 99.5% for unseen data.

Sreejith and Reghuraj [32] proposed a speaker independent system to recognize isolated spoken Malayalam word. They applied MFCC feature extraction technique and k-mean clustering for recognition purpose. The words are identified using the minimum quantization distance between train and test features. The system consists of six stored words and 100 speakers are asked to speak the stored word. They obtained average accuracy of the proposed system is almost 88 %.

Review of Speech Recognition System in Kannada language:

Hemakumar and Punitha [33] designed a system which recognized spoken Kannada words independent of speakers. The speech signal is analyzed and Linear-Predictive coding coefficients are extracted from isolated word and convert them into Real Cepstrum Coefficient. The authors collected 294 unique Kannada words from 10 speakers so the total sample data is 2940. Out of that 2352 data are used for training and 1176 data are used for testing. The proposed system obtained success rate of 98.29 % and 91.66 % for known and unknown speaker data respectively.

Kannadagu and Bhat [34] built an automatic phoneme recognition system based on Bayesian Multivariate Modeling which recognized Kannada character. The authors used 15 Kannada phonemes and each phoneme is recorded 500 times for training and 200 times for testing. Therefore, there are 7500 phonemes in the training database and 3000 phonemes in testing database. Features are extracted using MFCC technique and unknown phoneme is recognized using Bayesian classifier. The performance analysis of model is calculated in terms of Phoneme Error Rate (PER).

Hegde et. al. [35] proposed system to recognize isolated word for Kannada language. The authors collected 10 numerals from 5 speakers. Each numeral uttered 2 times by each speaker so there are 1000 records in the database. Out of that 900 records are used for training and 100 records are used for testing purpose. The performance of the proposed model is evaluated using Support Vector Machine (SVM) combined with MFCC feature extraction technique and obtained average success rate of the proposed system is 79%.

Muralikrishna et. al. [36] has implemented Kannada isolated digit recognition system. The authors have collected 50 utterances of each digit from same speaker and 25 samples for each digit from same speaker. The experiments are conducted on three different feature set: 12-dimensional MFCC, 26-dimensional MFCC (first order derivative or delta MFCC) and 39-dimensional MFCC (second order derivative or delta-delta MFCC) feature vector. HMM based classifier is used for testing purpose and achieved average accuracy rate 75.6%, 84.2% and 97.2% with MFCC, first order derivative and second order derivative respectively.

Review of Speech Recognition System in Tamil language:

Vimla and Radha [37] addressed isolated Tamil spoken word system. For experimental purpose, they collected Tamil digits (0-9) and 5 spoken names from speakers of age group of 20-35. The total size of dataset is 600. They perform comparative study of most popular feature extraction techniques like MFCC, LPC, PLP and GFCC and observed that GFCC technique provided best result than other techniques. Moreover, they also observed that better accuracy is obtained with HMM and DTW followed by MLP and SVM techniques for both training and test data.

Karpagavalli et. al [38] proposed a speech recognition system which recognized isolated Tamil spoken digit. They prepared a codebook for each word in the vocabulary created using Linde-Buzo-Gray (LBG) vector quantization algorithm. For training purpose, Tamil digits are collected from 20 speakers and each speaker utters each digit 6 times. The overall recognition rate of the proposed system is 91.8%.

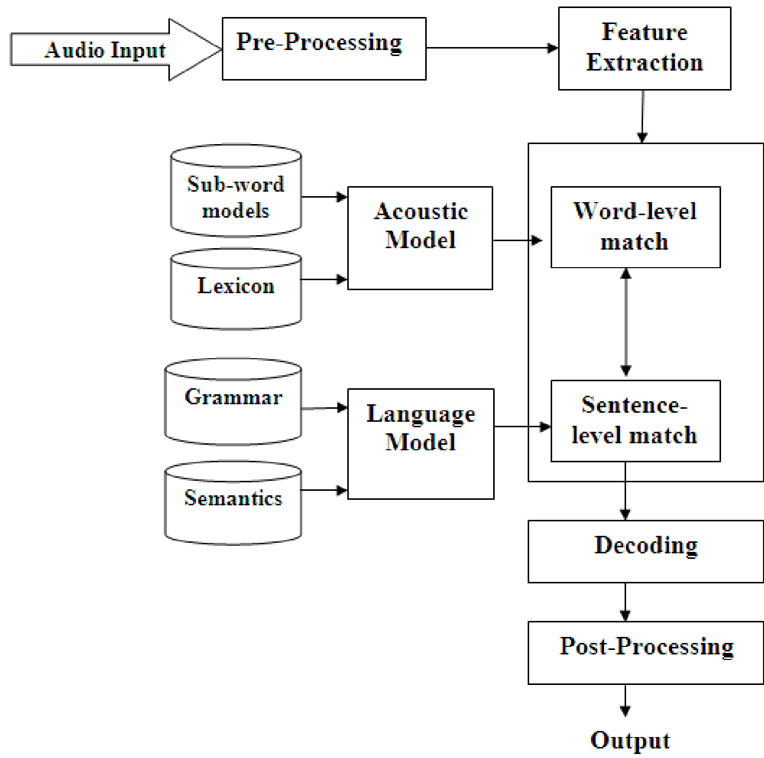
Vimla and Radha [39] described speaker independent isolated speech recognition system for Tamil language which is implemented using Sphinx4. The database size used for this research work is 2,500 words and feature set is created using MFCC technique. The proposed system produces average accuracy rate is 88% for Tamil isolated spoken word.

Kumar and Foo [40] proposed a bilingual speech recognition system for English and Tamil. The authors used NTIMIT corpora for English collected over telephone contained 4620 sentences for training; however Tamil corpora contained only 400 utterances for training. They used with or without adaptation approach during experiment and conclude that with adaptation approach improved result than without adaptation approach.

Thangarajan et. al. [41] addressed continuous speech recognition for Tamil language. They created corpus containing 22.5 hours of continuous speech of 50 speakers out of that 25 speakers are males and 25 females for training purpose. All the speakers spoke from the same set of unique 550 sentences so there are 27,500 utterances in the dataset. They provided solution of challenging features of Tamil language such as agglutination and morph phonology by using the syllable as a sub-word unit in an acoustic model. The author proposed an algorithm based on prosodic syllabus and perform two types of experiments. In the first experiment, modeling syllable as an acoustic unit is suggested and CI syllable models are trained and tested. The second experiment proposed integration of syllable information in the conventional triphone or CD phone modeling. Finally, they concluded that the accuracy of the proposed system is comparable to that of the baseline triphone system.

Architecture of Automatic speech recognition system:

The architecture of an automatic speech recognition (ASR) system typically consists of several interconnected components that work together to convert spoken language into written text. Fig 1 shows the block diagram of ASR. While there can be variations in the specific implementation, here is a general overview of the typical architecture of an ASR system:



[Fig. 1: Block diagram of Speech Recognition System]

1. Audio Input: The ASR system begins with the audio input, which can be captured from various sources such as microphones, recorded audio files, or streaming audio. The audio input is typically in the form of a continuous waveform.
2. Pre-processing: The audio input undergoes pre-processing to enhance its quality and extract relevant features. This stage may include steps such as noise reduction, filtering, and resampling to ensure optimal input for subsequent processing.
3. Feature Extraction: In the feature extraction stage, the pre-processed audio signal is transformed into a sequence of acoustic features that capture the linguistic content of the speech. Commonly used features include MFCCs or filterbank energies, which represent the spectral characteristics of the speech signal.
4. Acoustic Modeling: Acoustic modeling is a crucial component of an ASR system. It involves creating a

statistical model that represents the relationship between the acoustic features and the corresponding linguistic units, such as phonemes, sub-word units, or context-dependent units. Hidden Markov Models have been widely used in acoustic modeling, where each HMM represents a different speech sound or unit.

5. Language Modeling: Language modeling is responsible for capturing the statistical patterns and constraints of natural language. It helps the ASR system to assign higher probabilities to likely word sequences and make more accurate predictions. Language models can be based on n-gram models, recurrent neural networks (RNNs), or transformer-based models. These models utilize large amounts of text data to estimate the probability of word sequences.
6. Decoding: Decoding is the process of finding the most likely sequence of words given the acoustic and language models. It involves searching through a vast space of possible word sequences to find the best match for the observed acoustic features. Decoding algorithms, such as the Viterbi algorithm, beam search, or neural network-based approaches, are used to efficiently explore the search space and generate the most probable transcription.
7. Language Post-processing: Once the decoding stage provides the most probable word sequence, post-processing techniques can be applied to improve the output. This may involve correcting any recognition errors, applying grammatical rules, or adding punctuation to the transcribed text.
8. Output: The final output of the ASR system is the recognized text, representing the spoken input in written form. This output can be further processed or utilized for various downstream applications such as transcription services, voice assistants, voice-controlled systems, or any application that requires speech-to-text conversion.

It is important to note that the architecture may vary depending on the specific ASR system and the techniques used. Advanced ASR systems may incorporate additional components such as speaker diarization (identifying who is speaking) or acoustic adaptation (adapting the models to specific speakers or environments) to improve accuracy and robustness.

Overall, the architecture of an ASR system involves a combination of signal processing, statistical modeling, and language processing techniques to accurately convert spoken language into written text.

III. Conclusion

This paper provides a comprehensive overview of speech recognition techniques, explores the diverse application areas of speech recognition systems, and reviews the progress made in speech technology specifically for Indian scripts. By focusing on feature extraction techniques and classification methods, it serves as a valuable resource for newcomers in understanding the fundamentals of speech recognition and guides them in their pursuit of further research in this dynamic field.

References

[1]. Davis K. H., Biddulph R. and Balashek S., “Automatic recognition of spoken digits,” J. Acoust. Soc. Am., Volume 24, Issue 6, pp. 637-642, 1952.

[2] Huang X., and Acero A., “Spoken Language Processing: A Guide to Theory, Algorithm and System Development, Prentice Hall PTR, 2001.

[3] Wang Ye-Yi, Dong Yu, Yun-Cheng Ju and Acero A., “An introduction to voice search”, IEEE Signal Processing Magazine, 25(3), pp. 28–38,2008.

[4] Patel Bharat C and Desai Apurva A., “Recognition of spoken Gujarati Numerals Using Dynamic Time Warping” A.VNSGU Journal of Science and Technology, Volume 3, Issue 2, pp. 81-88, March 2012.

[5] Patel Himanshu N. and Virparia P.V., “A Small Vocabulary Speech Recognition for Gujarati”, International Journal of Advanced Research in Computer Science, Volume 2, Issue 1, pp. 208-210, Jan-Feb 2011.

[6] Patel Bharat C. and Desai Apurva, “Recognition of Spoken Gujarati Numeral and Its Conversion into Electronic Form”, International Journal of Engineering Research & Technology (IJERT), Volume 3, Issue 9, pp. 474-480, September- 2014.

[7] Pipalia D. S. and Dave Bhoomika, “An Approach to Increase Word Recognition Accuracy in Gujarati Language”, international journal of innovative research in computer and communication engineering (ijircce), vol. 3297, no. 9, pp. 6442–6450, 2007.

[8] Pandit Purnima and Bhatt Shardav, “Automatic Speech Recognition of Gujarati digits using Dynamic Time Warping”, International Journal of Engineering and Innovative Technology (IJETT), Volume 3, Issue 12, pp. 69-73, June 2014.

[9] Panwar Madhvi, Sharma R.P., Khan I. and Farooq O., "Design of Wavelet Based Features for Recognition of Hindi Digits", 2011 International Conference on Multimedia, Signal Processing and Communication Technologies, pp. 232-235, 2011.

[10] Kumar Kuldeep, Aggarwal R.K. and Jain Ankita, "A Hindi speech recognition system for connected words using HTK", Int. J. Computational Systems Engineering, Volume 1, Issue 1, pp. 25-32, 2012.

[11] Kumar Kuldeep and Aggarwal R. K., "Hindi speech recognition system using HTK", International Journal of Computing and Business Research, Volume 2 Issue 2, May 2011.

[12] Verma Sharmila and Mishra A.N., "Analysis of Speech Recognition Techniques on the Hindi Speech Digits Database", International Journal of Electronic Engineering Research, Volume 3, Issue 3, pp. 321-327, 2011.

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

[14] Kayte Charansing Nathoosing, "Isolated Word Recognition for Marathi Language using VQ and HMM", Science Research Reporter, Volume 2, Issue 2, pp. 161-165, April 2012.

[15] Patil P. P., and Pardeshi S. A., "Marathi connected word speech recognition system", 2014 First International Conference on Networks & Soft Computing, pp. 314-318, 2014.

[16] Khetri Gajanan Pandurang, Padme Satish L., Jain Dinesh Chnadra, Fadewar H. S., Sontakke B. R. and Pawar Vrushsen P., "Automatic Speech Recognition for Marathi Isolated Words", International Journal of Application or Innovation in Engineering & Management, Volume 1, Issue 3, pp. 69-74, November 2012.

[17] Nimbhore S. S., Ramteke G. D., and Ramteke R. J., "Pitch estimation of Marathi spoken numbers in various speech signals", 2013 International Conference on Communication and Signal Processing, pp. 405-409, 2013.

[18] Gaikwad Santosh, Gawali Bharti and Mehrotra S.C., "Polly Clinic Inquiry System using IVR in Marathi Language", International Journal of Machine Intelligence, Volume 3, Issue 3, 2011, pp-142-145.

[19] Katyal Anchal, Kaur Amanpreet and Gill Jasmeen, "Punjabi Speech Recognition of Isolated Words Using Compound EEMD and Neural Network", International Journal of Soft Computing and Engineering (IJSCE), Volume-4, Issue-1, pp. 150-154, March 2014.

[20] Ghai Wiqas and Singh Navdeep, "Continuous Speech Recognition for Punjabi Language", International Journal of Computer Applications, Volume 72, Issue 14, May 2013, pp. 23-28.

[21] Ravinder Kumar," Comparison of HMM and DTW for Isolated Word Recognition System of Punjabi Language", International Journal of Soft Computing, Volume 5, Issue 3, pp. 244-252, February 2010.

[22] Ghai Wiqas and Singh Navdeep, "Phone based acoustic modeling for automatic speech recognition for Punjabi language", Journal of Speech Sciences, Volume 1, Issue 3, pp. 69-83, 2013.

[23] Dua Mohit, Aggarwal R. K. , Kadyan Virender and Dua Shelza, "Punjabi Automatic Speech Recognition Using HTK", IJCSI International Journal of Computer Science Issues, Volume 9, Issue 4, No 1, pp. 359-364, July 2012.

[24] Ghanty Sumit Kumar, Shaikh Soharab Hossain and Chaki Nabendu, "On Recognition of Spoken Bengali Numerals", 2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM), pp. 54-59, 2010.

[25] Mandal Sandipan, Das Biswajit and Mitra Pabitra, "Shruti-II: A vernacular speech recognition system in Bengali and an application for visually impaired community", 2010 IEEE Students Technology Symposium (TechSym), pp. 229-233, 2010.

[26] Das B., Mandal S. and Mitra P., "Bengali speech corpus for continuous automatic speech recognition system", 2011 International Conference on Speech Database and Assessments (Oriental COCOSDA), pp. 51-55, 2011.

[27] Hossain M. A., Rahman M. M., Prodhan U. K. and Khan M. F., "Implementation of back-propagation neural network for isolated Bengali speech recognition", International Journal of Information Sciences and Techniques (IJIST) Volume 3, Issue 4, July 2013.

[28] Kurian Cini and Balakrishnan Kannan, "Speech Recognition of Malayalam Numbers", 2009 World Congress on Nature and Biologically Inspired Computing (NaBIC 2009), pp. 1475-1479, 2009.

[29] Mohamed Anuj and Nair K.N. Ramachandran, "Continuous Malayalam Speech Recognition Using Hidden Markov Models", A2CWiC '10: Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India, September 2010, Article No.: 59, pp. 1–4.

[30] Anand A. V., Shobana Devi P., Stephen J. and Bhadran V. K., "Malayalam Speech Recognition system and its application for visually impaired people", 2012 Annual IEEE India Conference (INDICON), pp. 619-624, 2012.

[31] Kurian C. and Balakrishnan K., "Connected digit speech recognition system for Malayalam language", Sadhana, Volume 38, Issue 6, pp. 1339–1346, 2013.

[32] Sreejith C and Reghuraj P C, "Isolated Spoken Word Identification in Malayalam using Mel-frequency Cepstral

Coefficients and K-means clustering", International Journal of Science and Research (IJSR), Volume 1, Issue 3, pp. 163-167, December 2012.

[33] Hemakumar G. and Punitha P., "Speaker Independent Isolated Kannada Word Recognizer", Multimedia Processing, Communication and Computing Applications, pp. 333–345, 2013.

[34] Kannadaguli Prashanth and Bhat Vidya, "Phoneme Modeling for Speech Recognition in Kannada using Multivariate Bayesian Classifier", SSRG International Journal of Electronics and Communication Engineering (SSRG-IJECE), volume 1, issue 9, pp. 1-4, Nov 2014.

[35] Hegde S., Achary K.K. and Shetty S., "Isolated Word Recognition for Kannada Language Using Support Vector Machine", In: Venugopal, K.R., Patnaik, L.M. (eds) Wireless Networks and Computational Intelligence. ICIP 2012. Communications in Computer and Information Science, vol 292. Springer, Berlin, Heidelberg, pp 262–269, 2012.

[36] Muralikrishna H., Ananthakrishna T., and Shama K., "HMM based isolated Kannada digit recognition system using MFCC", 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 730-733, 2013.

[37] Vimala.C and Radha.V, "Suitable Feature Extraction and Speech Recognition Technique for Isolated Tamil Spoken Words", International Journal of Computer Science and Information Technologies, Volume 5, Issue 1, pp. 378-383, 2014.

[38] Karpagavalli S., K. Usha Rani, R. Deepika and P. Kokila, " Isolated Tamil Digits Speech Recognition using Vector Quantization", International Journal of Engineering Research & Technology (IJERT) Volume 1, Issue 4, pp. 1-12, June – 2012.

[39] Vimla C. M. and Radha V., "Speaker Independent Isolated Speech Recognition System for Tamil Language using HMM", Procedia Engineering, 30, pp. 1097–1102, 2012.

[40] Kumar C. S., and Foo Say Wei, "A bilingual speech recognition system for English and Tamil", Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint, pp. 1641-1644, 2003.

[41] Thangarajan R., Natarajan A. M. and Selvam M., "Syllable modeling in continuous speech recognition for Tamil language", International Journal of Speech Technology, Volume 12, Issue 1, pp. 47–57, 2009.