

Review of Development of Speech corpora and speech recognition research in Hindi

Dr.Harshalata Petkar

ABSTRACT

Communication of information and interaction is advocating very progressive changes due to technological elevation spoken language is not only natural but also faster way of communication among human being. Speech recognition and natural language understanding is bottleneck in realization of conclusive man-machine interface speech recognition along with the wide applicability in domain of computer science, medical science, psychology sports, neurology linguistics has increased demand for speech and text data. This speech corpora is utilized for the development of acoustic and language models which can be used for training automatic speech recognition, synthesis or translation system applicability of automatic speech recognition is amplified and reach to the masses if only developed in native languages. Hindi is one of the most widely spoken languages in the world and is major language in India. Manifold expression of Hindi language and its utility deviates the scientist towards thought-provoking research. This paper review the speech corpora development developed by different agencies and researcher speech recognition research in Hindi languages. In this paper 26 speech corpora is surveyed and their status and utility is observed also Hindi speech recognition system are review which is developed by 20 different researchers.

I. INTRODUCTION

The possession of speech is the grand distinctive character of man. It involves invention of words and the development of the ability to speak from understand them, and think with them. [1] All of these thinks are quite unprecedented in the animal kingdom. Spoken language is a paramount importance because this is one of the vehicle for human beings to communicate. Transmission of though heritage is possible through diverse channels like spoken language [2]. It has significant links to both national identity and individual existence. The use of spoken language to constant identity has been explored in education. [3]Impact of information technology on language is widely acknowledged and is key to improve the economical, social environmental resources in the country. In the area of speech and language technology research development in speech and language include speech recognition, synthesis, speaker identification language identification and message understanding, speech forensic large amount of annotated speech data is needed to address the research quotations particularly speech recognition variability in speaker language. To modal the variability speech corpora is essential. Current recognition technologies, adopt HMM ANN for training and testing purpose. They require a large amount of training data to perform well [4].Looking to the importance of speech corpora in speech recognition research, this paper reviews both the development of Hindi speech corpora and research ventures in speech recognition system.

This paper is organized in 5 sections. Section 2 represents the significance of speech corpora and its development break through. section 3 reviews the speech corpora development for Indian languages in general and Hindi language in particular. Section 4 furnish details about ASR basics. Review of research accomplishment by academicians are presented in section 5 .Section 6 disseminate discussion and conclusion.

II. SIGNIFICANCE OF SPEECH OF CORPORA

Speech corpus is used to create the acoustic model pertaining to research task in speech recognition and speech synthesis. In linguistics spoken corpora is used to impart the research requirement in phonetic, conversation analysis, dialectology etc. Accordingly development of annotated speech corpora and acoustic-phonetic databases are prerequisite for research in spoken language domain.

These are several problems in studying spoken language. It is necessary to use the written form of language to study the speech. On the other hand transcription does not provide the phonetic and prosodic information and can not present the complexity of spoken language. The form of speech and writing are completely different and these inevitably contain features present in one form only, which cannot fully serve in another. Another problem reported is to recognize the speech unit and its emotional aspects. To attempt the problem in spoken

language, speech corpora can be developed with the following directions[5].

A.Selecting the text corpora:-For developing grammatically and phonetically rich speech corpora, text corpora can be selected carefully.Properly transcribed text selection proceeds with research in spoken language processing the type of speech recognition viz,isolated ,continuous,spontaneous.

B.Speaker selection:-The speech data collected are spoken by selected speaker. The speaker will be selected to cover the complete diversity of age, gender literacy group, language and emotional states.

C. Speech recording:-The speech recording can be performed in a quiet computer room/professional recording studio using a standard headset microphone connected to the computer / recorder. The distance of headset from microphone to mouth and recording level was kept constant.For recording PRAT,Gold Wave Editor,Audicity can be used which are open source. Speech can be recorded in realistic conditions like hotels/restaurant.

D. Data Collection:-The speaker is asked about speaking the words with the number of utterances of every word.Proper variation of the speakers can be chosen so that system will be speaker independent.

III. DEVELOPMENT OF SPEECH CORPORA IN HINDI

Hindi is fourth most widely spoken language in the world.Approximately 405 million people speak Hindi and Urdu worldwide[6].This fact gives chance to research community to unfold the challenges in spoken Hindi language. To model various issues speech corpora is principle requirement. This section reviews the speech corpora development in Hindi language.

CDAC Noida has developed the parallel text corpora Gyan-Nidhi for 12 Indian languages, including Hindi[7].

Corpora have 1000 phonetically rich sentences and 10,000 most frequent words set. Phonetically rich sentences [8] consist of words having C³oVC³o type monosyllables in maximum numbers. These syllables may occur at starting, middle or at end position in the word or it may occur in isolation. A vocabulary of monetary and temporal expression is created first covers words related to Digits, Days, Months, Years, Time, Quantitative units and currency etc. in addition words having most frequent cluster also have been recorded.

The database also contains prosodically representative sentences set of 1000 sentences for different categories. This set contains sentences, reflecting anger, joy, sadness question type sentences, negative command exclamation, etc. reaction is done in 16-bit PCM mono with sampling rate of 44.1KHz in noise free and echo canceled studies. Professional speakers (Male & Female)

where use of recording to maintain a constant pitch and prevent strong phenomenon. The speech units are tagged in a hierarchical manner consisting of sentences, word, syllable and phoneme levels.

EMILLEE-CIIL [9] speech corpus is developed using two approaches normally BNC and Asia-Radio programmer's recording for Hindi and other Indian language

In BNC approach, collection of speech data consists of recordings of everyday conversations among families and friends are taken. Under Asia-Radio programmers recording BBCUK and Asia Network broadcastings are recorded in 5 languages Bengali, Gujarati, Hindi, Urdu and Panjabi. It includes recordings of news, review, interview and telephonic interview. This gives a range in recordings in terms of number of speakers of gender differences. Recordings are digitized, edited and transcribed with time alignment. The recordings are done for 26 lakh words (about 5 lakh words of each language). The data has been annotated with information made available for speaker ID, sex, age, profession and relationships etc.

CDAC Noida is also working on development of Hindi speech database in collaboration with ELDA, France. The scope includes development of Annotated Speech Corpora for Hindi by 2000 persons. Recording of speech corpora in various environments like Home/Office, Moving vehicle, Public place, Street etc. by speakers having Hindi as first language using GSM network is being done. Speech Corpora covers Hindi speech prompts spoken by people from different dialectal regions and age groups. The scope includes designing a set of Machine readable Phonetic Alphabet Set, Development of phoneme rich & balanced sentences and words, Annotation/Transcription of recorded speech.

The LILA Hindi Belt database collected in India was recorded within the scope of the LILA project. It contains the recordings of 2,023 Hindi speakers (1,011 males and 1,012 females, all speakers with Hindi as first language) recorded over the Indian mobile telephone network. Five dialectal regions have been selected as best representing major Hindi variants in the Hindi belt, focusing on the capital New Delhi. The regions are: Delhi, Northern Uttar Pradesh, Awadh, Rajasthan and central India [14]. Various acoustic conditions were selected as representative of mobile user's environment. Several environments like, passenger in moving car, railway, bus, public places, stationary pedestrian by road side, home/office environment, passenger in moving car with hands-free kit. Speech files are stored as sequence of 8-bit-8KHz. A raw speech file is not compressed. Speaker utterance includes isolated digits, connected digits, natural numbers, currency,

yes/no question, time phrses, application words, phonetically rich word, phonetically rich sentence, person name etc.

CFSL, Chandigarh has developed Speaker Identification Database (SPID) for English and Hindi. Isolated and contextual speech samples generally used in conversation are taken for database creation. Contemporary and Non-Contemporary recording of speech samples in 10 Different modes of normal and disguised speaking is done Speaker Identification system in text independent mode is being developed for forensic applications.

ICS, Hyderabad developed speech database for Telugu and Hindi for development of Interactive Voice Response Systems. Telugu and Hindi words are used as Macro-syllabus. This is a general purpose database and they have used the same for synthesizing the content on their website called iKissan.com.

At Devi Ahilya University, Indore [10] a speech database has been developed that consists of 1000 Hindi phonetically rich sentences spoken by 100 native speakers. The wideband speech was recorded at Indore in 16-bit PCM format with a sampling frequency of 16,000Hz. This data is being used for studying the effect of noise enhancement techniques for speech recognition and for comparing differences in Hindi spoken in Hindi speaking states and other states of India.

At CDAC Noida Text and Speech corpora is being developed in collaboration with Scientific Analysis Group, DRDO for three Indian Languages namely Hindi, Bengali and Manipuri. The scope includes different sets of speech recorded and annotated at phoneme level, in multiple environments like quite, normal and noisy office and in moving conditions by 30 speakers (equal number of male and

female speakers) for 30 minutes duration by each speaker. Data in Hindi is also getting recorded by Bengali and Manipuri speakers. It will lead to the development of Language Identification, Speech Recognition and Speaker Identification system useful for defense applications.

A Continuous speech database of Hindi and Telugu languages has been created by IIT, Chennai [11]. The database consists of a large number of sentences from news bulletins. These sentences have been segmented at phoneme levels and other speech units. Labeling is also done. An algorithm was used for automatic segmentation and labeling of phonemes and syllables etc. These were judged manually and corrected. Each argument of the Corpora has been annotated with several features required for an application. The database has been used for developing speech systems such as TTS and speech recognition System for these languages. HP Labs India developed Hindi & Indian language database for TTS[12].

IIT Khargpur developed Speech database for Hindi including Marathi & Urdu, 180 spekers both male and female where use for recording which is done at 22,050Hz and 10 repetitions. Various environment conditions are exposed (Train/Collegs/Slums/remote village/Research labs/Farms) while speech data collection. This is used for text independent speaker identification in ASR[13].

IV. FUNDAMENTAL OF ASR

The basic structure of a speech recognition system is shown in figure 1.

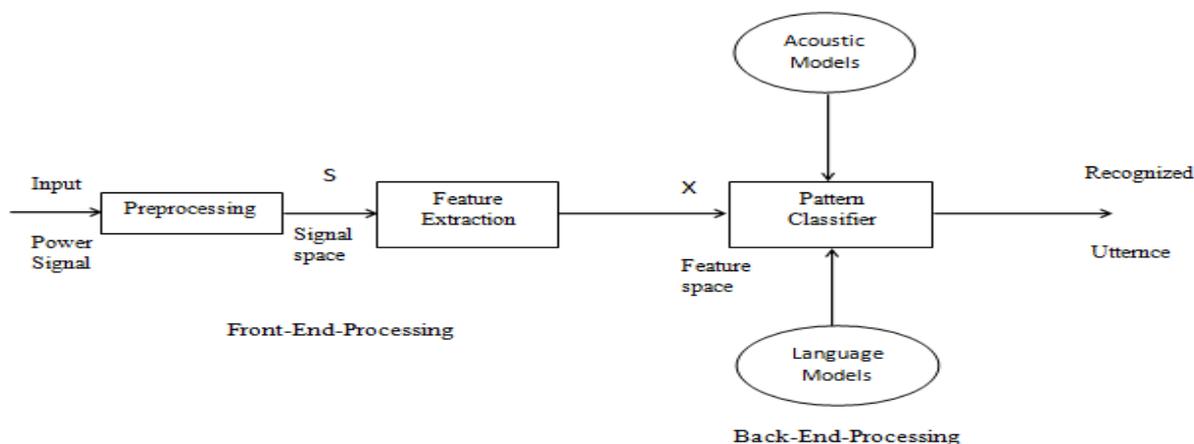


Figure1: Architecture of ASR

It covers mainly these tasks- A/D conversion, Background noise filtering, Pre emphasis, Blocking and Windowing. In first task

A. Preprocessing

analog electrical signals are digitized, i.e. these are converted into a discrete-time, discrete-valued signal. This process of analog to digital conversion has two steps: sampling and quantization. A signal is sampled by measuring its amplitude at a particular time; the sampling rate is the number of samples taken per second. In general, a sampling rate between 8 and 20 kHz is used for speech recognition application [15]. As a point of reference, perceptual studies generally indicate that frequencies up to about 10 kHz (10,000 cycles per second) occur in speech, but speech remains intelligible within a considerably narrower range. The second important factor is the quantization factor, which determines what scale is used to represent the signal intensity. Generally, it appears that 11-bit numbers capture sufficient information, although by using a log scale, we can get by with 8-bit numbers. In fact, most current speech recognition systems end up classifying each segment of signal into only one of 256 distinct categories.

So a typical representation of a speech signal is a stream of 8-bit numbers at the rate of 10,000 numbers per second clearly a large amount of data. The challenge for speech recognition system is to reduce this data to some manageable representation.

Once signal conversion is complete, background noise is filtered to keep SNR high. While speech capturing, background noise and silence (absence of speech) will also be quantized with speech data. An important problem in speech processing is to detect the presence of speech in a background of noise and silence from our recording speech. This problem is often referred to as the endpoint detection problem [16]. By accurately detecting the beginning and end of an utterance the amount of processing can be kept at minimum. According to literature [17], accurate determination of end points is not very difficult if the signal-to-noise ratio is high, say greater than 60 dB.

The next step is the pre-emphasis. The motivation behind it is to emphasize the important frequency components in the signal (i.e. amplify important areas of the spectrum) by spectrally flatten the signal. For example hearing is more sensitive in the 1 KHz-5 KHz region of the spectrum. It amplifies this area of spectrum, assisting the spectral analysis algorithm in modeling the most perceptually important aspects of the speech spectrum.

B. Feature Extraction/Parametric Transform

The goal of feature extraction is to find a set of properties of an utterance that have acoustic correlations in the speech signal, that is parameters that can somehow be computed or estimated through processing of the signal waveform. Such parameters are termed as features. Feature extraction is the

parameterization of the speech signal. It includes the process of measuring some important characteristic of the signal such as energy or frequency response (i.e. signal measurement), augmenting these measurements with some perceptually meaningful derived measurements (i.e. signal parameterization), and statically conditioning these numbers to form observation vectors. There are several ways to extract features from speech signal as given below:

- Linear Predictive Cepstral Coefficients (LPCC)
- Mel Frequency Cepstral Coefficients (MFCC) [18]
- Wavelet as feature extractor [19]
- Missing feature approach [20]

C. Acoustic Modeling

In this subsystem, the connection between the acoustic information and phonetics is established. Speech unit is mapped to its acoustic counterpart using temporal models as speech is a temporal signal. There are many models for this purpose like,

- Hidden Markov Model (HMM)
- Artificial Neural Network [21]
- Dynamic Bayesian Network (DBN) [22]

ANN is a general pattern recognition model which found its use in ASR in the early years. Rabiner (1991), first suggested the HMM approach leading to substantial performance improvement. Current major ASR systems use HMM for acoustic modeling. Since then researchers have tried to optimize this model for memory and computation requirements. In the current state, it seems that HMM has given the best it could and now we need to find other models to go ahead in this domain. This leads to consideration of other models in which Dynamic Bayesian Network seems a promising direction. Still our experiments and verification has been done on a HMM based system.

D. Language Modeling

The goal of language modeling is to produce accurate value of probability of a word W , $Pr(w)$. A language model contains the structural constraints available in the language to generate the probabilities. Intuitively speaking, it determines the probability of a word occurring after a word sequence. It is easy to see that each language has its own constraints for validity. The method and complexity of modeling language would vary with the speech application. This leads to mainly two approaches for language modeling. Generally, small vocabulary constrained tasks like phone dialing can be modeled by grammar based approach where as large applications like broadcast news transcription require stochastic approach.

V. SURVEY OF HINDI SPEECH RECOGNITION

In 1950, the earliest attempt in ASR research exploited fundamental ideas acoustic phonetics to try to device system for recognizing phonetics (Try & Denes 1959) and recognizing of isolated digits from a single speaker (Davis Etal 1952). These attempt continued in the 1960's by the entry of several Japanese laboratories such as Radio Research Lab, NEC and Kyoto University to the arena. Research and development activity for India languages study early in 2000. This section attempt to review of speech recognition research for Hindi language.

Tarun Pruthi et al. [23] described a speaker-dependent, isolated word recognizer for Hindi. Features are extracted using LPC and system was developed using HMM. He recorded the sounds of two male speakers. The vocabulary used by him consists of Hindi digits (0, pronounced as "shoonya" to 9, pronounced as "nau"). However the system is giving good performance, but the design is speaker dependent and have very small vocabulary [21].

Gupta [24] worked for an isolated word speech recognition tool for Hindi language. He made use of continuous HMM. Word acoustic model was used for recognition. Again the word vocabulary contains digits of Hindi language. Results was good when tested for speaker dependent model. The results are satisfactory for other sounds too. Main drawback is that vocabulary size is too small.

Anup Kumar Paul, Dipankar Das et al [25] developed Bangla Speech Recognition System using LPC and ANN (automatic neural network). This paper presents recognition system of the Bangla speech. There are two major parts of Bangla speech recognition system. Signal processing is the first part and the second part is speech pattern recognition technique. Starting and end point are detected during the speech processing stage. Pattern recognition is done in the second part Artificial Neural Network (ANN). Speech signals are recorded using an audio wave recorder in the normal room environment.

Al-Qatab et al. [26] implemented an Arabic automatic speech recognition engine using HTK. The engine recognized both continuous speech as well as isolated words. The developed system used an Arabic dictionary built manually by the speech-sounds of 13 speakers and it used vocabulary of 33 words. R.L.K. Venkateswarlu, R. Ravi Teja et al [27] Developed Efficient Speech Recognition System for Telugu Letter Recognition. In this research, both MLP (Multilayer Perceptron) and TLRN (Time Lagged Recurrent Neural Network) models were trained and tested on a dataset that consists of Four different speakers (2 Male and 2 Female) are allowed to utter the letters for 10 times. Speaker dependent mode is used for recognition of the Telugu letters.

Tested data presented to the network are same as the trained data in this mode.

R. Kumar [28] implemented a system which involves only one speaker and recognize isolated word for the Punjabi language and further extended its work to compare the performance of speech recognition system for small vocabulary of speaker dependent isolated spoken words using the Hidden Markov Model (HMM) and Dynamic Time Warp (DTW) technique. The presented work emphasized on template-based recognizer approach using linear predictive coding with dynamic programming computation and vector quantization with Hidden Markov Model based recognizers in isolated word recognition tasks.

Ahmad A. M. Abushariah [29] worked for English speech recognition. This paper aims to design and implement English digits speech recognition system using MATLAB (GUI). This work was based on the Hidden Markov Model (HMM), which provides a highly reliable way for recognizing speech. Mel Frequency Cepstral Coefficients (MFCC) technique was used to extract the features. This paper focuses on all English digits from (Zero through Nine).

M Singh et al. [30] described a speaker independent, real time, isolated word ASR system for the Punjabi language was developed by The Vector Quantization and Dynamic Time Warping (DTW) approaches were used for the recognition system. The database of the features (LPC Coefficients or LPC derived coefficients) of the training data was created for training the system and for testing the system the test pattern (features of the test token) was compared with each reference pattern using dynamic time warp alignment. The system was developed for small isolated word vocabulary.

Bharti W. Gawali, Santosh Gaikwad et al [31] presents a Marathi database and isolated Word recognition system. Mel-frequency cepstral coefficient (MFCC), and Distance Time Warping (DTW) are used for extraction of features. Marathi speech database has been designed by using the Computerized Speech Lab. For the extraction of the feature. The vocabulary consists of the vowels of Marathi and isolated words which start from a vowel and simple Marathi sentences. In this paper voice of 35 speakers was recorded and each word was repeated 3 times. The comparative recognition accuracy of DTW and MFCC was presented in this paper.

Kuldeep kumar [32] worked for ASR for Hindi language. This paper aims to build a speech recognition system for Hindi language. System is developed using Hidden Markov Model Toolkit (HTK). Acoustic word model is used to recognize the isolated words. The system is trained for 30 Hindi words. Training data was collected from eight speakers. The overall accuracy of the presented

system is 94% K. Kumar et al. [33] worked for connected-words speech recognition system for Hindi language. HiddenMarkov model toolkit (HTK) was used to develop the system and the system was trained to recognize anysequence of words. R Kumar et al[30] worked for connected words speech recognition system Sharmila et al [35] Perform with isolated digits in Hindi Implementation of system has done using Hidden Markov model (file) Annu et al[36] worded for automatic speech recognition system for isolated and connected words of Hindi language by using Hidden Markov Model Toolkit Gaurav et al [37] Worked for development of application specific continuous speech recognition system in Hindi work was limited to 29 phonemes of Hindi

PreetiSaini et al [38] work to develop and implement speech recognition system for Hindi language using the HTK open source toolkit system is sensitive to changing spoken methods and changing scenario. Ganesh Sivaraman et al [39] worked for Hindi speech recognition and online speaker adaptation speaker adaptation is performed using maximum likelihood linear regression (MLLR) transformation approach and ASR is train using narrowband speed result shows improvement in result by minimizing the difference between test and train condition system work for just one sentence spoken by the speaker provided sufficient computing resources are made available. ShwetaSinha et al [40] Worked for speech processing for Hindi dialat recognition. Author have used 2-layer feed forward neural network performance is observed with spectral features and prosodic features as input recognition score is 79% corpora is called from 15 speakers from each dileat.

VI. CONCLUSION

A future of speech recognition is enchanting although technology has been around for over 50 years there are hundreds potential applications and improvements to be involved in research. To promote the research and development activities in India domain, one must focus on dramatic possibilities of speech recognition. To analyses and exploit the hidden pots speech corpora, is basic requirements. In this paper speech corpora development activities for Hindi language is reviewed speech corpora can be developed further with the Hindi delects and multimodal environment. Paper explores the speech recognition research with their shortcomings speech corpora development and speech recognition research work need to reinforce for the Hindi language where the possibility to interpret human through automatically by computer can not be rejected.

REFREANCE

- [1] Huxley, T. H. (1863/2005). *Collected Essays of Thomas Huxley: Man's Place in Nature and Other Anthropological Essays*. WhiteWsh, MT: Kissinger Publishing. (1871). 'On the relations of man to the lower animals.' *Collected Essays, Volume 7*. <http://aleph0.clarku.edu/huxley/CE7RelM-L-A.html>
- [2] Hudson, R. A. (1980). *Sociolinguistics*. Great Britain: Cambridge University Press. Ch. 2, 3.
- [3] Adger, C. T. (1998). Register shifting with dialect resources in instructional discourse. In S. Hoyle & C. T. Adger (Eds.), *Kids talk: Strategic language use in later childhood* (pp. 151-169). New York: Oxford University Press.
- [4] Dan El lisy and Nelson Morgany ,size matters: An empirical study of neural network training for large Vocabulary continuous speech recognition,;International Computer Science Institute, 1947 Center St, Berkeley, CA 94704,University of California at Berkeley, EECS Department, Berkeley, CA 94720
- [5] E. Douglas-Cowie, N. Campbell, R. Cowie, and P. Roach, "Emotionalspeech: towards a new generation of databases," *Speech Communication*, vol. 40, pp. 33-60, April 2003.
- [6] H. Sil (1999), *Ethnologue: Languages of the World*. <http://www.ethnologue.com/web.asp> (visited 23, October, 2012)
- [7] SS Agrawal, SunitaArora&KaruneshArora, *Towards design, development and standardization of speech Corpora for developing Indian language TTS System*, COCOSDA-2005, Dec 6-8, 2005, Jakarta , Indonesia , pp 29-33 6.
- [8] KaruneshArora, SunitaArora, KapilVerma, S SAgrawal, *Automatic Extraction of Phonetically Rich Sentences from Large Text Corpus of Indian Languages*, INTERSPEECH2004 -ICSLP Jeju, Korea
- [9] Paul Baker, Andrew Hardie, Tony McEnery, and Sri B D Jayaram, *Constructing Corpora of South Asian Languages*.
- [10] Samudravijaya K, P.V.S.Rao, and S.S.Agrawal, *Hindi Speech Database*, Proc. Int. Conf. on Spoken Language processing (ICSLP00), Beijing, China, October 2000.
- [11] *Database for Indian Languages , Speech & Vision Lab. IIT Madras, Chennai-2001*
- [12] Shweta Sinha, Pooja Singh, Shyam Agarwal ,JesperOlsen,"Development of Text And Speech Databse For Hindi And Indian English Specific to Mobile Communication Environment", *The Language Resources and*

- Evaluation Conference, LREC, Istanbul, Turkey (2012).
- [13] KSreenivasa and Shashidhar G Koolagudi, "Identification of Hindi Dialects and Emotions using Spectral and Paosodic features of Speech", *Systematic Cybernetics and Informatics*, Volume 9, 4 Nov. 2011.
- [14] http://catalog.elra.info/product_info.php?products_id=1170 on dated 17 May 2014.
- [15] James Allen. Third Edition 2005. *Natural Language Understanding*, Pearson Education.
- [16] Lawrence R. Rabiner and S. Atal Bishnu. 1976. "A Pattern Recognition Approach to Voice-Unvoiced- Silence Classification with Applications to Speech Recognition", *IEEE Transaction ASSP-24*(3).
- [17] D. Raj Reddy. 1976. "Speech Recognition by Machine: A Review", *IEEE* 64(4): 502-531.
- [18] M. D. Skowronski and J.G.Harris. 2002. Increased MFCC Filter Bandwidth For Noise-Robust Phoneme Recognition. *IEEE*.
- [19] A.E. Biem. February 2001. Discriminative feature extraction applied to speech recognition In *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 9, pp. 96-108.
- [20] B. Raj and Stern. 2005. Missing-Feature pproaches in Speech recognition. *Signal Processing magazine, IEEE* Volume 22, Issue 5, pp. 101-116.
- [21] K.S. Rao, B. Yegnanarayana. May 2004 Modelling Syllable Duration in Indian Languages Using Neural Networks. In proceeding of ICASSP Montreal, Qubic, Canada, pp 313-316.
- [22] Q. Huo, H. Jiang, & C.H. Lee. 1997. A Bayesian predictive classification approach to robust speech recognition, in proc. *IEEE ICASSP*, pp.1547-1550.
- [23] Pruthi T, Saksena, S and Das, P K (2000) Swaranjali: Isolated Word Recognition for Hindi Language using HMM and VQ. *International Conference on Multimedia Processing and Systems (ICMPS)*, IIT Madras.
- [24] Gupta, R (2006) *Speech Recognition for Hindi*, M. Tech. Project Report, Department of CSE, IIT, Mumbai.
- [25] Anup Kumar Paul, Dipankar Das et al (2009) *Bangla Speech Recognition System using LPC and ANN*
- [26] B.A.Q Al-Qatab (2010), "Arabic Speech Recognition Using Hidden Markov Model Toolkit (HTK)", Paper presented at *International Symposium in Information Technology (ITSim)*. Kuala Lumpur, June 15-17, 2010
- [27] R.L.K Venkateswarlu, R. Ravi Teja et al (2010) "Developing Efficient Speech Recognition System for Telugu Letter Recognition"
- [28] R. Kumar (2010) "Comparison of HMM and DTW for Isolated Word Recognition of Punjabi Language"
- [29] Ahmad A. M. Abushariah (2010) *English Digits Speech Recognition System Based on Hidden Markov Models*.
- [30] R. Kumar and M. Singh (2011), "Spoken isolated Word Recognition of Punjabi Language Using dynamic time Warping Technique"
- [31] Bharti W. Gawali¹, Santosh Gaikwad et al (2011) "Marathi isolated word Recognition System using MFCC and DTW Features"
- [32] Kuldeep Kumar and R. K. Aggarwal (2011) "Hindi Speech Recognition System using HTK" *International journal of Computing*
- [33] K. Kumar, R. K. Aggarwal, and A. Jain (2012) "A Hindi speech recognition system for connected words using HTK" *International Journal Computational System Engineering*.
- [34] "Hidden Markov Model Toolkit" HTK, available at <http://htk.eng.cam.ac.uk>, 2012.
- [35] Sharmila, Dr. Neeta Awasthy, Dr .R.K.Singh "Performance of Hindi Speech Isolated Digits In HTK Environment" *IOSR Journal of Engineering* May. 2012, Vol. 2(5) pp: 1020-1023.
- [36] Annu Choudhary, Mr. R.S. Chauhan, Mr. Gautam Gupta, 2013, *Automatic Speech Recognition System for Isolated & Connected Words of Hindi Language by Using Hidden Markov Model Toolkit (HTK)*, AETS. 2013. 3. 234, Proc. of Int. Conf. on Emerging Trends in Engineering and Technology.
- [37] Gaurav, Devanesamoni Shakina Deiv, Gopal Krishna Sharma, Mahua Bhattacharya, 2012 "Development of Application Specific Continuous Speech Recognition System in Hindi" *Journal of Signal and Information Processing*, 2012, 3, 394-401.
- [38] Preeti Saini, Parmeet Kaur, Mohit Dua "Hindi Automatic Speech Recognition Using HTK," *International Journal Of Engineering Trends And Technology*, vol 4, issue-June 2013.
- [39] G. Sivaraman and K. Samudravijaya, "Hindi Speech Recognition and Online Speaker Adaptation," *Proceed-ings of ICTSM 2011*, Vol. 145, 2011, pp. 233-238.

- [40] ShwetaSinha, Aruna Jain, Shyam S. Agrawal” Speech Processing for Hindi Dialect Recognition” ,Advances in Intelligent Systems and Computing Volume 264, 2014, pp 161-169

International Journal of Engineering Research and Applications (IJERA) is UGC approved Journal with Sl. No. 4525, Journal no. 47088.

Dr.Harshalata Petkar. " Review of Development of Speech corpora and speech recognition research in Hindi " **International Journal of Engineering Research and Applications (IJERA)** 7.7 (2017): 12-19