International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 06 | June - 2021 ISSN: 2582-3930

HINDI DIALECTS CONVERT INTO PURER HINDI LANGUAGE WITH TRANSLATION TECHNIQUES FOR INTERACTION

Ms. Dimpal Sahu[1], Mrs. Ati Jain[2]

[1]Computer Science & engineering & SAGE University, Indore
[2]Computer Science & engineering & SAGE University, Indore

Abstract - As we know that the more than 19,500 languages or dialects are spoken in India as mother tongues. This will be worked out as software, based on speech processing fundamentals and language translation techniques; it will be a research based software system. Technically, the input would be a speech signal and text form from the country India i.e. villagers (illiterate) and cities in his different-different dialects i.e. Nimadi, Malawi, Bhagheli, Bhundelkhandi etc. and accent in 'Hindi'. We shall apply dialect and accent filtering to generate normal form Hindi input signal. This purer Hindi signal will be translated to different- different country mother tongues i.e. American, Chinese, Germany etc. Additional computational efforts will be done to enhance understandability of input and accuracy of output. Our aim is convert different-different Hindi dialects into purer Hindi Language. Under this all, we wish to study effect of speech processing techniques, pertaining to noise reduction, dialect and accent processing etc., on the continuous speech input in terms of attaining higher accuracies of computer understandings. One of the techniques from speech processing namely, Natural Language Understanding and machine translation are candidates for this work. Present flow of proposed solution is : input (in 'Hindi with regional dialect'), speech processing module Hindi to English translator, ambiguity and accuracy enhancer module, output generator in the form as Input languages.

***Key Words*:** Ambiguity remove, Slang remove, Speech Processing, Hindi Dialects, noise reduction, NLP, Machine Translation, ASR.

1.INTRODUCTION

Current research in the area of speech recognition does not only concentrate on the correct evaluation of the linguistic information embodied in the speech signal, it also works towards identifying variations naturally present in speech. Several variability due to speaker related characteristics can be observed that influences the success of speech recognition system[1]. Next to gender, dialect used by speaker is one of the major factor that influences the result of an analysis and speech processing. Characterization of the effect of dialects, together with related techniques to achieve ASR robustness is a major research topic. Characterization of the effect of dialects, together with related techniques to achieve ASR robustness is a major research topic.

Dialect of a given language is a pattern of pronunciation or vocabulary of a language used by community of native speakers belonging to same geographical area. A language when used by people from different regions can be analyzed to see the usage of words with different lexis and even if they speak some standard form of the word the difference in spectral properties of sound produced can be observed. Every individual develops a style of speaking at an early age. This style is dependent upon his native language or the dialect he speaks. When speaking some other language or even the standard form of his native language the speaker will carry traits of this style into it. This style may be influenced by sociolinguistic or regional environment of the speaker and impacts into speaking variations referred as accent. Different accents give rise to several differences in the realization of an utterance. Incorporating knowledge about dialects in pronunciation dictionary, acoustic training can increase efficiency of an ASR. Few studies have been carried out for automatic dialect identification. Linear discriminants successfully classify dialects of American English using average duration and average cestrum of phonemes. Arslan et al. proposed algorithm for English accent classification usingmel-cepstrum coefficients and energy as speech features. J C Wells in his study emphasized that accent variation does not only stretch out in phonetic characteristics but also in prosodic characteristics of speakers. Influences of acoustic, prosodic and contextual information on dialects were emphasized by Kumpf et al. in their work on Australian English accent classification. Huang et al.studied the difference among English dialects and proposed Gaussian Mixture Model(GMM) for identification of dialects. Mahnoosh et al. showed efficiency of pitch pattern for dialect recognition on Arabic dialects. A few studies on the analysis of dialects of Indian language have been carried in recent past. Most of these are based on phonological approach. Some accent based classification approach for Hindi based on acoustic characteristics has been proposed in recent times. These work uses speech samples from non native Hindi speakers and the system performance is also not noteworthy[1].

In this study our focus is on estimating parameters for dialect specific information at segmental and supra segmental level and exploring spectral and prosodic features for identification of dialects of Hindi Language. Four major dialects of Hindi; Khari Boli (KB), Bhojpuri (BP), Haryanvi (HR) and Bagheli (BG) Hindi Dialects i.e. Nimadi , Malawi have been considered for studying these differences in the spoken utterance[2]. The features so obtained are further used as input

to a feed forward neural network to show their sufficiency for mechanical dialect classification. The paper is arranged as: section 2 describes speech material and database , section three discusses the features and their analysis for different dialects, Section four presents the implementation of FFNN for dialect classification. Next section discusses the results and observations and the last section presents the conclusion[2].

2.MACHINE LEARNING

Machine learning teaches computers to do what comes naturally to humans and animals: learn from experience. Machine learning algorithms use computational methods to "learn" information directly from data without relying on a predetermined equation as a model. The algorithms adaptively improve their performance as the number of samples available for learning increases.

REAL-WORLD APPLICATIONS

With the rise in big data, machine learning has become particularly important for solving problems in areas like these:

• Computational finance, for credit scoring and algorithmic trading

• Image processing and computer vision, for face recognition, motion detection, and object or diseases detection

• Computational biology, for tumor detection, drug discovery, and DNA sequencing

• Energy production, for price and load forecasting • Automotive, aerospace, and manufacturing, for predictive maintenance

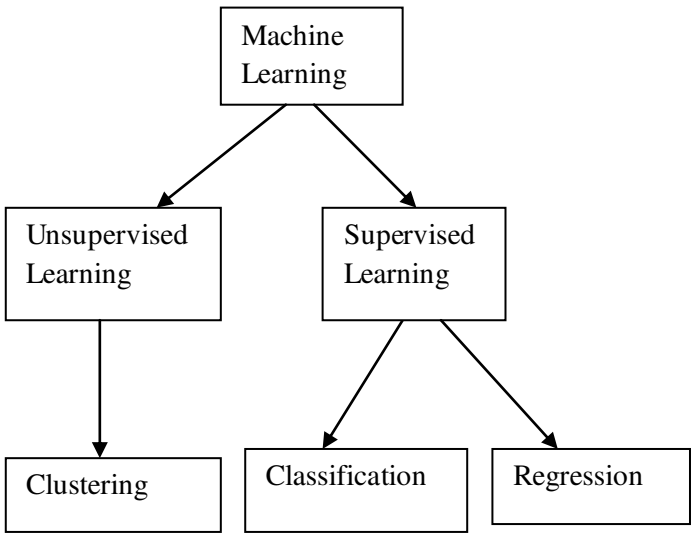
• Natural language processing

HOW MACHINE LEARNING WORKS

Machine learning uses two types of techniques: supervised learning, which trains a model on known input and output data so that it can predict future outputs, and unsupervised learning, which finds hidden patterns or intrinsic structures in input data.

Machine Translation(MT)

Machine translation (MT) is the application of computers to the task of translating texts from one natural language to another. One of the very earliest pursuits in computer science, MT has proved to be an elusive goal, but today a number of systems are available which produce output which, if not perfect, is of sufficient quality to be useful for certain specific applications, usually in the domain of technical documentation. In addition, translation software packages which are designed primarily to assist the human translator in the production of translations are enjoying increasingly popularity within professional translation organizations.



Comprehending the enormous complexity of translating human language and the inherent limitations of the current generation of translation programs is essential to understanding MT today. MT systems are designed according to one of the following parameters: coverage and Reliability. An MT system can either be designed to reproduce for a small language segment i.e. a sub-language or a controlled language with high fidelity and precision or it may be designed to perform informative, general purpose translations. In the former case, the system will have high reliability, whereas in the latter case, its coverage will be high.

However, both properties are, to a certain extent, mutually exclusive.

- Coverage refers to the extent to which a great variety of source language texts can successfully be translated into the target language. A successful translation can be described as to be informative in the sense that allows a user to understand more or less the content of the source text.

- Reliability refers to the extent to which an MT system approaches an "ideal" translation (of a restricted domain) for a given purpose or for a given user. A reliable translation is user-oriented and correct with respect to text type, terminological preferences, personal style, etc.

PRINCIPLE OF SPEECH PROCESSING

Imagine you are sitting relaxed on the sofa and just ordering your computer or laptop or cell phone to carry out simple tasks like typing a letter or carrying out a few commands. Is it possible? Of course, it is, that's where Voice recognition comes into the picture. Going by the definition it is the process of recognizing human speech and decoded it into text form. The basic principle of voice recognition involves the fact that speech or words spoken by any human being cause vibrations in the air, known as sound waves. These continuous or analog waves are digitized and processed and then decoded to appropriate words and then appropriate sentences[2].

Translation categories

Translation is categorized into four types where a computer and a man can collaborate[3]:

1.1. Machine aided human translation (MAHT)

The MAHT translation consists of using word processing software completed by electronic dictionaries, which can be improved during the translation. Translations are human-made.

1.2. Human aided machine translation (HAMT)

This category of translation requires a human assistance before and after the automatic translation (pre-edition of the source text and post edition of the target text). The Canadian Metro system is classified into this category of translation.

1.3. Interactive translation (IT)

In this category of translation, the system translates with an interactive human assistance. For each ambiguity problem during the translation process, the system asks for a human disambiguation. Alps is one of the interactive translation systems.

1.4. Machine translation (MT)

Theoretically the machine translation aims to completely avoid the human assistance to the system. Nowadays, no one of the existing machine translation systems can be qualified as being a MT system.

Natural language processing

(NLP) is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.

3. RESEARCH AND METHODOLOGY[4]

Machine translation techniques

In this section we introduce a descriptive presentation of the different machine translation techniques. These techniques are based on different models: bilingual, transfer, interlingual and corpus-based model which includes the memory-based, statistical-based and example based models.

3.1. Bilingual-based machine translation

A bilingual machine translation system is dedicated only to a pair of languages and can not be adapted to other languages. Indeed the translation process is built according to specific characteristics of the two languages. A source text in one language is analysed to be specifically generated to another language. The transfer phase is minimised to bijective lexical and syntactical relations. It is understandable that the programs may be dependant on the language pairs making difficult their adaptation to new languages. The Systran system is a collection of bilingual sub-systems dedicated to different language pairs.

3.2. Transfer-based machine translation

The transfer translation model is built on three modules:

- Analysis module that transforms the source text into a source structural description.

- Transfer module that transforms the source structural description into a target one.

- Generation module that transforms the target structural description into a target text.

3.3. Interlingual-based machine translation

The interlingual translation model is built on two main modules:

- Analysis module that transforms the source text into an interlingual description.

- Generation module that transforms the interlingual description into a target text.

3.4. Memory-based machine translation

Machine translation based on the "translation memory" is a corpus-based approach. It is dedicated to professionals or experts in the translation services. The system does not really analyses the source text to translate but just reuses possible translations previously stored by the professional translator. For the parts of text that have not been previously translated, terminology (dictionary) support is used to help the expert to translate them. This "new" translation concept offers a

computer-assisted translation that automates repetitive tasks, freeing the professional translator to attend to the finer points of translation that require the judgment of an expert.

The IBM Translation Manager [IBMTM.99] is one of the systems based on that concept. As an example, the following figure [IBMTM.98] shows the translation environment of IBM Translation Manager with the translation editor, the window for the translation memory proposals and the window for terminology support.

3.5. Statistical-based machine translation

Statistical-based machine translation is a corpus-based approach. Statistical concepts are among the first techniques for machine translation. They was proposed by Warren Weaver in the early 1949's but that theories foundered on the rocky reality of the limited computer resources of the day. In the late 1980's IBM researchers felt that the increase in computer power made reasonable a new look at the applicability of statistical techniques to translation. Statistical machine translation was re-introduced by the Candide group at the IBM Watson Research Center [CAND.94]. The principle of this translation concept is that the computer inspects large collection of translated data and, from the collection, "learns" how to translate. However the statistical techniques are no longer encouraged in the machine translation domain.

3.6. Example-based machine translation

Example-based machine translation allows to rich systems. Translation examples are stored as feature annotated and sometimes structured representations. Translation templates are generated which contain (weighted) connections in those positions where the source language and the target language equivalences are strong. In the translation phase, a multi-layered mapping from the source language into the target language takes place. Sentences are more finely decomposed into phrases and linguistic constituents e.g. NPs, PPs, subject, object, etc. The example-based approach can make use of morphological knowledge and relies on word stems as a basis for translation. Translation templates are generalised from aligned sentences

by substituting differences in sentence pairs with variables and leaving the identical sub-strings un substituted. An iterative application of this method generates translation examples and translation templates which serve as the basis for an example-based MT system.

4. LITERATURE REVIEW

Speech recognition is a term of computer science term famously known as automatic speech recognition. It is a feature that turns speech into text. It is a big advantage to people who may suffer from disabilities like blind people. The overall advantage of this is the time management.

A.1920s and 1960s[5]

As we know speech recognition came into existence in the 1920.The first speech recognition systems that were developed could only understand digits .It was Bell Laboratories that designed in 1952 the Audrey system, which identified digits spoken by a single person at a time. Then it was after ten years when IBM launched Shoebox machine that could understand 16 words at a time. Researches were also going on in the laboratories of the countries like Japan and United States and they developed other hardware that were mainly based on words that are spoken. In the beginning steps were taken for inventing machine based on automatic speech recognition system. It was then in the 1960s when several Japanese laboratories showed their capability and excitement for building a particular hardware that perform a speech recognition task. Most noticeable was the vowel recognizer of Suzuki and Nakata at laboratories in Tokyo.

B. 1970s: Speech Recognition Takes Off

Speech recognition technologies have made important developments in the 1970s .There was an program known as Speech Understanding Research program, from 1971 to 1976, it was one of the largest of its kind in the history of speech recognition.

In 1970s there were also marked some other important milestones in the technology of speech recognition. First milestone was that the area of isolated word became a usable technology which were based on fundamental studies done by Velichko and Zagoruyko in Russia,Chiba in Japan. The study in Russia helps in the advanced use of pattern recognition. Itakura research work showed the ways ideas of linear predictive coding are used in low bit rate speech coding. .Another milestone in the 1970s in the field of speech recognition was the beginning of a one of the highly successful group effort vocabulary speech recognition at IBM; researchers have studied three different tasks over time of nearly two decades.

C.1980-1990

As we know recognition was a main focus of research in the 1970s, the connected word recognition was a focus in the 1980s.Here the main motive was to develop a robust system that is able to recognize a fluently spoken string of words. Moshey J. Lasry has developed a speech recognition system based on feature in the beginning of 1980. Speech research in the 1980s was characterized by the shift in technology from template based approaches to statistical modeling methods[6]. One such approach was hidden Markov approach, although this approach was well known but not widespread.

1) Hidden Markov Model*:* It is one of the major technologies that were developed in the 1980s.It is doubly stochastic process which can only be observed through another stochastic process. It was well known in some laboratories but was not wide spread. In the 1970s Lenny of Princeton University developed a mathematical approach for this. This was concluded by Shweta Sinha, S S Aggarwal and Aruna Jain. With the technological advancement and voice based communication advantage it has driven attention towards mechanical speech recognition. Literature survey of this shows that mostly system uses Gaussian Mixture Hidden

Markov model .The evaluation of Gaussian dominates the total computational load in this approach.

The correct selection of Gaussian mixture is very crucial. In this author estimates the number of Gaussian mixture estimates[7].

2) Neural Net: This was another new technology that was launched in the 1980s.There was idea of applying neural networks to the speech recognition system. The concept of Neural Networks was first introduced in the 1950 but initially it was not proved to be useful as they had many problems in practical life. Although in 1980s a deep knowledge was introduced.

3) DARPA Program: At last in 1980s, defense Advanced Research Project Agency community sponsored a large program which aimed at getting high accuracy for about thousand words. Major research contributions were the results of efforts of CMU, BBN,SRI etc.. The DARPA program has continued up to the 1990s.

D. Segmentation for improving speech performance: The Emerging growth of information technologies has greatly influenced the trends in research to focus on speech technologies .The process of preprocessing of any speech signal serves many applications .This was provided by B. Sudhakar and R Bens Raj. It includes the following like noise removal, endpoint detection, pre-emphasis, framing, windowing. Here automatic sentence boundary detection is basic step for many applications like speech synthesis and recognition. Here this study proposes an algorithm for automatic segmentation of voice speech of Indian languages. Here entropy based method is also used that yield good performance.

E. Speech recognition under noisy conditions: The foundation approach on this system was given by Darryl Stewart, Rowan Seymour, Adrian Pass, and Ji Ming .They presented the robust and efficient weighted stream posterior .An important advantage of MWSP is that it does not need any particular measurements of the signal in any of the stream to calculate the stream weights. This has one more advantage that it can be effectively used alongside any other approach. For the evaluation we have used the large XM2VTS database. Here extensive tests are done that include both clean and corrupted utterances with noise added in both the audio and video streams. The experiments have been conducted that show this approach gives excellent performance in comparison to another dynamic approach.

F. Multimodal speech recognition: Humans make use of multimodal communication when they communicate with each other. Studies of speech have shown that having visual and audio information enhance the rate of successful transfer of information, even when the message to be transferred is complicated .Here we can use visual information like lip reading. Jerome R. have contributed lot in speech recognition.

5.TECHNOLOGIES USED LITERATURE REVIEW[6]

There are various technologies used for speech recognition techniques are as follows in Table 1.

Technologies	Type	Recognition type	Vocabulary size allowed	Recognition quality	Common usage	Free
Microsoft Speech API	Windows COM API	Dictation Complex Phrase Recognition	Large	High	Desktop Application Development	NO
Microsoft .Net System. Speech namespace	Windows .NET API	Dictation Complex Phrase Recognition	Large	High	Desktop Application Development	NO
Microsoft Speech Platform	Windows API	Command Isolated word recognition	Large	NA	Server Application Development	NO
Microsoft Unified Communication System	Windows API	Command Isolated word recognition	Large	NA	Server Application Development	NO
Sphinx 4	Complete Framework for SR	Dictation Complex Phrase Recognition	Large Depend on implementation	Depends on Implementation	Speech application Development	YES
HTK	Complete Framework for SR	Dictation Complex Phrase Recognition	Large Depend on implementation	Depends on Implementation	Speech Application Development	No, Licence needed
Jlius	Decoder system	Dictation Complex Phrase Recognition	Large	Depends on Implementation	Speech Application Development	YES
Java Speech API	Specification	NA	NA	Depends on Implementation	NA	YES
Google Web Speech API	JavaScript API for Chrome	Dictation Complex Phrase Recognition	Large	Average	Web Application Development	YES
Nuance Dragon SDK	API for desktop and mobile application	Dictation Complex Phrase Recognition	Large	High	Client, server and mobile Application Development	NO

6. PROPOSED ARCHITECTURE OF SYSTEM TO BE BUILT

This architecture is presented to implementation approach of our system –The solution of the above problem is to make a software or website. We can understand by the following system architecture of the given problem.

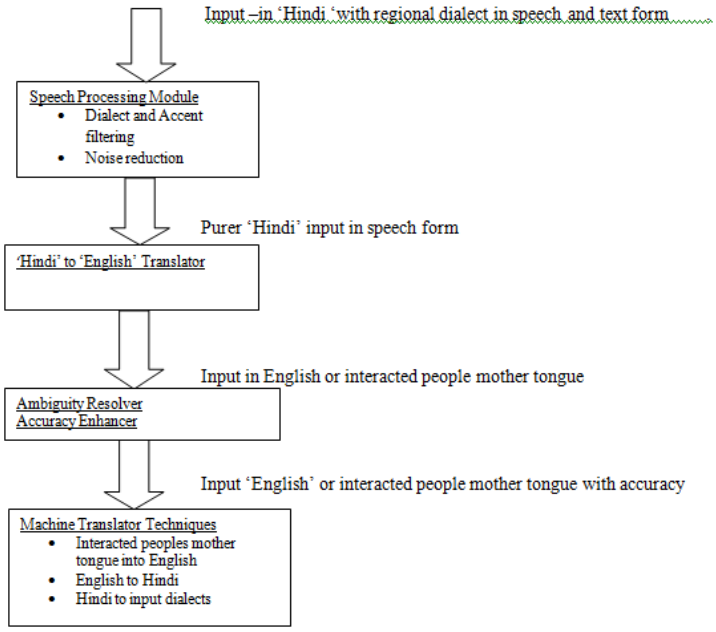


Fig 1: System Architecture

7. IMPLEMENTATION APPROACH

We will try to implement our system to convert Hindi Dialects(Nimadi, Malawi) into purer Hindi and Hindi to English Language for interaction between more than one person from different-different countries.

There are various steps to be follows to implement our system:-

Step 1: Give the input in the form of Speech signal and text form.

Step 2: Dialects accent filtering and noise reduction module used.

Step 3: Identify to Hindi Dialects and compare to implementation dictionary.

Step 4: Convert Hindi Dialects into purer Hindi Language with the help of Machine Translation and NLP techniques.

Step 5: Finally Purer Hindi language convert into different-different countries mother tongues with the help Google Translation API's

Communication Channel: We Will try to understanding the interaction between different-different countries people through communication channel.

Channel 1: Indian People (sender) try to speak in Hindi dialects in the form of input signal.

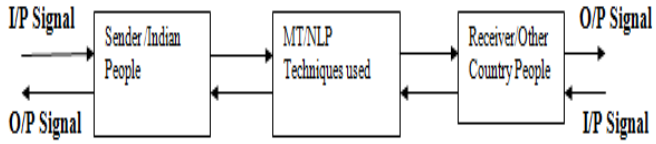


Fig. 2: Communication Channel 1 & 2

Channel 2: Other Country People (receiver) receive signal in their mother tongue and send the signal in their mother tongue language.

Our system based on following Levels are as follows:-

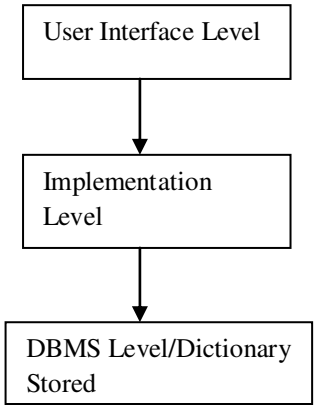


Fig.3: Architecture Level

8. DISCUSSION

The given table contains some of techniques used along with their authors and method given by them[7].

Problems in automatic speech recognition:-

There are many problems that are faced in automatic speech recognition some of them are as follows:

1) Human beings comprehension of speech: As here we only have speech signals and hence we cannot model the world knowledge

2) Spoken and written language is different: In case of spoken language we face more errors as compared to written one

3) Noise: It is unwanted information that combines with the required information .For example we hear different noises like background noise and television noise etc.

4) Speaking style: As we know every person is unique in its own way so different persons have different speaking style. Examples are happy sad, fear, anger etc.

5) Variability of channel: It is the context in which there is utterance of acoustic wave. Different types of noise present, microphones affect the variability of the channel.

6) Realization: If a person pronounces the same the same set of words again and again, each time the result will not be the same

7) Different Dialects: There are different dialects of different regions, for example regional and social dialects.

9. CONCLUSION

Conventional Speech Processing techniques along with conventional language translation techniques have been reported to provide accuracies of their respective works in a limited fashion. They are many applications these days which need higher amount of accuracies, through our studies of machine translation techniques, specifically with reference to our proposed the work of speech translation software development, are expected to provide higher accuracies.

REFERENCES

[1] Sinha Shweta Research Scholar (CS), Birla Institute of Technology, Mesra, Ranchi," Analysis and Recognition of Dialects of Hindi Speech", in International Journal of Scientific Research in Computer Science and Engineering.

[2] Peacocke Richard D. and Graf Daryl H. Components of Speech Recognition System by An Introduction to Speech and Speaker Recognition.

[3]Https://alselectro.wordpress.com/2014/01/31/voice-recognition-board-hm2007-how-to-control-a-motor-using-voice-commands/

[4] Shivakumar G. ,Gupta Ripul "Speech Recognition for Hindi", in Department of Computer Science and Engineering,Indian Institute of Technology ,Bombay Mumbai ,India

[5]Https://www.reliancedigital.in/solutionbox/how-does-smart-devices-understand-what-you-say/.

[6]Prikladnicki Rafael , Calefato Fabio, Lanubile Flippo "Speech Recognition for voice-based Machine Translation", in IEEE software in 2014.

[7] Kale K.V., Gawali Bharti, Gaikwad Santosh "Accent Recognition for Indian English using Acoustic Feature Approach", in International Journal of Computer Applications (0975 – 8887) Volume 63– No.7, February 2013

[8]Dhanvijay Nikita, Badadapure P.R." Hindi Speech Recognition Technique Using Htk", In International Journal Of Engineering Sciences & Research Technology, December 2016 volume 3.0 ISSN: 2277-9655.

[9] Agrawal R.K., Dave Mayank "Discriminative Techniques for Hindi Speech Recognition System", in Department of Computer Engineering, National Institute of Technology, Kurukshetra, Haryana, India, in January 2011DOI: 10.1007/978-3-642-19403-0_45In book: Information Systems for Indian Languages (pp.261-266)

[10] Sinha Shweta, Jain Arun, Agrawal S.S. "Acoustic-Phonetic Feature Based Dialect Identification In Hindi Speech", In International Journal On Smart Sensing And Intelligent Systems Vol. 8, No. 1, March 2015.

[11] Abdul Kadir K, (2010), "Recognition of Human Speech using q-Bernstein Polynomials," International Journal of Computer Application, Vol.2 – No.5, pp.22-28.

[12] Mari Ostendorf et.al., From HMM s to segment Models: A unified view of stochastic Modeling for Speech Recognition , IEEE Transactions on Audio, Speech and Language processing Vol.4,No.5, Septemeber1996.

[13] Yaxin Zhang et.al., Phoneme-Based Vector Quantization in a Discrete HMM Speech Recognizer,I EEE Transactions On Speech And Audio Processing, Vol. 5, No. 1,January

[15] Lawrence R. Rabiner, AT&T Labs Florham Park, New Jersey07932, APPLICATIONS OF SPEECH RECOGNITION IN THE AREA OF TE

[16] K.S. Rao, S. Maity, and V. R. Reddy, "Pitch synchronous and glottal closure based speech analysis for language recognition," International Journal of Speech Technology, vol. 16, no. 4, pp. 413–430, 2013.

[17] L. Mary and B. Yegnanarayana, "Extraction and representation of prosodic features for language and speaker recognition," Speech communication, vol.50, no. 10, pp. 782–796, 2008. LECOMMUNICATI0NS, 1997 IEEE.