Implementation of Speech Recognition in Web Application for Sub Continental Language

Dilip Kumar[1], Abhishek Sachan[2], Malay Kumar[3]

[1]Research Scholar, Department of Computer Science & Engineering, S.R.M.U. Lucknow, Uttar Pradesh India

2Assistant Professor Department of Computer Science and Engineering, S.R.M.U. Lucknow, Uttar Pradesh India

3Assistant Professor Department of Computer Science and Engineering, S.R.M.U. Lucknow, Uttar Pradesh India

***Abstract*— Speech recognition is the aptitude of a mechanism in the web application to identify voice instructions agree with the pattern stored in the glossary. Mainly two concepts are summarized in this paper: Hindi voice conversion into text and searching that text in web application like Google. Currently, Hidden Markov Model (HMMs) is used for Hindi voice recognition and its Toolkit is accustomed to identify Hindi language. It recognizes the isolated words using acoustic word model. The system is trained for many Hindi words. Training data has been collected from nine speakers. The experimental results show that the overall accuracy of the presented system with few states in HMM. This paper aims to build a speech recognition system for Hindi language and search Hindi text in web search engine. In addition the text can be search from the database server using application server. Such a design makes it truly practical to use text conversion and its searching over the internet.**

***Keywords*— HMM, HTK, MFCC, Automatic Speech Recognition.**

I. INTRODUCTION

The input through the keyboard is not an easy task, it require certain level of skill. On the other hand mouse requires a good hand-eye co-ordination. It is not very effortless for entering large amount of text data, so it require supplementary medium . Linking the verbal language is not enough but it requires recognization and implementation in web application. P. Saini et al. [1]. My goal is to recognize Hindi words and implements in voice search on internet. Physically challenged people find computers difficult to use. Linking the verbal communication to the computer science is a big step towards common users.

Now a day's computer interfaces are assume to a certain level of literacy. Users must have skill in English. In India at state level literacy is about 50% but at grass root stage these constraints are about to eliminate. Speech interface can help us to removing these issues. Speech synthesis and speech recognition together form a speech interface. T. Masuko et al [2]. A speech synthesizer converts text into speech. Due to which it read textual word. The ability to know the oral words and translate in to text is speech recognizer. M.B.Shinde et al. [3]. For implementing Indian languages to the web application we require such type of software.

Many times keyboard acts as a barrier between computer and the user. This is true especially for rural areas. It plays an important role towards reducing this gap by providing the use of rural languages. This language identification promoting the technology in rural areas. The communication medium occur in the machines through the speech is very convenient in comparison of keyboard. S.R.Mankala et al. [4]. In opposition speech reorganization is the technique of converting vocal sounds to their perspective content. Between these tasks speech reorganization is difficult one excluding it has mixture of applications like as application for handicapped persons. R.K. Aggarwal and M.Dave. [5]. Numbers of civic field software apparatus are available for the research work related to speech recognition like as large vocabulary continuous speech recognition (LVCSR) engine Julius from Japan (julius, 2011), Hidden Markov Model Toolkit (HTK , 2011), sphinx from Carnegie Mellon university (sphinx, 2011). S.Mandal et al. [6]. This paper aims to develop and implement speech recognition system for Hindi language using the HTK open source toolkit and the language is used in web application for making easy access. Hindi language is not only used by Indians but also by peoples of outer countries. There are many peoples who understand Hindi language only. It is a common language used over Internet. Due to efficient recovery on search engines Unicode faces problem by enabling read and write. Hindi language is supported by many search engines but it is bounded up to sample matching. Any type of searching requires documents, digital objects. So it is important to understand the language of the customers of the Internet.

II. LITERATURE REVIEW

Review of literature of speech recognition in general rises the anxiety towards the first innovation of Alexender graham bell's of converting sound waves into electrical impulses and Vrinda et al. [14] discovered the very first speech recognization system. The approaches to speech recognition, evolved thereafter, had a major stress on result language sounds and providing appropriate labels to these sounds. Most of the concepts and types of speech recognition systems exist in application since last many decades gradually. This revolution has lead to a remarkable impact on the success of speech recognition systems for many languages worldwide. Basically, a speech recognization system is a converter of speech to text. The output of this converter is the text referenced to the speech. Speech feature vectors were extracted by MFCCS. For speaker dependent and independent

environment this system proves helpful, through implementation of Hindi language on internet. M. Dua , X.H.Lee et al. [7][8].Consecutive transformation in acoustic micro system configuration of verbal communication into amplified phonetic macro configuration is done by speech recognization. Speech recognization system is developed for languages like Japan, Tamil, Chinese, Russian, Bengali, Portuguese, Spanish, Hindi are well known among these. Maximum work is done by English language in speech recognization system.P.Saini et al. [1].

Web helps the adult peoples to carry out their studies which may base on observations, research, surveys and expert opinions. The w3c Web Accessibility Initiative (WAI) first released its guidelines in 1999 referenced some of the studies mainly forgetfulness related to the work. Vrinda et al. [14]. This section presents some of the related works available in the literature that are similar to the presented work. Some of the work provides ASR system to south Asian languages pruthi et al. (2000) have developed a speaker-dependent, real-time, isolated word recognizer for Hindi. Using detached HMM model in linear predictive Cepstral feature extraction and recognization were carried out. Two male speaker used in the system each with Hindi vocabulary system. Gupta in 2006 developed a speech recognization system of isolated word in Hindi. This system consists of word based acoustic units using CDHMM (Continuous Density Hidden Markova Model). It produces good result on testing with training model sounds and for other type sounds also. Development and implementation of Arabic languages using HTK discuss in al-qatab and ainon in 2010. In this words and speech are equally recognize. This system uses an Arabic dictionary contains 33 words with 13 speakers. P. Saini et al. [1].

III. Research Methodology

Used methodology contain following components: Sound Recording and word detection component, feature extraction component, speech recognition component, acoustic and language model.

1) Feature extraction component: Feature vectors are generated from the input sound signals. To uniquely identify sound signal Mel frequency spectrum coefficient and normalised energy feature are generates. D.M. Kuansan et al. [9].

2) Recognition component: This is the best component for finding the best match to the stored knowledge. It is based on continuous multi dimensional Hidden Markov Model.

3) *Knowledge model:* It comprises of acoustic component. This model is basically for word representation. It is used by recognition system.

4) *Sound Recording and Word Detection Component*: It is based upon taking input through microphone after that it find out words in that sound. This word detection uses energy and zero crossing rates. Output produces in the form of wave. Vrinda et al. [14].

Hindi Phone Set: - Devanagari is a hand writing form of Hindi language. This Hindi language divided in vowels and consonants in velthuis 2011.

Vowels: In Hindi vowels represents through symbols called Matras. It is about 12 in Hindi, Table II shows this. N.Awasthy et al. [10].

Consonants: On the basis of way of communication and place it divides into different categories. There are about 5 Vargs and 9 Non Vargs. Vargs is further divided in to 5 consonants in which one nasal. Primary and secondary pairs are made with the first 4 consonants which are voiced and unvoiced. 5 semivowels, 3 sibilants and 1 aspirate are obtained from the remaining 9 non Vargs. R.K.Aggrawal et al. [11]. The complete table of consonants is given below-

TABLE I
CONSTANTS

Phonetic Property	Primary Consonant (Unvoiced)		Secondary Consonants (Voiced)		Nasals
Category	Un-Aspirated	Aspirated	Un-Aspirated	Aspirated	
Gutturals(कवगर्)	क	ख	ग	घ	ङ
Patatals(चवगर्)	च	छ	ज	झ	ट
Cerebrals(टवगर्)	ट	ठ	ड	ढ	ण
Dentals(तवगर्)	त	थ	द	ध	न
Labials(पबगर्)	प	फ	ब	भ	म
Semivowels	य. र. ल. व				
Sibilants	श. ष. स				
Aspirates	ह				

TABLE II
VOWELS

Vowels	अ	आ	इ	ई	उ	ऊ	ए	ऐ	ओ	औ	ऋ	ऋ ☼
Sibilants	-	ा	ि	ी	ु	ू	े	ै	ो	ौ	ृ	?

IV. Creating Initial Phoneme Model For Hindi Language

The speech recognition in English is used in a phone set ∏ that consist of 52 phones words. Since the speech recognition system in English has been built completely, we can settle the phonetic space of by using labeled data in English language[a]. But in Hindi language to populate the space represented by r, we either need a Hindi recognition system that can make labeled data or record only the isolated phoneme data. It occurs in the different contexts of language. In this part we represent a map of English phonetic space to Hindi in arrange to produce the initial models for the Hindi phonetic space recognition.

In this model 60 dimensional Cepstral vector generated from English speech is labeled with the equivalent phone using viterbi alignment and the truth. Each 60 dimensional space we form models for by on behalf of each phone model by a set of mixture Gaussians. S.Mandal et al. [6]. I set up a map m that rearrange the English data into 61 cluster, that specifying a phone from . The real mapping is the one important that extract a model forlike that the arrangement rate in this rebuild space i.e. maximum. If we represent <φ> phone model of φ, then the map m provide such that

$$< \prod > \xrightarrow{m} < t >$$

We map m from the acoustic model of knowledge of the phones in Γ and $\prod$. Each segment in Γ is identified by a single element or a combination set models $\prod$.

$$\gamma_i = u_{л n},\ i = 1,2,3\ldots,64,\ \gamma \in \Gamma,\ _{л n} \in$$

If the element $\gamma \in \Gamma$ is acoustically equal to an element of , then the phone set for γ is mainly produced from the vectors that develop the matching model of in the English phonetic space. Though if the best acoustic model is very close to phone in then this is achieved by collecting more sound from model, the model for that γ is identified from the vectors which form models for all those set in. We settle the Hindi phonetic model in the following way.

Using m, identified the phones set model in which produce the very closeness of acoustic model sound for γ by taking vectors from the subset of phones set model in generate the sound γ. if we find more than one γ time, erratically separate the vectors by labeled into many $\gamma \in \Gamma$. It generates a Gaussian mixture model from these analysis vectors. But it produces the number of space and collectively creates models with many overlaps, in this way finally produced result have less arrangements. By identifying this phonetic space model maps. Here we have created initial phone models for Hindi language with the help of English speech recognition system. M. Kumar [12].

Searching of text in Google: When a keyword is converted into text it entered in Google search bar, Google match that keyword in available sites and generate results. Let keyword Amitabh Bachchan name enter, it get searched in Google where the matching names resides. If this word written in Hindi is enter the then it generates the list of the sites where this name reside in Hindi. It shows that language is the main factor of searching in web applications [13].

V. Proposed Model

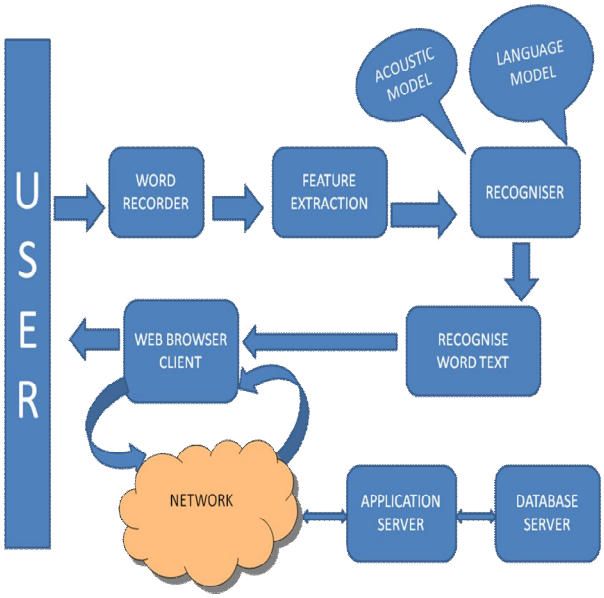

Fig.1 Proposed Model For Speech Recognition For Hindi In Web Application.

1) Feature Mining: It is the method of extracting properties from the speech signals that have acoustic correlations. All the process is done on waveform. This type of parameter is named as feature extraction. Feature extraction is the process of removal of unwanted information from the signal prescribed on some specified parameters and this help for statically condition observation (jain et al, 2010).

*2) Hidden Markov Model and HTK:*This model consists of finites number of states with the set of random function. One function is associated with each state. It is a doubly stochastic process produces using two inter related mechanism. The two process variability and flexibility builds the ASR system. Only observed sequences are known in ASR and transistor is unknown due to which it is known as Hidden Markov Model. This model consists of a tool called Hidden Markov Toolkit. This toolkit is used to build and manipulates HMM. It is also used for speech recognise.

VI. Implementation

In this part, the implementation of the speech system mainly based upon the developing system architecture has been shown.

1) *System Description*- Hindi Speech Recognition System is developed by using HTK toolkit on the Linux platform(Ubuntu12.04). HTK v3.4 and ubuntu12.04 are used. Firstly, through unknown utterances and transcriptions parameters of hmm is calculated. Secondly, these utterances' are described in written form using HTK, 2011. The system used in this process is friendly with the 113 Hindi words. To recognize words model is used. N.Awasthy et al. [10].

2) *Data Interpretation*- For performing any type of operations we require some sets of utterances system mainly

the set of 113 words. The data used in such type of operations is speech which is recorded using one directional micro phone Sony f-v120. The approximately distance between mouth and speaker is about 5 to 10 cm. All types of operation are carried in specified environment. The rate of sampling data is 16000 Hz. This sampling data is consisting of 5 males and 4 females. Each person completes each words three times. That is 3051(113*3*9) speech files. These files are stores in .wav format (velthuis, 2011).

4) *Performance Evaluation*- The performance evaluation[14] of the conversion from Hindi voice to English text system is tested against speaker which have parameters by accessing two types of speakers: one who contained in training data and tested on both and the other involved in only testing. The other parameter is used for testing system performance by checking of states in hmm tools. Here 6 different people are used for testing and everyone asked to utter some words in the system.

5) *Web Based Search Implementation*- After finding the text from the speech recognition model, it is searched in web application by key features of World Wide Web (WWW). Without search engine it will be very hard to imagine web application. I have implemented the searching procedure using java application. The java application provides the platform independent to the different speakers. In the implementation process primarily I used jdk1.6, apache-tomcat server and eclipse tools as IDE JDK (java development kit) is used for providing run environment to java program and the apache server provide the application server where the web programs will run. When the speaker speaks the voice converted into text then that text will entered automatically in the place of search area. This text sent to the web server and the web server search the similar text from the server and provides the result on the browser in the form of text.

VII. CONCLUSION

These days the use of internet users is increasing day by day and they want an easy access of application. To making more efficient of web application and fast accessing for sub continental language like Hindi is the need of hour. The work implemented in the paper is a step toward the development of Hindi language accessing on web very fastly by sending voice and getting the result in the text format. The work may further be extended to send voice and get result in the form of same as given input i.e. voice.

ACKNOWLEDGMENT

This research paper is being provided possible through the helping and supporting for everyone, including: parents, teachers, family, friends, and in essence, all sentient beings. Especially, please allow me to dedicate my acknowledgment of gratitude toward the following significant advisors and contributors:

Firstly, I would like to thank Er. Abhishek Sachan for his most support and encouragement. He kindly read my research paper time to time and told invaluable detailed advices on grammar, organization, and the theme of the my paper.

Second, I would like to thank Mr.Malay Kumar, Long to read my paper and to provide valuable advices, Dr.NeerajTiwari to reproof the paper.

Finally, I sincerely thank to my parents, family, and friends Er.Pradeep Kr. Yadav who provide the advice and financial support. The product of this research paper would not be possible without all of them.

REFERENCES

[1] Preeti Saini, Parmeet Kaur, Mohit Dua *"Hindi Automatic Speech Recognition Using HTK,"* International Journal Of Engineering Trends And Technology, vol 4, issue-June 2013.

[2] T. Masuko, K. Tokuda, T. Kobayashi, *"Speech synthesis using HMMs with dynamic features"*IEEE vol. 1, May 1996, Page(s):389 – 392.

[3] Mayur Babaji Shinde, Dr. S. T. Gandhe "Speech processing for isolated Marathi word recognition using MFCC and DTW features ," International Journal of Innovations in Engineering and Technology, Vol. 3 Issue 1 October 2013

[4] SaiSujith Reddy Mankala, Sainath Reddy Bojja, V. Subba Ramaiah3 & R. Rajeswara Rao4 *"Automatic Speech Processing Using HTK for Telugu Language,"* International Journal of Advances in Engineering & Technology, Vol. 6, Issue 6, pp. 2572-2578, Jan. 2014.

[5] R.K. Aggarwal and M. Dave, "Performance Evalution of Sequentially Combined Heterogenous Feature Streams for Hindi Speech Recognition System," Telecommunication System, Volume52, Issue 3 pp1457-1466, 01 sept 2011

[6] S. Mandal, B. Das, P. Mitra," *Shruti-II: A vernacular speech recognition system in Bengali and an application for visually impaired community*" Students' Technology Symposium (TechSym), 2010 IEEE Date of Conference:3-4 Page(s):229 – 233,April 2010

[7] Mohit Dua, R.K. Aggarwal, Virender Kadyan and Shelza Dua, *"Punjabi Automatic Speech Recognition Using HTK,"* IJCSI International Journal of Computer Science Issues, vol. 9, issue4, no.1, july 2012.

[8] Xuedong Huang Lee, K.-F. "On speaker-independent, speaker-dependent, and speaker-adaptive speech recognition" IEEE Volume:1 , Issue: 2, Page(s):150 – 157, Apr 1993.

[9] D.M.Kuansan Wang Goblirsch, "Extracting dynamic features using the stochastic matching pursuit algorithm for speech event detection"IEEE Workshop on Date of Conference:14-17 Page(s):132 – 139, Dec 1997.

[10] 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.

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

[12] M. Kumar "A large-vocabulary continuous speech recognition system for Hindi" IBM Journal of Research and Development 2004 Volume:48 , Issue: 5.6

[13] How to Use Google Search for Hindi Websites http://www.findurlaptop.com/tech/2013/07/08/google-search-for-hindi-websites

[14] Ms. Vrinda, Mr. Chander *Shekhar "speech recognition system for english language,"* International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 1, January 2013.