

Web Recognition of Spoken Hindi

Kamlesh Sharma^{1*} and Suryakanthi Tangirala²

¹Department of CSE, FET, MRIU University, Faridabad – 121004, Haryana, India; Kamlesh0581@gmail.com

²Faculty of Business, University of Botswana, Gaborone, Botswana; ch.kanthi@gmail.com

Abstract

Technology has evolved and computers have but still Indian communities are far from the use of computers, only 37% [13] user of Indian society like persons from academics, health, engineering and research make use of the benefits of computer. Although there is a revolution in development of operating systems in the past two years, there are no operating systems that support Indian languages like Hindi, they only support English language and people who know English can work on these. This work shows the development of speaker independent ASR system for Indian speaker of a particular type i.e. spoken Hindi. **Objective:** An automated system is developed which accept Hindi query from the user and search the keywords from the query on web and shows the result accordingly. A bilingual lexicon of 3145 words was built. The techniques to convert acoustic speech signals into strings of words were also designed. The dictionary maintains the POS of source to target language and CFG grammar based model for Hindi language was developed. **Methods:** A speech database consisting of speech files is created at the rate of 16 KHz. Interviews were conducted to record speech files from users for developing the speech database for training the system. The system takes spoken Hindi as input and generates Hindi sentences. Hindi sentences are parsed with the help of Top-Down Parser and identify the keywords and form a exact query with the help of grammar rules which are given as input to Google search engine to search the result as per query and display on screen. The lexical categories of Hindi language in grammar are done on the bases of verb, noun, preposition, adjective, adverb phrases. **Result:** A total of 101 speakers from Faridabad district, comprising of both males and females took part in data collection process. The system was tested in a sound proof room and achieved 88.05% accuracy. It is observed that utilization of more training data together with detailed modeling of speech signals can improve the system performance to a level adequate for actual deployment. **Application/Improvements:** This system provides access to a computer system or to digital content over the internet to illiterate, vision-impaired, urban and semi-urban Indian people who are not able to read/write English language. The same speech interface can be enhanced to work in other regional languages of India like Punjabi, Marathi and Telugu, etc. To enhance or to extend it for other languages, the transfer or translation rules of grammar is needed, which can be generated with great ease by using the same dataset with different target languages.

Keywords: Corpus, Hindi, Recognition, Speech, Web

1. Introduction

Human-computer interaction plays a major role for literate/illiterate and visually challenged people to interact with a computer system and access the information from web. The interaction with the system can be done with the help of speech, text, gestures, symbols or a mixture of

these¹. An interaction could be in the form of a dialogue, including commands, statements, questions, answers and expressions².

A lot of research has been carried out on speech recognition systems. Speech recognition systems basic principle is to identify the acoustic signals and convert them into set of words³. Technology advance-

*Author for correspondence

ments make life easy and simplify many complex day to day activities. The speech technology has evolved from Text-to speech (Interactive Voice Response (IVR) and now the latest advancement is Automatic Speech Recognition/Generation (ASR/G). ASR has emerged as a very powerful speech recognition technique and has its applications in fields of engineering and science. Some applications being voice based navigation assistants for blind, speech enabled GPS, live subtitling on television, dictation tools in medical and legal profession, expert systems. The desire to develop an ASR system dates back to 1950. The early attempts failed which were based on the direct conversion of speech signal into a sequence of phoneme-like units. After embracing the failure for 20 years, in 1970 the attempt was successful in spoken word recognition. The technique used was the general pattern matching technique. An input pattern, speech signal is matched with as set of stored patterns and exact match is identified. This technique has a high failure rate as speech signal varies on several parameters like different speakers, their speaking rates, acoustic conditions and others. To overcome this statistical techniques are applied on ASR⁴. Speech recognition systems available now are able to recognize thousands of words. Over past 10 years many high performance algorithms and systems were developed in speech recognition area. This led to commercialization and deployment of speech recognition software's like CMU Sphinx, Agnitioand Julius.

With the advent of Cloud technology and digital convergence of computers and telecommunications, it is easy to get remote access of information stored in a central location for distribution across the world. A machine can access the information and can interpret it. Expert systems are examples of such applications. Pouring formalized and specific knowledge into a system makes it an expert in that field. The advantage will be of multifold if that machine is speech enabled. The concept of machine being able to interact with people in a mode that is natural as well as convenient for human beings is very appealing⁵. Knowledge based ASR is giving spoken commands to a system and getting the required response is a long held dream. This has motivated research in speech recognition as well as speech synthesis.

This paper discusses about the development of a web recognition system of spoken Hindi. The structure of this paper is as follows, the next section outlines about the speech database developed for the system. Then the

corpus details are elaborated. Subsequently the data collection methods are discussed, followed by the results section where in the results obtained were shown in tabular format and accuracy was given in a column chart representation. The last section summarizes the conclusions and future scope.

2. Speech Database

In speech recognition applications, the system needs to be trained with a large amount of speech data corresponding to sentences containing all words and phonemes of the language in all valid phonetic. Thus, it is necessary to have databases, which comprise of appropriate sentences spoken by the typical users in realistic acoustic environment. This necessitates creation of sets of phonetically rich sentences that provide a good coverage of pairs of phones of the language⁶.

Speech recognition involves recognition and matching of spoken words to text by the system. The technique followed here is pattern matching⁷. A large corpus or training database is required for identifying and matching the acoustic speech signals to text. The input patterns (speech signals) should be classified by matching it with one of the stored patterns. But as understood speech signals are highly variable as they are dependent on parameters like, acoustic conditions, speakers, speaking rate and many others. To develop a speaker independent speech recognition system a database with spoken data, tagged with their phonetic identity is developed for training purpose. It is necessary that speech data need to be collected from a large number of speakers. It is not practical and viable to record many input sentences from speakers so that all phonemes can be recorded. It is therefore required to construct databases with sentences that are phonetically rich⁸.

The speech recognition system developed for Hindi to English takes the input as the Hindi spoken text⁹, which is recognized by system. Once the spoken text is recognized, Roman Hindi text is generated which is then converted to English text with the help of mapping from the database. After the mapping of individual English words is done, reordering of the words is done to make it grammatically correct English sentence as shown in Table 4. This is done by the grammar rules developed for the software. The grammar rules make the reordering of Hindi grammar to English grammar^{10,11}.

3. Corpus Detail

Most of the information in IT world is accessible to a few who read or understand English language. Language translators provides solutions in the form of interfaces which can translate English to regional languages. These make the digital content reach to the masses and facilitate the exchange of information across various people speaking various languages. These technologies play a crucial role in multi-lingual societies such as India

The speech database is a collection of speech files created at 16 KHz rate¹². Each sentence is stored in a separate file. In this paper, the implementation of a bilingual corpus is presented. This corpus was specifically recorded from urban and semi-urban users, basically farmers. It was collected for system testing and the purpose of duration modeling⁶. It is meant only for demonstration purpose of development of a query processing system for Hindi Language.

The Hindi language queries are questions asked on web which are redirected to Google search engine, an extract of such queries in Hindi which are recorded/obtained from different users are shown in Table 1. These sentence templates were formed in terms of word categories in the database.

There are three tables in the database helping in the conversion of Hindi words to English and then reorder-

ing them properly to form the grammatically correct sentence in English¹³ as illustrated below in Table 2, Table 3 and Table 4.

Table 2 is for mapping the parts of speech of English words such as noun, verb and pronoun. And also identifies the subject. This is necessary for the conversion of Hindi grammar to English Grammar

Table 4 is for the grammar conversion rules of Hindi to English to form the complete sentences of English which are grammatically correct.

4. Data Collection

The large vocabulary ASR system for Hindi requires the construction of a domain specific and a balanced speech corpus for recognition of continuous and spontaneous speech in Hindi¹⁴. Domain specific means that the corpus should cover all the word related to farmer, student, housewife and people from urban and semi urban areas who speak Hindi Language.

The system is a continuous speech system where in the words are connected together and not separated by pauses it is affected by the rate of speech. Therefore it is required that the training corpus also maintains that balance. The balance refers to the property that the words should occur in the corpus with the same relative frequency distribution as in naturally spoken Hindi. The

Table 1. Hindi language queries

S.No	Hindi	Transliteration	English
	सोनेकाक्याभावहै	sonekakyabhavhai	What is the price of Gold
	भारतवर्षमेंकतिनेराज्यहै	BhAratvarshmeinkitnerajYahai	How many state are in India
	आगालाओलंपिककिसिदेशमेंहोगा	Aaga la olympicskisdeshmeinhoga	In which country next Olympic will be held
	इंटरनेटकाआविष्कारकसिनेकियाथा	internet kaavishkarkissnekiyatha	Who invented Internet
	हवाईजहाजकाआविष्कारकसिनेकियाथा	airplane kaaVishKarKissnekiyatha	Who invented the airplane
	डब्ल्यूएचओक्याहै	W H O kyahain	What is WHO
	चंद्रमापरजानेवालापहलाआदमीकौनथा	ChandRama par JaaNewalaPehlaAadmiKountha	Who was the first man go to moon
	चीनकीराजधानीक्याहै	china kirajdhanikyahain	What is the capital of china
	इंटरनेटक्याहोताहै	internet kyahotahai	What is internet
	जीववज्जिज्ञानवषियकीखोजकसिनेकी.	Biology subject kikhojkissneki	Who invented the Biology

Table 2. Hindi to English word mapping table

HINDI WORD	ENGLISH WORD
Yeh	This
Hai	Is
Ghar	House
Mera	My
Dukan	Shop
Mein	I
Mera	Mine
Aapka	Your
Aap	You
Tum	You
Tumhara	Your
Tu	You
Wo	He
Mein	I

Table 3. English words to its type mapping table

ENGLISH WORD	TYPE OF WORD
He	subject
She	subject
You	subject
Your	subject
Go	Verb
Eat	Verb
Don't	negative
To	preposition
There	Adverb
Here	Adverb
Now	Adverb
Quickly	Adverb
Come	Verb
India	Noun

Table 4. Hindi grammar to English grammar conversion table

HINDI GRAMMAR	ENGLISH GRAMMAR
subject adjective verb	subject verb adjective
adjective verb	verb adjective
subject verb	subject verb
subject question verb	question verb subject
question verb	question verb
noun question verb	question verb noun
noun question noun verb	question noun verb noun
subject pronoun preposition noun	subject noun preposition pronoun
noun subject question verb	noun question verb subject
noun1 question noun2 verb	question noun2 verb noun1
noun1 subject noun2 question verb	question verb subject noun2 noun1
subject noun question verb	question verb subject noun
noun adverb verb	verb adverb noun
subject noun1 noun2 verb	subject noun1 verb noun2

first task is to develop a text corpus with these attributes¹⁵. In order to develop a balance corpus, recording of the actual interviews, questionnaires, group discussion and using everyday speech was done and transcribed. Figure 1 shows recording of a person interview with a farmer.



Figure 1. Recording of interview in Shahapur Kalan village.

The Hindi corpus that was used for this purpose has been developed which consists 3145 words of Hindi. The

data was gathered from various people who live in different areas of Faridabad District of Haryana state, (e.g. Shahapur Kalan village, Kheri village, Faridabad Urban, etc.) and is tabulated in Table 5.

The corpus then is made to read out by a native speaker of Hindi and is recorded to produce the speech corpus for Hindi language. This task of developing sentences queries and commands serve to provide the baseline for acoustic model for Hindi web search. The fundamental criterion remains to cover all possible words that are used by urban and semi-urban people in Hindi.

5. Experimental Detail

The system performance was tested against speaker independent parameter with the help of a mix different types of speakers¹⁶, who were involved in training and testing. Some were involved in only testing. The second parameter for evaluating system performance was various environments. The system was tested in a sound proof room, classroom, laboratory room, office and in open areas. A total of 101 distinct speakers covering farmers, students, laborers, teachers and office staff were involved for forming speech based queries and each one was asked to speak out arbitrarily 100-150 words covering 100-200 queries as shown in Table 6 and Figure 2. The results reveal that the implemented system performed well with various

speakers and in various environments¹¹. The average performance of the system in ideal condition was 88.05%.

The success of each trial was based on whether the system was able to correctly retrieve the required information to the user or not.

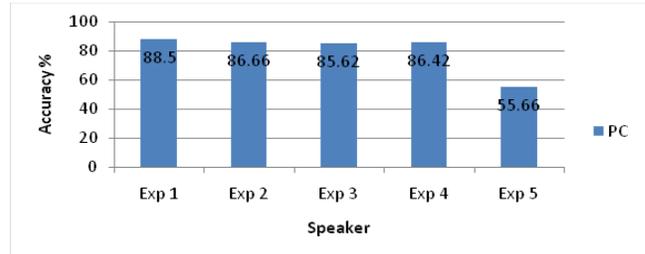


Figure 2. Reliability of recognition of queries in different environment by different untrained speakers.

6. Conclusion and Future Scope

There is an urgent need for development of a flexible, multi-modal human computer interface for web that enables a wider population of the country to get the benefits of the massive amount of data which are available on the internet¹⁰. Here, the role of speech interface can be emphasized as it is the most natural and convenient mode of communication among human beings. Satisfactory speech recognition accuracy can be obtained

Table 5. Speaker information on different parameters

S. no	Village/city	Number of speakers	Male	Female	% age	% Educated
1	KheriKalan	39	18	21	18 - 58	15 %
2	ShahapurKalan	32	19	13	20 - 52	5 %
3	Faridabad Urban	30	14	16	18 - 58	72%

Table 6. Reliability of recognition of queries in different environment by different untrained speakers

Exp. No	Environment	Speakers	Age	Query spoken	Correct	Error	CP	EP
Exp. 1	Sound proof	Untrained speakers	18-58	200	177	23	88.5	11.5
Exp. 2	Lab room	Untrained speakers	18-58	180	156	14	86.66	13.34
Exp. 3	Class room	Untrained speakers	18-58	160	137	23	85.62	14.38
Exp. 4	Office	Untrained speakers	18-58	140	121	19	86.42	13.58
Exp. 5	Market	Untrained speakers	18-58	120	62	38	55.66	44.34

using sophisticated statistical model such as HMM¹² that adequately characterize the temporal aspect of the speech signal in addition to its spectral properties. Utilization of more training data together with detailed modeling of speech signal can raise the system performance to a level adequate for actual deployment in appropriate task domains.

The same system can be configured for other languages with different grammar rule like Punjabi, Tamil, Telugu and Gujarati etc. There are several challenges the system will need to deal with in the future. First, the overall robustness of the system must be improved to facilitate implementation in real life applications involving telephone and computer systems. Second, the system must be able to reject irrelevant speech that does not contain valid words.

7. Acknowledgement

I would like to thank my guides Dr. T.V. Prasad and Dr. S.V.A.V Prasad for their continuous support and guidance in doing the research and publishing mywork.

8. References

1. Kamlesh S, Suryakanthi T, Prasad TV. Exploration of Speech enabled System for English. Proceedings of International Conference on System Modeling and Advancement in Research Trends (SMART); India. 2012. PMid: 22529636 PMCID: PMC3326968.
2. Mantena GV, Rajendran S, Gangashetty SV, Yegnanarayana B. Development of a Spoken Dialogue System for accessing Agricultural Information in Telugu. Proceedings of 9th International Conference on Natural Language Processing; India. 2011.
3. Rajendran S, Yegnanarayana B. Word boundary hypothesization for continuous speech in Hindi based on patterns. *Speech Communication*. 1996 Jan; 18(1):21–46. Crossref
4. Jackson M. Automatic Speech Recognition: Human Computer Interface for Kinyarwanda Language. A Project Report Submitted in Partial Fulfillment of the Requirements for the Award of the Degree Master of Science in Computer Science. Makerere University; 2005 Aug.
5. Kamlesh S, Prasad TV. CONATION: English Command Input/output System for Computer. Proceedings of International Conference on Science, Engineering and Spirituality; India. 2010. PMid: 21938195, PMCID: PMC3002773.
6. Somnath M, Dutta AK. Automatic recognition of isolated Bengali words. Proceedings of International Workshop on Speech technology for man-machine interaction. Rao PVS and Kalia BB, editors. Tata McGraw Hill. 1990. p. 283–88.
7. Ramana Rao GV, Yegnanarayana B. Word boundary hypothesization in Hindi speech. *Computer Speech and Language*. 1991 Dec; 5(4):379–92. Crossref
8. Vishal Chourasia S, Samudravijaya K, Manohar C. Phonetically Rich Hindi Sentence Corpus for Creation of Speech Database. Proceedings of International Conference on Speech Databases and Assessment; Indonesia. 2005.
9. Kamlesh S, Prasad TV. A Text-to-Hindi Speech Interface for Web Browsing. National Conference on Advances in Knowledge Management (NCAKM'10); India. 2010. PMid: 21938195, PMCID: PMC3002773
10. Kamlesh S, Prasad SVAV, Prasad TV. A Hindi Speech Actuated Computer Interface for Web Search. *International Journal of Advanced Computer Science Applications*. 2012; 3(10): 146–52.
11. Kamlesh S, Prasad TV, Prasad SVAV. Hindi Speech enabled windows application using SAPI. *International Journal of Computer Engineering and Technology*. 2013 Apr; 4(2):425–36.
12. Samudravijaya K, Ahuja R, Bondale N, Jose T, Krishnan S, Poddar P, Rao PVS, Raveendran R. A feature-based hierarchical speech recognition system for Hindi. *Sadhana*. 1998; 23(4):313–40. Crossref
13. Kamlesh S. Designing and developing speech actuated computer interface for Hindi Language. [Doctral Thesis]. India. 2014.
14. Mobin A, Agrawal SS, Anil Kumar, Pavate KV. A voice input-output system for isolated words. Proceedings of International Workshop on Speech technology for man-machine interaction. Rao PVS and Kalia BB, editors. Tata McGraw Hill. 1990. p. 283–8.
15. Samudravijaya K, Rao PVS, Agrawal SS. Hindi speech database. Proceedings of International Conference on Spoken Language Processing; China. 2010. P. 456–9.
16. Ganesh SP. Realization of hidden Markov model for English Digit Recognition. *International Journal of Computer Applications*. 2014 Jul; 98(17):37–40.
17. Rabiner LR. A tutorial on hidden Markov models and selected Applications in Speech Recognition. Proceedings of IEEE. 1989 Feb; 77(2):257–86. Crossref