

## **Research Issues in ASR: A leading edge to Punjabi Language**

Navneet Kaur Walia<sup>1</sup>, Sharanpreet Kaur Tiwana<sup>2</sup>,

<sup>1</sup>PG student, Punjabi University, Patiala, Punjab.

<sup>2</sup>Assistant Professor, Department of Computer Science, Mata Gujri College Fatehgarh Sahib, Punjab.

**ABSTRACT:** Speech recognition is a field of computer science that deals with designing human interactive systems that recognize spoken words. It is a process of converting an acoustic signal, captured by a microphone to a set of words. The major consideration for developing an efficient ASR is maintaining high recognition accuracy rate and low word error rate. The paper is divided into seven sections where introduction is followed by the process of ASR and further by related work. From the vast literature survey various research gaps are identified. The need for an efficient Email system in Punjabi language is demanded by the research gaps. However there are certain limitations to the validity of proposed work which are mentioned in the paper. The paper basically aims at lime lighting the major issues which demand future research work in the field of speech recognition, yet the major concern of the paper emphasizes on proposing an Email system in Punjabi language.

**Keywords -** ASR (Automatic speech recognition), ADHD (Attention Deficit Hyperactivity Disorder), RSI (Repetitive Strain Injury).

### **I. INTRODUCTION**

Most of the software's used today are accessible via keyboard and mouse. Interaction with computer systems using keyboard and mouse is not a tough job. It is quite easy for those people who are physically fit and mentally sound. But it reaches to a level of difficulty for the ones who are disable, visually impaired and suffering from certain disabilities like Dyslexia, Attention Deficit Hyperactivity Disorder (ADHD), Cognitive Disability, Vision Impairment, Repetitive Strain Injury (RSI), etc. Though computers are the originators of knowledge yet this knowledge is still inaccessible to many people across the globe. The reasons of this inaccessibility could be:

- a) Individuals lack the skills to use keyboard and mouse finding it difficult to access the computers.
- b) Persons who are physically handicapped and mentally less sound.
- c) People who are not familiar with the use of English Language.

The solution to above mentioned problems is Automatic Speech Recognition system where a person will be using his voice for interaction with the system instead of keyboard strokes and mouse's cursor. This would be a great assistance for them to use their voice to control the computer system in their native language. Even for the normal individuals, it is more convenient to work with speech oriented applications using their mother tongue like Telgu [10], Hindi [8] and Punjabi [7]. In the last few decades, the researchers across the globe have worked for the development of Speech Recognition Systems for the most commonly used languages such as English, French, Japanese and many more. Since Punjabi is commonly used language in major parts of the world. Yet no appreciable work has been done in this area despite of the fact that it has a large number of speakers. Punjabi language has gained stature in publishing and communication and hence it demands a footstep in the growing field of automatic speech recognition. The proposed work promises significant advantages in the areas where the individuals are not much comfortable in English language and feels more relaxed in Punjabi Language. The paper focuses on the persons who find it difficult to share their messages due to some physical problems like Dyslexia, Vision Impairment, Repetitive Strain Injury (RSI), etc.

### **II. PROCESS OF ASR**

The process of speech recognition involves various constraints like speaker style, vocabulary, speakers, style of speech etc. All these constraints sometimes lead to improper functionality of speech recognition systems. Thus the validation suffers in such extreme cases. The following diagram depicts multiple parameters on the basis of which speech is recognized in the ASR systems.

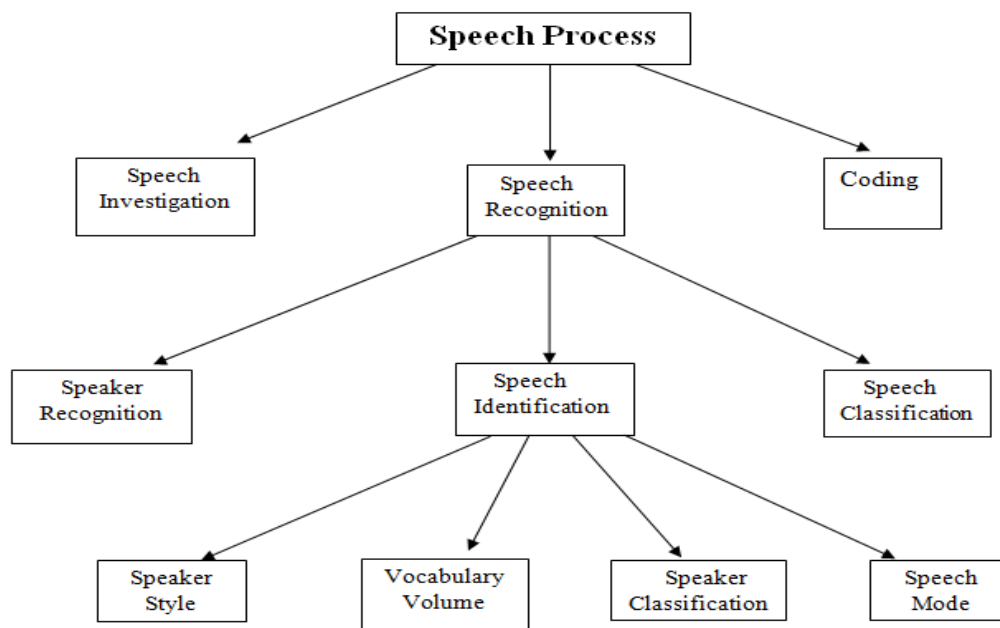


Fig.1 Process of speech recognition

### III. RELATED WORK

In the past few decades, a lot of work has been done in the field of Speech Recognition and Speech Synthesis for communication with the computer system. Some of the suggested approaches are discussed below.

Bill Manaris, Renée McCauley, Valanne MacGyvers [1] provided an interface to access all available PC functionality of a keyboard and mouse using SUITKeys for motor disabled computer users. The users simply read aloud menu entries and dialog box elements to select them.

Mu-Chun Su, Ming-Tsang Chung [2] designed a voice controlled human computer interface which enabled severely handicapped individuals to operate a computer. This system consisted of six main blocks: the keyboard and the mouse layouts; the D6106 speech recognizer; the headset incorporated with a microphone and two mercury switches; the keyboard circuit and the mouse circuits; the mouse control circuit; and the microcontroller. A prototype system was built and tested.

Masataka Goto, Katunobu Itou, Tetsunori Kobayashi [3] described speech interfaces using nonverbal speech information. They made the use of non-verbal speech by focusing on certain parameters like hesitation (filled pauses) and prosody (voice pitch). Four Speech Interface Functions were created - Speech Completion, Speech Shift, Speech Starter, and Speech Spotter.

Frank Loewenich and Frederic Maire [4] developed a system with the web camera and microphone which fully emulated the functionality normally associated with a mouse device. The speaker issued the relevant commands such as click-drag, drag-drop operations using voice and the resultant functionality was performed.

M. Abdeen et al. [5] provided the architecture for multi lingual hands free desktop control system. This approach was based on PC windows (Microsoft windows XP) and it is language independent. The framework was tested for English (a Latin language) and Arabic (a Semitic language). It was based on DLL's and is capable of supporting many languages depending on the available DLLs.

M.A.Anusuya and S.K.Katti [6] presented a survey on the speech recognition and elaborated the advances in the last 60 years in the fields of ASR. Various approaches and algorithms were discussed and compared with some of the well known methods used in various stages of speech recognition system.

Bhupinder Singh and Parminder Singh [7] used the concept of Hidden MARKOV Model for speech command recognition (speaker dependent) using Punjabi Language and turned a machine to intelligent one which responded to the instructions in Punjabi Language. However the system shall be extended to speaker independent model and vocabulary could be increased to follow multiple commands.

Kuldeep Kumar and R. K. Aggarwal [8] provided a Hindi Speech Recognition System Using HTK to build a speech recognition system for Hindi language. The system recognized 30 isolated Hindi words using acoustic word model. The overall accuracy of the presented system was 94.63%. The future works involved the development of system for more vocabulary size and to improve the accuracy of the system

Md.Sipon Miah, Tapan Kumar Godder [9] proposed a voice Control Keyboard Systems, fully controlled by a computer, and display output on the display device with predefined time. In this project voice was the input using sound application programming interface (SAPI 4.1) recognized the voice and transferred the command to the microprocessor according to the programming code. The displayed device generated the output. This system was able to be implemented in other system for example voice control car system. For a mobile person, the same solution could be used in connection with a portable system.

P. Vijai Bhaskar and Dr. S. Rama Mohan Rao, A.Gopi [10] introduced a speech recognition system for Telugu language for continuous speech. Hidden Markov Model Toolkit (HTK) was used to develop the system. The data had been taken from male speakers. However the accuracy of the system was a challenging area to work upon. Different speech enhancements and noise reduction techniques were applied for making system more efficient, accurate and fast.

Wiqas Ghai and Dr. Navdeep Singh [11] performed a case study on the Analysis of Automatic Speech Recognition Systems for the family of Indo-Aryan Languages: Punjabi, Hindi, Marathi, Gujarati, Sindhi, Bengali, Nepali, Sinhala, Oriya, Assamese, and Urdu. In this paper the applicability of Punjabi language had been discussed so that a concrete work can be initiated for Punjabi language. Speech recognition tools HTK and SPHINX are revealed. Advanced feature extraction techniques such as MFCC and Extended MFCC were discussed.

F. Reena Sharma and S. Geetanjali Wasson [12] introduced assistive technology application using which two physically disabled persons blind and deaf could communicate to each other in US English Accent via US B Serial Adaptor. The technology used was Microsoft SAPI on Windows Vista.

Manish Raverkar et al. [13] proposed a system that controlled the machines and environment with speech. A prosthetic device was controlled by the voice commands delivered by the users. The stand-alone isolated word speech recognition system was designed and developed to control the electronic hand by voice command which was taken as input through a microphone.

Kamlesh Sharma et al. [14] developed a Hindi Speech Actuated Computer Interface for Web Search through which the users could provide the queries to the web browser in Hindi Speech. Then the browser returned the searched result on the console. The use of Continuous Speech recognition was done with the large vocabulary size for searching the result.

Ameera Al-Rehili et al. [15] demonstrated a desktop application which was able to convert English text to Arabic text, pronounce English and/or Arabic text, and recognize the English speech to convert it then into a corresponding English text. This application could also be embedded to make the application more beneficial to the learners.

Kamlesh Sharma et al. [16] introduced Hindi Speech Enabled Windows Application in which the basic operations were performed using Hindi Speech like MS Word, Calculator, and MS Outlook. The application was designed in target keeping in view the persons who were either physically handicapped or more comfortable in their native language.

Wiqas Ghai and Dr. Navdeep Singh [17] demonstrated Continuous speech recognition for Punjabi Language. No work had been done so far for Punjabi language in Continuous Speech. They build an automatic speech recognizer to recognize continuous speech sentences by using Tri-Phone based acoustic modeling approach on HTK 3.4.1 speech engine on the Linux Platform. Sentence recognition accuracy for all the three phases of experiment was found to be 82.18% at sentence level and 94.32% at word level.

S.Ganesh et al. [18] introduced the role of Information Communication Technology (ICT). It enabled the physical disable persons to access the information along with others. This product aimed at capturing the lip movements of the dumb and converting these into vibrations and consequently into sound signals which were later transmitted to the receiver. A device needed to be fitted in their mouth to sense the lip movements and then the process of recognizing the speech from vibrations takes place. Hence this product proved to be a great relief for conversing with dumb people and thereby understanding and sharing their thought and feelings.

Sourav De Arup Kumar Das et al. [19] calculated the feature extraction of voice using the MFCC technique and the output was examined with three different algorithms to calculate the efficiency. The accuracy level achieved in this format was almost 80% for both the cases of male and female. In the next phase the integration of LPCC with MFCC technique was done. The authors discussed that if the efficiency of this method could be increased over 90% then method could be taken further for different biometric tests, like retina, finger prints.

Michael D. Goller et al. [20] developed the Speech Interface for Search Engines which receives voiced utterances from a user, converted then into patterns to perform a search and responds back the result.

Parneet Kaur et al. [21] revealed various techniques of Speech Recognition in detail like Dynamic Time Wrapping, Hidden Markov Models, and Artificial Neural Networks which could be easily used for the development of ASR System.

Akansha Madan and Divya Gupta [22] clarified the various feature extraction techniques like LPC MFCC, their merits demerits along with the tools and techniques needed to implement ASR System with LPC and MFCC.

Dr E.Chandra et al. [23] justified the following feature extraction techniques like LPC,LPCC, PLP-MFCC,FFT,MEL, RASTA,DELTA The two most important pattern recognition techniques HMM and Neural Network were explained. The author mentioned that MFCC and PLP was the logarithmically spaced filter bank of human auditory system and hence the better Report compared to LPC parameters.

Mr. Anand Mantri et al. [24] interpreted the various techniques which were evaluated by scientist for voice recognition and available in daily use. In this paper the authors designed and developed MFCC (Mel-Frequency Cepstral Coefficients) feature and HMM classifier based human recognition system, which identify the human based on their voice input.

Bhupinder Singh et al. [25] introduced area of speech recognition process to build a perfect speaker independent ASR system. Commonly used techniques in ASR like Background Noise Removal, End Point Detection, Pre-emphasis, Framing in Preprocessing were discussed in detail.

#### IV. RESEARCH GAP

The paper is based on the difficulties faced by the individuals dealing with the English language for making the use of ICT Tools. The aim of this paper is to introduce the areas where ASR could work for the Punjabi language e.g. the development of Punjabi Speech Empowered Email System. It is providing the solution for the Punjabi speaking community. Various issues are mentioned below:

1. **Speech recognition for mobile phones**-ASR is a challenging field proved for mobile phones manufacturers which offers inbuilt ASR applications like:-

- a) **Voice Dialing** -It will enable the user to deliver the command in Punjabi to make a call to a specific user.
- b) **Manage applications**-This feature will assist the user in managing the applications which are present in the phones. The instructions will be made in Punjabi language for simplicity.

2. **Speech recognition in Transportation**-This application is basically meant for four wheeler vehicles, where the user will deliver the commands in Punjabi in order to play music player, select radio channels etc.

3. **Speech recognition in Computers**-Speech will act as a bridge between the human and computer. A few applications are discussed:

a) **User interface using speech**-The applications are available in English and other languages for interacting with the computer through speech [14], [18],[12].These applications will enable physically challenged persons to communicate. A Punjabi enabled system is desired for efficient usage of computer aimed at Punjabi speaking individuals.

b) **Email system**-This application will facilitate the user to deliver the speech in Punjabi and the resultant basic email operation will be performed like opening mail , closing mail, and sending email (whose contact address is saved in the address book of the mail) to the specified person. The user will be delivering the words in Punjabi and the resultant operation will be performed. .To makes the system more interactive a fast GUI will be developed using the .NET platform.

#### V. OBJECTIVE

Speech Recognition is the developing assistive technology by which all kind of people either able or physically disable can access the information. The main focus of the paper is to highlight the usage of Punjabi enabled speech email system. Different aspects for the work are given below in Table 1. Some of the perspectives are:

a) **Commanding and Dictating** - Commands allow controlling the system and applications through speech as shown in Table 1. Dictating allows writing emails using speech, make corrections etc.

S No.	List of related actions performed	Instructions in Punjabi
1.	Open Email Account	ਖੁੱਲ੍ਹੋ ਈਮੇਲ ਖਾਤੇ
2.	Open Email (provided the name of sender saved in the contact list) e.g. Navneet Walia's email open	ਨਵਨੀਤ ਵਾਲੀਆ ਦਾ ਈਮੇਲ ਓਪਨ
3.	Compose Email	ਈਮੇਲ ਲਿਖੋ
4.	Attach File	ਫਾਇਲ ਨੱਥੀ ਕਰੋ

5.	Send Email to recipient name e.g. Preet	ਪ੍ਰੀਤ ਨੂੰ ਈਮੇਲ ਭੇਜੋ
6.	Send Email	ਈਮੇਲ ਭੇਜੋ
7.	Sign Out Email	ਈਮੇਲ ਖਾਤੇ ਸਾਈਨ ਆਊਟ

**Table 1 Commands in Punjabi and their related actions in English**

The proposed work has mentioned only few aspects for designing the email system in Punjabi language. However the vocabulary could be further increased like managing the contacts in email account, managing folders and sub folders with their relevant operations. The suggested work has practical and methodological significance. The speech corpus developed will be very useful to any researcher who may wish to venture into Punjabi language. By developing and training the system a speech recognition system in Punjabi language the semi illiterates would be able to use IT Tools. Since speech technology is a technology of today and tomorrow the result of this work will help many indigenous Punjabi language speakers. The technology will find applicability in software supporting Punjabi language as an embedded application. By developing and training speech recognition for Email system in Punjabi language it would be a big step towards persons with disabilities.

## VI. THREATS TO VALIDITY

1. **Noise**-Whenever you speak you do this in the environment of many other background voices which are called noise. The major difficulty is that the ASR has to filter out these unwanted sounds from the speech. Sometimes due to this correct inputs are not identified by the systems.
2. **Rapid Emergence of new Languages**- To anticipate the growing needs some language resources and infrastructures need to be established.
3. **Humans comprehend easily than ASR**-Human being usually grasp the meaning by predicts words which are not spoken but this is not the case with ASR. The prediction is limited to an extended only.
4. **Spoken Vs written language**-The spoken language is usually a two way communication process but this is contrasting with respect to ASR.
5. **Speaker style and inconsistency**-As the speakers speak at different pitch. Not only their voices are different, there is wide variability in speech. The speaker speaks at different frequency and their pronunciation styles may differ.
6. **Privacy**-With ASR we have to speak loudly than your normal voice which is sometimes not feasible. Further females speak at a pitch which is almost double than that of males.
7. **Amount of data and search of data** -To make communication possible with the computer, the machine needs to match the speech with the phones. Due to the quality of sound sometimes the words are undetected and the problem of out-of-vocabulary arises.

## VII. CONCLUSION & FUTURE DIRECTIONS

Speech is the most effective and promising medium of sharing feelings between human beings. However this approach has been extended from humans to computer interaction via Automatic Speech Recognition. The paper focused on the discussion of the vast application areas of ASR and the future recommendations for the research community. Various research areas are discussed but email system for Punjabi language is the foremost focus. The paper basically provides the directions to the uncovered research gaps in the field of speech recognition. The issues are open and demands future investigation. Further improvement in the mentioned areas will lead to a more powerful and efficient speech recognition systems.

## REFERENCES

- [1] Bill Manaris, Renée McCauley and Valanne MacGyvers, "An Intelligent Interface for Keyboard and Mouse Control – Providing Full Access to PC Functionality via Speech" Proceedings of 14th International Florida AI Research Symposium May 2001, pp. 182-188.
- [2] Mu-Chun Su 1 and Ming-Tsang Chung, "Voice-controlled human-computer interface for the disabled" Computing & Control Engineering Journal, Vol. 12, Issue 5, October 2001, pp. 225 – 230.

- [3] Masataka Goto, Katunobu Itou and Tetsunori Kobayashi, "Speech Interface Exploiting Intentionally-Controlled Nonverbal Speech Information" Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology pp. 35-36, October 2005.
- [4] Frank Loewenich and Frederic Maire, "Motion-Tracking and Speech Recognition for Hands-Free Mouse-Pointer Manipulation" ISBN 978-953-7619-29-9, November 2008, pp. 550.
- [5] M. Abdeen, H. Mohammad, M. C.E. Yagoub "An Architecture For Multi-Lingual Hands-Free Desktop Control System For Pc Windows" IEEE 2008, Canadian Conference on, vol., no, pp.001747-001750, 4-7 May 2008.
- [6] M.A.Anusuya and S.K.Katti, "Speech Recognition by Machine: A Review" "International Journal of Computer Science and Information Security, Vol. 6, No. 3, 2009
- [7] Bhupinder Singh and Parminder Singh, "Voice Based user Machine Interface for Punjabi using Hidden Markov Model" "IJCS Vol. 2, Issue 3, September 2011.
- [8] Kuldeep Kumar and R. K. Aggarwal, "Hindi Speech Recognition System Using HTK", International Journal of Computing and Business Research Volume 2 Issue 2 May 2011.
- [9] Md. Sipon Miah, and Tapan Kumar Godder, "To Design Voice Control Keyboard System using Speech Application Programming Interface" IJCSI Vol. 7, Issue 6, November 2010
- [10] P. Vijai Bhaskar, Prof. Dr. S. Rama Mohan Rao, A.Gopi, "HTK Based Telugu Speech Recognition" International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 12, December 2012
- [11] Wiqas Ghai and Navdeep Singh, "Literature Review on Automatic Speech Recognition" International Journal of Computer Applications (0975 – 8887) Volume 41– No.8, March 2012
- [12] F. Reena Sharma and S. Geetanjali Wasson, "Speech Recognition and Synthesis Tool: Assistive Technology for Physically Disabled Persons" "International Journal of Computer Science and Telecommunications Volume 3, Issue 4, April 2012.
- [13] Manish Raverkz, Shubhangi and Davinder Pal Singh, "Audio Signal Synthesis for Detection of Speaker Dependent Voice Commands for Controlling a Prosthetic Device" SGI Reflections International Journal of Science, Technology and Management, ISSN 0976-2140 Vol. - 3 Issue - 1 May-2012.
- [14] Kamlesh Sharma, Dr. S.V.A.V. Prasad and Dr. T. V. Prasad", "A Hindi Speech Actuated Computer Interface for Web Search" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No. 10, 2012
- [15] Ameera Al-Rehili, Dalal Al-Juhani, Maha Al-Maimani and Munir Ahmed, "A Novel approach to convert speech to Text and Vice-Versa and Translate from English to Arabic Language" International Journal of Science and Applied Information Technology Volume 1, No.2, May – June 2012 ISSN No. 2278-3083.
- [16] Kamlesh Sharma, Dr. T.V. Prasad and Dr. S. V. A. V. Prasad, "Hindi Speech Enabled Windows Application Using Microsoft Sapi" " International Journal Of Computer Engineering & Technology (IJCET) ISSN 0976 – 6367(Print) Volume 4, Issue 2, March – April (2013), pp. 425-436 .
- [17] Wiqas Ghai and Navdeep Singh, "Continuous Speech Recognition for Punjabi Language" International Journal of Computer Applications (0975 – 8887) Volume 72– No.14, May 2013.
- [18] S.Ganesh, Dr.Saravana Kumar, Dr. Shankar, Samuel D. Raj and R.Kartik, "A Novel Voice Recognition System For Dumb People" ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195.
- [19] Sourav De, Arup Kumar Das Brotin Biswas and Arindam Biswas, "Processing and Recognition of Voice" International Journal of Computer Applications (0975 – 8887) Volume 85 – No 5, January 2014.
- [20] Michael D. Goller and Stuart E. Goller, "Speech Interface for Search Engines" United State Patent Application Publication Pub. No.: US 2008/0114747 A1, March 2008.
- [21] Parneet Kaur, Parminder Singh and Vidushi Garg, "Speech Recognition System; Challenges and Techniques" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (3) , 2012,3989-3992 ISSN No. 0975-9646.
- [22] Akansha Madan and Divya Gupta, "Speech Feature Extraction and Classification: A Comparative Review" "International Journal of Computer Applications (0975 – 8887) Volume 90 – No 9, March 2014.
- [23] Dr E.Chandra, K.Manikandan and M. Sivasankar, "A Proportional Study on Feature Extraction Method in Automatic Speech Recognition System" International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering Vol. 2, Issue 1, January 2014.
- [24] Mr. Anand Mantri, Mr. Mukesh Tiwari and Mr. Jaikaran Singh, "Performance Evaluation of Human Voice Recognition System based on MFCC feature and HMM classifier" "Int. Journal of Engineering Research and Applications www.ijera.com ISSN: 2248-9622, Vol. 4, Issue 2, February 2014, pp.715-719.
- [25] Bhupinder Singh, Vanita Rani and Namisha Mahajan, "Preprocessing In ASR for Computer Machine Interaction with Humans: A Review" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 3, March 2012 ISSN: 2277 128X.