

Speech Recognition System: Overview of the State-Of-The-Arts

Roseline Oluwaseun Ogundokun^{1,*}, Oluwakemi Christiana Abikoye²,
Adekanmi Adeyinka Adegun³, Joseph Bamidele Awotunde⁴

^{1,3}*Department of Computer Science, Landmark University Omu Aran, Nigeria.*

^{2,4}*Department of Computer Science, University of Ilorin, Ilorin, Kwara State.*

Abstract

Human-Computer Interaction (Human Factor) is described as an approach in which individuals communicates with facts, news, computers, tasks mainly in the areas of the profession, administrative, legislative as well as in cultural contexts. This study aimed to deliver a summary and survey on the concept of human-computer interaction (HCI) also called human factor (HF) technologies focusing on one of the audio based HCI which is speech recognition (SR). The study also examined the history of HCI, the HCI research trends, processes of HCI, theories and principles of HCI, HCI models and architecture of HCI. The authors also discussed the overview of speech recognition (SR) which includes the definition of SR, phases of SR, application areas of SR, literature review of past researches on speech recognitions, a table showing the comparative study of the speech recognition system approaches and lastly the study was concluded.

Keywords: Human Computer Interaction, Information Technology, Speech to text recognition System, HCI Model, HCI Domain, Phases, Application area, Literature Review

1 INTRODUCTION

This study discusses about the overview and state-of-art of the Human Computer Interaction as well as Speech Recognition System. Basic characteristics and structure of HCI, component of HCI, HCI domain relationship, HCI Theories and Principles, HCI Models, Fragment of HCI, overview of the speech to text system, various kinds of speech recognition systems were also examined. This overview will include definitions and terminology of speech recognition, a survey of existing technologies and some application areas. The study also introduced progress in research and development of human-computer interaction (HCI), HCI Research Fields and also the preprocessing techniques in automatic speech recognition were discussed. The study considered a literature review process that consists of three key stages which are planning the review, conducting the review and lastly reporting the review. Initial research of 40 articles was downloaded and more than 25 out of them were selected for the review using an inclusion and exclusion criteria. The findings gotten from this state of art review are an overview of HCI and Speech to text recognition based system, existing development of HCI, HCI and speech to

text recognition research fields, the procedure of preprocessing in automatic speech recognition.

2 BACKGROUND

Human computer interaction had remained well-thought-out as computer associated millstone field of study which is intensely correlated with the composition of information, collaboration, communiqué and expertise (Park & Mckilligan, 2018). The techniques of human interacting with computer system has been for decades and the journey still continues up till now as new technologies come to being each day. The study in this field had continued rising extremely rapidly and is still growing (Nande, Patel & Yadav, 2014). Human factor deals with the approaches and devices utilized for the evolution of human computer interfaces, accessing the functionality of the mainframe systems (Preece & Rombach, 1994).

Human Computer System is referred to as the research of collaboration connecting individuals and computers (Hciprojectcit06, 2010). Since human computer interaction investigates individuals and computer devices in conveying with each other, which moves from contributing to information together with machines and the individual side (Hciprojectcit06, 2010). Approaches in computer graphics, operating systems, programming languages and development environments are significant on the machine point of view while communication theory, graphic and industrial design disciplines, linguistics, social sciences, cognitive psychology, and individual executions are relevant on the human HCI point of view (Hciprojectcit06, 2010).

3 LITERATURE REVIEW

3.1 History

Human Computer Interaction is a domain of exploration together with application, which materialized sometimes 1970s and immediate 1980s and it was originally as a domain in Computer Science. For three decades now, HCI had long-drawn-out speedily and progressively and it is currently drawing experts from several other field of study and encompassing different thoughts and methodologies (Hciprojectcit06, 2010).

Personal Computer emerged in the year 1970s. Broad programme of reasoning learning that incorporated artificial intelligence, linguistics, cognitive anthropology and principles of the intellect were fashioned at the finishing of the year 1970s (Hciprojectcit06, 2010). There was internet and combined works in the year 1990s. The internet commenced expedition this year (1990s) and this led to easy communication among people. In the year 2000s, GSMs, personal development assistance and Android phones are using existing realm. SMS, MMS, Multimedia, games, email, internet, chatting, video conference, GPS and so on where used for servicing people at the era (Hciprojectcit06, 2010). Today, the HCI is employed in the field of Reasoning Learning such as interactive interfaces, internet medical facilities and so on.

3.2 Research Trends

Human computer interaction research fields.

According to Hciprojectcit06, 2010, research fields in Human Computer Interaction are discussed below:

1. Ubiquitous Communication: Conversation with one and other is accomplished by means of rapid local networks which is an over wide area nationwide and conveniently through the use of infrared, ultrasonic, cellular plus additional technologies.
2. Extreme Functionality System: The key goal of investigating on HCI is create a more usable and beneficial system which will provide clients with capabilities appropriate for their certain contextual awareness and goals.
3. Bulky Accessibility of Computer Visuals: Research interest such as image processing, graphics transformations, rendering and interactive animations are computer visuals competencies.
4. Mixed Media: This is also an additional essential research area of interest which encompasses HCI systems handling images, speech, resonances, cinematographic, text and structured records.
5. Excellent bandwidth Interface: This is an area of research which examines the proportion at which individuals and

devices relates. It is an area accumulating substantively owing variation speed, computer visuals, latest media and latest input/output devices.

6. Large and Thin displays: New display technologies are becoming matured by enabling very large displays and also displays that are thin, light-weight, and have low power consumption.
7. Embedded Computation: This is an area of research which deals with the environment of computations has changed to entrenched totaling and this is getting larger daily.
8. Cluster Boundaries: Currently, diverse boundaries are required for summits, manufacturing schemes including authoring collaborative documents, hence development of such kinds of interfaces requires suitable investigation.

3.3 Process of HCI.

HCI is described below according to Dix, Finlay, Abowd, Beale, (2004) and Sharp, Rogers, Preece, (2007)

- a. Requirement: This concentrated on considering consumer's need including their existences to offer perceptions on the way HCI creators can produce collaborative results to be equivalent to the consumer's requirements.
- b. Analysis: This step emphasizes synthesizing the main issues coming from the first step and provides directions to the next step, design.
- c. Design: This phase solves problems while bringing usability factors and practices into the process.
- d. Implement and deploy: This step involves integrating physical device and software where HCI researchers rely on guidelines to assess design violations while users interact with the solutions in their own environment
- e. Prototype: This often treated as restricted presentations of a design and used for testing the solutions effectiveness with the users.

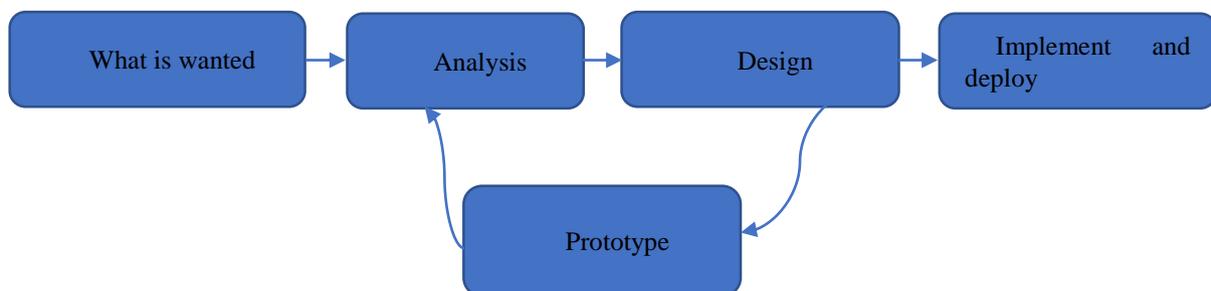


Fig. 1. Human Computer Interaction Scheme Process (Source: Dix, Finlay, Abowd, Beale, 2004)

3.4 Theories and Principles of Human Computer Interface

Nielsen, 1993, Baker, Greenberg & Gutwin, 2002) presented 10 instructions which he called usability heuristics and which were considered to elucidate enormous percentage of glitches detected in the interface projected, which was suggested ought to be monitored by entirely all the operators of the interface design as they are listed below

1. Simple and natural dialogue: There should be compelled attempt in order to avoid irrelevant information.
2. Speak the Users' Language: Every communication which are recognizable to the client should be conveyed in perceptions rather than the system been familiar.
3. Minimize the Users' Memory Load: It is important that the user should not have to remember information from one part of a dialogue to another
4. Consistency: Words and actions circumstances should continually signify the equivalent thing as subject to where they happened in the system
5. Feedback: Whatever is currently going on in the system should be significant and appropriate also clients should as well constantly be updated
6. Clearly marked Exits: There is the necessity for a speedy crisis outlet to arrival to preceding state exclusive of engaging in stretched conversation whenever an error(s) are made in choosing functions.
7. Shortcuts: This is unnoticed by a inexperienced user but been requisite to a professional user and this aid in fast-moving communication with the scheme
8. Worthy fault notes: This is necessary to be comprehensible and be in simple grammar to user and has to be precise to single out difficulty then recommend a result.
9. Counteract Faults: A cautious plan would intercept a hitch from happening.
10. Help and documentation: A system that is of highest standard can be made use of without engaging in any documentation

3.5 HCI Models.

Various different models have been projected which are intended to offer an HCI theory a certain framework. The model includes

- a. Norman's Model
- b. Abowd and Beale's Model (Interaction Model)
- c. Audience Participation Model

a. **Norman's Model:** This approach is centered on the consumer mapping out a scheme and then undertaking it at the

interface. This approach is classified into 7 stages

- i. Launching the purpose
- ii. Making the plan
- iii. Detailing the attainment structure
- iv. Implementing the act
- v. Identifying the scheme state
- vi. Construing the scheme state
- vii. Appraising the scheme state with reference to the purpose and plan prepared or set down

b. **The Interaction Model:** This model although established on Norman's model was considered to be further achievable and this model also produced an interaction framework (Dix et al, 1992). This consist of 4 key constituents which are listed as

- i. The method
- ii. The operator
- iii. The idea
- iv. The production

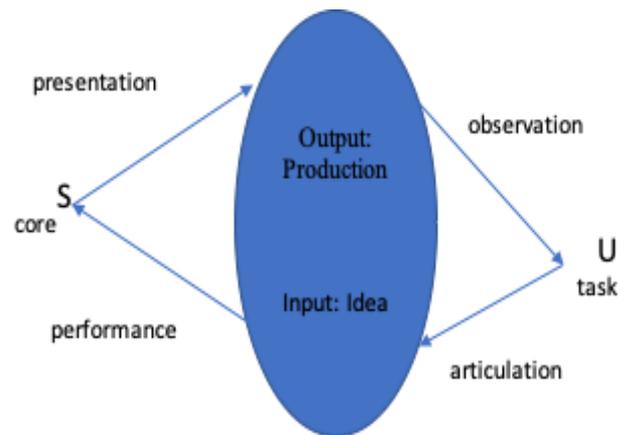


Fig. 2. The Overall Communication Structure (Source: Hinze-Hoare V, 2007)

Task (idea) is being put together by an operator in order to accomplish a purpose, the device is then influenced by means of hinged idea through the means of idea language. The idea language is thereafter interpreted for the system primary linguistic to carry out the processes then the method is attributed as the production. The production is consequently conveyed to the client by observation (Hinze-Hoare V, 2007).

c. **Audience Participation Model:** As taken into consideration by Nemirovsky (2003) that the old approach is that of computers as deterministic boxes irrationally following the computer's directives while handlers are incompetent of altering the progress of the program consecutively on the computer (Hinze-Hoare V, 2007).

3.6 HCI Systems Architecture

Human Computer Interaction System Architecture illustrates whatever it is that the inputs and outputs have being together with how they work with each other. The arrangement and designs upon which an interface is established is as follows (Karray, Alemzadeh, Saleh & Arab, 2008)

1. **Unimodal HCI Systems:** Every one of the distinct independent definite medium is called modality. Unimodal refers to a method which is centered on simply one modal and they are divided into three categories namely (Nande, 2014, Karray et al, 2008)

- a. **Visual Based:** Research field in this category includes Body movement tracking (Large Scale), Facial Expression Analysis, Eye-movement tracking (Gaze Detection), Gesture Recognition and so on (Karray et al, 2008, Nande, 2014))
- b. **Audio Based:** Some of the research area here includes Speaker recognition, Speech recognition, Musical interaction, Human-Made noise/Sign detections (Gasp, Sigh, Laugh, Cry etc) and so on (Karray et al, 2008, Nande, 2014)
- c. **Sensor Based:** Areas includes Joysticks, Mouse and keyboard, Pen-Based interaction (mobile devices, pen

gesture & handwriting recognition), Pressure sensors, Haptic sensors (Robotics and Virtual reality), motion tracking sensors and Digitizers, Taste/Smell sensors and so on (Karray et al, 2008, Nande, 2014)

2. **Multimodal HCI Systems:** This refers to the amalgamation of several modalities (Karray et al, 2008, Nande, 2014). These modalities mostly refer to the ways the system response to the input that is communication channels. Despite the improvement made in multimodal HCI, the modalities are still attended to independently and at the end only is when the outcomes of the separate modalities are being merged with each other (Karray et al, 2008).

According to Karray et al, 2008, application Areas of Multimodal Systems are as follows

- a. Multimodal Systems for Disabled people
- b. Emotion Recognition Multimodal Systems
- c. Map Based Multimodal Applications
- d. Multimodal Human Robot Interface Applications
- e. Multimodal HCI in Medicine

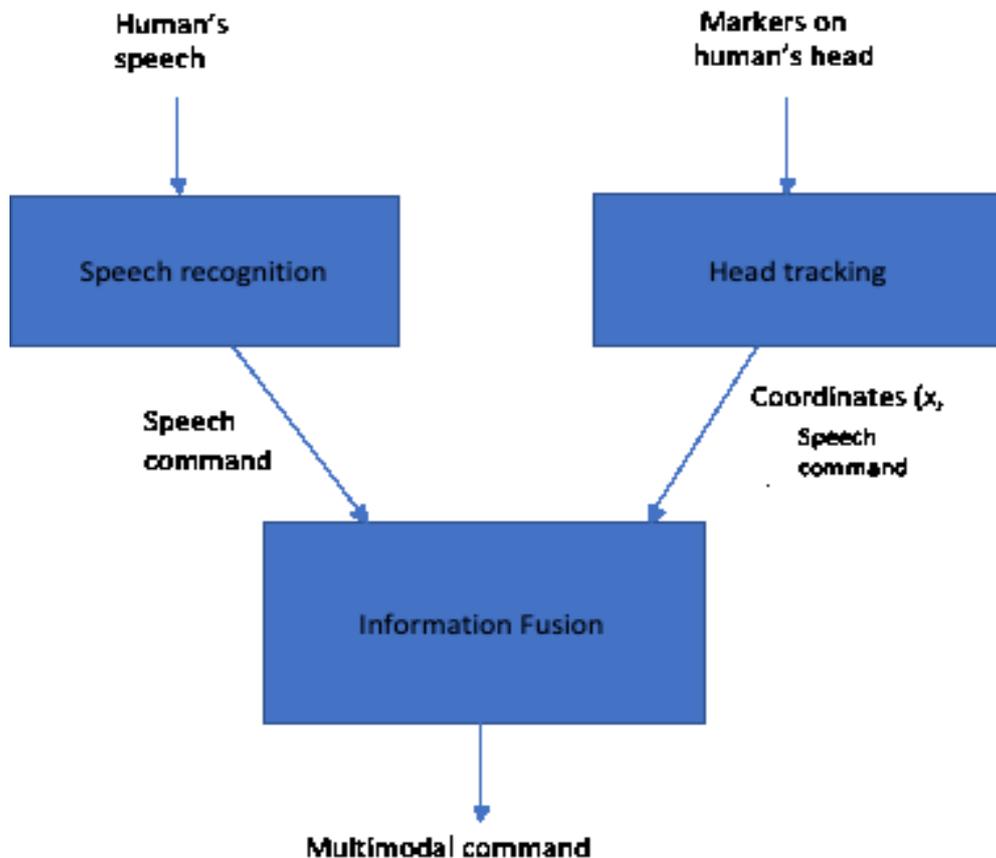


Fig. 3. Diagram of a bimodal structure (Source: Ronzhin & Karpov, 2005)

3.7 Speech Recognition.

Speech recognition refer to the process of transforming whichever speech signal into its orthographic depiction (Chandra & Akila, 2012). Speech Recognition is a process of transcribing speech to text (Sajjan & Vijaya, 2015). Automatic Speech Recognition (ASR) system allows a computer to identify the words a person speaks into a microphone or telephone and then convert it into text (Choudhary, Chauhan & Gupta, 2013).

3.8 Phases of Speech Recognition.

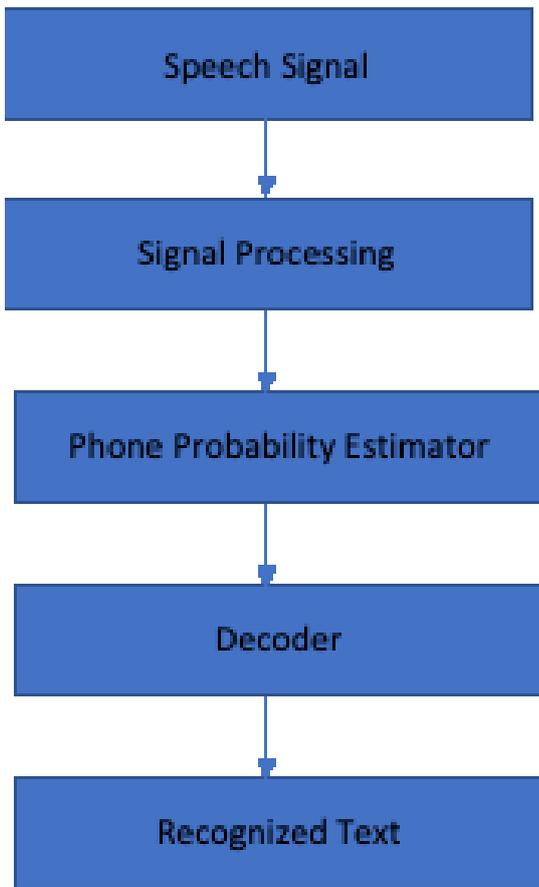


Fig. 4. Phases of Speech Recognition

- a. **Speech Signal:** The words spoken by humans are taken in as resonance and digitalized with the application of microphone. The digitalized signal is then transferred to signal processing unit.
- b. **Signal Processing:** In this phase, feature extraction is being executed. It transforms linear amplitude signal into insubstantial description (Janin, 2004).
- c. **Phone Probability Estimation:** Probability that the specified attributes denote a certain sound or composition of sounds in the language is been estimated in this phase (Janin, 2004).

- d. **Decoding:** The decoder accepts the succession of approximations from the prior phase and likens them in opposition to models of each potential speech in the language and thereafter it produced the furthestmost prospective speech (Chandra & Akila, 2012).

3.9 Applications of Speech Recognition

Some of the applications of speech recognition outline by Chandra & Akila, 2012, are stated below

1. Data Entry Enhancements in an Electronic Patient Care Report (ePCR)
2. Dictation
3. Command and Control
4. Telephony
5. Wearable
6. Medical/Disabilities
7. Embedded Applications
8. Agricultural application to get farmer queries.

3.10 Review of Literature

Chandra & Akila, 2012 described some speech synthesis and speech recognition algorithms and a comparative analysis of their performance which is based on accuracy and quality was conducted. Dynamic Time Warping (DTW) and Hidden Markov Model (HMM) algorithms were compared in speech recognition based on accuracy while CELP and MBROLA algorithms of speech synthesis were compared based on quality.

Sajjan & Vijaya, 2015 study investigated speech recognition using Monophone and Triphone based continuous density hidden Markov models and it was perceived that recognition accuracy intensify the sum of mixture constituents is increased and it functioned satisfactorily for the tied-state Triphone centred Hidden Markov Model (HMM) for huge languages. For the implementation of the system, Continuous Speech Corpus was utilized and the recognition accuracy was examined for the documented speeches.

Ibrahim, Odiketa & Ibiyemi, 2017, examined a research on preprocessing of a speech recognition established on human computer interaction system. The author discussed the procedures involved in preprocessing of speech recognition which includes Noise removal, Voice Activity Detection, Pre-emphasis Framing and Windowing. Applications being related to disable individual that could not be able to make use of the computer system through mouse and keyboard. could be able to make use of the system developed. They can use the computer which is enabled with an automatic speech recognition features and they operate the system with speech commands.

Gawali¹, Gaikwad, Yannawar, Mehrotra, 2011, postulated a Marathi Database and a confined characters recognition method. Mel-frequency Cepstral Coefficient (MFCC) and Distance Time Warping (DTW) were applied for the feature extraction method. Vocabulary comprises of Vowels of Marathi and restrained words that begins from a vowel and simple Marathi sentences. Marathi speech database used in this study was designed by using the computerized Speech lab for the feature mining. The authors made use of thirty-five speakers and they were documented and all of the words happened to be recurred thrice. The study also offered the evaluation of recognition accuracy for DTW and MFCC.

Pruthi, Saksena, & Swaranjali, 2010. discussed a speech maker related, unique word identifier for Hindi. LPC was used for the feature extraction and the technique was established by applying HMM. The sounds of two speakers (Men) were documented and Hindi digits were the vocabulary applied by the authors. Digit 0 is pronounced as 'Shoonya' to 9 and they are pronounced as 'nau'. The system is speaker subjected and has very little vocabulary.

Gupta, 2006, suggested a remote word speech detection tool for Hindi language. A continuous HMM was executed and word acoustic model was utilized for the system implementation. The speech vocabulary comprises of digits of Hindi languages. The system was verified with speaker reliant model and the outcome indicated that the system is of high standard and acceptable. The major shortcoming of this study is that the size of the vocabulary is very few.

Kumar & Aggarwal, 2011 research was based on ASR for Hindi language. The authors developed a speech recognition method for Hindi language and they used Hidden Markov technique device kits (HTK). Audio word approach was utilized to identify the segregated speeches. 30 Hindi words were instructed for the implementation and the instructed data were gathered from 8 speakers. The authors concluded that the general accuracy of the investigated method was 94%

Paul, Das & Kamal, 2009, established Bangla Speech Recognition System applying Automatic neural network and LPC. Two main parts were implemented. The first is signal processing while the second is speech pattern recognition approach. Here at the speech processing phase, the starting and end point were discovered and the speech signals were documented by means of an acoustic wave stereo in the natural room atmosphere. Artificial Neural Network was used for the implementation of pattern recognition.

Al-Qatab 2010, executed an Arabic automatic speech detection device by exploiting HTK. Both the steady and remote words were accepted by the device. The system been brought into existence make use of an Arabic vocabulary built manually by the speech sounds of thirteen speaker and 33 words vocabulary were applied.

Vyas & Suthar, 2017, provided short reports concerning speech recognition application by incorporating several speech parameters, diverse techniques to speech identification, essentials of acoustic methods, language methods, complex algorithms and feature extraction approaches. The study mainly focused on speech processing along with its relevance in technologies like Human Computer Interface (HMI) and numerous daily life applications

Venkateswarlu, Teja & Vasantha, 2012, established a competent speech identification scheme for Telugu letter identification. Both multilayer perception and time lagged recurrent neural network methodologies were trained and tested (both are same) on a dataset that consists of both sex individuals (2 each of both sex), they were then permitted to put into words the letters in 10 places. The study made use of speaker dependent mode for identification of the Telugu characters.

Kumar, Aggarwal & Jain 2012, research was based on connected-words speech identification scheme for the language Hindi and Hidden Markov methodology toolkit (HTK) was utilized to implement the technique. The scheme was being prepared to identify whichever way the structure of the words is represented.

Kumar, 2010 executed a scheme that required simply one speaker and identify remote word for the Punjabi language. In addition, the author compared the performance of speech identification scheme for little vocabulary of speaker reliant remote spoken letters by means of Hidden Markov Model (HMM) along with Dynamic Time Warp (DTW) approaches. The study examined emphasized on template-based identifier method applying linear forecasting programming accompanied by dynamic programming computation and vector quantization with Hidden Markov Model cantered identifiers in remote word identification jobs.

Abushariash, 2010, research was founded on English speech recognition. This study seeks to project and execute English digits speech identification scheme making use of MATLAB (GUI). Hidden Markov Model was the bases of the study and this provided a highly way for identifying speeches. Mel frequency Cepstral Coefficients (MFCC) method was utilised to obtain the features. This study focused on every one of the remote letters recognition jobs.

Kumar & Singh 2011, designated a speaker independent, real time, remote word. Vector Quantization and Dynamic Time Warping techniques were applied in executing ASR scheme for Punjabi language and the both techniques VQ as well as DTW were utilised for the system identification. LPC coefficients of the training data was generated for training and the testing of the scheme and they were likened with one and all reference pattern by means of dynamic time wrap alignment. The method was established for small remote word vocabulary.

3.11 Speech Recognition Technique

SRT were discussed and summarized in the table 1 below.

Table 1. Table 1. Speech Recognition Techniques

S/N	Speech Recognition Techniques	Advantages	Disadvantages
1.	Acoustic Phonetic Recognition	<ul style="list-style-type: none"> It decreases method time for associated words 	<ul style="list-style-type: none"> Applications for commercial purposes don't usually use it due to the high implementation of each isolated characters
2.	Pattern Recognition Technique	<ul style="list-style-type: none"> It identifies pattern speedily, effortlessly and automatically as a reason that matching of word to word exist. 	<ul style="list-style-type: none"> The process is slow, and it doesn't identify speeches if novel variant of pattern transpires. Another drawback is the template storing.
3.	Template Centered Technique	<ul style="list-style-type: none"> Minimal fault transpires owed to subdivision and categorization of little variable constituents and it is enhanced for isolated words 	<ul style="list-style-type: none"> It is costly owing to enormous vocabulary volume; template corresponding and formulation demands extra time, when input diverges from templates it break down fast
4.	Dynamic Time Wrapping	<ul style="list-style-type: none"> Consistent time alliance between reference and test pattern and steadiness is not so much essential 	<ul style="list-style-type: none"> It necessitates extreme time for complicated computational work templates are restricted, unidentified variation speaking indicator for training is been necessitated
5.	Vector Quantization Technique	<ul style="list-style-type: none"> Advantageous for resourceful data lessening 	<ul style="list-style-type: none"> It is being contingent on text
6.	Statistical Based Techniques	<ul style="list-style-type: none"> The size of the vocabulary is extremely great hence, can train huge quantity of data, the mathematical context is extremely correct instructed algorithms are effortlessly accessible can execute effortlessly and extra enhanced possess competent acquiring algorithm, can unsupportively discover variable 	<ul style="list-style-type: none"> Huge volume of data is needed and there is a high weighty rise in complication time
7.	Artificial Neural Network Technique	<ul style="list-style-type: none"> It doesn't take time is solving difficult computational jobs efficiently it can also take care of noisy and less quality data It also demands slightest training data vocabulary 	<ul style="list-style-type: none"> The outcome for huge vocabulary is not proficient it is costly require extra training time during architecture of neural network additional fault difference transpires, entire nature of neural network is not completely comprehended till now

4 CONCLUSION

Human Computer Interaction is an essential portion involves in designing systems. The way a client makes use of the system depict the quality of the system that had been designed. Hence, vast proportion of consideration had been discharged to

improve designs of HCI. This study endeavor to present an overview of matters such as and came up with a study on current researches in the HCI by means of an across-the-board literature review. The study also examined the overview of speech recognition techniques, the techniques, applications and even past research articles in this area of research.

REFERENCES

- [1] Nande S. B., Patel R. & Yadav V., 2014. Study of Human Computer Interaction (HCI) with its Application. Global Journal of Engineering Science and Research Management, 1(16), 47-52
- [2] Park H. & Mckilligan, 2018. A Systematic: literature Review for Human Computer Interaction and Design Thinking Process Integration. LNCS 10918. 725-740. https://doi.org/10.1007/978-3-319-91797-9_50
- [3] Preece, J., Rombach, H.D.: A taxonomy for combining software engineering and human-computer interaction measurement approaches: towards a common framework. Int. J. Hum.-Comput. Stud. 41(4), 553–583 (1994)
- [4] Hciprojectcit06, 2010. Human Computer Interaction (HCI) History and Research Trends. Hciprojectcit06.wordpress.com retrieved from <https://hciprojectcit06.wordpress.com/2010/03/14/human-computer-interactionhci-history-and-research-trends/amp/>
- [5] Dix, A., Finlay, J., Abowd, G.D., Beale, R.: Human-Computer Interaction. Pearson Prentice, New York (2004). *
- [6] Sharp, H., Rogers, Y., Preece, J.: Interaction Design: Beyond Human-Computer Interaction, 2nd edn. Wiley, Hoboken (2007). *
- [7] Dix, A., Finlay, J., Abowd, G. D., and Beale, R., (1992), Human Computer Interaction, 2nd ed., Prentice Hall, ISBN 0-13- 239864-8.
- [8] Hinze-Hoare, V. (2006) CSCR: Computer Supported Collaborative Research, <http://arxiv.org/ftp/cs/papers/0611/0611042.pdf>
- [9] Hinze-Hoare V., 2007. Review and Analysis of Human Computer Interaction (HCI) Principles
- [10] Nemirovsky, P. (2003) Redefining Digital Audience: Models and Actions, *Human Computer Interaction INTERACT 03*, Published by IOS Press, © IFIP, 2003, pp 391-398 <http://web.media.mit.edu/~pauln/research/publications.html>
- [11] Karray F., Alemzadeh M., Saleh J. A. and Arab M. N., 2008. Human-Computer Interaction: Overview on State of the Art. INTERNATIONAL JOURNAL ON SMART SENSING AND INTELLIGENT SYSTEMS,1(1), 137-159
- [12] A. Ronzhin and A. Karpov, “Assistive multimodal system based on speech recognition and head tracking”, *Proceedings of 13th European Signal Processing Conference*, Antalya (2005).
- [13] Speech Recognition on Vector Architectures by Adam Louis Janin, University of California, Berkeley, pg 21-27, December 2004.
- [14] Chandra E. & Akila A., 2012. An Overview of Speech Recognition and Speech Synthesis Algorithms. *Int. J. Computer Technology & Applications (IJCTA)*, 3(4), 1426-1430
- [15] Sajjan S. C. & Vijaya C., 2015. Speech Recognition Using Monophone and Triphone Based Continuous Density Hidden Markov Models. IJRSI, 2(10), 30-35
- [16] Choudhary A., Chauhan R. S. & Gupta G., 2013. Automatic Speech Recognition System for Isolated & Connected Words of Hindi Language by using Hidden Markov Model Toolkit (HTK). *Proc. of Int. Conf. on Emerging Trends in Engineering and Technology*, 847-853. DOI: 03. AETS.2013.3.234
- [17] Pruthi T, Saksena, S and Das, P K Swaranjali, “Isolated Word Recognition for Hindi Language using HMM and VQ”. International Conference on Multimedia Processing and Systems (ICMPS), IIT Madras (2010).
- [18] Gupta, R Speech Recognition for Hindi, M. Tech. Project Report, Department of CSE, IIT, Mumbai (2006). [3] Anup Kumar Paul, Dipankar Das, Md. Mustafa Kamal, “Bangla Speech Recognition System using LPC and ANN”, seventh international conference on advances in pattern recognition(ICAPR) on 4-6 Feb 2009, page no. 171-174.
- [19] B.A.Q Al-Qatab, “Arabic Speech Recognition Using Hidden Markov Model Toolkit (HTK)”, Paper presented at International Symposium in Information Technology (ITSim). Kuala Lumpur, June 15-17, 2010
- [20] R.L.K Venkateswarlu, R. Ravi Teja and R. Vasantha kumari , “Developing Efficient Speech Recognition System for Telugu Letter Recognition” computing communication and applications (ICCCA), 2012 International conference on 22-24 Feb , 2012.
- [21] R. Kumar, “Comparison of HMM and DTW for Isolated Word Recognition of Punjabi Language” In Proceedings of Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Sao Paulo, Brazil. Vol. 6419 of Lecture Notes in Computer Science (LNCS), pp. 244– 252, Springer Verlag, November 8-11, 2010.
- [22] Ahmad A. M. Abushariah et al, “English Digits Speech Recognition System Based on Hidden Markov Models”, international conference on computer and communication engineering (ICEEE ,2010), 11-13 May 2010, Kuala Lumpur, Malaysia.
- [23] R. Kumar and M. Singh, “Spoken isolated Word Recognition of Punjabi Language Using dynamic time Warping Technique” Demo in Proceedings of Information System for Indian Languages, Punjabi University, Patiala, India, March 9 - 11, 2011. Vol. 139 of Communication in Computer and Information Science (CCIS), Page 301, Springer Verlag.
- [24] Bharti W. Gawali1, Santosh Gaikwad, Pravin Yannawar, Suresh C. Mehrotra, “Marathi isolated word Recognition System using MFCC and DTW Features”, ACEEE Int. J. on information technology, vol. 01, no. 01.March 2011.

- [25] K. Kumar and R. K. Aggarwal “Hindi Speech Recognition System using HTK” International journal of computing and Business research ISSN vol. 2 issue 2 May 2011.
- [26] K. Kumar and R. K. Aggarwal and A. Jain, “A Hindi speech recognition system for connected words using HTK” International Journal Computational System Engineering, vol. 1, no. 1, 2012.
- [27] Ibrahim Y. A., Odiketa J. C. & Ibiyemi T. S., 2017. Preprocessing Technique in Automatic Speech Recognition for Human Computer Interaction: An Overview. Anale. Seria Informatica (Annals. Computer Science Series), 15 (1)
- [28] Vyas J. V. & Suthar A. C., 2017. Study of Speech Recognition Technology and its Significance in Human-Machine Interface. International Journal of Science Technology & Engineering, 3(10), 416-422
- [29] Mehmood F. & Khan S. An Overview on Speech Recognition System and Comparative Study of its Approaches