Review on Voice Recognition Techniques

[1]Krutika Chakor, [2]Mital Gaikwad, [3]Manali Haldar , [4] Abhishek Kothawale

[1]Student, [2]Student, [3]Student, [4] Student

[1]Information Technology,

[1]NDMVPS's KBT COE, Nashik, India

Abstract: The Speech is most essential & foremost means of communication among human beings. Today, speech technologies are commercially available with an unlimited but interesting range of functionalities. These technologies enable machines to respond to the voice commands and provide valuable services to users. This has become one of the latest technologies that are being used in mobile phones and computers to enhance the speed of execution of a function through voice input. This paper gives an overview of major technological perspective and appreciation of the fundamental progress of speech recognition. It gives an overview on technical development in each stage of speech recognition and also compares different speech recognition systems.

Index Terms - Artificial Intelligence, Automatic voice /speech recognition, Feature Extraction, Machine learning, Natural language processing, Speaker Identification.

I. INTRODUCTION

Now-a-days the world and the technologies are moving at the faster pace. As humans tend to spend less time on manual work and wants to increase the production more and more, the techno world is also building technologies to satisfy the budding demands. Speech being the most basic, common and efficient form of communication method for people to interact with each other. People are comfortable with speech therefore they would prefer to find it more convenient to interact with computers via speech, rather than using primitive interfaces such as keyboards and pointing devices.

More than anyone, world of computer is having the most advance technologies to build the environment which is suitable for all kinds of working and business. In era of digitization we are trying to reduce as much load as we can which requires less and less physical work and time. Hence when it comes to project management in IT sector there is demand in tools and techniques which could increase the development process and this could be achieved by automated systems with voice recognition systems.

II. VOICE RECOGNITION

A voice recognition system requires a processing unit which accepts and interprets the human voice. It consists of:
i.   A microphone: to send voice command
ii.   Software for voice recognition
iii.   A processing unit for interpreting the voice and performing the task
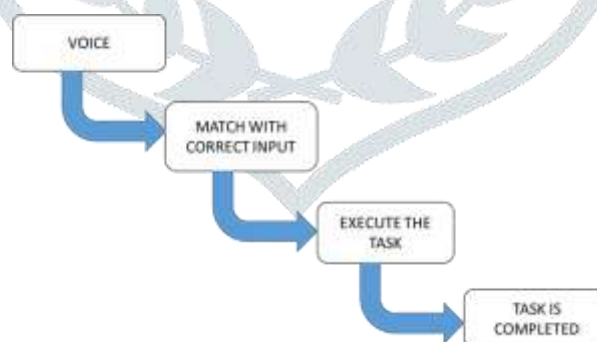The general process of voice recognition is mentioned in following figure:

fig.: steps involved in voice recognition

The steps involved in voice recognition are as follows:
1.   VOICE: In this step the user voice is accepted and converted into text.
2.   MATCH WITH CORRECT INPUT: In this phase the system checks if command is given accurately, also it matches it with the dictionary.
3.   EXECUTE THE TASK: When the second step is successfully completed the required task is performed by the system.
4.   TASK IS COMPLETED: In this phase the final result is displayed.

As there are so many spoken languages exist. We require the processing unit which will understand and translate voice commands in native languages [1]. Voice recognition involves capturing sequence of words, breaking them down and finding a best suitable match for the word. Then the software interprets the possible meaning. Automated voice recognition has a variety of

applications such as automatic call processing [2], Smart Mobile Attendance [3], Wireless room automation [4], Laughing voice recognition[5]and many more.

Voice recognition is based on Natural Language Processing[6]. Human voice is captured and processed using some techniques. In context of Natural Language Processing or Semantic modelling used in Semantic and Linguistic Grammars both define a formal way of how a natural language sentence can be understood. Linguistic grammar deals with linguistic categories like noun, verb, etc. while Semantic grammar, on the other hand, is a type of grammar whose non-terminals are not generic structural or linguistic categories like nouns or verbs, but rather semantic categories like person or a company.

Semantic Modelling[7] has undergone the renaissance, and now it is the basis of almost all commercial NLP systems such as Google, Cortana, Siri, Alexa, etc. Using supervised self-learning, the Semantic Model learns more with every curation and ultimately can know more than it was taught at the beginning. Hence, the model can start small and learn up through human interaction which the process that is not unlike many modern AI applications.

Semantic Modelling has gone through several peaks and valleys in the last 50 years. With the recent advancements of real-time human curation interlinked with supervised self-learning, this technique has finally grown up into a core technology for the majority of today's NLP/NLU systems[8]. Hence, we can say that in making of Siri or Alexa there is a Semantic Model working on the answer in backend of the system.

III. VOICE RECOGNITION TECHNIQUES

Many voice recognition techniques exist. In this paper some of them are explained.

Error Correction in voice input:

The authors Keith Vertanen, Per Ola Kristensson, of have focused on the creation of a dictionary for common spoken web search queries in the paper[9]. The main purpose of this project was to recognize and give output for the common web search queries. A corpus was created and how it was used to give accurate results. It uses search specific vocabulary which automatically generates superior pronunciation. This is better than the dictionary with fixed pronunciation. The main purpose of this work is to easily retrieve the web search queries and give the result faster.A formative user study was conducted to investigate recognition and correction aspects of voice search in a mobile context. This dictionary is built by using the data's recorded by the spy pages on the different search engines. Generally only one pronunciation for a particular word is stored. This creates problem, i.e. even if the user does a single mistake in pronunciation the output may not be displayed properly. This gives special attention to what are the different pronunciations for the same word the user may use. Though there was considerable amount of error rate, the user were able to get the output for the search queries in 18sec. The user was able to do this while walking about with the help of a touch screen mobile. The time for retrieval of output is considerably reduced by the use of this dictionary. The drawback of this system is that if a specific word is absent in the dictionary its output retrieval will be quite difficult.

Automatic Speech Recognition system:

In the paper author Douglas O'sShaughnessy [10] stated that speech is the most basic, common and efficient form of communication method for people to interact with each other. This can be accomplished by developing an Automatic Speech Recognition (ASR) system. This is the process of converting a speech signal to a sequence of words by means of an algorithm implemented as a computer program. It has the potential of being an important mode of interaction between humans and computers. The main goal of speech recognition area is to develop techniques and systems for speech input to machine. The research in ASR by machines has attracted a great deal of attention for about sixty years and ASR today finds widespread application in tasks that require human machine interface, such as automatic call processing, and also computer which can speak and recognize speech in native language.

Voice Recognition Based Model for Inputting:

Voice recognition technology is the method of identifying and understanding the voice signals of a user, which is converted into text or commands for a computer program. In this paper Teenu Therese Paul [11] stated that, the voice recognition technology is applied into a lab information system for identifying each of the technician's voice using the user's voice sample, and developing a secure authentication system, where the unique features of the user's voice are extracted and stored at the time of registration into the system. Further during the login stage, unique features of the user's new voice sample are extracted and then the system compares the features with all the stored features rather than the just previous one. For this, system sets a unique username to all the users. The comparison operation is performed with all voice samples under that particular user name and the voice feature comparison process is done by method of Fast Fourier Transform techniques. After a successful login the user can enter the results of smear test through his voice rather than typing into the system. The system mainly consists of two parts, a client system and a server system. The client system is developed using Android and the server system is implemented in Java.

 This system has a basic module for voice recognition and authentication which has been the most important contribution in the field of machine learning and will be useful in the implementation of a module which could accept voice command for various other functionalities.

Hybrid Genetic Algorithm:

The Hybrid Genetic model consists of two parts. The first one is concerned with the creation of feature patterns by fetching spoken words from speech by means of the discrete Fourier transform and frequency analysis.

The second part of model constitutes the working of design of the supervised learning and classification system. The technique used for the classification task is based on the simplest classifier, which is *K*-nearest neighbour algorithm. GAs, which have been considered a good optimization and machine learning technique, are applied to the feature extraction process for the pattern vectors in the algorithm. The main motive is to adapt such a hybrid approach to the task of voice command recognition, developing an implementation and to test and verify its performance. Following figure[12] shows the working:
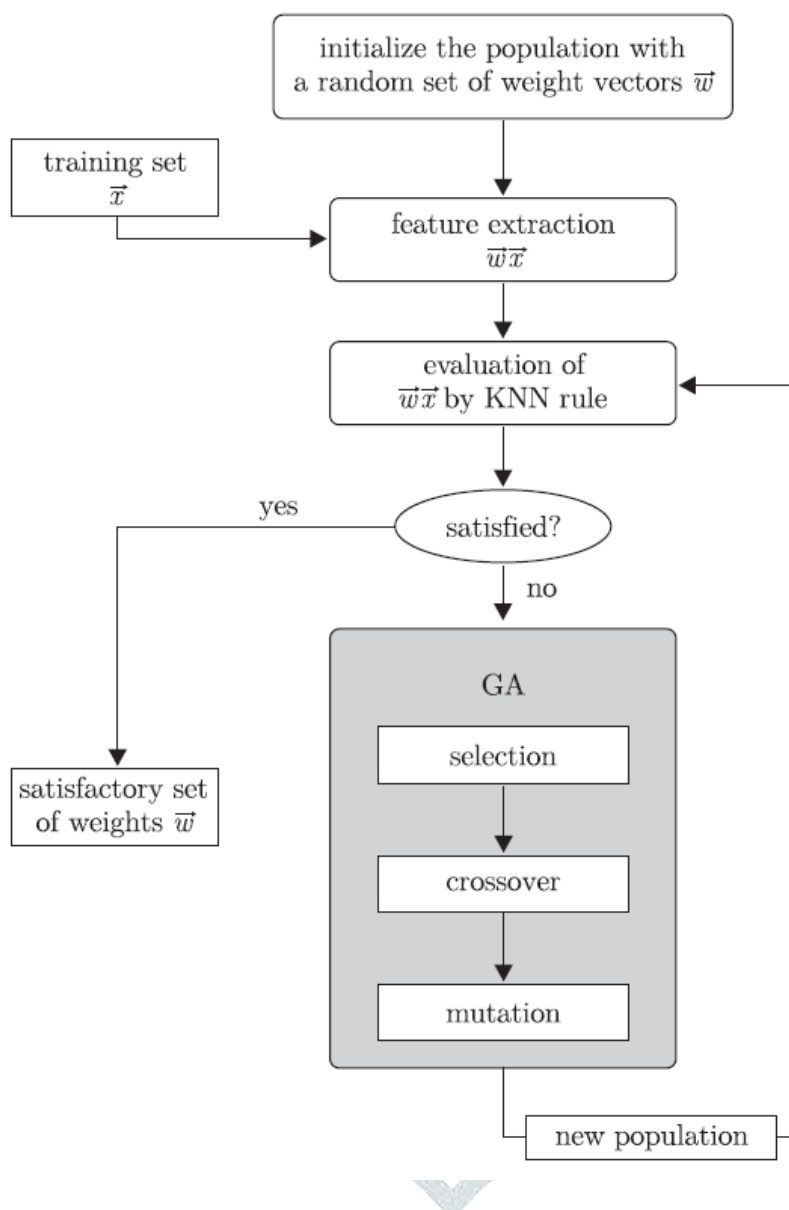

fig. : mode of action of the hybrid ga/knn approach

Cuckoo Search Algorithm:

Cuckoo search algorithm for voice recognition has been found to be one of the recent optimization algorithms in the league of nature based algorithm whose results are better than the PSO and ACO optimization algorithms. Cuckoo Search (CS) is heuristic search algorithm which is inspired by the reproduction strategy of cuckoo birds. Cuckoo search algorithm via Lévy flights by Xin-She Yang and Saush Deb for optimizing a nonlinear function uses generation of random numbers with symmetric Lévy distribution obtained by Mantegna's algorithm. The applications of Cuckoo includes optimizing weights of neural networks, parameters of Support vector machines and Radial basis function, job scheduling, finding optimal cluster head in wireless sensor networks, finding shortest path and clustering and is aimed to understand the breeding behaviour of the cuckoo bird.

This can be applied in the field of Speaker Recognition systems and voice to optimise is by use of a fitness function which will be used to match voices using features produced by the Cuckoo Search algorithm.

Speaker recognition can be optimized using Swarm Intelligence algorithm, Cuckoo Search. This algorithm aims at finding and short listing the features from voice which can be uniquely identified. After feature extraction using a threshold to remove the unwanted signal, noise or disturbance and considering only the voice sample, a Fitness Function is applied based on mean of the individual sample. The function will fetch few of the unique and best features and will discard the remaining ones. Hence, there is no need to match a speaker's voice through all the features present. To increase the security in authentication, a threshold value is

added to correlation in the matching phase. Since only the optimized features are extracted from voice command, this model will optimize this technique and but also save resources. Following flowchart[13] shows the working:
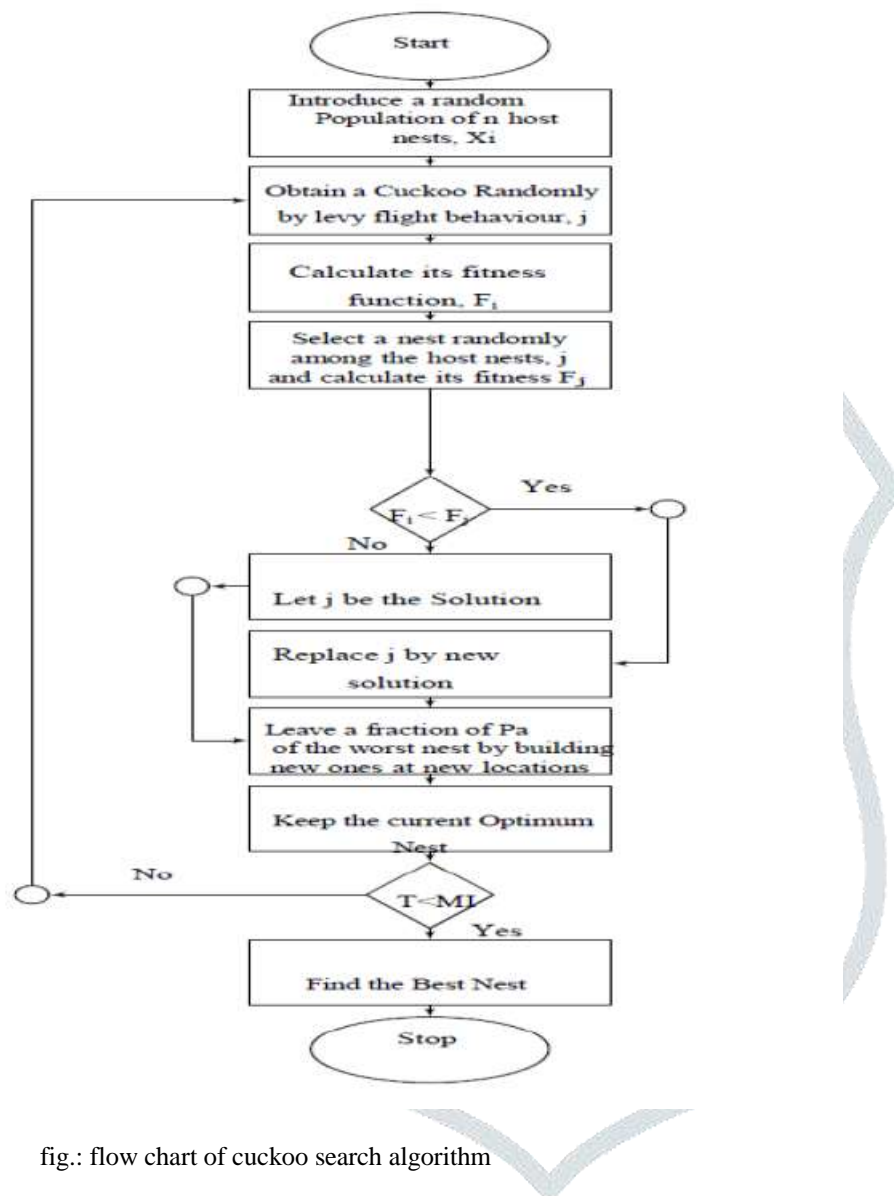

fig.: flow chart of cuckoo search algorithm

IV. CONCLUSION

From all the above papers we can conclude that among other methods voice recognition method is least explored and have a great prospect to be used for various purposes and have a great scope in field of access control, Law enforcement, transaction authentication and personalization. Hence, we are using the voice recognition technology of a home automated voice recognition device like Google Home for the purpose of personalizing and easing the work in field of database creation and package installation through voice commands.

REFERENCES

[1] Samudravijaya K. Speech and Speaker recognition tutorial TIFR Mumbai 400005.
[2] R.KlevansandR.Rodman, "Voice Recognition", ArtechHouse, Boston, London 1997.
[3] BenfanoSoewito; Ford Lumban Gaol ; Echo Simanjuntak; Fergyanto E. Gunawan, "Smart mobile attendance system using voice recognition and fingerprint on smartphone", ISITIA 2016.
[4] Avishek Paul ; MadhurimaPanja; MonalisaBagchi; Nairit Das ; RudrabrataMitraMazumder, "Voice recognition based wireless room automation system" ICICPI, 2016.

[5] TaisukeSakano; Takahiro Kigawa; Masanori Sugimoto ; FusakoKusunoki; Shigenori Inagaki, Hiroshi Mizoguchi,"Laughing voice recognition using periodic waveforms and voice-likeness features", ROBIO, 2016.

[6] A. Gelbukh,"Natural language processing", IEEE, 2005.

[7] Anatoly Anisimov, OleksandrMarchenko, VolodymyrTaranukha, andTarasVozniuk, "Semantic and Syntactic Model of Natural Language Based on Tensor Factorization", Springer International Publishing Switzerland 2014.

[8] EvgeniyGabrilovich, ShaulMarkovitch, "Wikipedia-based Semantic Interpretation for Natural Language Processing", Journal of Artificial Intelligence Research 34 (2009) 443-498.

[9] Keith Vertanen, Per Ola Kristensson, "Recognition and Correction of VoiceWeb Search Queries", University of Cambridge, Cavendish Laboratory, United Kingdom 2009.

[10] DOUGLAS O'SHAUGHNESSY, "Interacting With Computers by Voice: Automatic Speech Recognition and Synthesis", Proceedings of the IEEE, VOL. 91, NO. 9, September 2003, 0018-9219/03$17.00 © 2003 IEEE.

[11] Teenu Therese Paul, Shiju George,"Voice Recognition Based Secure Android Model for Inputting Smear Test Results", International Journal of Engineering Sciences & Emerging Technologies, Dec. 2013.

[12] Madhumita Dash, Ricky Mohanty, "Cuckoo Search Algorithm for Speech Recognition", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 10, October 2014.

[13] MARTA WRONISZEWSKA AND JACEK DZIEDZIC, "VOICE COMMAND RECOGNITION USINGHYBRID GENETIC ALGORITHM", Revised manuscript received 16 December 2010.