

**VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF
TECHNOLOGY**

Department of Computer Engineering



Project Report on

Virtual Assistant using Speech Recognition in Hindi

In partial fulfillment of the Fourth Year, Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai Academic Year 2017-2018

Submitted by

Mrudali Birla (D17 - A, Roll no - 09)

Ronit Dhonde (D17 - A, Roll no - 22)

Prashant Dombale (D17 - B, Roll no - 15)

Shreesh Rao (D17 - B, Roll no - 58)

Project Mentor

Asst. Prof. Ms. Kajal Jewani

(2017-18)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

Department of Computer Engineering



Certificate

This is to certify that *Mrudali Birla, Ronit Dhonde, Prashant Dombale, Shreesh Rao* of Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on “*Virtual Assistant using Speech Recognition in Hindi*” as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor *Asst. Prof. Ms. Kajal Jewani* in the year 2017-2018 .

This thesis/dissertation/project report entitled *Virtual Assistant using Speech Recognition in Hindi* by *Mrudali Birla, Ronit Dhonde, Prashant Dombale, Shreesh Rao* is approved for the degree of *B.E.*

Programme Outcomes	Grade
PO1, PO2, PO3, PO4, PO5, PO6, PO7, PO8, PO9, PO10, PO11, PO12, PSO1, PSO2	

Date:

Project Guide: Asst. Prof. Ms. Kajal Jewani

Project Report Approval

For

B. E (Computer Engineering)

This thesis/dissertation/project report entitled *Virtual Assistant using Speech Recognition in Hindi* by *Mrudali Birla, Ronit Dhonde, Prashant Dombale, Shreesh Rao* is approved for the degree of *B.E.*

Internal Examiner

External Examiner

Head of the Department

Principal

Date:

Place:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Signature)

Mrudali Birla (D17 - A, Roll no - 09)

(Signature)

Ronit Dhonde (D17 - A, Roll no - 22)

(Signature)

Prashant Dombale (D17 - B, Roll no - 15)

(Signature)

Shreesh Rao (D17 - B, Roll no - 58)

Date:

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Ms. Kajal Jewani** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair** , for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement at several times.

Computer Engineering Department
COURSE OUTCOMES FOR B.E PROJECT

Learners will be able to,

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solution for the problem.
CO 4	Able to interpret the data and datasets to be utilized.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one's competency.
CO 11	Able to develop professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

Abstract

Speech is one of the most common and widely accepted mode of communication between humans. It has been working efficiently well and a number of attempts have been made to use it for human computer interaction.

Hindi is a widely used language in India that is written in Devanagari script. It is spoken by about 260 million speakers, and yet the input methods for Hindi are obscure and are difficult to learn. Hence, a speech-based system in Hindi would be very useful typically for people in rural areas.

Our project provides assistance to the user in different forms. The two main components of the system are: 1. The Virtual Assistant and 2. The Chatbot. User can easily perform various tasks like open apps, perform complex calculations, get information, send emails and so on. The user can also have a conversation with the system Chatbot. The Chatbot can also solve the user's queries, if any.

One of the major purpose of this project is to provide a link between the illiterate or not very well-educated people and technology. By abridging this gap, things would be much easier for them and they'll move one step closer to modern future. Another possible use is that, this system could also be used by some business to expand its reach across the globe.

Table of Contents

Chapter no.	Title	Page no.
1	Introduction	1-4
1.1	Introduction to the project	1
1.2	Motivation	2
1.3	Problem Definition	3
1.4	Relevance of the project	3
1.5	Methodology employed for development	3
2	Literature Survey	5-15
2.1	Literature survey and its various sources	5
2.2	Newspapers referred	14
3	Requirement Gathering	16-17
3.1	Functional Requirements	16
3.2	Non-Functional Requirements	16
3.3	Constraints	16
3.4	Hardware and Software Requirements	16
3.5	Selection of the Hardware, Software, Technology and tools	17
4	Proposed Design	18-22
4.1	Block Diagram of the system	18
4.2	Modular design of the system	19
4.3	Detailed Design	19
4.4	Project Scheduling & Tracking	22

5	Implementation Details	23-25
5.1	Algorithms for the respective modules developed	23
5.2	Comparative Analysis with the existing algorithms	24
5.3	Evaluation of the developed system	24
6	Testing	26
7	Result Analysis	27-32
7.1	Simulation Model	27
7.2	Parameters Considered	27
7.3	Screenshots of UI	27
8	Conclusion	33
8.1	Limitations	33
8.2	Conclusion	33
8.3	Future Scope	33
	References	34
	Appendix	37
	CD	

List of figures

Figure 1	Language Distribution Pie-Chart
Figure 2	Virtual Assistant Users Bar-Graph
Figure 3	Block Diagram
Figure 4	Modular Diagram
Figure 5	DFD level 0

Figure 6	DFD level 1
Figure 7	DFD level 2
Figure 8	Flowchart
Figure 9	Gantt chart (1)
Figure 10	Gantt chart (2)
Figure 11	WER testing
Figure 12	User Interface Screenshot
Figure 13.1	Search
Figure 13.2	Email
Figure 13.3	Opening Application (1)
Figure 13.4	Opening Application (2)
Figure 13.5	Performance of Task (1)
Figure 13.6	Performance of Task (2)
Figure 13.7	Conversation with Chatbot (1)
Figure 13.8	Conversation with Chatbot (2)

List of tables

Table 1	Unit Testing
Table 2	Plagiarism report (1.1)
Table 3	Plagiarism report (1.2)
Table 4	Plagiarism report (1.3)
Table 5	Plagiarism report (2.1)
Table 6	Plagiarism report (2.2)

1.Introduction

1.1 Introduction to the project

In modern civilized societies, for communication between humans, speech is one of the most common methods. Different ideas formed in the mind of the speaker are communicated by speech in the form of words, phrases, and sentences by applying some proper grammatical rules. Speech is the primary mode of communication among human being and also the most natural and efficient form of exchanging information among humans.

Hindi is the native language of most people living in Delhi, Uttar Pradesh, Uttarakhand, Chhattisgarh, Himachal Pradesh, Chandigarh, Bihar, Jharkhand, Madhya Pradesh, Haryana, and Rajasthan. In Hindi, there is one-to-one correspondence between the sound and the character that represents that sound. So, developing a virtual assistant hindi speech recognition system. In rural areas, people are not very well educated but most of them can communicate in Hindi. So, to make their lives easier and give them an experience of technological luxury, we need to develop the system in hindi.

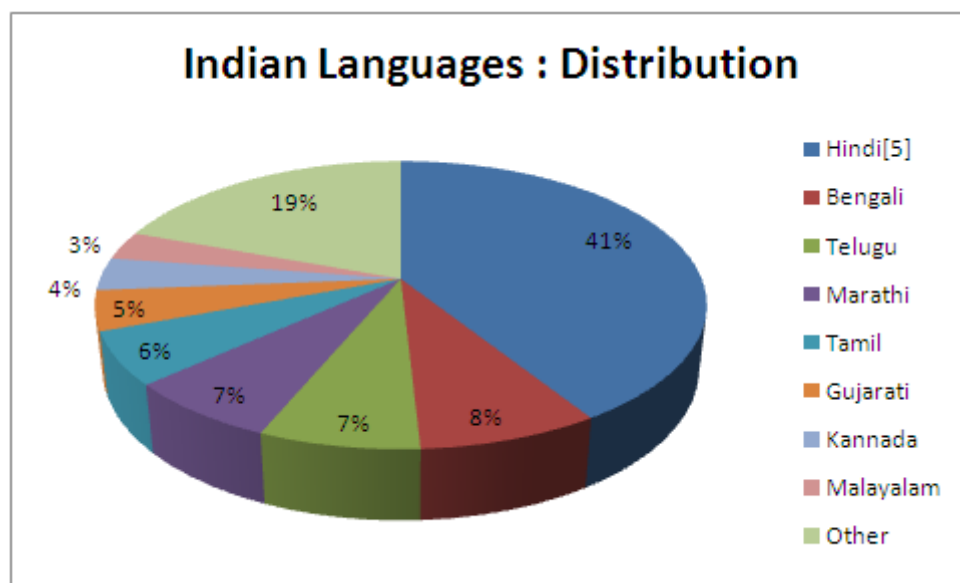


Fig 1: Language Distribution

People in rural parts of India, involved in some small scale or large scale industry often come across certain hurdles and need solution for their problems. A Chatbot is a good example of conversational partner which can have a normal conversation and at the same times resolve various problems. It takes in user queries and performs the task assigned or gives the solution to a problem.

In today's world, technology has made human lives more comfortable by providing various features. Speech recognition is one of the features which makes it easier to communicate with a device like computers or mobile phones. Virtual assistants makes its more efficient by doing tasks for the user. Speech recognition is

the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-readable format. A speech recognition system consists of a microphone, for the person to speak into; speech recognition software; a computer to take and interpret the speech; a good quality sound card for input and/or output.

A virtual assistant is a software agent that can perform tasks or services for an individual.

Existing virtual assistants which are most popular are:

Apple's Siri

Google's Ok Google

Amazon Alexa

Microsoft Cortana

1.2 Motivation

Virtual personal assistant should be accessible by everyone irrespective of the language people use. In some rural parts of India, people still cannot speak or write in English. Due to this issue, people don't tend to use VPAs on a larger scale even though they make functioning of tasks extremely easy and effortless. Our system mainly caters to this section and allows them to experience the ease as well as thrill of using a VPA.

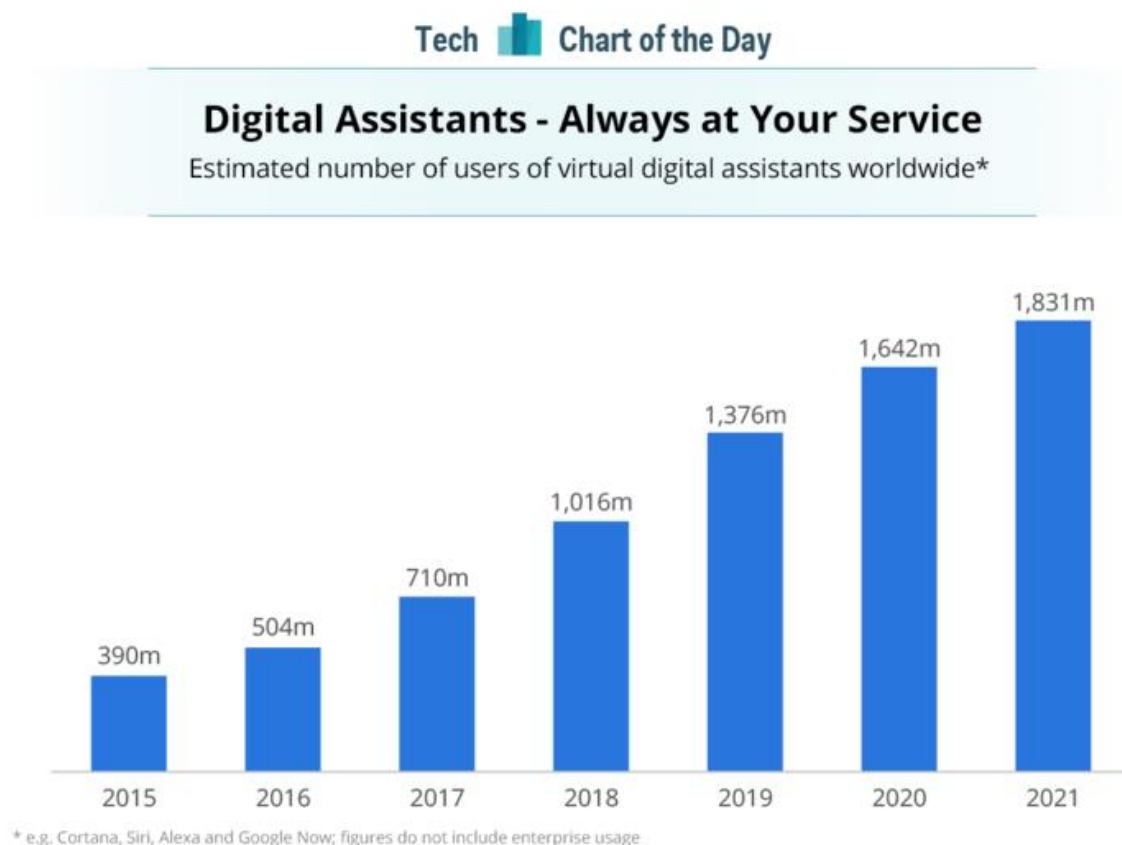


Fig 2: Virtual Assistant Users worldwide

1.3 Problem Definition

Our motive is to create a system which will take input in the form of speech in Hindi, understand it using speech recognition and perform the task specified, in addition to answering the multiple queries that the user may ask. In case of chatbot, give feasible solution or at least guide as to how to go about a particular problem.

1.4 Relevance of the Project

Speech recognition allows converting speech into text, making it easier both to create and to use information. Speech is easier to generate, it's intuitive and fast, but listening to speech is slow, it's hard to index speech, and easy to forget. Text is easier to store, process and consume, both for computers and for humans, but writing texts slow and requires some intention. In the past people mostly imagined speech recognition directly producing the end-result, e.g. a dictated document or a computer performing a command. This is a limited perspective, as availability of speech recognition is likely to make possible much more varied applications. Also, with the inclusion of a Chatbot with the system broadens the scope of the project greatly. The project is most efficient due to its user friendly nature.

1.5 Methodology employed for development

For our system, we have used the hybrid of Waterfall model and Adaptive Project Framework (APF).

1. Waterfall:

The waterfall model is the most common design process used in software development. In the waterfall model, the phases of development do not overlap. Each stage must be completed before the next “level” of the waterfall is begun. This makes the waterfall model most appropriate for our system, especially for the initial stages of the system. With this model, if a revision is needed, it is easier and cheaper to make that revision early. When using the waterfall model, the emphasis lies on making sure the requirements and design fit the needs of the project before moving on to the later stages of development.

The six life cycle stages of the original waterfall model are:

1. Requirements - In this phase, the expectations and goals of the project are defined, and risks are analysed. So, for our system, we start with the predetermined and pre-acquired set of requirements.
2. Design - After identifying the goals of the project, we start commencing on designing the product. We create a blueprint for the members with a plan for meeting the requirements.
3. Construction - Also known as development, implementation or coding, where the software for the project is written. Main coding is done here.

4. Testing and Debugging - In this stage, the software is stress-tested and combed over for errors, and to make sure the product meets the client's requirements.
5. Installation - After testing and debugging are completed, the product is implemented according to the agreed-upon requirements. Another round of testing and verification will follow after implementation.
6. Maintenance - After the delivery of the product to the client, we will ensure the product continues to work as designed.

2. Adaptive Project Framework:

Adaptive project management is a structured and systematic process that will allow us to gradually improve our decisions and practices, by learning from outcomes of the decisions that we take at each stage of the project. Adaptive Project Framework (APF) adjusts scope at each iteration. APF is more like creating a recipe more than following one. When fused with waterfall model, it is both systematic as well as improvising, thus providing the best of both models.

Main characteristics of Adaptive Framework are:

- a. Thrive on change
- b. Client driven

What distinguishes APF from other methodologies is the fact that it places the client as the central figure who decides on project's next step. In our system, after deploying the project our task is not over. The product changes according to the needs of the client for eg.: If system is developed for schools in rural areas and the users come across certain wanting aspects of our system or identify some functionality that might be useful for them, then we need to update our system according to our clients needs.

2. Literature Survey

2.1 Literature Survey and its various sources

1) Paper Name: A review on Speech recognition techniques

Published in: International journal of computer applications Vol 10-No 3

Date: November 2010

Abstract:

The Speech is most prominent & primary mode of Communication among of human being. The communication among human computer interaction is called human computer interface. Speech has potential of being important mode of interaction with computer .This paper gives an overview of major technological perspective and appreciation of the fundamental progress of speech recognition and also gives overview technique developed in each stage of speech recognition. This paper helps in choosing the technique along with their relative merits & demerits. A comparative study of different technique is done as per stages. This paper is concludes with the decision on feature direction for developing technique in human computer interface system using Marathi Language.

Inference:

The paper mentioned, discusses the various kinds of techniques that can be used in the process of human speech recognition. Each of these techniques can be specified for a particular stage of the speech recognition system. According to the paper there are many methods by which efficient feature extraction of speech can be done and analysis can be performed on the recognized speech. Out the mentioned methods, the Mel-frequency cepstrum (MFCC's) turns out to the most fundamental method as it has a strong property of computing power spectrum by performing Fourier Analysis. As far the modeling techniques are concerned the Hidden Markov Model (HMM) can be considered to as the best option because in this, the system being modeled is assumed to be a Markov process with unobserved states.

2) Paper Name: An efficient speech recognition system

Published in: An International Journal (CSEIJ), Vol. 3, No. 4

Date: August 2013

Abstract:

This paper describes the development of an efficient speech recognition system using different techniques such as Mel Frequency Cepstrum Coefficients (MFCC), Vector Quantization (VQ) and Hidden Markov Model (HMM). This paper explains how speaker recognition followed by speech recognition is used to recognize the speech faster, efficiently and accurately. MFCC is used to extract the characteristics from the input speech signal with respect to a particular word uttered by a particular speaker. Then HMM is used on Quantized feature vectors to identify the word by evaluating the maximum log likelihood values for the spoken word.

Inference:

The MFCC and Distance Minimum techniques have proved out to be very efficient for the speaker identification system. The speech recognition phase uses the most fundamental HMM (Hidden Markov Model) algorithm. From the experimental results considered in this paper, it is found out that speaker recognition module improves the efficiency of speech recognition scores. MATLAB is a crucial tool that is responsible for the coding of all the techniques which are mentioned in the paper. It has been found that the combination of MFCC and Distance Minimum algorithm gives the best performance and also accurate results in most of the cases with an overall efficiency of 95%. The study also reveals that the HMM algorithm is able to identify the most commonly used isolated word. As a result of this, speech recognition system achieves 98% efficiency.

3) Paper Name: A REVIEW ON SPEECH TO TEXT CONVERSION METHODS

Published in: International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 7

Date: July 2015

Abstract:

Speech is the first important primary need, and the most convenient means of communication between people. The communication among human computer interaction is called human computer interface. This paper gives an overview of major technological perspective and appreciation of the fundamental progress of speech to text conversion and also gives overview technique developed in each stage of classification of speech to text conversion. A comparative study of different technique is done as per stages. This paper concludes with the decision on future direction for developing technique in human computer interface system in different mother tongue and it also discusses the various techniques used in each step of a speech recognition process and attempts to analyze an approach for designing an efficient system for speech

recognition. However, with modern processes, algorithms, and methods we can process speech signals easily and recognize the text. In this system, we are going to develop an on-line speech-to-text engine. However, the transfer of speech into written language in real time requires special techniques as it must be very fast and almost 100% correct to be understandable. The objective of this review paper is to recapitulate and match up to different speech recognition systems as well as approaches for the speech to text conversion and identify research topics and applications which are at the forefront of this exciting and challenging field.

Inference:

This paper represents the topics relevant to the development of STT systems. The speech to text conversion may seem effective and efficient to its users if it produces natural speech and by making several modifications to it. This system is useful for deaf and dumb people to Interact with the other peoples from society. Speech to Text synthesis is a critical research and application area in the field of multimedia interfaces. In this paper gathers important references to literature related to the endogenous variations of the speech signal and their importance in automatic speech recognition. A database has been created from the various domain words and syllables. The desired speech is produced by the Concatenative speech synthesis approach. Speech synthesis is advantageous for people who are visually handicapped. This paper made a clear and simple overview of working of speech to text system (STT) in step by step process. The system gives the input data from mice in the form of voice, then preprocessed that data & converted into text format displayed on PC. The user types the input string and the system reads it from the database or data store where the words, phones, diaphones, tri-phone are stored. In this paper, we presented the development of existing STT system by adding spellchecker module to it for different language. There are many speech to text systems (STT) available in the market and also much improvisation is going on in the research area to make the speech more effective, and the natural with stress and the emotions.

4) Paper Name: Speech Recognition Using Vector Quantization through Modified K-meansLBG Algorithm

Published in: ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online) Vol 3, No.7

Date: October 2012

Abstract:

In the Vector Quantization, the main task is to generate a good codebook. The distortion measure between the original pattern and the reconstructed pattern should be minimum. In this paper, a proposed algorithm called Modified K-meansLBG algorithm used to obtain a good codebook. The system has shown good performance on limited vocabulary tasks.

Inference:

This paper mainly represents the statements that the Vector Quantization Techniques can be efficiently applied in the development of speech recognition systems. In order to increase the performance of the speech recognition systems, a novel vector quantization algorithm called K-meansLBG is proposed. This new algorithm is a combination of the K-means and the LBG algorithm. Also, the recognition accuracy of the K-meansLBG algorithm is much higher as compared to the other two. From the analysis mentioned in the paper, the average recognition accuracy of K-meansLBG algorithm is more than 2.55% using K-means algorithm while the average recognition accuracy of K-meansLBG algorithm is more than 1.41% using LBG algorithm.

5) Paper Name:

Survey on Chatbot Design Techniques in Speech Conversation Systems

Published in: (IJACSA) International Journal of Advanced Computer Science and Applications,

Vol.6.

Date: November 2015

Abstract:

Human-Computer Speech is gaining momentum as a technique of computer interaction. There has been a recent upsurge in speech-based search engines and assistants such as Siri, Google Chrome and Cortana. Natural Language Processing (NLP) techniques such as NLTK for Python can be applied to analyze speech, and intelligent responses can be found by designing an engine to provide appropriate human like responses. This type of program is called a Chatbot, which is the focus of this study. This paper presents a survey on the techniques used to design Chatbots and a comparison is made between different design techniques from nine carefully selected papers according to the main methods adopted. These papers are representative of the significant improvements in Chatbots in the last decade. The paper discusses the similarities and differences in the techniques and examines in particular the Loebner prize-winning Chatbots.

Inference:

In this paper, the literature review has covered a number of selected papers that have focused specifically on Chatbot design techniques in the last decade. A survey of nine selected studies that affect Chatbot design has been presented, and the contribution of each study has been identified. In addition, a comparison has been

made between Chatbot design techniques in the selected studies and then with the Loebner Prize winning Chatbot techniques. From the survey above, it can be said that the development and improvement of Chatbot design is not grow at a predictable rate due to the variety of methods and approaches used to design a Chatbot. The techniques of Chatbot design are still a matter for debate and no common approach has yet been identified. Researchers have so far worked in isolated environments with reluctance to divulge any improved techniques they have found, consequently, slowing down the improvements to Chatbots. Moreover, the Chatbots designed for dialogue systems in the selected studies are, in general, limited to particular applications. General-purpose Chatbots need improvements by designing more comprehensive knowledge bases.

6) Paper Name: Modular Conversational Bot

Published in: (IJESC) International Journal of Engineering Science and Computing

Volume 7 Issue No.3

Date: March 2017

Abstract:

Human-Computer Speech is picking up energy as a strategy of computer interaction. There has been a late upsurge in discourse-based web crawlers and assistants, for example, Siri, Google Chrome and Cortana. Natural Language Processing (NLP) methods, for example, NLTK for Python can be connected to break down discourse, and smart reactions can be found by outlining an engine to give suitable human like reactions. This sort of project is known as a Chatbot, which is the center of this project. Main Objective of this project is to create a Conversational interface to common services like weather, location, government services. User can interact either by text or by speech with the bot. The user can also interact with the bot in their own regional language.

Inference:

The Modular conversational bot system makes use of the voice recognition, natural language processing and other technologies to make a simpler interface to common web services and information on the World Wide Web and other online sources. We are seeing that types of interfaces to the web are changing now a day. Characteristics of browsers in different devices are different. There are different types of devices such as hand-held personal digital assistants, smart digital telephones, and television set-up boxes. Input output format of each of these devices are different. The Modular conversational system provides multiple ways to access the web makes it flexible and universal over different devices and different languages. Because of these reasons and user friendliness this will reach up to a much larger audience. In this paper, the literature

review has covered a number of selected papers that have focused specifically on Chatbot design techniques in the last decade. In this paper, the literature review has covered a number of selected papers that have focused specifically on Chatbot design techniques in the last decade.

7) Paper Name: An intelligent web-based voice chat bot.

Published in: (IEEE) Institute of Electrical and Electronics Engineers.

Date: June 2009

Abstract:

This paper presents the design and development of an intelligent voice recognition chatbot. The paper presents a technology demonstrator to verify a proposed framework required to support such a bot (a web service). While a black box approach is used, by controlling the communication structure, to and from the web-service, the web-service allows all types of clients to communicate to the server from any platform. The service provided is accessible through a generated interface which allows for seamless XML processing; whereby the extensibility improves the lifespan of such a service. By introducing an artificial brain, the web-based bot generates customized user responses, aligned to the desired character. Questions asked to the bot, which is not understood is further processed using a third-party expert system (an online intelligent research assistant), and the response is archived, improving the artificial brain capabilities for future generation of responses.

Inference:

Using modular design for all its components a distributed environment facilitating transparent and high performance of the overall system has been created. The performance is relative to the processing capacity of the systems involved. Since all the modules are not running off one system the possible load has been decreased and further decreased by delegating the voice processing to the chat client communicating with the service. The use of an expert system (Ultimate Research Assistant) allows unlimited and autonomous intelligence improvements. Conventional implementations of the ALICE-bot engine required an administrator to update the AIML files manually to increase the intelligence. All content received back from the expert system is processed minimally since the information has already been processed for its relevance. This can be somewhat subjective considering that the intelligence reliance was shifted to another party and when such a third-party system is decommissioned this system would also fail.

8) Paper Name: A review on Speech recognition techniques

Published in: International journal of computer applications Vol 10-No 3

Date: November 2010

Abstract:

The Speech is most prominent & primary mode of Communication among of human being. The communication among human computer interaction is called human computer interface. Speech has potential of being important mode of interaction with computer. This paper gives an overview of major technological perspective and appreciation of the fundamental progress of speech recognition and also gives overview technique developed in each stage of speech recognition. This paper helps in choosing the technique along with their relative merits & demerits. A comparative study of different technique is done as per stages. This paper is concluding with the decision on feature direction for developing technique in human computer interface system using Marathi Language.

Inference:

The paper mentioned, discusses the various kinds of techniques that can be used in the process of human speech recognition. Each of these techniques can be specified for a particular stage of the speech recognition system. According to the paper there are many methods by which efficient feature extraction of speech can be done and analysis can be performed on the recognized speech. Out the mentioned methods, the Mel-frequency cepstrum (MFCC's) turns out to be the most fundamental method as it has a strong property of computing power spectrum by performing Fourier Analysis. As far the modeling techniques are concerned the Hidden Markov Model (HMM) can be considered to as the best option because in this, the system being modeled is assumed to be a Markov process with unobserved states.

9) Paper Name: An efficient speech recognition system

Published in: An International Journal (CSEIJ), Vol. 3, No. 4

Date: August 2013

Abstract:

This paper describes the development of an efficient speech recognition system using different techniques such as Mel Frequency Cepstrum Coefficients (MFCC), Vector Quantization (VQ) and Hidden Markov Model (HMM). This paper explains how speaker recognition followed by speech recognition is used to recognize the speech faster, efficiently and accurately. MFCC is used to extract the characteristics from the input speech signal with respect to a particular word uttered by a particular speaker. Then HMM is used on Quantized feature vectors to identify the word by evaluating the maximum log likelihood values for the spoken word.

Inference:

The MFCC and Distance Minimum techniques have proved out to be very efficient for the speaker identification system. The speech recognition phase uses the most fundamental HMM (Hidden Markov Model) algorithm. From the experimental results considered in this paper, it is found out that speaker recognition module improves the efficiency of speech recognition scores. MATLAB is a crucial tool that is responsible for the coding of all the techniques which are mentioned in the paper. It has been found that the combination of MFCC and Distance Minimum algorithm gives the best performance and also accurate results in most of the cases with an overall efficiency of 95%. The study also reveals that the HMM algorithm is able to identify the most commonly used isolated word. As a result of this, speech recognition system achieves 98% efficiency.

10) Paper Name: A REVIEW ON SPEECH TO TEXT CONVERSION METHODS

**Published in: International Journal of Advanced Research in Computer Engineering
& Technology**

(IJARCET) Volume 4 Issue 7

Date: July 2015

Abstract:

Speech is the first important primary need, and the most convenient means of communication between people. The communication among human computer interaction is called human computer interface. This paper gives an overview of major technological perspective and appreciation of the fundamental progress of speech to text conversion and also gives overview technique developed in each stage of classification of speech to text conversion. A comparative study of different technique is done as per stages. This paper concludes with the decision on future direction for developing technique in human computer interface system in different mother tongue and it also discusses the various techniques used in each step of a speech recognition process and attempts to analyze an approach for designing an efficient system for speech recognition. However, with modern processes, algorithms, and methods we can process speech signals easily and recognize the text. In this system, we are going to develop an on-line speech-to- text engine. However, the transfer of speech into written language in real time requires special techniques as it must be very fast and almost 100% correct to be understandable. The objective of this review paper is to recapitulate and match up to different speech recognition systems as well as approaches for the speech to text conversion and identify research topics and applications which are at the forefront of this exciting and challenging field.

Inference:

This paper represents the topics relevant to the development of STT systems. The speech to text conversion may seem effective and efficient to its users if it produces natural speech and by making several modifications to it. This system is useful for deaf and dumb people to Interact with the other peoples from society. Speech to Text synthesis is a critical research and application area in the field of multimedia interfaces. In this paper gathers important references to literature related to the endogenous variations of the speech signal and their importance in automatic speech recognition. A database has been created from the various domain words and syllables. The desired speech is produced by the Concatenative speech synthesis approach. Speech synthesis is advantageous for people who are visually handicapped. This paper made a clear and simple overview of working of speech to text system (STT) in step by step process. The system gives the input data from mice in the form of voice, then preprocessed that data & converted into text format displayed on PC. The user types the input string and the system reads it from the database or data store where the words, phones, diaphones, tri-phone are stored. In this paper, we presented the development of existing STT system by adding spellchecker module to it for different language. There are many speech to text systems (STT) available in the market and also much improvisation is going on in the research area to make the speech more effective, and the natural with stress and the emotions.

11) Paper Name: Speech Recognition Using Vector Quantization through Modified K-meansLBG

Algorithm

Published in: ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online) Vol 3, No.7

Date: October 2012

Abstract:

In the Vector Quantization, the main task is to generate a good codebook. The distortion measure between the original pattern and the reconstructed pattern should be minimum. In this paper, a proposed algorithm called Modified K-meansLBG algorithm used to obtain a good codebook. The system has shown good performance on limited vocabulary tasks.

Inference:

This paper mainly represents the statements that the Vector Quantization Techniques can be efficiently applied in the development of speech recognition systems. In order to increase the performance of the speech recognition systems, a novel vector quantization algorithm called K-meansLBG is proposed. This new algorithm is a combination of the K-means and the LBG algorithm. Also, the recognition accuracy of

the K-meansLBG algorithm is much higher as compared to the other two. From the analysis mentioned in the paper, the average recognition accuracy of K-meansLBG algorithm is more than 2.55% using K-means algorithm while the average recognition accuracy of K-meansLBG algorithm is more than 1.41% using LBG algorithm.

2.2 Newspapers referred

1. Newspaper Name: The Economic Times

Article Name: All you need to know about Voice Recognition

Date: 17 October 2011

Contents:

A BRIEF HISTORY

Speech recognition was first introduced to personal computing around 10 years back — around the time Windows 98 was introduced. However, you may be surprised to know that research on this technology started way back in 1936.

WHAT IS VOICE RECOGNITION?

Voice recognition and speech recognition are two different terms. Voice recognition relates to identifying an individual voice — along the same lines as a biometric scanner. Speech recognition, on the other hand, relates to identifying spoken words in the correct sense and then translating them into computer language.

HOW IT WORKS

Both speech and voice recognition work on the principal of translating 'analog' spoken words into 'digital' signals that a machine can understand. As simple as this may sound, it requires a lot of back-end processing, all the while compensating for differences in dialect, volume levels, tempo and pronunciation. Translated analog signals from speech, once converted, are then sent back to the device in digital format which in turn executes a command. Because speaking out a line takes mere seconds, translating, conversion and execution needs to be done on the fly — thus the need of a fast data connection to transfer the data to and fro.

WHAT CAN YOU USE IT FOR?

Speech to text and controlling a machine using your voice is obvious. But the technology holds promise for those with disabilities. Applications like DriveSafe.ly for your phone can read out text messages and emails for you — helpful for the visually impaired. Various apps also allow you to search the web or type out messages by speaking — helpful for those with limited motor control.

WHAT COMES NEXT?

The biggest challenge that any speech recognition system faces today is deciphering the various dialects and accents that people may have. Plus, in natural speech, we often tend to use a lot of slang, which automated systems find hard to understand. The first step would be to build a system that looks beyond any of these current issues. A possible application then, would be a universal, real-time voice translator, often seen in sci-fi movies — simply speak and a device will be able to instantly speak out the same in any language with 100% accuracy. Going forward, there are also going to be major developments in speech understanding — true artificial intelligence, when a machine can truly grasp the context of what you're saying and talk back, rather than just recognizing the words.

2. Newspaper Name: Times of India

Article Name: Your voice new key to phone banking

Date: 28 July 2016

Contents:

Phone banking is cumbersome at the best of times, with users required to punch in number after number for authentication. "First you type in your account or card number, then your T-PIN or date of birth, followed by the expiry date or CVV on your debit card. It is long, cumbersome and expensive for both Indian and NRI customers on international calls," says ICICI Bank executive director Rajiv Sabharwal.

A quicker method of authentication would help, especially in emergencies, such as when a panicky customer is trying to report loss or theft of the card. Enter voice recognition — it promises instant authentication, and is already being rolled out by ICICI Bank, Kotak Mahindra Bank and a few others. ICICI Bank, which started using voice recognition in May 2015, has already got 3 million customers on board and hopes to add another million by the year-end.

Data security experts, however, say voice recognition alone does not provide adequate protection. "Banks need to have systems that are smart enough to challenge the user, and not stop at voice recognition.

Voice is a relatively cheap option for banks compared with iris recognition or biometrics that require a special device. Voice is only effective when it is used along with other security measures," a security professional told TOI.

Experts say protecting a customer's voice sample is crucial. "If a hacker gets his hands on the voice sample, it will be very easy for him to fool the bank's system," an expert said.

3. Requirement Gathering

3.1 Functional Requirements

1. Efficaciously record speech and understand correctly- The recognized words should match with what the user said and should not be interfered by a certain level of noise.
2. Recognize speech in Hindi- The main goal is to recognize speech said in hindi.
3. Precision in performing tasks specified by the user- The system should understand and perform the command given by the user correctly.
4. Response time should be minimum - The system should be quick with the tasks assigned.
5. Given query should be solved or user should get an insight on how to solve.

3.2 Non-functional Requirements

1. Celerity in Performance- The system should be swift.
2. Security of data in operation- Personal data should be kept safe.
3. Scalability of system
4. Meet new requirements of the user- The system should be adaptable and improvable to meet the requirements of the user.

3.3 Constraints

1. A specific range of lexicons should be given as input.
2. The system should be given one task at a time to perform. Eg. “Email Mr.Rao” and not “Email Mr. Rao and send a text message to Mr. Dombale”.
3. Noise should be as low as possible while speaking into the microphone.
4. Only specific queries should be asked to assistant and not any random thing.
5. Chatbot is restricted to certain fields and is not completely exhaustive.
6. VA is also constrained to a fixed set of tasks which can be expanded later.

3.4 Hardware and Software Requirements

- Hardware Requirements: Computer, Microphone and Soundcard.
- Software Requirements: Python (3.6), Speech Recognition APIs like speech_recognition, ChatterBot, nltk (natural language toolkit), tkinter for UI etc for Python, APIs for wikipedia and WolframAlpha. Built-in libraries of python.
- Technology:

- Speech processing - The input in the form of speech is accepted by the system and processed using techniques used in cloud speech API.
- Machine learning techniques - Supervised learning methods like Bayesian classification are used in training chatbot.
- Tools: NLTK, tkinter

3.5 Selection of Hardware and Software

The primary task of our system is to recognize speech which cannot be done without a microphone. Since our project is a computer application the programming language we have decided to use is Python. Python shrinks the gap between capturing valid requirements, writing code, and designing automated tests. Rooted in strong daily cross-functional collaboration, Python ensures speedy, qualitative, and high-value applications. It is easy to customize neural networks and NLP with the help of various libraries in Python. Python supports many speech recognition engines and APIs, including Google Speech Engine, Google Cloud Speech API. The speech recognition API helps in converting speech to text whereas the tkinter API is used to provide a UI for our system. Tkinter is an easy to use API for building UI for Python applications.

4. Proposed Design

4.1 Block Diagram of the system

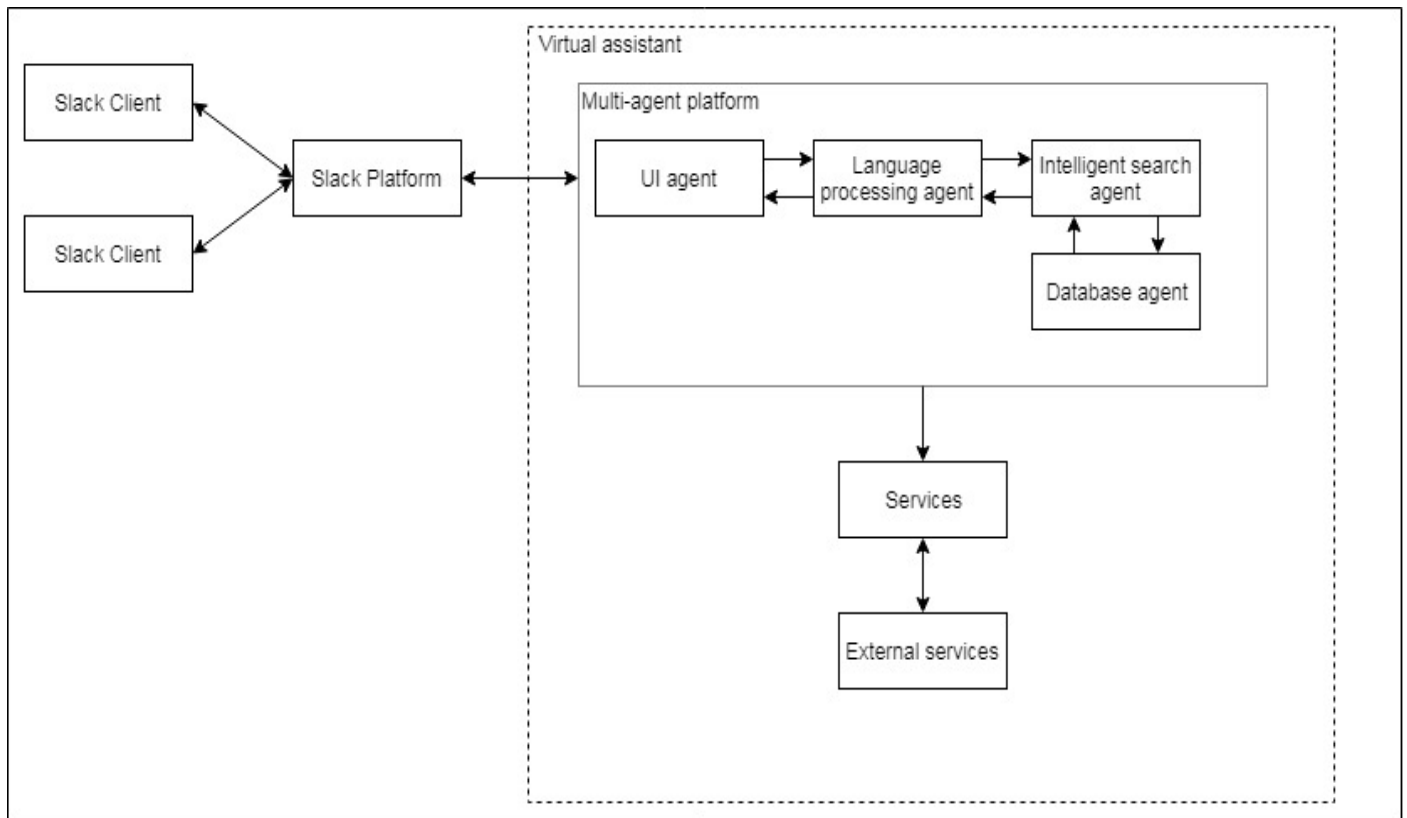


Fig 3: Block diagram

In the block diagram, client interacts with the system via the interface platform. The main system consists of multiple agents: UI, processing, intelligence and database. These agents communicate with each other and fetch data from internal or external services and provide it to the client or use it to perform task specified by the client.

4.2 Modular Diagram of the system

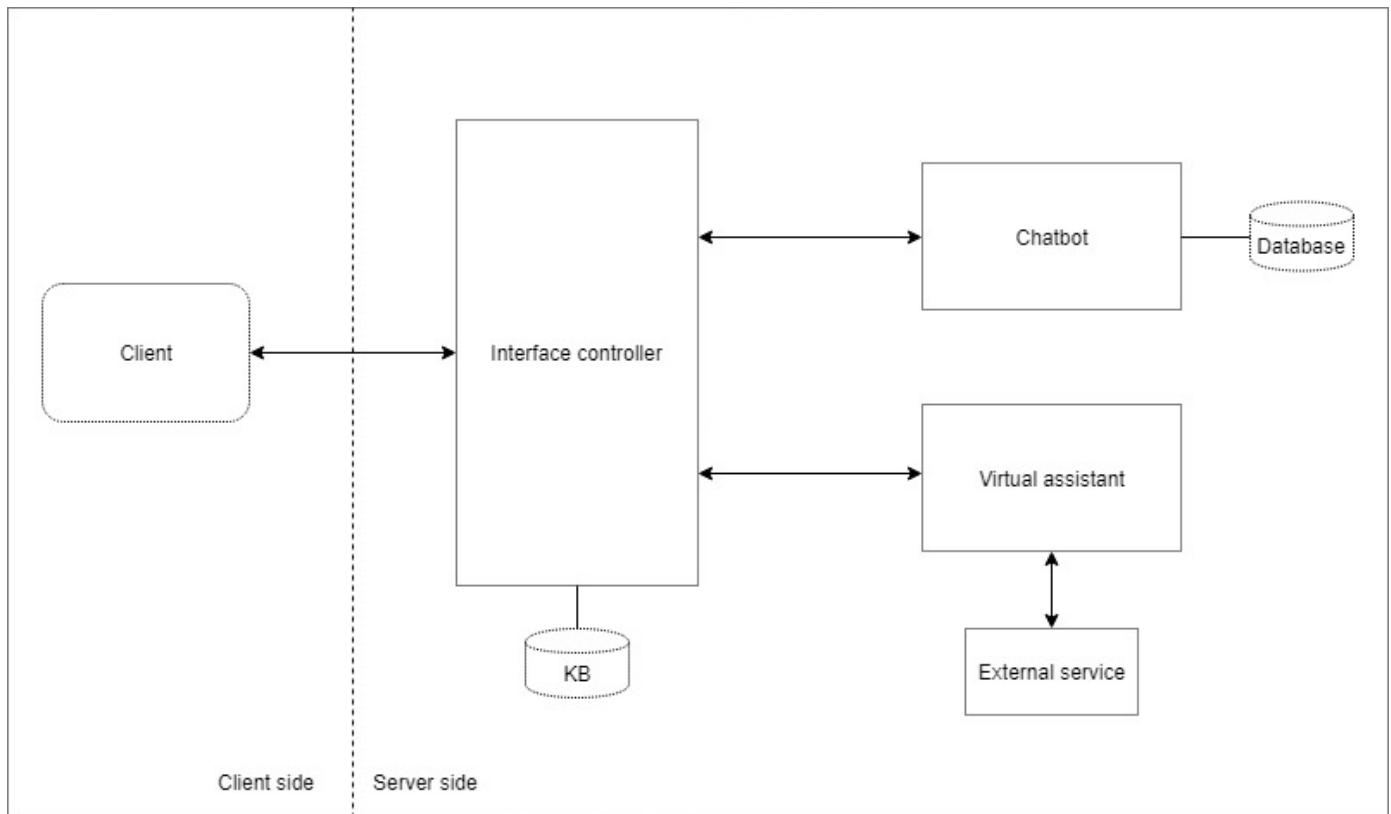


Fig 4: Modular diagram

The modular diagram represents our system as 2 different parts - Virtual assistant and Chatbot. The system has an interface which is the link between the client and the core of the system. The knowledge base stores the information that is available to draw on. The chatbot uses the database provided to it for training and learning purposes. Virtual assistant performs tasks assigned to it and uses external services for some specific tasks.

4.3 Detailed Design

- **DFD level 0**

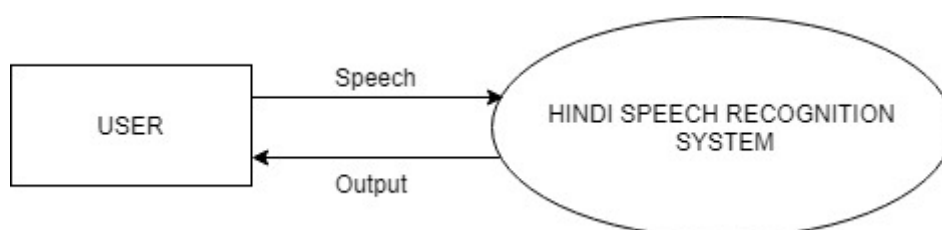


Fig 5: DFD level 0

- **DFD level 1**

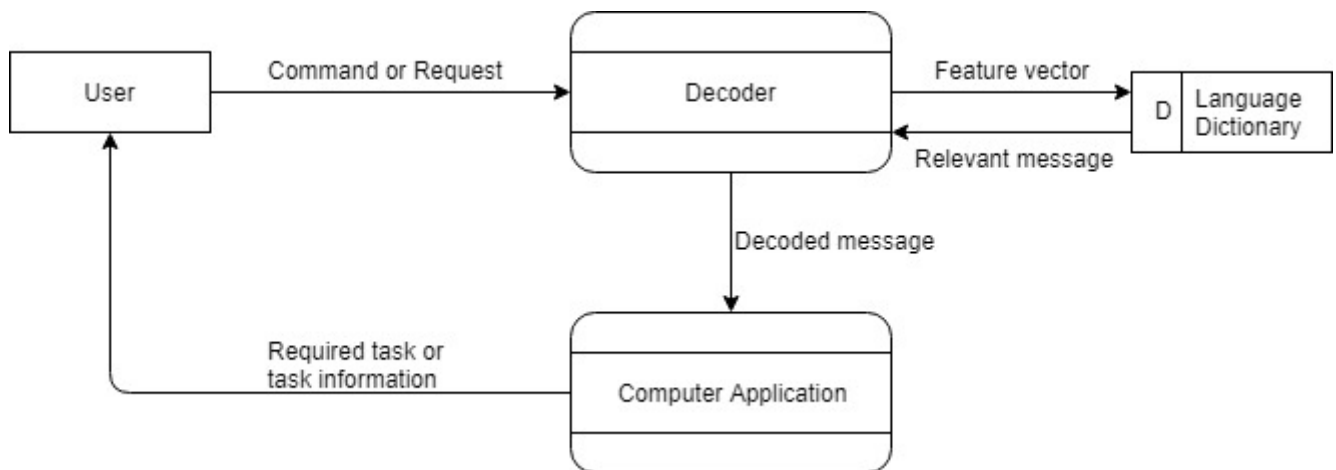


Fig 6: DFD level 1

- **DFD level 2**

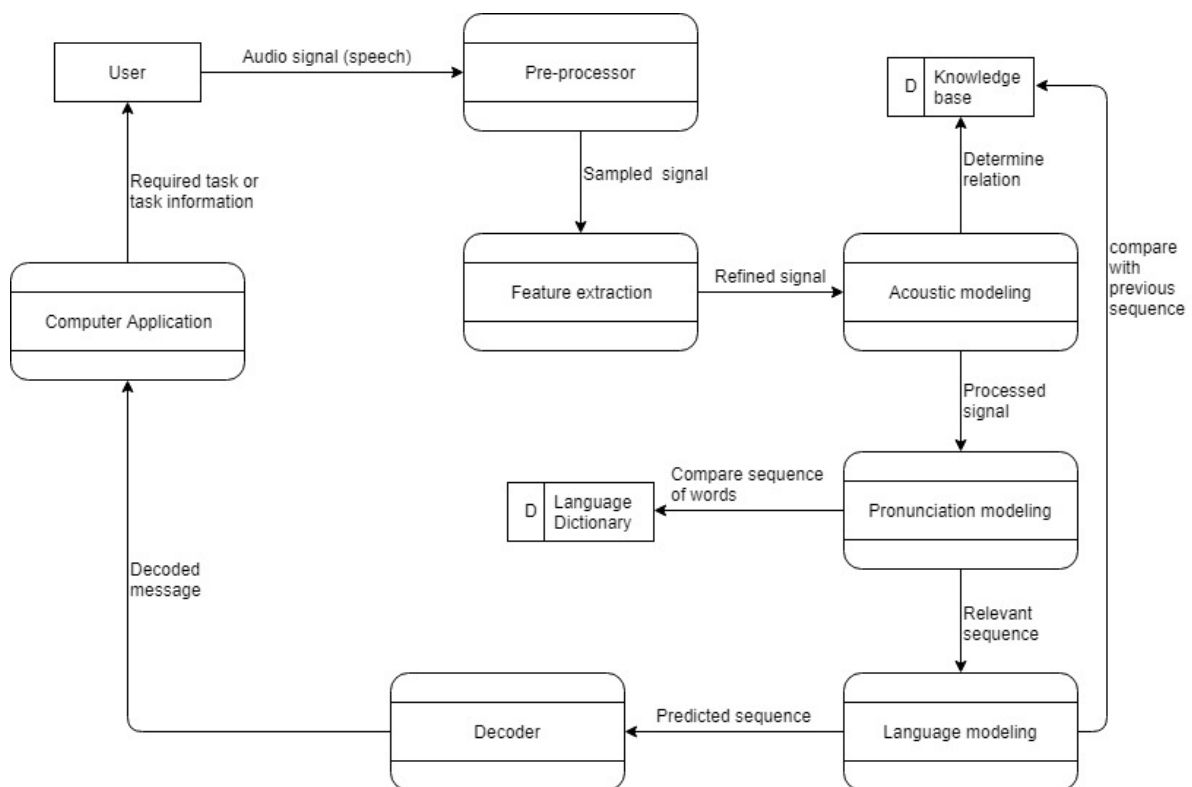


Fig 7: DFD level 2

- **Flowchart for the proposed system**

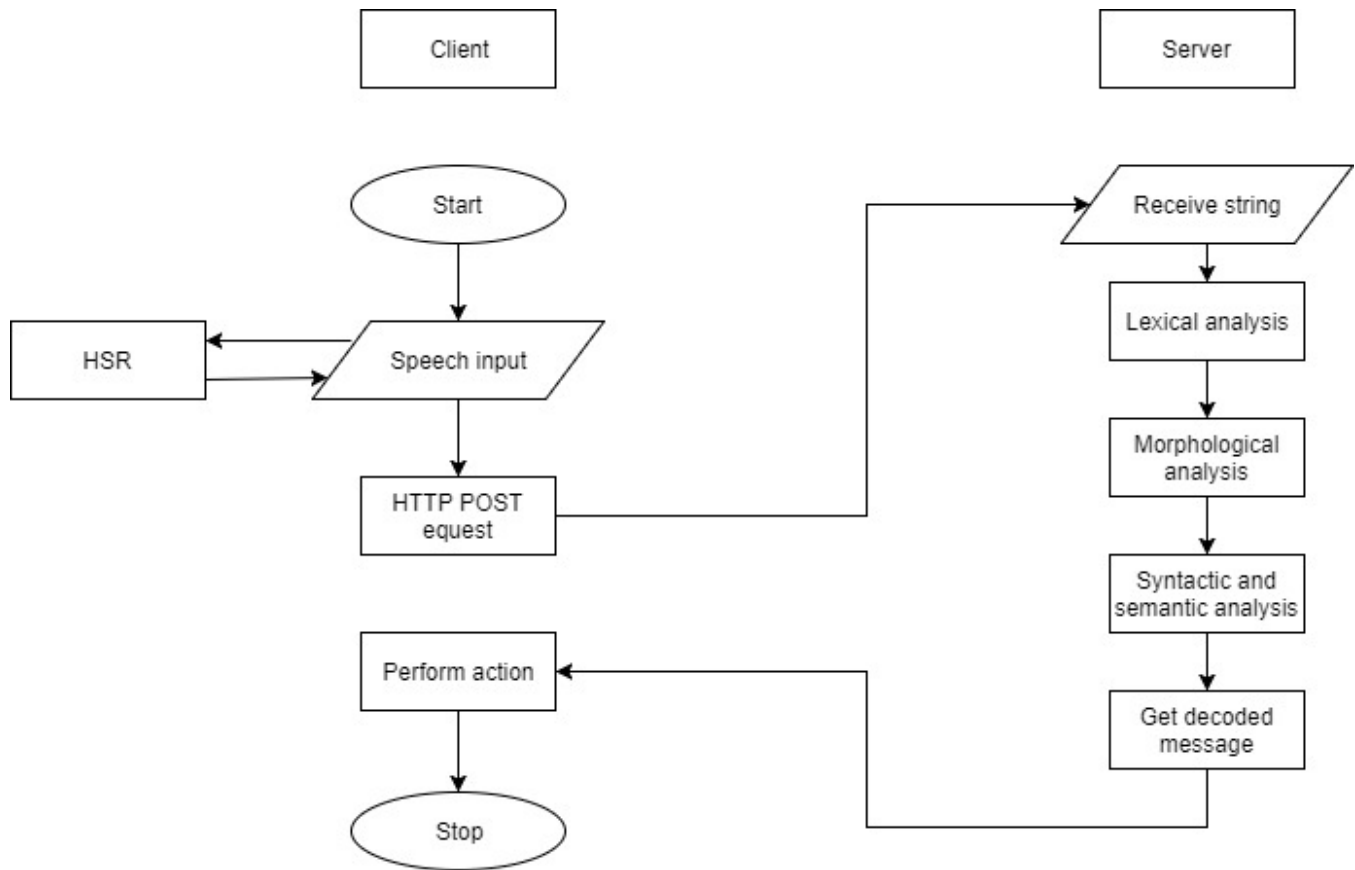


Fig 8: Flowchart

4.4 Project Scheduling and tracking

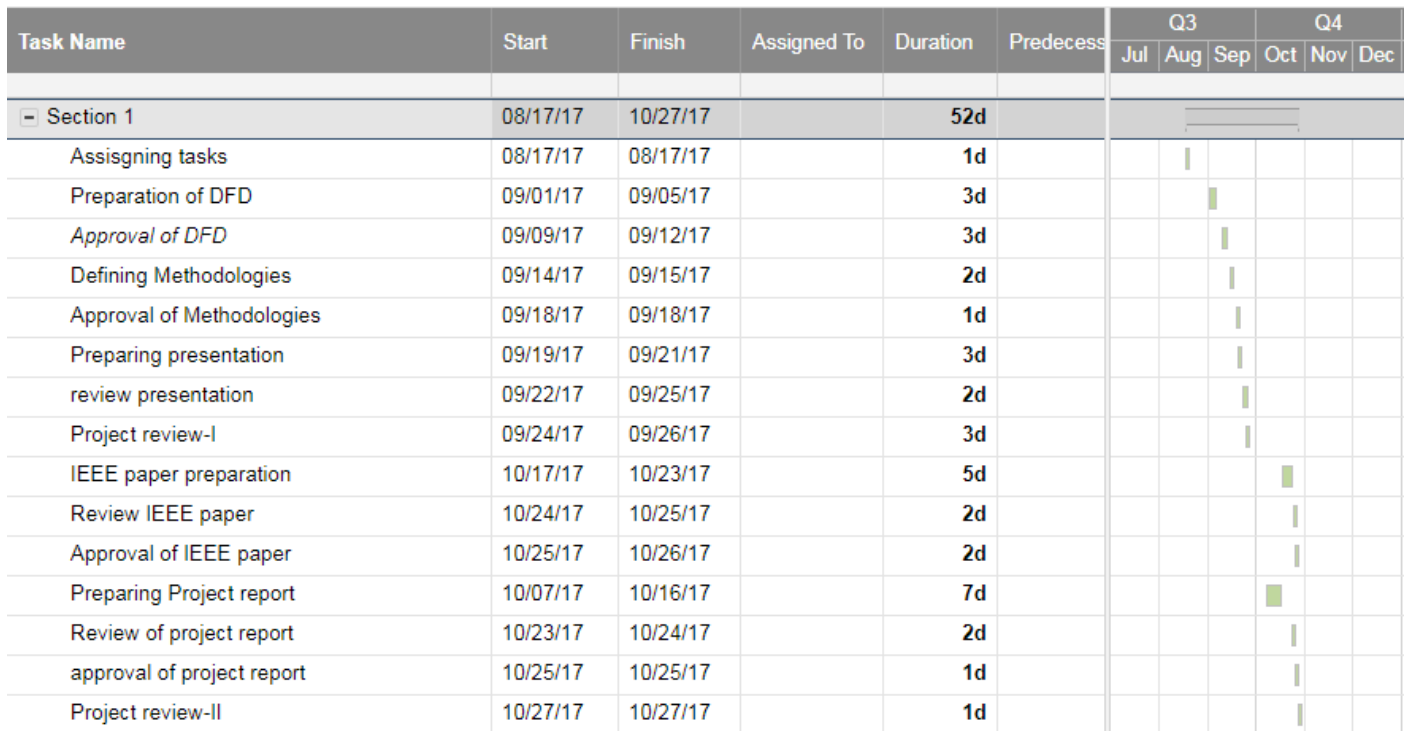


Fig 9: Gantt Chart (1)

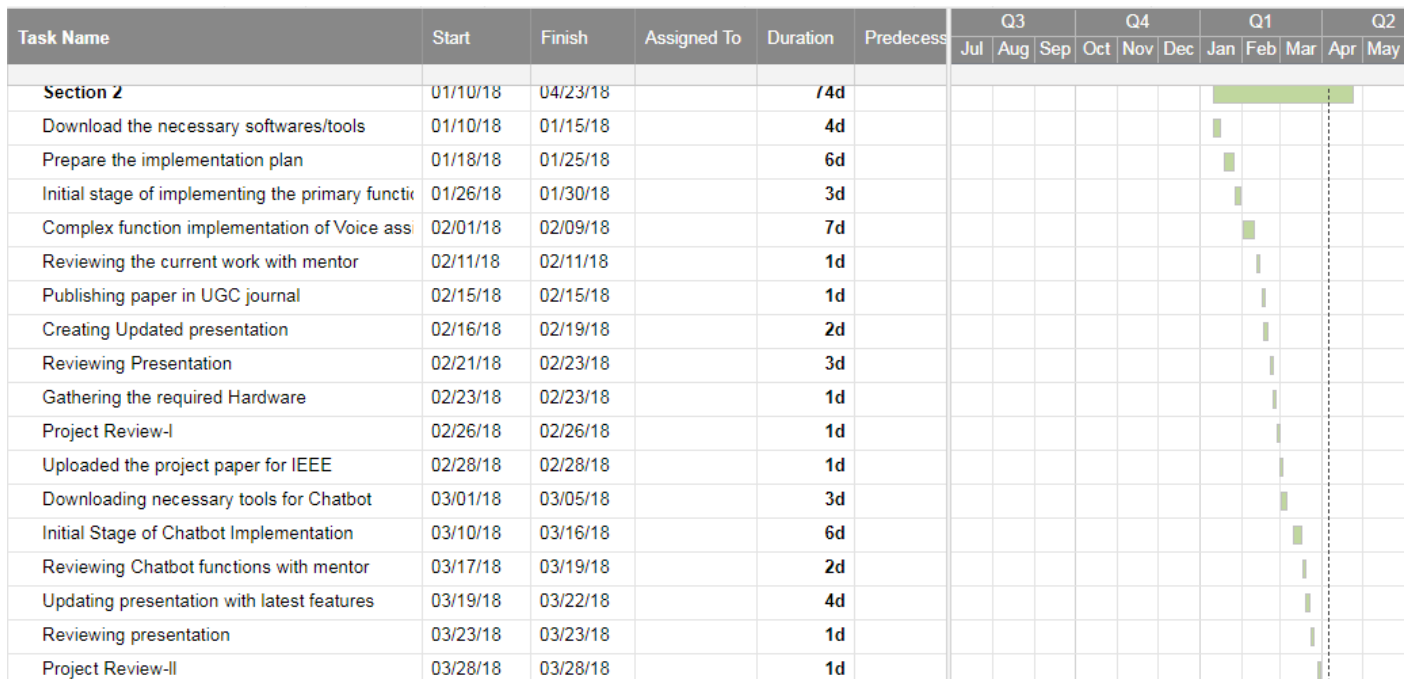


Fig 10: Gantt Chart (2)

5. Implementation Details

5.1 Algorithms for the respective modules developed

PLP (Perceptual Linear Prediction)

- This technique uses three concepts from the psychophysics of hearing to derive an estimate of the auditory spectrum: (1) the critical-band spectral resolution, (2) the equal-loudness curve, and (3) the intensity-loudness power law.
- In comparison with conventional linear predictive (LP) analysis, PLP analysis is more consistent with human hearing. PLP analysis is computationally efficient and yields a low-dimensional representation of speech.
- These properties are found to be useful in speaker-independent automatic-speech recognition.

Viterbi algorithm

- The Viterbi Algorithm determines the most probable sequence of hidden states very efficiently. For instance in speech recognition, it is used in the language model to find the most probable word out of a phonemes sequence.

Deep Neural Networks

- A deep neural network (DNN) is an artificial neural network (ANN) with multiple hidden layers between the input and output layers.
- DNNs can model complex non-linear relationships, in which data can flow in any direction, are used for applications such as language modeling.

Discriminative training

- Discriminative training is used to optimize the model parameters to minimize the recognition error rate on training data.
- In discriminative training, an objective function with respect to the model parameters is used to express the recognition error.
- That is, the larger the value of the objective function is, the smaller the recognition error is.

Naive-Bayes classification

- In machine learning, naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

- This is a classic algorithm for text classification and natural language processing (NLP). It works is relatively simple, common and surprisingly effective. This classifier is “naive” because it assumes independence between “features”.
- This algorithm proves to be fast, reliable and accurate along with providing simplicity.

5.2 Comparative analysis with the existing algorithms

- The Microsoft API uses context-dependent deep neural network hidden Markov model (CD-DNN-HMM). The Speech API and the CD-DNN-HMM models, they determined that the large-vocabulary speech recognition achieves better results.
- Our system makes use of the technology that is deep learning neural network. Our system has low word error rate as compared to other systems.
- We have gathered data as follows: The Sphinx-4 (37% WER), Our system (9% WER) and Microsoft Speech API (18% WER). (WER - Word Error Rate)
- The acoustic modeling and language model of our system is superior as compared to CMU sphinx and Microsoft API.

5.3 Evaluation of the developed system (Accuracy, Effectiveness & Efficiency)

The application and all its modules were evaluated by calculating the Word Error Rate(WER). Word error rate (WER) is a common metric of the performance of a speech recognition or machine translation system.

Word error rate can then be computed as:

$$WER=(I+D+S)/N$$

where,

S is the number of substitutions,

D is the number of deletions,

I is the number of insertions,

C is the number of the corrects,

N is the number of words in the reference ($N=S+D+C$)

The intuition behind 'deletion' and 'insertion' is how to get from the reference to the hypothesis. So, if we have the reference "This is Sparta" and hypothesis "This _ Sparta", we call it a deletion.

The result obtained was as follows-

File	S	N	I	S	D	CW	EW	WA	WER
TSX223	1	8	0	0	0	9	0	1.0	0.0
TSX293	1	11	0	1	1	9	2	0.82	0.18
TSI1894	1	9	0	0	0	9	0	1.0	0.0
TSI1400	1	14	0	1	0	13	1	0.93	0.07
TSX188	2	6	0	0	0	6	0	1.0	0.0
TSI1628	2	12	0	2	0	10	2	0.83	0.17
TSX314	2	12	0	0	0	12	0	1.0	0.0
DIG001	3	15	0	0	0	15	0	1.0	0.0
TSX216	1	9	0	0	0	9	0	1.0	0.0
TSX209	1	7	0	0	0	7	0	1.0	0.0
TSI1584	2	13	0	5	2	6	7	0.46	0.54
TSX371	1	11	0	0	0	11	0	1.0	0.0
TSI1373	1	14	0	0	0	14	0	1.0	0.0
TSX233	1	7	1	0	0	6	1	0.71	0.14
OGE003	1	8	0	2	1	5	3	0.63	0.38
AENG08	1	9	0	0	0	9	0	1.0	0.0
AENG08	1	9	0	2	0	7	2	0.78	0.22
AENG07	1	6	0	0	0	6	0	1.0	0.0
AENG02	1	7	0	0	0	7	0	1.0	0.0
Mean									0.09

Fig 11: WER testing

According to this testing the WER was calculated as 9%. Therefore, it can be concluded that the system accuracy is high.

6. Testing

6.1 Unit testing

Test case no.	Test case	Description	Input	Expected Output	Actual Output	Pass or fail
1.	Speech to text	Understanding the speech and converting it to text	Speech	Spoken words as text on screen	Spoken words as text on screen	Pass
2.	Opening applications with speech	Open different types of applications such as calculator, PDF, Word, etc. when said	Speech command to open an application	The application runs on listening the command	The application runs on listening the command	Pass
3.	Classify sentence according to groups	Input sentence should be analyzed and placed in suitable cluster	Sentence (in speech or text)	Sentence classified into respective group according to score	Sentence classified into respective group according to score	Pass

Table 1: Unit testing

6.2 Integration Testing

The speech to text conversion modules takes speech as input and converts it into text for further processing. The translation module translates Hindi text to English and finally the required task is performed according to the instruction given by the user. All the modules seem to be working correctly.

6.3 User acceptance testing

The application was tested out by several users from students to teachers in order to understand the user acceptance of the application. All the users were easily able to use the application. The application took proper input even with varying user accents. The users found the application to be very useful and convenient to use.

7. Result analysis

7.1 Simulation Model

A Virtual Assistant (VA) computer application is developed in order to help users who cannot speak or understand English. The user can speak in Hindi as well as English. The speech command is taken as input by the application. The user can request the virtual assistant to do a task and it will be carried out by the VA. The user can ask the VA to send a mail, open a calculator, open a word, find the meaning of a word, etc. In case the command is not recognized by the system, it redirects the user to an article related to that command. This application also includes a chatbot which will further help the user in solving various queries by talking to a trained bot.

7.2 Parameters Considered

7.2.1 Noise interference

While giving a command the noise around the device plays a major part in determining the output. This means that less noise would increase the accuracy of the application. A lot of noise would result in the system taking in wrong input which would cause incorrect results. For increasing the accuracy even further we can use external microphones.

7.2.3 Word Error Rate

Word error rate (WER) is a common metric of the performance of a speech recognition or machine translation system. WER should be as low as possible. It depends on the no. of words substituted, deleted and added.

7.3 Screenshots of UI

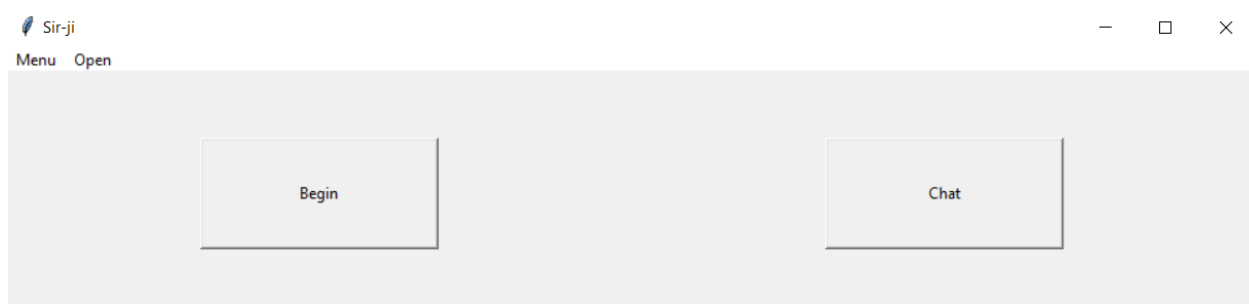


Fig 12: User Interface

How may I help you?

पानीपत की लड़ाई

पानीपत का तृतीय युद्ध अहमद शाह अब्दाली और मराठों के बीच हुआ। पानीपत की तीसरी लड़ाई (95.5 किमी) उत्तर में मराठा साम्राज्य और अफगानिस्तान के अहमद शाह अब्दाली, दो भारतीय मुस्लिम राजा Rohilla अफगान दो आब और अवध के नवाब Shuja-उद-Daula (दिल्ली के सहयोगी दलों) के एक गठबंधन के साथ अहमद शाह अब्दाली के एक उत्तरी अभियान बल के बीच पर 14 जनवरी 1761, पानीपत, के बारे में 60 मील की दूरी पर हुआ। लड़ाई 018 वीं सदी में सबसे बड़े, लड़ाई में से एक माना जाता है और एक ही दिन में एक क्लासिक गठन दो सेनाओं के बीच लड़ाई की रिपोर्ट में मौत की शायद सबसे बड़ी संख्या है।

मुगल राज का अंत (१६८०-१७७०) में शुरू हो गया था, जब मुगलों के ज्यादातर भू भागों पर मराठाओं का अधिपत्य हो गया था। गुजरात और मालवा के बाद बाजी राव ने १७३७ में दिल्ली पर मुगलों को हराकर अपने अधीन कर लिया था और दक्षिण दिल्ली के ज्यादातर भागों पर अपने मराठाओं का राज था। बाजी राव के पुत्र बाला जी बाजी राव ने बाद में पंजाब को भी जीतकर अपने अधीन करके मराठाओं की विजय पताका उत्तर भारत में फैला दी थी। पंजाब विजय ने १७५८ में अफगानिस्तान के दुरोनी शासकों से टकराव को अनावश्यक कर दिया था। १७५९ में दुरोनी शासक अहमद शाह अब्दाली ने कुछ पसतून कबीलों के सरदारों और भारत में अवध के नवाबों से मिलकर गंगा के दोआब क्षेत्र में मराठाओं से युद्ध के लिए सेना एकत्रित की। इसमें रोहिलेआ अफगान ने भी उसकी सहायता की। पानीपत का तीसरा युद्ध इस तरह सम्मिलित इस्लामिक सेना और मराठाओं के बीच लड़ा गया। अवध के नवाब ने इसे इस्लामिक सेना का नाम दिया और बाकी मुसलमानों को भी इस्लाम के नाम पर इकट्ठा किया। जबकि मराठा सेना ने अन्य हिन्दू राजाओं से सहायता की उम्मीद की।

थी (राजपूतों और जाटों) जो कि उन्हें न मिल सकी। इस युद्ध में इस्लामिक सेना में ६००००- 100000 सैनिक और मराठाओं के ओर से ४५०००-६०००० सैनिकों ने भाग लिया।

१४ जनवरी १७६१ को हुए इस युद्ध में भूखे ही युद्ध में पहुँचे मराठाओं को सुरवती विजय के बाद हार का मुख देखना पड़ा . इस युद्ध में दोनों पक्षों की हानियों के बारे में इतिहासकारों में भारी मतभेद है। फिर भी ये माना जाता है कि इस युद्ध में १२०००० लोगों ने सक्रिय रूप से हिस्सा लिया था जिसमें अहमद शाह अब्दाली विजय हुई थी और मराठाओं को भारी हानि उठानी पड़ी.

Fig 13.1: Search

How may I help you?

ईमेल भेजना है

किसको भेजना चाहते हैं: shreesh.rao@ves.ac.in

प्रोजेक्ट पर काम कर रहे हैं

sirji.beproject@gmail.com

(no subject) - प्रोजेक्ट पर काम कर रहे हैं

Fig 13.2: Email

How may I help you?
कैलकुलेटर खोलो

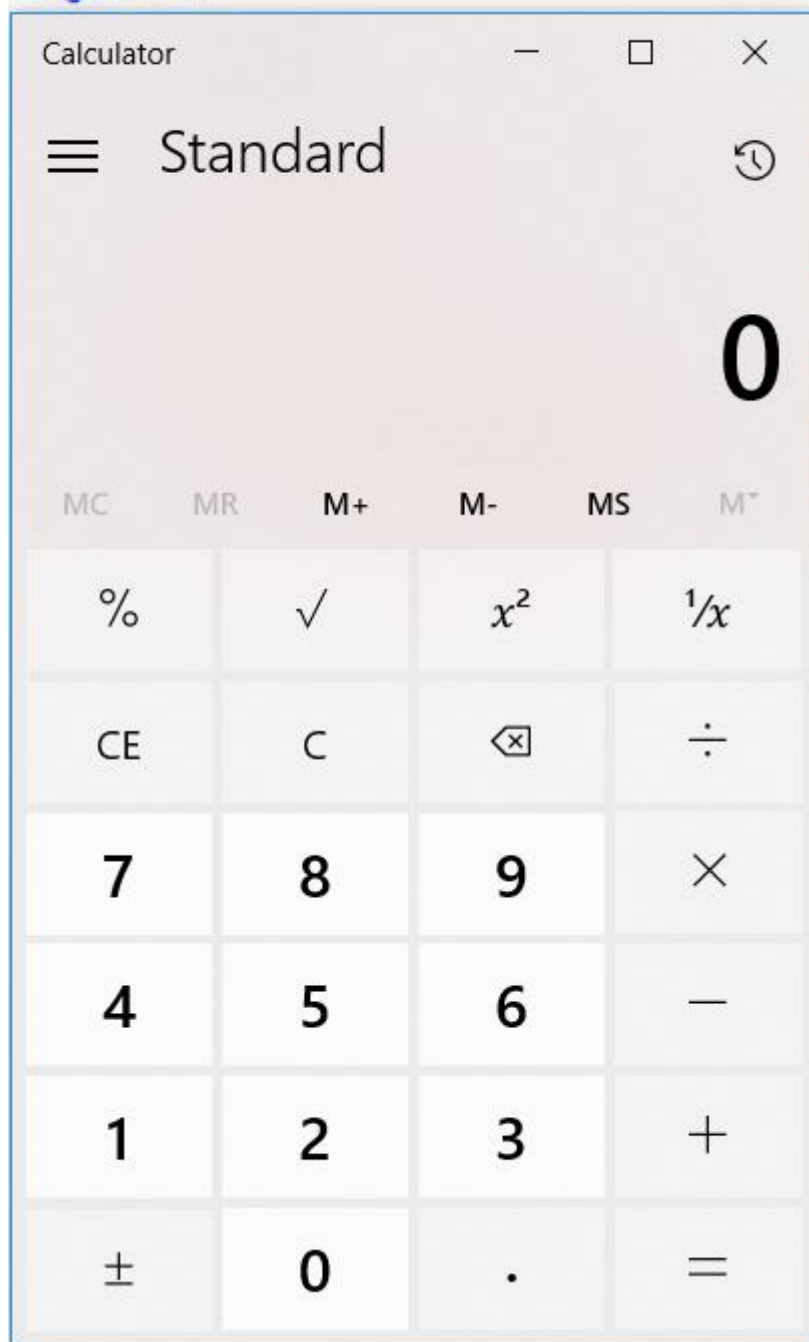


Fig 13.3: Opening application (1)

How may I help you?
माइक्रोसॉफ्ट वर्ड खोला

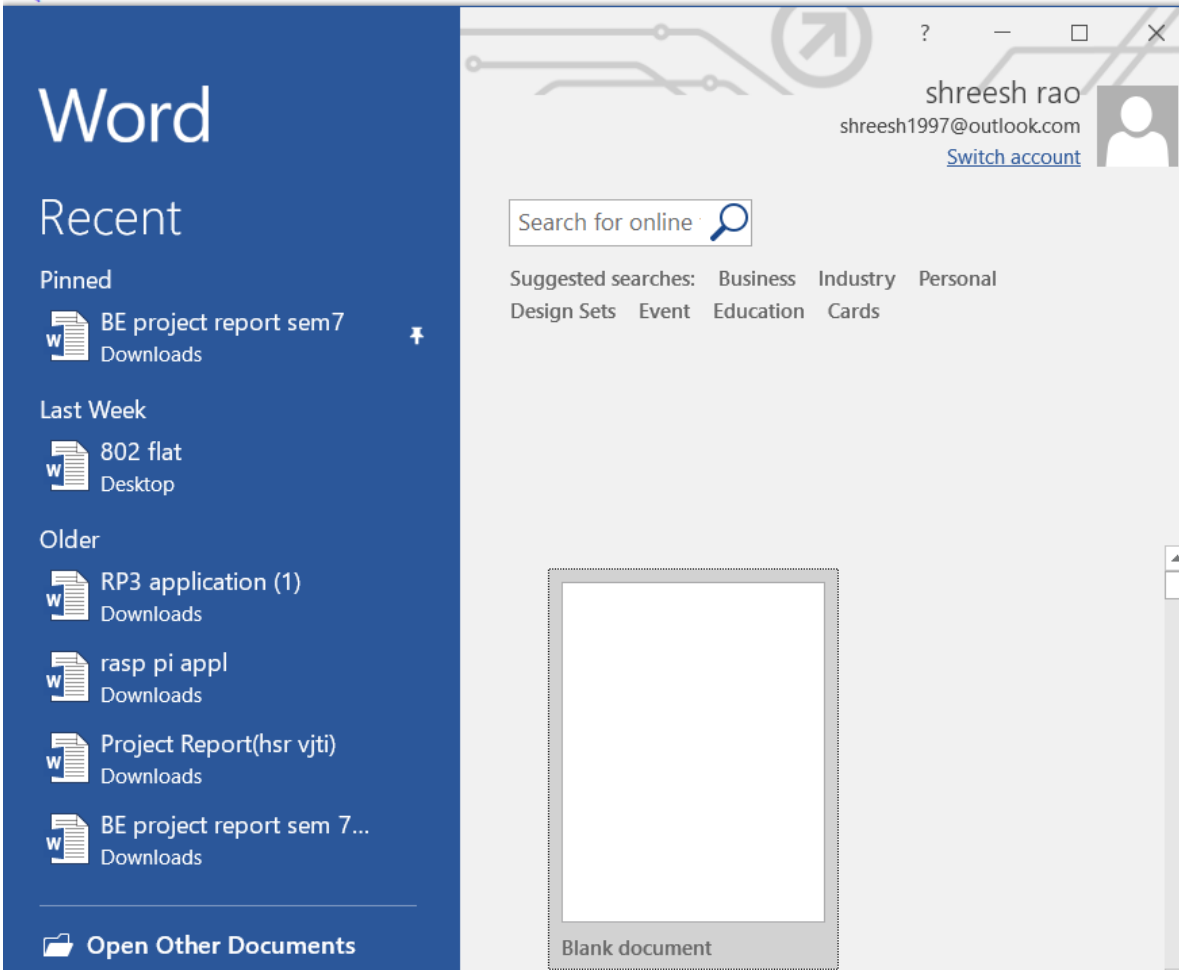


Fig 13.4: Opening application (2)

```
How may I help you?
father's mother's sister's son
first cousin once removed | (no other relation possible)
How may I help you?
now
7:48:49 am IST | Saturday, April 21, 2018
How may I help you?
X square + 2 X + 1
X = -1

How may I help you?
motor
1 | noun | machine that converts other forms of energy into mechanical energy and so imparts motion
2 | noun | a nonspecific agent that imparts motion
3 | adjective | conveying information to the muscles from the CNS
4 | adjective | causing or able to cause motion
5 | verb | travel or be transported in a vehicle
(5 meanings)

How may I help you?
पिता के बेटे का बेटा
पोता
```

Fig 13.5: Performance of Task (1)

How may I help you?
गाने के लिरेक्स दो
गायक का नाम बताएं
Imagine Dragons
गाने का नाम बताएं
Thunder

Just a young gun with a quick fuse
I was uptight, wanna let loose
I was dreaming of bigger things
And wanna leave my old life behind
Not a yes-sir, not a follower
Fit the box, fit the mold
Have a seat in the foyer, take a number
I was lightning before the thunder

Thunder, thunder
Thunder, thun-, thunder
Thun-thun-thunder, thunder, thunder
Thunder, thun-, thunder
Thun-thun-thunder, thunder

Thunder, feel the thunder
Lightning then the thunder
Thunder, feel the thunder
Lightning then the thunder
Thunder, thunder
Thunder

Kids were laughing in my classes
While I was scheming for the masses

Fig 13.6: Performance of Task (2)

Start conversation:
Done!
You: I am hungry
Sir-ji: Burgers are my favourite
Done!
You: what would you like to eat
Sir-ji: I would love to eat something
Done!
You: nice to meet you
Sir-ji: Bye
Done!
You: good morning
Sir-ji: Hello
Start conversation:
Done!
आप: क्या आपको भूख लगी है
सर-जी: मुझे कुछ खाना पसंद आएगा

Fig 13.7: Conversation with Chatbot (1)

Start conversation:
Done!
You: hello how are you
Sir-ji: have a good day
Done!
You: laptop mosquito
Sir-ji: I did not get you
Start conversation:
Done!
आप: मुझे गुस्सा आ रहा है
सर-जी: हमम ... भावनाओं को समझना मुश्किल है
Start conversation:
Done!
You: hey good morning
Sir-ji: Good day pal
Done!
You: what do you think about football
Sir-ji: sports always gets me pumped up
Start conversation:
Done!
You: how are you feeling
Sir-ji: Always stay happy no matter what

Fig 13.8: Conversation with Chatbot (2)

8. Conclusion

8.1 Limitations

- ‘SIR-JI’ currently works as a voice assistant cum chatbot who can take inputs in the form of speech in Hindi & English language.
- So as of now our system supports only two languages. People who are not fluent in these languages will have difficulty in using our system.
- The system is programmed in Windows as a result it cannot be implemented in android smart phones and apple phones.
- For accurate results, there should be minimum noise in the background.

8.2 Conclusions

- Virtual Personal Assistants (VPA) chatbots are the most effective way to obtain information and do tasks in a simpler way.
- The limitation which is encountered while using VPAs is the fluency in English language. However, in our project, we are developing VPA which can understand and process human commands given in Hindi language.
- The complexity and the accent problems of various chatbots is resolved in our project thereby making even illiterate audience use the system with ease.
- Since Hindi is the most spoken language in India, we think most of the Indian population can make efficient use of the VPA. Also, with the help of chatbot they can resolve their doubts easily.
- VPAs are also more reliable than Human Assistant because, VPAs are easily portable and one can use them anytime. Our project has a simple and easy to understand interface.

8.3 Future Scope

- In future, we are planning to increase the versatility of ‘SIR-JI’ by adding more input languages so that more amount of people from different parts of the world can access our system.
- Also, currently due to some constraints ‘SIR-JI’ only functions on Windows, so in the upcoming years we’ll try to make it platform independent so that android and apple users can also access our system to make their daily life simpler.

References

- A REVIEW ON SPEECH TO TEXT CONVERSION METHODS, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 7, July 2015.

<http://ijarcet.org/wp-content/uploads/IJARCET-VOL-4-ISSUE-7-3067-3072.pdf>

- SPEECH RECOGNITION USING VECTOR QUANTIZATION THROUGH MODIFIED K-meansLBG ALGORITHM, ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online) Vol 3, No.7, October 2012

<https://pdfs.semanticscholar.org/907b/2aa3add619b0d357a9295325e3e783d91a31.pdf>

- EFFICIENT SPEECH RECOGNITION SYSTEM, An International Journal (CSEIJ), Suma Swamy 1 and K.V Ramakrishnan, August 2013

<https://pdfs.semanticscholar.org/a361/2073b4a6666817ad0fd74849198c9d275ef2.pdf>

- A REVIEW ON SPEECH RECOGNITION TECHNIQUE, International journal of computer applications Vol 10-No 3, November 2010.

https://www.researchgate.net/publication/49587837_A_Review_on_Speech_Recognition_Technique

- A SURVEY ON CHATBOT DESIGN TECHNIQUES IN SPEECH CONVERSATION SYSTEMS, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 6, No. 7, 2015

https://thesai.org/Downloads/Volume6No7/Paper_12-

[Survey on Chatbot Design Techniques in Speech Conversation Systems.pdf](#)

- A STUDY OF TODAY'S A.I. THROUGH CHATBOTS AND REDISCOVERY OF MACHINE LEARNING, International Journal of u- and e- Service, Science and Technology, Vol.8, No. 7 (2015)

http://www.sersc.org/journals/IJUNESST/vol8_no7/28.pdf

Project progress review sheet 1

GROUP NO.: 26

Project Evaluation Sheet 2017 - 18

Title of Project: Virtual Assistant Using Speech Recognition in Hindi
Group Members: Ronit Dhonde, Nandali Birla, Paashant Dombale, Shreshth Rao

Engineering Concepts & Knowledge	Interpretation of Problem & Analysis	Design / Prototype	Interpretation of Data & Dataset	Modern Tool Usage	Societal Benefit, Safety Consideration	Environment Friendly	Ethics	Team work	Present ation Skills	Applied Engg & Mgmt principles	Life - long learning	Professional Skills	Innovative Approach	Total Marks
(5)	(5)	(5)	(3)	(5)	(2)	(2)	(2)	(2)	(3)	(3)	(3)	(5)	(5)	(50)
4	4	2	2	3	2	1	1	2	3	1	2	4	3	34

Review of Project Stage 1

Comments:

• Need to test atleast one application by own algorithm (Not by Google API)

• Not shown chatbot integration

Vidya S. Wase
Name & Signature Reviewer1

Engineering Concepts & Knowledge	Interpretation of Problem & Analysis	Design / Prototype	Interpretation of Data & Dataset	Modern Tool Usage	Societal Benefit, Safety Consideration	Environment Friendly	Ethics	Team work	Present ation Skills	Applied Engg & Mgmt principles	Life - long learning	Professional Skills	Innovative Approach	Total Marks
(5)	(5)	(5)	(3)	(5)	(2)	(2)	(2)	(2)	(3)	(3)	(3)	(5)	(5)	(50)
3	4	4	3	2	2	1	1	1	2	1	1	3	4	32

Review of Project Stage 1

Comments:

Need to test atleast one application by own algorithm & Compare with existing API

Date: 26/02/2018

Shubha DC
Name & Signature Reviewer2

Project progress review sheet 2

Inhouse/ Industry:

Project Evaluation Sheet 2017 - 18

Class: D17 A/B/C
Group No.: 26

Title of Project: Virtual assistant using speech recognition in hindi.

Group Members: Mrudali Birla (D17A), Ronit Dhorde (D17A), Prashant Dombale (D17B)

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation on Skills (3)	Applied Engg & Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	05	05	03	04	02	02	02	02	03	03	02	04	04	45
Comments:	Chatbot done ... translation in Hindi required.														

Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation on Skills (3)	Applied Engg & Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	4	3	3	3	02	02	02	02	3	02	02	3	2	43
Comments:	Chatbot done ... Conversation translation in Hindi Required.														

Date: 15th March, 2018

Name & Signature Reviewer2

Appendix

Sr. no.	Content
1	Paper 1 (Published at IJARSE, March 2018)
2	Plagiarism report (1.1)
3	Plagiarism report (1.2)
4	Plagiarism report (1.3)
5	Certificates
6	Paper 2
7	Plagiarism report (2.1)
8	Plagiarism report (2.2)

Hindi Speech Recognition

Ms. Kajal Jewani Asst Prof.¹, Shreesh Rao², Prashant Dombale³,
Ronit Dhonde⁴, Mrudali Birla⁵

^{1,2,3,4,5}Department of computer science, V.E.S.I.T., Mumbai University
Sindhi Society, Chembur, Maharashtra, (India)

ABSTRACT

Speech is one of the most common and widely accepted mode of communication between humans. It has been working efficiently well and a number of attempts have been made to use it for human computer interaction. Most of these system work with English language. This paper has tried to describe plausible techniques to recognize speech in Hindi. It has limned some of the previously existing speech recognition systems briefly along with the techniques used and has compared them in terms of flexibility, accuracy and celerity. Another purpose of this paper is to open the door for research in speech recognition using other vernacular languages as well.

Keywords— *Speech recognition, Hidden Markov Model, Vector quantization, MFCC.*

I.INTRODUCTION

Speech is the fundamental method of modern day communication. Various ideas and thoughts in the minds of different individuals are shared with the help of different languages. A person's speech may comprise of various words, phrases and sentences. In India, since Hindi is the primary language, people exchange their opinions and thoughts by making use of Hindi language.

Hindi language belongs to the Devanagiri script and hence there is a one-to-one correspondence between the sound and the syllable that represents that sound.

Speech recognition is the process of analyzing the input speech from the user, followed by identifying the words and phrases present within the input and then converting them into a machine-readable format. Comparison between the input speech is done with the in-built vocabulary present within the system. The task is to make the computer understand the Hindi speech and then react appropriately by converting the speech into another medium. A Microphone, computer and a good quality sound card are some of the rudimentary apparatus for a speech recognition system.

II.SYSTEM CLASSIFICATION

Speaker dependent models are developed for individual users. A particular user can operate this model with his voice more accurately than other users. It is accurate but not as flexible as speaker independent system. So, if any other user tries to give input with his voice, the system will not respond in the same way. Hence, as this model is for an individual speaker, it is not adaptive to any other user. But, this type of models is more accurate

as there is no variation in the voice and overall pattern remains the same. Speaker independent model can take voice of variety of users. Also, they are very adaptive and flexible as different users can operate on it. But, as it takes voice of different users the overall accuracy reduces due to the changing pattern in voices. They are more expensive as compared to speaker dependent models and also more difficult to develop as they are complex.

There are different types of speech which can be classified in different ways. Isolated word takes single utterances which are not connected to each other. It can analyze individual words at a time. In Connected word, there is combination of two or more isolated words and allows the utterances which are different to run with a slight pause. When the speaker speaks continuously without any gap, it is known as continuous speech. A special method is needed in order to capture such a kind of speech. Lastly, there is also a kind of speech which is not practised or rehearsed, called as spontaneous speech. It allows variety of words and speeches to be taken.

There are different types of word matching techniques in our system.

Whole word matching: In this type of matching technique, we take the entire sentence and match it with a prerecorded template of words. It requires a lot of memory and takes less processing as compared to sub word matching technique. The template is stored in form of recognition vocabulary in the system and from there words are recognized.

Sub-word matching: In this type of matching technique, sub-words are taken and then pattern is recognized on basis of these words. It usually takes more processing as compared to whole word matching

III.LITERATURE SURVEY

A recognition system basically follows an algorithm which tries to match the input syllable with the corresponding alphabet present in the dictionary that has been already present in the computer database. A variety of methods can also be used to perform the analysis procedure, but among all of them Mel-Frequency Cepstrum (MFC) turns out to be the most efficient and effective method as it comprises of a strong property where it can compute power spectrum by performing Fourier Analysis. As far as the modelling techniques are concerned the Hidden Markov Model (HMM) proves out to be the most feasible option, where the system being modelled is assumed to be a Markov process with unobserved states.

A speech recognition system is always followed by a speaker recognition system, as it increases the accuracy and efficiency of the software. As discussed earlier, the MFC can extract the characteristics from the input speech signal with respect to a particular word uttered by a particular speaker. Identification of a word with the help of HMM on Quantized feature vectors is done in order to maximize the log likelihood values for the spoken word. Formulation of the techniques mentioned above can be executed by the coding the techniques in MATLAB. For best performance and accurate results, MFC and distance minimum algorithm can be combined. According to the experiments performed with this combination the overall efficiency level has turned out to be of 95 percent.

There are a few systems available in the market for speech recognition in Hindi. Shrutlekhan-Rajbhasha is one of the Hindi speech recognition software application available in the market. It is developed by C-DAC and IBM. One drawback of this software is that if mixed English–Hindi dictation is given, it can recognize Hindi words but it is not successful to recognize English words. Also, another variant of this software is available in

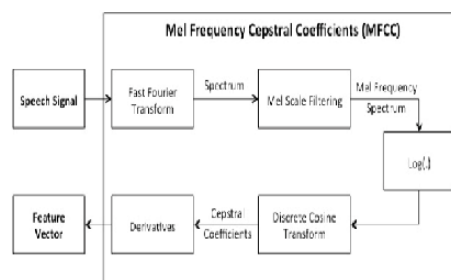
the market that is Vachantar-Rajbhasha, which takes English sound as input, converts it to English text and then translates it to Hindi. It uses MANTRA-Rajbhasha translation engine. The major task done by both “Shrutlekhan-Rajbhasha” and its variant “Vachantar-Rajbhasha” is conversion of Hindi to a particular type of script and translation from Hindi to English but it does not perform other tasks and its use is limited.

Some of the existing systems very popular in market like Microsoft’s cortana, Apple’s Siri, Google Now and Amazon Alexa. These can be operated using English language only. Few of them support languages like Hindi to some extent but they show very little accuracy and have various problems like accent of the user and its interpretation by the system. Even though these systems have great features and show high accuracy but are not very accessible to the people who do not know English. Our system mainly aims to cater needs of such population in India. Google has a Hindi Input app which lets Android users type in Hindi. It works like a keyboard extension. Once you install the app, you can type in Hindi. You can compose text messages, type emails and create word documents in Hindi. The basic testing of app has shown less degree of accuracy. Also, one should be able to write to access the app whereas in our system a person not able to write can also access it. Baidu known as "Google of China" is the country's biggest search engine and at 96 percent. Its voice recognition is better than most humans at identifying spoken words. The system understands both English and Mandarin. In China, voice commands are more popular as it takes up a lot of their time for typing in Mandarin. Similarly, our system will be useful in India to take voice commands using Hindi.

IV.METHODOLOGY USED

1. Feature Extraction: Feature extraction is the most fundamental step of the speech recognition system. Recording of various speech samples of each word from the vocabulary is done by different speakers. The samples so far collected are then transformed from analog to digital form with the help of sampling frequency of about 16Hz. Filtering of background noise is crucial so that the input speech signal can be processed correctly and thus producing error free outputs. Quantization is a process which can assist in the filtering step. The incoming sound is transformed into an internal representation by feature extraction and also reconstruction of the original signal is possible. MFCC, PLP, RAST, LPCC are some of the techniques which assist to extract features, however MFCC is widely used.

2. Mel Frequency Cepstral Coefficients (MFCC): The design and formulation of MFCC is done in accordance with the human auditory system and hence can be used in every state of speech recognition system or art speech. MFCC includes 5 basic steps which perform the feature extraction task. Framing, Windowing, DFTH, Mel filter bank algorithm and computing the inverse of DFT are the MFCC steps.



3. Hidden Markov Model(HMM): The model comprises of two stochastic inter-related processes similar to Markov Chain. The only difference is that, output symbol and the transitions are probabilistic. Every HMM state comprises of output symbols set known as output probabilities with finite number of states $Q = \{q_1, q_2, \dots, q_n\}$. One process is related to the transitions among the states which are controlled by a set of probabilities named as transition probabilities which are used to model the temporal variability of speech. Other process is concerned with the state output observations $O = \{o_1, o_2, \dots, o_n\}$ regulated by Gaussian mixture distributions $b_j(o_t)$ where $1 \leq j \leq N$, to simulate the spectral variability of speech. Another possibility that every sequence of states that has the same length as the symbol sequence is also evident. The word 'Hidden' in the model name signifies the denial of viewership of the sequence of states from the observer. The Markov nature of the HMM i.e. the probability of being in a state is dependent only on the previous state, admits use of the Viterbi algorithm to generate the given sequence symbols, without having to search all possible sequences. At a particular instance, one process is assumed to be in some state while the observation is formed by the other process representing the same current state. $a_{ij} = P[Q_{t+1}=j | Q_t=i]$ represents the underlying Markov chain that alters states in accordance to its transition from state i to state j .

4. Dynamic time warping(DTW): It is a master technique to find an optimal alignment between two given time dependent sequences under certain restricted conditions. Matching of the two sequences is performing in a non-linear fashion by warping them intuitively. Normally different speech patterns are compared with the help of DTW. In fields such as data mining and information retrieval, DTW has been successfully applied to automatically cope with time deformations and different speeds associated with time-dependent data.

The objective of DTW is to compare two (time-dependent) sequences $X = (x_1, x_2, \dots, x_N)$ of length $N \in \mathbb{N}$ and $Y = (y_1, y_2, \dots, y_M)$ of length $M \in \mathbb{N}$. These sequences may be discrete signals (time-series) or, more generally, feature sequences sampled at equidistant points in time. In the following, we fix a feature space denoted by F . Then $x_n, y_m \in F$ for $n \in [1: N]$ and $m \in [1: M]$. To compare two different features $x, y \in F$, one needs a local cost measure, sometimes also referred to as local distance measure, which is defined to be a function.

One of the most primitive approaches in an isolated speech recognition system was to compare the incoming speech with the prototypical version of the word present within the vocabulary, and considering the closest match for the next step. A number of issues arise on factors considering the form of templates and how are they compared to the incoming signals. In order to avoid these problems, feature vectors can be used. The problem with this approach is that if a constant window spacing is used, the lengths of the input and stored sequences is unlikely to be the same. Moreover, within a word, there will be variation in the length of individual phonemes: Cassim might be uttered with a long /A/ and short final /i/ or with a short /A/ and long /i/. The matching process needs to compensate for length differences and take account of the non-linear nature of the length differences within the words.

The Dynamic Time warping algorithm eliminates this issue by finding an optimal match between two sequences of feature vectors which allows compressed and stretched sections of the sequence. Vector quantization technique performs the process of modelling probability density functions by distribution of prototype vectors.

V.APPLICATIONS

1. Agricultural development: The rural population of India can search for better solutions for improving the growth of crops and can also get information about new and efficacious techniques for growing particular crops. They can easily search for information using voice recognition which makes it much easier and user-friendly for them.
2. Source of learning for the blind people: People who cannot see can use voice recognition feature for performing their tasks and for searching relevant information, all this in Hindi. It will be convenient for them to operate the device using voice and language will not be a barrier for Indians as Hindi language is known to most of the people in our country.
3. Travel: If a person who is comfortable with Hindi and not articulate with his diction in English is travelling to another country and wants to communicate with the aborigine of that place or read a board written in vernacular language, he can use our system and get his answer in Hindi.
4. Source of communication: People can communicate with each other using voice SMSs or calls as directed to the system. They can convert voice-to-text and even email the messages. It will be beneficial for people who cannot read or write to communicate vocally.
5. Education: This system can also be used in schools located in remote rural areas for educational purposes. Both teachers and students can benefit from this system.

VI.CONCLUSIONS

In this paper, development of an effective Speech to text conversion system was limned along with the techniques used for development. This system particularly focuses on Hindi language. The main techniques in our system are chosen after an exhaustive research and comparison with other available methods and algorithm. Taking into consideration every known technique that are used in currently existing systems, we came up with the most effective ones for our system which include HMM, MFCC and DTW. Speech recognition is one of the most integrating areas of machine learning. Abridging the language barrier and increasing the human-computer interaction for people with less or no knowledge of English language will be our main goal.

VII.ACKNOWLEDGEMENT

We are thankful to our college, Vivekanand Education Society's Institute of Technology, for considering our project and providing us help at every stage of the research. A special thanks to Ms. Kajal Jewani for mentoring and guiding us in the right direction. We would like to thank Dr.(Mrs.) Nupur Giri and our principal Dr. (Mrs.) J.M. Nair for giving this valuable opportunity. We are immensely grateful to everyone for their help with the research without which it would have been difficult to get our job done. Also, thanks to our families for the moral support and encouragement. It is a great pleasure to acknowledge the help we received from the Department of Computer Engineering.

REFERENCES

- [1] Preeti Saini, Parneet Kaur, Automatic Speech Recognition: A Review, in International Journal of Engineering Trends and Technology- Volume4Issue2- 2013
- [2] Suma Swamy, K.V. Ramakrishnan, An Efficient Speech Recognition System, in An International Journal (CSEIJ), Vol. 3, No. 4, August 2013.

PLAGIARISM REPORT

Plagiarism Scan Report	
Summary	
Report Genrated Date	27 Oct, 2017
Plagiarism Status	100% Unique
Total Words	986
Total Characters	6067
Any Ignore Url Used	

Table 2: Plagiarism report (1.1)

Plagiarism Scan Report	
Summary	
Report Genrated Date	27 Oct, 2017
Plagiarism Status	88% Unique
Total Words	810
Total Characters	4979
Any Ignore Url Used	

Table 3: Plagiarism report (1.2)

Plagiarism Scan Report	
Summary	
Report Genrated Date	27 Oct, 2017
Plagiarism Status	100% Unique
Total Words	471
Total Characters	2893
Any Ignore Url Used	

Table 4: Plagiarism report (1.3)

CERTIFICATES



IJARSE
INTERNATIONAL JOURNAL of
Advance Research in Science
and Engineering
ISSN(O) : 2319-8354, ISSN(P) : 2319-8346

Certificate

This is to certify that
Shreesh Rao
has published a paper title
Hindi Speech Recognition
in
International Journal of Advance Research in Science and Engineering
Volume No. 07, Special Issue No. (01), March 2018
This paper can be downloaded from the following link: www.ijarse.com

Impact Factor 2.83

Editor in Chief
International Journal of Advance Research in Science and Engineering
website: www.ijarse.com
E-mail: submission@ijarse.com
IJARSE Team wishes all the best for your bright future

BVP-109



IJARSE
INTERNATIONAL JOURNAL of
Advance Research in Science
and Engineering
ISSN(O) : 2319-8354, ISSN(P) : 2319-8346

Certificate

This is to certify that
Prashant Dombale
has published a paper title
Hindi Speech Recognition
in
International Journal of Advance Research in Science and Engineering
Volume No. 07, Special Issue No. (01), March 2018
This paper can be downloaded from the following link: www.ijarse.com

Impact Factor 2.83

Editor in Chief
International Journal of Advance Research in Science and Engineering
website: www.ijarse.com
E-mail: submission@ijarse.com
IJARSE Team wishes all the best for your bright future

BVP-109



IJARSE
INTERNATIONAL JOURNAL of
Advance Research in Science
and Engineering
ISSN(O) : 2319-8354, ISSN(P) : 2319-8346

Certificate

This is to certify that
Ronit Dhonde
has published a paper title
Hindi Speech Recognition
in
International Journal of Advance Research in Science and Engineering
Volume No. 07, Special Issue No. (01), March 2018
This paper can be downloaded from the following link: www.ijarse.com

Impact Factor 2.83

Editor in Chief
International Journal of Advance Research in Science and Engineering
website: www.ijarse.com
E-mail: submission@ijarse.com
IJARSE Team wishes all the best for your bright future

BVP-109



IJARSE
INTERNATIONAL JOURNAL of
Advance Research in Science
and Engineering
ISSN(O) : 2319-8354, ISSN(P) : 2319-8346

Certificate

This is to certify that
Mrudali Birla
has published a paper title
Hindi Speech Recognition
in
International Journal of Advance Research in Science and Engineering
Volume No. 07, Special Issue No. (01), March 2018
This paper can be downloaded from the following link: www.ijarse.com

Impact Factor 2.83

Editor in Chief
International Journal of Advance Research in Science and Engineering
website: www.ijarse.com
E-mail: submission@ijarse.com
IJARSE Team wishes all the best for your bright future

BVP-109

Virtual assistant using speech Recognition in Hindi

Ms. Kajal Jewani Asst Prof., Shreesh Rao, Prashant Dombale, Ronit Dhonde, Mrudali Birla

*Department of computer science, V.E.S.I.T., Mumbai University
Sindhi Society, Chembur, Maharashtra, India*

kajal.jewani@ves.ac.in

shreesh.rao@ves.ac.in

prashant.dombale@ves.ac.in

ronit.dhonde@ves.ac.in

mrudali.birla@ves.ac.in

Abstract— Speech has been an integral part of human communication since a very long time, hence using speech in order to get things done plays a huge role in our life. That being the central idea we have created a system to understand and process the speech given by a user and perform the task specified. There are certain problems that can arise which would make it unable to understand the spoken sentences such as language barrier, unclear voice, background noise etc. Our system ‘Sir-ji’ aims to eliminate these issues. Furthermore a chatbot system can be added in order to solve user queries.

Keywords—component; formatting; style; styling; insert (key words)

I. INTRODUCTION

Speech is one of the most commonly used form of communication. It can be used in various fields due to the advent in fast computing. Speech recognition is widely accepted as the future of interaction with various mobile as well as computer devices. A particular task such as opening an application, setting a reminder can be done with the help of voice commands, the use of keyboards and mouse can be avoided in such cases.

II. MODULES

A. Virtual Assistant

A virtual assistant is a software agent that can perform tasks or services for an individual. Virtual personal assistant should be accessible by everyone irrespective of the language people use. In some rural parts of India, people still cannot speak or write in English. Due to this issue, people don't tend to use VPAs on a larger scale even though they make functioning of tasks extremely easy and effortless. Our system mainly caters to this section and allows them to experience the thrill of using a VPA.

B. Chatbot

Chatbots can assist the human computer interaction by answering queries and doubts from the users. It is an intelligent computer program which analyzes the user's input and gives an appropriate reply. The chatbot should give intelligent and accurate responses to the user. A Chatbot can be divided into three parts: Responder, Classifier and Graphmaster :

1) Responder: This part plays the role of interfacing between the user and the bot's routine. The responder's job is to transfer data from the user to the Classifier and also controlling the input and output.

2) Classifier: Classifier works between the Responder and the Graphmaster. Its functions are filtering and normalising the input, segmentation of the input given by the user into logical components, transferring the normalised sentence into the

Graphmaster, processing the output from the Graphmaster, and handling the instructions of the database syntax.

3) Graphmaster: The main job of Graphmaster is pattern matching. It organizes the brain's contents, holds the pattern matching algorithm and storage.

Fundamental design techniques for a chatbot:

1. Parsing: This technique analyses the input and manipulates it using a number of NLP functions; for example, trees in Python NLTK.

2. Pattern matching: Pattern matching is the process of checking and locating of specific sequences of data based on some pattern in a raw data.

3. SQL and relational database: it is a technique recently used to remember previous conversations in a chatbot.

4. Markov Chain: Markov Chain is used to build efficient chatbots which are more applicable probabilistically and consequently. Markov chain is based on the idea that the probability of occurrences of each letter or word in the same textual data is fixed..

5. Language tricks: In this technique of language tricks there are sentences, phrases, or even paragraphs available in Chatbots in order to add variety to the knowledge base and make it more convincing.

III. SPEECH ANALYSIS AND RESPONSE

Speech analysis can be divided into three stages: (i) voice recognition and conversion to text, (ii) text processing, and (iii) response and action taking. These stages are explained as follows:

First, speech independent of the speaker passes through a microphone to a computer which converts it into a stream of pulses that contain speech information. Certain instructions can be used to read input speech then to convert it into text. This stage provides speech text for processing in the next stage.

Next, the remaining text is split into separate words for tagging with parts-of-speech labels based on their positions of the words in the sentence. Different types of grammar can be used in this stage to break the individual words in a particular order to form phrases. Keywords are extracted by eliminating unwanted words or phrases in operations. If incorrect these keywords can be checked and corrected.

Existing systems and proposed system:

- The proposed system aims to receive input and provide output using a voice assistant in Hindi Language. A chatbot can be added to further assist the user.
- Google has introduced its chatbot “Allo” for interfacing with google search and assist the user, but this chatbot only responds to English sentences.
- The proposed system would enable the user to perform tasks with commands given in Hindi.
- Most of the existing systems are designed to respond to certain commands only which limits the possibilities of a virtual assistant.
- The proposed system will be recognizing commands and cater to the needs of user with the available information.
- It will have an interface which will be simple and so easy to use that anyone can use it. Compared to the existing systems it is a bit simpler to understand and interact with.

IV. RESULT

The system is a Virtual assistant system functional in both Hindi and English language. The outputs are produced in Hindi or English depending on the input received. Execution speed currently depends on the speed and efficiency of internet connection used. The Chatbot takes the queries in form of words or phrases and responds accordingly.

V. ACCURACY

Using modular design for all its components the system delivers high performance. The performance is relative to the processing capacity of the systems involved. Since all the modules are not running off one system the possible load has been decreased and further decreased by delegating the voice processing to the chat client communicating with the service. The use of an expert system (Ultimate Research Assistant) allows unlimited and autonomous intelligence improvements.. All content received back from the expert system is processed minimally since the information has already been processed for its relevance. This can be somewhat subjective considering that the intelligence reliance was shifted to another party and when such a third-party system is decommissioned this system would also fail.

The application and all its modules were evaluated by calculating the Word Error Rate(WER).Word error rate (WER) is a common metric of the performance of a speech recognition or machine translation system.Word error rate can then be computed as:

$$WER=(I+D+S)/N$$

where,

S is the number of substitutions,

D is the number of deletions,

I is the number of insertions,

C is the number of the corrects,

N is the number of words in the reference ($N=S+D+C$)

The main idea behind 'deletion' and 'insertion' is how to get to the hypothesis from the reference. So if we have the reference "This is Sparta" and hypothesis "This _ Sparta", we call it a deletion.

VI. CONCLUSION

In this paper, development of an effective virtual assistant along with a chatbot was shown. This system particularly focuses on Hindi language. Virtual Personal Assistants (VPA) chatbots are the most effective way to obtain information and do tasks in a simpler way. The limitation which is encountered while using VPAs is the fluency in English language. However in our project, we are developing VPA which can understand and process human commands given in Hindi language. The complexity and the accent problems of various chatbots is resolved in our project thereby making even illiterate audience use the system with ease. Since Hindi is the most spoken language in India, we think most of the Indian population can make efficient use of the VPA. Also, with the help of chatbot they can resolve their doubts easily.VPAs are also more reliable than Human Assistant because, VPAs are easily portable and one can use them anytime. Our project has a simple and easy to understand interface.

VII. ACKNOWLEDGMENT

We are thankful to our college, Vivekanand Education Society's Institute of Technology, for considering our project and providing us help at every stage of the research. A special thanks to Ms. Kajal Jewani for mentoring and guiding us in the right direction. We would like to thank Dr.(Mrs.) Nupur Giri and our principal Dr. (Mrs.) J.M. Nair for giving this valuable opportunity. We are immensely grateful to everyone for their help with the research without which it would have been difficult to get our job done. Also, thanks to our families for the moral support and encouragement. It is a great pleasure to acknowledge the help we received from the Department of Computer Engineering.

REFERENCES

- [1] Preeti Saini, Parneet Kaur, Automatic Speech Recognition: A Review, in International Journal of Engineering Trends and Technology- Volume4Issue2-2013
- [2] Suma Swamy, K.V. Ramakrishnan, An Efficient Speech Recognition System, in An International Journal (CSEIJ), Vol. 3, No. 4, August 2013
- [3] An intelligent web-based voice chat bot. Published in: (IEEE) Institute of Electrical and Electronics Engineers. Date: June 2009
- [4] A review on Speech recognition techniques. Published in: International journal of computer applications Vol 10-No 3.Date: November 2010
- [5] Survey on Chatbot Design Techniques in Speech Conversation Systems Sameera A. Abdul-Kader School of Computer Science and Electronic Engineering/University of Essex Colchester/ UK Diyala University/ Diyala/ Iraq Dr. John Woods School of Computer Science and Electronic Engineering/University of Essex Colchester/ UK

PLAGIARISM REPORT

Plagiarism Scan Report	
Summary	
Report Genrated Date	21 Apr, 2018
Plagiarism Status	100% Unique
Total Words	938
Total Characters	5788
Any Ignore Url Used	

Table 5: Plagiarism report (2.1)

Plagiarism Scan Report	
Summary	
Report Genrated Date	21 Apr, 2018
Plagiarism Status	100% Unique
Total Words	399
Total Characters	2478
Any Ignore Url Used	

Table 6: Plagiarism report (2.2)