REVA UNIVERSITY

Yelahanka, Bengaluru-560064

A Mini Project Report on

"SPEECH RECOGNITION"

Submitted
In partial fulfilment of the requirement for the VI Semester of Bachelor of Technology in Computer Science and Engineering during the academic year 2017-18

SUBMITTED BY:

J BHARAT	R15CS140
L ADITYA	R15CS184
MAHAVEER PINCHA	R15CS193
SK SHAREEF	R15CS547

Project Group ID : C15

Under the guidance of :

Prof. VANI K, Sr.Asistant Professor,
School of C & IT

REVA UNIVERSITY
School of Computing and Information Technology
Bengaluru-560064
2017-18

REVA UNIVERSITY
Bengaluru, India

SCHOOL OF COMPUTING AND INFORMATION TECHNOLOGY

This is to certify that the mini-project entitles "**Speech Recognition**" is bonafide work carried out by **J Bharat ,L Aditya ,Mahaveer Pincha ,Sk Shareef** bearing **R15CS140 ,R15CS184 ,R15CS193 ,R15CS547** respectively in partial fulfilment of 6th semester of Computer and Engineering program of Bachelor of Technology, REVA University during the academic year 2018-19. It is certified that all the corrections/suggestions indicated for internal assessment have been incorporated in the report deposited in the school library. The mini-project report has been approved as it satisfies the academic requirements in respect of mini-project prescribed for the 6th semester of CSE program.

Signature of the Guide

Signature of the Director

(Dr. Sunilkumar S. Manvi)

ACKNOWLEDGEMENT

Any given task achieved is never the result of efforts of a single individual. There are always a bunch of people who play an instrumental role is leading a task to its completion. Our joy at having successfully finished our mini project would be incomplete without thanking everyone who helped us out along the way. We would like to express our sense of gratitude to REVA UNIVERSITY for providing us the means of attaining our most cherished goal.

We cannot express enough thanks to our respected Director, Dr. Sunilkumar S. Manvi for providing us with a highly conducive environment and encouraging the growth and creativity of each student.

We would like to take this opportunity to express our gratitude to our mini project Guide, Prof. Vani K, Sr. Assistant Professor for continuously supporting and guiding us in our every endeavor as well for taking a keen and active interest in the progress of every phase of our project. Thank you for providing us with the necessary inputs and suggestions for advancing with our mini project. We deeply appreciate the wise guidance that sir has provided.

Finally, we would like to extend our sincere thanks to all the faculty members, staff from School of Computing and Information Technology.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Speech recognition technology is one from the fast growing engineering technologies. Various interactive speech aware applications are available in the market. Speech recognition systems are the efficient alternatives for such devices where typing becomes difficult. It has a number of applications in different areas and provides potential benefits. Nearly 20% people of the world are suffering from various disabilities; many of them are blind or unable to use their hands effectively. The speech recognition systems in those particular cases provide a significant help to them, so that they can share information with people by operating computer through voice input.

This project is designed and developed keeping that factor into mind, and a little effort is made to achieve this aim. Our project is capable to recognize the speech and convert the input audio into text using python script; it also enables a user to perform operations such as "save, open, exit" a file by providing voice input. It also helps the user to open different system software applications such as opening Ms-paint, notepad and calculator.

At the initial level effort is made to provide help for basic operations as discussed above, but the software can further be updated and enhanced in order to cover more applications.

1. INTRODUCTION

1.1 SPEECH RECOGNITION:-

Have you ever talked to your computer? I mean, have you really, really talked to your computer? Where it actually recognized what you said and then did something as a result? If you have, then you've used a technology known as speech recognition.

Speech recognition allows you to provide input to a system with your voice. Just like clicking with your mouse, typing on your keyboard, or pressing a key on the phone keypad provides input to an application, speech recognition allows you to provide input by talking. In the desktop world, you need a microphone to be able to do this.

1.2 MOTIVATION:-

We believe the best is yet to come. Improvements in speech recognition, including new architecture and optimization strategies, have been explored in a variety of machine learning disciplines, including face recognition.

We planned to make some common tasks that every user does on his/her computer (opening/ closing programs, editing texts, calculating) possible not only by mouse, keyboard, but also by voice.

1.3 OBJECTIVE:-

To understand the speech recognition and its fundamentals.
its working and applications in different areas
Its implementation as a desktop Application
Development for software that can mainly be used for:

- Speech Recognition
- Speech Generation
- Tool for operating Machine through voice

2.LITERATURE SURVEY:-

2.1 AN OVERVIEW OF SPEECH RECOGNITION:-

Speech recognition is a technology that able a computer to capture the words spoken by a human with a help of microphone [1] [2]. These words are later on recognized by speech recognizer, and in the end, system outputs the recognized words .The process of speech recognition consists of different steps that will be discussed in the following sections one by one.

An ideal situation in the process of speech recognition is that, a speech recognition engine recognizes all words uttered by a human but, practically the performance of a speech recognition engine depends on number of factors. Vocabularies, multiple users and noisy environment are the major factors that are counted in as the depending factors for a speech recognition engine [3]

2.2 HISTORY

The concept of speech recognition started somewhere in 1940s [3], practically the first speech recognition program was appeared in 1952 at the bell labs, that was about recognition of a digit in a noise free environment [4], [5].

1940s and 1950s consider as the foundational period of the speech recognition technology, in this period work was done on the foundational paradigms of the speech recognition that is automation and information theoretic models [6].

In 1990s the key technologies developed during this period were the methods for stochastic language understanding, statistical learning of acoustic and language models, and the methods for implementation of large vocabulary speech understanding systems.

After the five decades of research, the speech recognition technology has finally entered marketplace, benefiting the users in variety of ways. The challenge of designing a machine that truly functions like an intelligent human is still a major one going forward.

2.3 TYPES OF SPEECH RECOGNITION:-

Speech recognition systems can be divided into the number of classes based on their ability to recognize that words and list of words they have. A few classes of speech recognition are classified as under:

2.3.1 Isolated Speech

Isolated words usually involve a pause between two utterances; it doesn't mean that it only accepts a single word but instead it requires one utterance at a time [4].

2.3.2 Connected Speech

Connected words or connected speech is similar to isolated speech but allow separate utterances with minimal pause between them.

2.3.3 Continuous speech

Continuous speech allow the user to speak almost naturally, it is also called the computer dictation

2.3.4 Spontaneous Speech

At a basic level, it can be thought of as speech that is natural sounding and not rehearsed. An ASR system with spontaneous speech ability should be able to handle a variety of natural speech features such as words being run together, "ums" and "ahs", and even slight stutters.

2.4 VARIOUS APPROACHES TO SPEECH RECOGNITION:-

The three broad approaches to automatic speech recognition are the acoustic-phonetic, pattern recognition and artificial intelligence (AI) approaches [7]. The acoustic phonetic approach to speech recognition has not been very successful in practical speech recognition systems. Both the pattern recognition and AI approach to speech recognition have achieved higher success rates than the acoustic-phonetic approach.

2.4.1 Acoustic-Phonetic Approach:-

In this speech recognition algorithm, the system tries to decode the speech signal in a sequential manner based on the observed acoustic features of the speech waveform and the known relations between acoustic features and phonetic symbols. Figure 1 shows a block diagram of the acoustic phonetic approach to speech recognition. The first step in the process is the parameter measurement process, which provides an appropriate spectral representation of the speech signal. The next step in the processing is the feature detection stage where the spectral measurements are converted to a set of features that describe the acoustic properties of the various phonetic units. Finally, the recognizer tries to determine the best matching word or sequence of words.

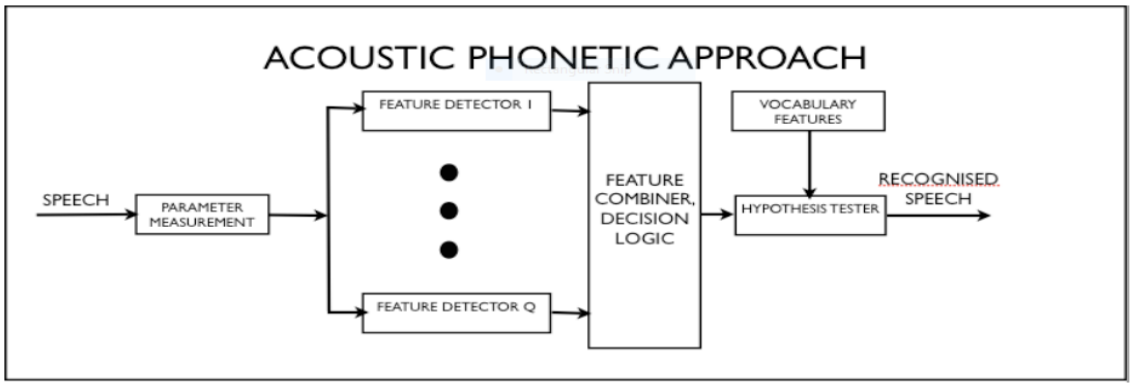


FIGURE [1]: Block diagram of Acoustic-Phonetic Approach

2.4.2 Pattern Recognition Approach:-

In this approach, the speech patterns are used directly without explicit feature determination and segmentation. The method has two steps-namely, training of speech patterns, and recognition of patterns by way of pattern comparison. Figure 2 shows a block diagram of the pattern-recognition approach. In the parameter measurement phase, a sequence of measurements is made on the input signal to define the "test pattern". The unknown test pattern is then compared with each sound reference pattern and a measure of similarity between the test pattern and reference pattern is computed. Finally the decision rule decides which reference pattern best matches the unknown test pattern based on the similarity scores from the pattern classification phase.

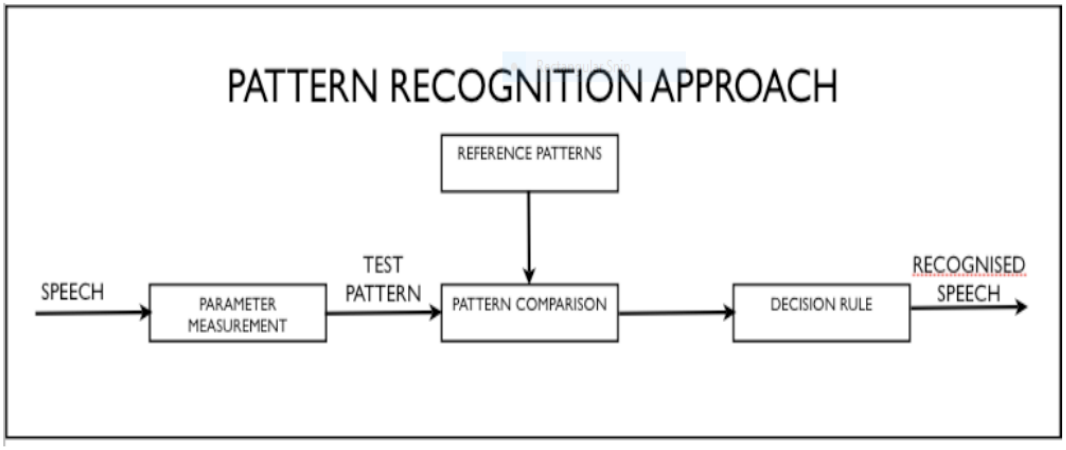


FIGURE [2] : Block diagram of Pattern Recognition Approach

2.4.3 Artificial Intelligence Recognition Approach

This approach is a hybrid of the acoustic-phonetic approach and the pattern recognition approach. In the artificial intelligence approach (AI), an expert system or self-organizing (learning) system, implemented by neural networks is used to classify sounds. The basic idea is to compile and incorporate knowledge from a variety of knowledge sources with the problem at hand.

2.5 Speech Recognition flaws :-

Besides all these advantages and benefits, yet a hundred percent perfect speech recognition system is unable to be developed. There are number of factors that can reduce the accuracy and performance of a speech recognition program.

Speech recognition process is easy for a human but it is a difficult task for a machine, comparing with a human mind speech recognition programs seems less intelligent, this is due to that fact that a human mind is God gifted thing and the capability of thinking, understanding and reacting is natural, while for a computer program it is a complicated task, first it need to understand the spoken words with respect to their meanings, and it has to create a sufficient balance between the words, noise and spaces. A human has a built in capability of filtering the noise from a speech while a machine requires training, computer requires help for separating the speech sound from the other sounds

Few factors that are considerable in this regard are:-

Homonyms: Are the words that are differently spelled and have the different meaning but acquires the same meaning, for example "there" "their" "be" and"bee". This is a challenge for computer machine to distinguish between such types of phrases that sound alike.

Overlapping speeches:

A second challenge in the process, is to under stand the speech uttered by different users, current systems have a difficulty to separate simultaneous speeches form multiple users.

Noise factor:

the program requires hearing the words uttered by a human distinctly and clearly. Any extra sound can create interference, first you need to place system away from noisy environments and then speak clearly else the machine will confuse and will mix up the words.

3 DESIGN AND IMPLEMENTATION:-

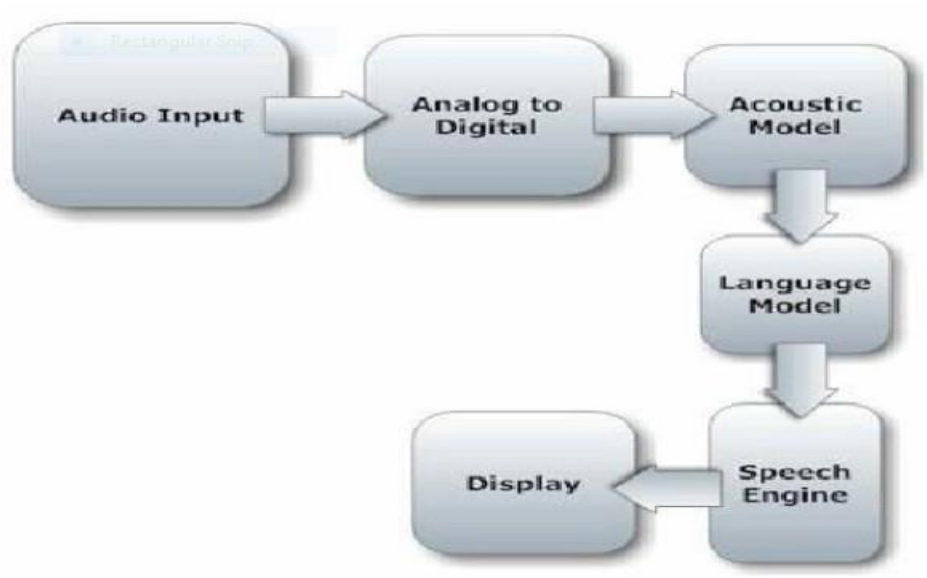


FIGURE [3] : Block diagram for processes of speech recognition

3.1 Components of Speech recognition System Voice Input

With the help of microphone audio is input to the system, the pc sound card produces the equivalent digital representation of received audio [8] [9] [10].

Digitization

The process of converting the analog signal into a digital form is known as digitization [8], it involves the both sampling and quantization processes. Sampling is converting a continuous signal into discrete signal, while the process of approximating a continuous range of values is known as quantization.

Acoustic Model

An acoustic model is created by taking audio recordings of speech, and their text transcriptions, and using software to create statistical representations of the sounds that make up each word. It is used by a speech recognition engine to recognize speech [8].The software acoustic model breaks the words into the phonemes [10].

Language Model

Language modelling is used in many natural language processing application such as speech recognition tries to capture the properties of a language and to predict the next word in the speech sequence [8]. The software language model compares the phonemes to words in its built in dictionary [10].

Speech engine

The job of speech recognition engine is to convert the input audio into text [4]; to accomplish this it uses all sorts of data, software algorithms and statistics. Its first operation is digitization as discussed earlier, that is to convert it into a suitable format for further processing. Once audio signal is in proper format it then searches the best match for it. It does this by considering the words it knows, once the signal is recognized it returns its corresponding text string.

4. SYSTEM REQUIREMENTS:-

4.1 Software Requirements:

OS: Windows 7/8/10/Mac OS X 10.0 or higher/Linux: GNOME or KDE Desktop

PYTHON IDLE: 3.6.5+

PROGRAMMING LANGUAGE: Python

LIBRARY: Speech Recognition,GTTS,Web Browser etc.

4.2 Hardware Requirements:

RAM: At least 1 Gigabytes

STORAGE: At least 10 Gigabytes

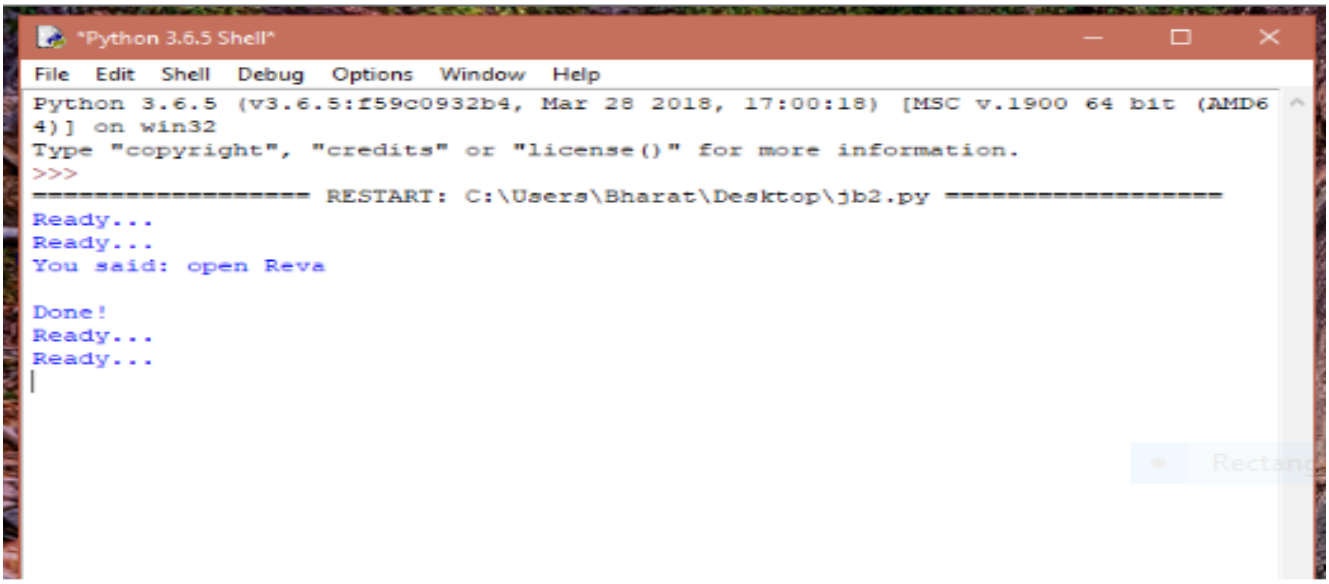
PROCESSOR: Intel Core i3 or above

EXTERNAL MIC

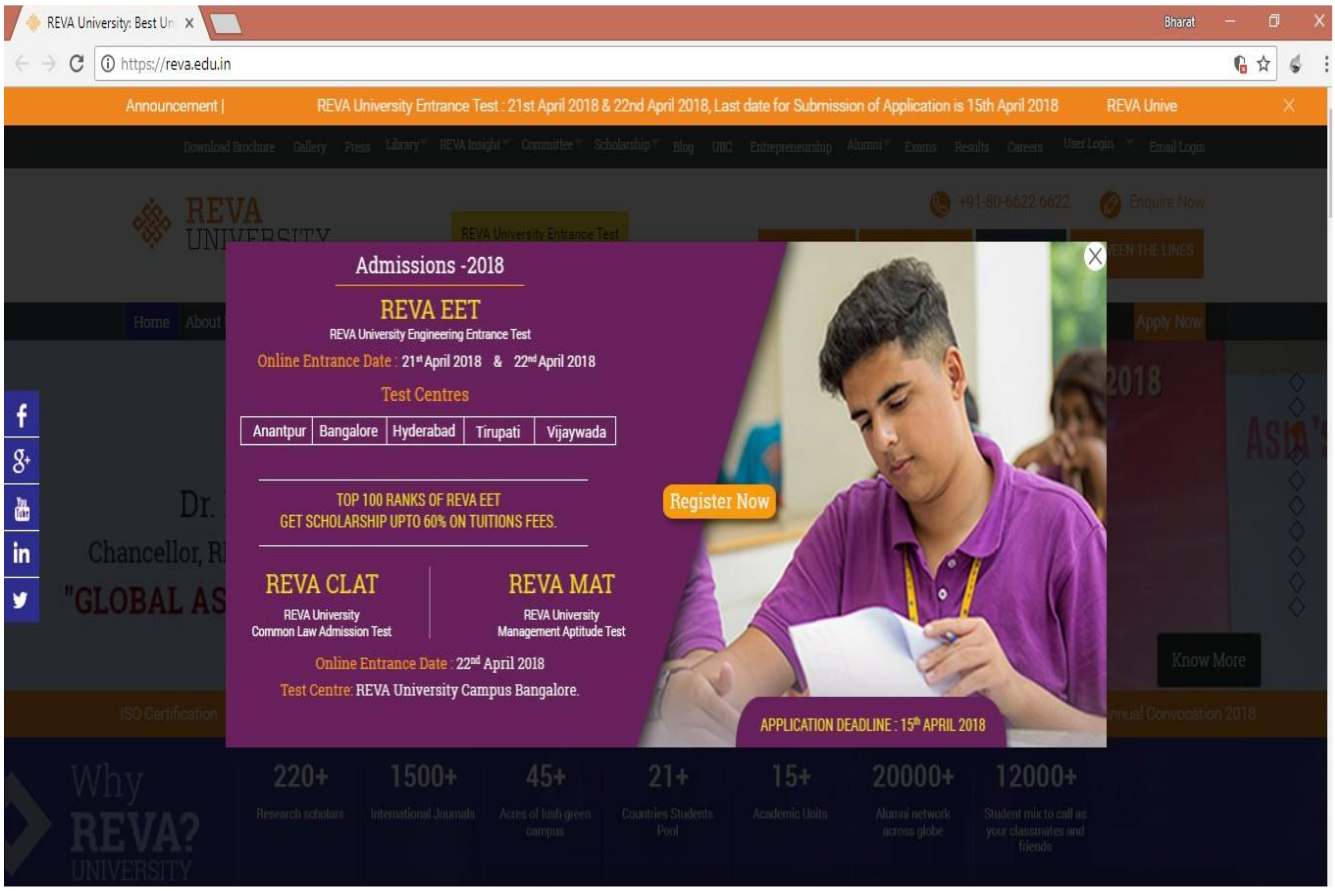
5. RESULT :-

RESULT- 1

INPUT : We gave command to open reva webpage .

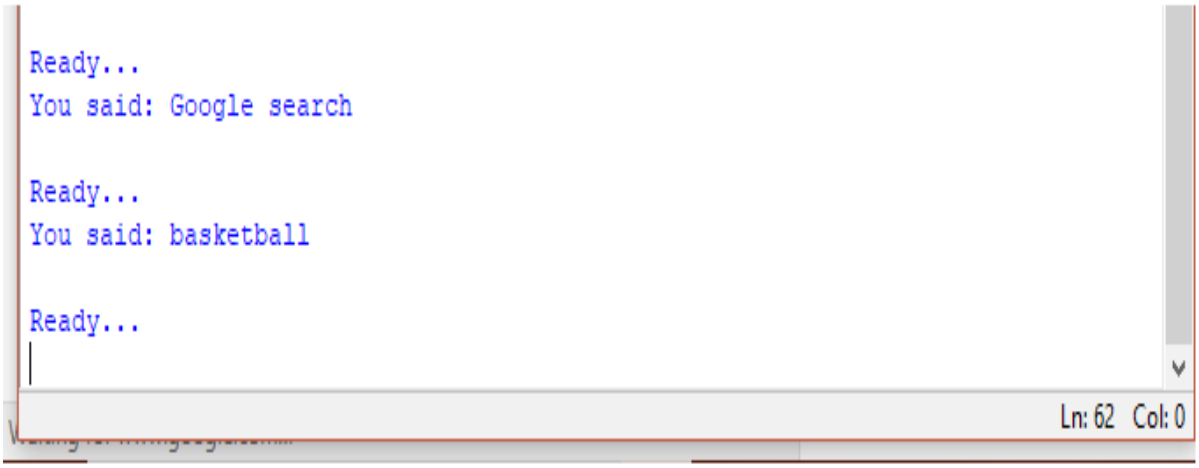


OUTPUT

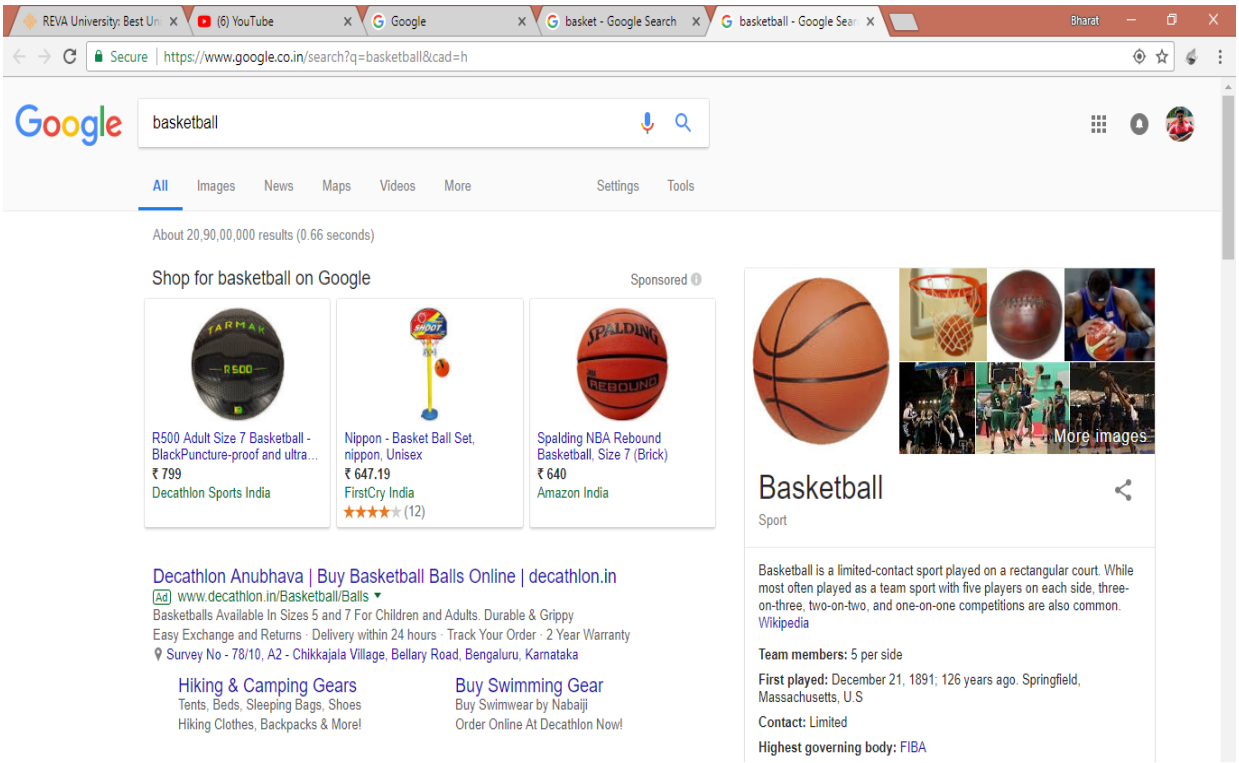


RESULT-2

INPUT :- we gave command to google search basketball



OUTPUT



RESULT-3

INPUT :- We gave command to get current weather in Bengaluru, Kochi, Tirupati.

OUTPUT

```
Ready...
Ready...
You said: current weather

Ready...
You said: Bangalore

Current temperature in Bangalore : 25 celsius
Ready...
Ready...
You said: current temperature

Ready...
You said: current weather

Ready...
You said: Kochi

Current temperature in Kochi : 29 celsius
Ready...
Ready...
You said: current weather

Ready...
You said: Tirupati

Current temperature in Tirupati : 27 celsius
Ready...
```

6 APPLICATIONS :-

6.1 From medical perspective

People with disabilities can benefit from speech recognition programs. Speech recognition is especially useful for people who have difficulty using their hands,in such cases speech recognition programs are much beneficial and they can use for operating computers. Speech recognition is used in deaf telephony, such as voicemail to text.

6.2 From military perspective

Speech recognition programs are important from military perspective; in Air Force speech recognition has definite potential for reducing pilot workload. Beside the Air force such Programs can also be trained to be used in helicopters battle management and other applications.

6.3 From educational perspective

Individuals with learning disabilities who have problems with thought-to-paper communication (essentially they think of an idea but it is processed incorrectly causing it to end up differently on paper) can benefit from the software.

7.CONCLUSION

This Thesis/Project work of speech recognition started with a brief introduction of the technology and its applications in different sectors. The project part of the Report was based on software development for speech recognition. At the later stage we discussed different tools for bringing that idea into practical work. After the development of the software finally it was tested and results were discussed, few deficiencies factors were brought in front. After the testing work, advantages of the software were described and suggestions for further enhancement and improvement were discussed.

8.FUTURE ENHANCEMENTS

This work can be taken into more detail and more work can be done on the project in order to bring modifications and additional features. The current software doesn't support a large vocabulary, the work will be done in order to accumulate more number of samples and increase the efficiency of the software. The current version of the software supports only few areas of the notepad but more areas can be covered and effort will be made in this regard.

9.REFERENCES

[1] "Speech recognition- The next revolution" 5th edition.

[2] Ksenia Shalonova, "Automatic Speech Recognition" 07 DEC 2007

[3] gilman, d. (2018). what is speech recognition and voice [online]Abilityhub.com.Available http://www.abilityhub.com/speech/speech-description.html

[4] "Fundamentals of Speech Recognition". L.Rabiner & B. Juang. 1993. ISBN: 0130151572.

[5]D.Jurafsky, J .Martin. "Speech and Language processing: An introduction to natural language processing, Computational Linguistics and Speech Recognition". 2000. ISBN:0130950656

[6] B.H. Juang & Lawrence R. Rabiner, "Automatic Speech Recognition – A Brief History of theTechnology Development," 08 AUG 2004.

[7] L.R.Rabiner, B.H.Juang. Fundamentals of Speech Recognition, Prentice-Hall, Inc.,Upper Saddle River, NJ. 1993

[8]John Kirriemuir "Speech recognition technologies" March 30th 2003

[9] Anon, (2018). [online] Available at: http://electronics.howstuffworks.com/gadgets/high-tech-gadgets/speechrecognition.htm/printable

[10]Anon, (2018).[online]Available at: http://www.jisc.ac.uk/media/documents/techwatch/ruchi.pdf