Assiting Voice Input Recognition Application

Snehalatha N
Assistant professor: dept. of CSE, JSSATE
line 2-name of organization, acronyms acceptable
Bengaluru, India

Rajesh Adiga P
Student: dept. of CSE, JSSATE
Bengaluru, India

Pavan Kumar T G
Student: dept. of CSE, JSSATE
line 2-name of organization, acronyms acceptable
Bengaluru, India

Rakshith V
Student: dept. of CSE, JSSATE
Bengaluru, India

***Abstract*—Assisting Voice Input Recognition Application (in short, AVIRA) is a multifunctional software programming application capable of managing computer's basic operations. It employs a user interface that is capable of receiving user's voice as its input, process it and perform a necessary operation. It communicates back with the user via synthesized speech. The application requests users input in English and responds in the same. Also this system will be very much useful to the differently abled persons and the visually challenged persons who cannot read and, wants to know the details in their education and in any other aspects of the real world. Normally, the visually challenged persons acquire knowledge and exchange information with others mainly through the speech and writing. This system will help those persons in their education and the computer will be made user friendly to achieve this purpose. This paper is based on the fact that the computer will be able to interact with the user and fulfill their needs in the computer world. In this, we have introduced a new approach of using Speech recognition and Speech synthesis to have a two way communication. This makes the user feel as if they are getting a reply from another person. A visually challenged person can easily interact with the computer systems without the help of other persons. Thus we have made an effort in making this an intelligent environment in the speech processing with the computer system by getting the Input from the user as speech input and artificially generate the synthesized voice which makes the process easy.**

Keywords*—Speech Recognition, Speech Synthesis and Artificial Intelligence.*

I. INTRODUCTION

Speech processing is the study of speech signals and the processing methods of these signals. The signals are usually processed in a digital presentation, so that speech processing can be regarded as a special case of digital signal processing, applied to speech signal. It is also closely tied to natural language processing (NLP). E.g. text-to-speech synthesis might use a syntactic parser on its input text and speech recognition's output may be used by (say) information extraction techniques. Speech processing technology enables natural interaction with all kinds of computer systems, from cell phones, PDAs, and PCs.

This application will be useful for physically challenged persons and visually challenged persons who have less or no knowledge about how to read or to get the information from the computer. This will help them to know what's happening in the real world and grasp the effect of the information they receive, so that they can choose to acknowledge or ignore it.

Normally, the visually challenged persons acquire knowledge and exchange information with others mainly through the speech and writing. This application will help those persons in their pursuit of knowledge and the computer will be a tailor-made to quench their thirst for knowledge acquisition. The process of speech based application will allow the user interaction to be easy and smooth which will help the persons who lack in theknowledge of using computers and thereby, not becoming a burden to other people to help them access it.

II. RELATED WORK

One of the systems proposed by E. Matusov, S. Kanthak, and H. Ney [1] focusses on the interface between speech recognitionand machine translation in a speech translation system. Based on a thorough theoretical framework, we exploit word lattices of automatic speech recognition hypotheses as input to our translation system which is based on weighted finite-state transducers. We show that acoustic recognition scores of the recognized words in the lattices positively and significantly affect the translation quality. In experiments, we have found consistent improvements on three different corpora compared with translations of single best recognized results. In addition we build and evaluate a fully integrated speech translation model.

Another advancement in the speech recognition is used in many fields like Voice Recognition System for the Visually Impaired [2] highlights the Mg Sys Visi system that has the capability of access to World Wide Web by browsing in the Internet, checking, sending and receiving email, searching in the Internet, and listening to the content of the search only by giving a voice command to the system.

Many surveys have been carried out regarding the speech recognition and one such attempt is made by M.A.Anusuya and S.K.Katti [3]. They discuss the major themes and advances made in the past 60 years of research, so as to provide a technological perspective and an appreciation of the fundamental progress that has been accomplished in this important area of speech communication. After years of research and development the accuracy of automatic speech recognition remains one of the important research challenges (eg., variations of the context, speakers, and environment).The design of Speech Recognition system requires careful attentions to the following issues: Definition of various types of speech classes, speech representation, feature extraction techniques, speech classifiers, database and performance evaluation. The problems which exist in ASR and the various techniques to solve these problems constructed by various research workers have been presented in a chronological order.

From AUDREY to SIRI [4], is speech recognition a solved problem? Addresses the issues faced by speech recognition technology over the past 60 years. This study introduces us to various speech recognition applications over the years which has made the users to access the contents of the computer using their voice. And with the improvements in the HMMs [5] and IVRs, many efforts have been made to make the speech recognition technology a hit among the masses.

III. ARCHITECTURE

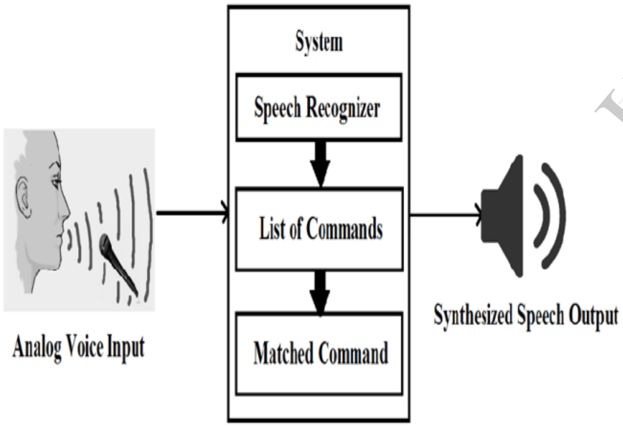


Fig 1: System’s Architecture

The Proposed system makes use of Windows Speech API to accept input and process it. The API has been built to handle parsing which makes the developer’s job easier. A microphone of suitable sensitivity is needed to catch the commands uttered by the user.

Next step is to send the voice note captured by the microphone to the processing stage. This stage is dividedinto a multiple phases. Each Phase has its own functionality that toys with the voice signals passed from the microphone.

Speech Recognizer will recognize the spoken words, processes it, isolating segments of sound that are probably speech and converting them into a series of numeric values that characterize the vocal sounds in the signal.
This recognition process can be thought of as having front end and a back end. The front end part is as explained in the preceding paragraph. Back end is a specialized search engine that takes the output produced by the front end and searches across three databases: an acoustic model, a lexicon, and a language model. In our case, Windows Speech API will perform the tasks of front end and back end.

After the speech recognizer identifies the word(s) spoken by the user and converts it in to a textual form, we match it with our list of commands. When there is a match, we execute that particular operation.

The matched command will be in textual form. We’ll convert it in to a synthesized voice. This will be heard by the user which confirms that the operation has been successfully carried out by our application.

IV. PROPOSED SYSTEM

The proposed system mainly focuses on making the system an interactive one via voice. Although, it does not delve deep into the semantics of the working of a speech recognizer and synthesizer, it efficiently makes use of the available resources to make the system work like a doll to the user’s command.

Our application directly accesses the Windows API to recognize the voice note of the user parses it and converts it into a textual form. And then, the synthesizer will artificially generate the voice response to the user.

The system will enable a person to get the basic details in any website that is displayed in the screen through voice. It will also act as a personal assistant wherein it’ll reply to you as though you are getting a reply from a known person. This application will be most useful for the visually challenged people as they can access the information they need with the help of their voice.

The proposed system will be able to communicate with the user via synthesized voice. This will be useful as they can know when a command has been executed or not. The commands given by the user, as mentioned earlier, will be processed using Windows SAPI and performs the operations as defined.

There are several programming languages available to build these kinds of systems but this is developed underC# with .NET 4.0 frameworks. It will be easy to deploy and to port among Windows family of operating systems.

The various operations that can be performed are:

- Pre-defined Commands: These are the commands that are needed for basic usage of the system, which we have defined.
- Web Commands: These can be used to access some websites. They have been assigned during the development phase.
- Auto Read: This will be able to read out the contents of a file. If a user wants the systemto read out the particular section of a file, that can also be achieved.
- Auto Type: This was a major challenge that we had to tackle as the system had to recognize the words that were never defined in SAPI. We tried to fill as many words as possible to increase the accuracy in recognition.
- Custom Commands: We have provided the users an option of adding their own commands based on the scenario.
- P-P (Pause and Play): This will make the system to temporarily stop listening and resumes when the user activates it.

Our ultimate goal is making Assisting Voice Input Recognition Application (AVIRA) with the help of this library and Natural Language Processing for visually challenged and physically challenged people.

A detailed study is carried out to check the work ability of the proposed system. The feasibility study is a test of system proposal regarding its work ability, impact on the organization, ability to meet user needs and effective use of resources [6]. The various tests are:

- Economic Feasibility: Economic justification includes abroad range of concerns that includes cost benefit analysis. Thissystem will be developed and operated in the existing hardware and software infrastructure. Hence there is no need of procuring additional hardware and software for the proposed system. The proposed will give theinformation within minutes, hence the performance is improved.Economic feasibility is checked by whether the financial benefits are exceeds the cost. This system uses onlythe Open source software which is economically feasible. The proposed system will minimize the time and efforts involved in processing, hence it is economically feasible.
- Feasibility analysis: During system analysis the feasibility study of the proposed system was carried out to see, how far itwould be beneficial to the organization [6]. A feasibility analysis is to test the system proposal according to its workability, impact on the organization, ability to meet user needs and effective use of resources for the present day online messaging reliable resources of data transfer is not available and instant message is possible [7].
- Technical Feasibility: The proposed system is said to be technically feasible if the necessary hardware and software areavailable for the system. The technical feasibility issues during the feasibility stage of the investigation include, (i) The apt technology is adopted to develop the system (ii) The system can be expanded and the organization should provide the sufficient equipment to develop thesystem. But, the proposed system requires only Windows 7 or above and an advanced domain as Natural Language Processing Should be used.
- Operational Feasibility: Operational Feasibility is the ability and desire of the management, users and others to use and supportthe proposed system.
- Operational Feasibility: The proposed system offers greater levels of user friendliness combined with effective services to theusers. After feasibility study the analyst has to find if there any faults or error still existing. When a problem is defined clear it can be easily solved using appropriate solutions. Each module is developed using the appropriate technology and the system is developed [8]. Thus the proposed system increases the efficiency and provides more benefits to the organization.
- Social Feasibility: Social Feasibility is the ability to make changes in the society or the country that the research will bevery helpful in its development of the education to the differently abled persons who cannot be able to complete their basic education and their knowledge without other persons help for their needs. Thus the proposed system increases the benefits to the people and provides more benefits to the differently abled persons by accessing information from the computers.

V. IMPLEMENTATION

The system was implemented using C# with .NET frameworks in Microsoft Visual Studio. Below is the data flow diagram of our system:

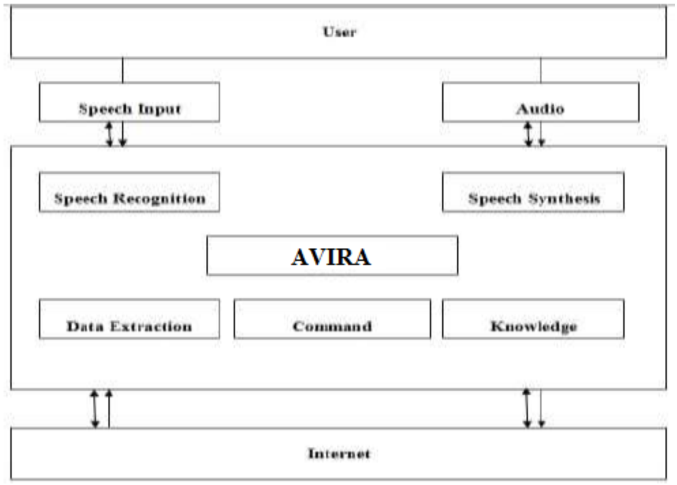


Fig 2: Dataflow diagram

The first and foremost is to get the user’s voice with the help of a microphone. Then it undergoes a series of transformations to produce an apt reply with the help of the synthesized voice.

Speech Recognition is a technology that can translate spoken words into text. Some speech recognition systems use "training" where an individual speaker reads sections of text into the speech recognition system. These systems analyze the person's specific voice and use it to fine tune the recognition of that person's speech, resulting in more accuratetranscription. Systems that do not use training are called "Speaker Independent" systems. Systems that use training are called "Speaker Dependent" systems [9]. Our application belongs to the former category.

Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech synthesizer, and can be implemented in software or hardware. A text-to-speech (TTS) system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech. These features of the speech recognition and synthesis combined together to form a library for perform both recognition and synthesis process in a single machine. The user can search the content by giving the work likewise the human can give the keyword through the speech sequence the system can analysis and understand the human need and search the content and parse the raw content to extract the exact knowledge data.

The speech recognition and synthesis operations are handled by the Windows Speech API in an efficient manner. We extract the recognized voice in the textual form and match it with our pre-defined commands to give a particular reply. And then, that reply is produced by a synthesized voice.

As soon as the application is started, the first form will ask how it would like to call us. The users should enter the particular content. The entered content is sent to the content which starts the main process. This includes the initialization of speech recognition engine, speech synthesizer as well as the grammar. The loading of grammar to the application is nothing but sending the text file that has the commands to match with that of users’.

Since the application should be capable of responding for a series of commands, the recognition engine is set so that it executes the operations that are defined by us. We have incorporated some social commands as well, which would help a normal person to feel that he isn’t speaking to a brainiac.

It is capable of giving information about the weather outside. The users can change their location and get the weather report of that particular city. To accomplish this, we have made use of Yahoo! API. All it needs is the Where On Earth IDentifier (WOEID) of the city. It generates an exception resulting in an error message if the users’ are not connected to a network.

Another feature that has been implemented is it gives information regarding the status of the battery; it gives suitable reply corresponding to the battery level. This feature is exclusively for portable computers. It gets the information about the battery from the Windows’ Management Class. If there are multiple batteries connected, this class gets all the instances about the battery.

Another major enhancement was the capability of performing variety of Google search operations for websites as well as images and getting the contents in Wikipedia about anything. This is performed in another windows form. This form inactivates the main program as there should not be any ambiguity. Once the users are done with this, they need to activate the program by pressing the “Wake Up Button”.

Auto Read is implemented as the users might get tired of reading on their own. This copies the selected contents on to another form and starts reading it. Once the contents are readout, this form is closed.

In case if a 3rd party wants to add his/her own command, it is also provided. This form includes three textboxes, 1st for command, 2nd for its reply and 3rd for the location of that particular program. Once they are done adding their own commands, they need to update the library. This is nothing but unloading the library from the program, writing the contents of these three textboxes on to the particular text files and loading them on to the program. This feature will cease to work if it encounters an empty line.

We have given a feature that will allow the system to type a text document without the need of keyboard. The Auto Type feature will contain two textboxes. First, for the name of the file (not mandatory) and second for the contents. Listen button should be pressed to start the recognizer; the user can cancel this operation and can view the text document after it has been created. This form inactivates the main program as it should not get ambiguous regarding the pre-defined commands.

All these features have been implemented with the help of these 4 modules:

- Speech Recognition:

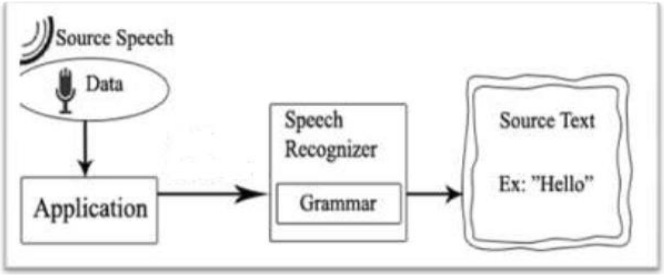


Fig 3: Speech Recognition

i. Speech recognition is the process of converting spoken language to written text or some similar form.
ii. To get the command or voice from the user through microphone.

iii. Then it will check the clarity of voice and pass to the Extraction part.

iv. We use the grammar database using "Windows Speech API" to perform Speech Recognition.

v. The recognized words can be the final results for linguistic processing in order to reply the persons.

- Speech Synthesis:

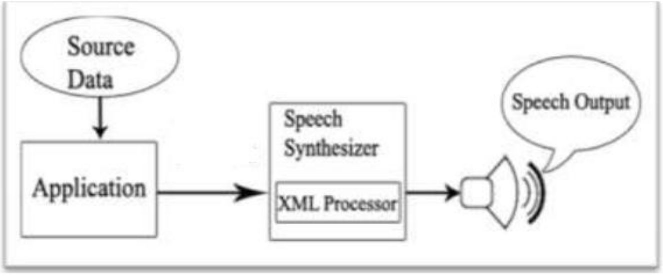


Fig-4: Speech Synthesis

i. The text in any file or a source can then converted to the voice through synthesizer.

ii. It will convert the digital string to the respective voice through the US or UK Dictionary which we wereused.

iii. Then it will dictate the related word in the dictionary.

- DATA EXTRACTION AND KNOWLEDGE UNDERSTAND

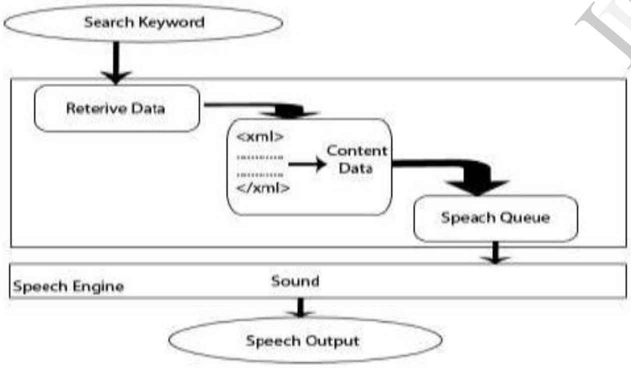


Fig 5: Data Extraction and Knowledge Understand

i. The voice signal is then converted to the digital format based on the Parser by the Windows Speech API.

ii. It will convert the digital string to word.

iii. Then it will process the command given by the user and sends to the synthesizer.

iv. It redirects to synthesizer.

v. Data Extraction performs the role of getting the raw factor of the user input from the internet (if necessary). These rawfactor (XML Based data) could be mined and get the exact knowledge. Knowledge Understand can understand the user keyword to perform both searching process as well as command processing Knowledge.

- Command Processing

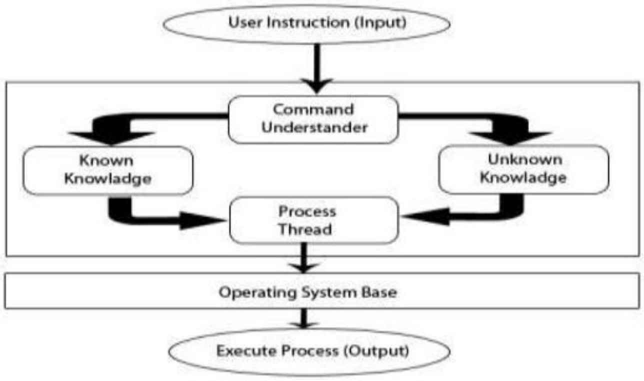
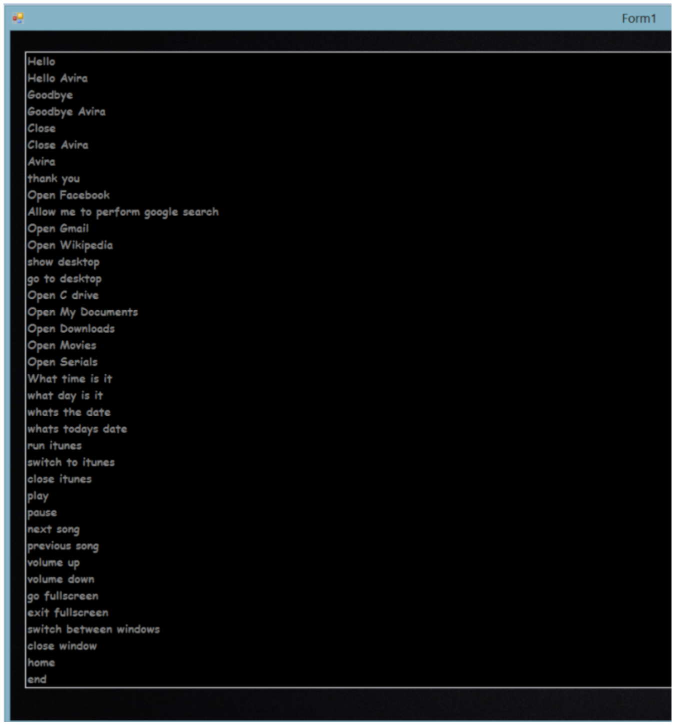


Fig 6: Command Processing

i. After the data extraction and knowledge understand, the command is processed to perform the task correctly, as indicated by the user.

ii. Then the interaction processes simultaneously speaks with user for effective communication.

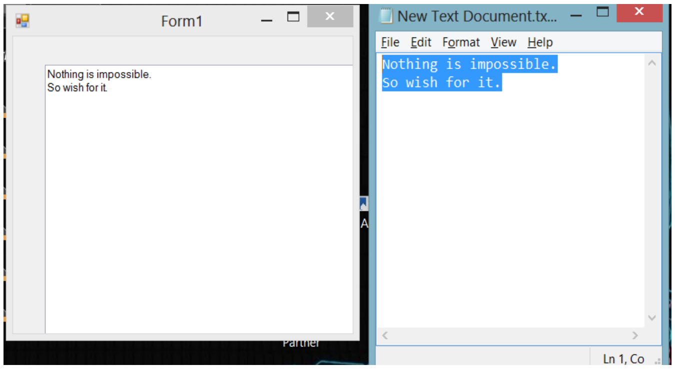
As a result of implementing these modules, we were able to come up with our application and here are a few snapshots of the same.



```
Hello
Hello Avira
Goodbye
Goodbye Avira
Close
Close Avira
Avira
thank you
Open Facebook
Allow me to perform google search
Open Gmail
Open Wikipedia
show desktop
go to desktop
Open C drive
Open My Documents
Open Downloads
Open Movies
Open Serials
What time is it
what day is it
whats the date
whats todays date
run itunes
switch to itunes
close itunes
play
pause
next song
previous song
volume up
volume down
go fullscreen
exit fullscreen
switch between windows
close window
home
end
```

Fig 7: List of Commands

These are the list of pre-defined commands that are necessary for the basic operation of proposed system.

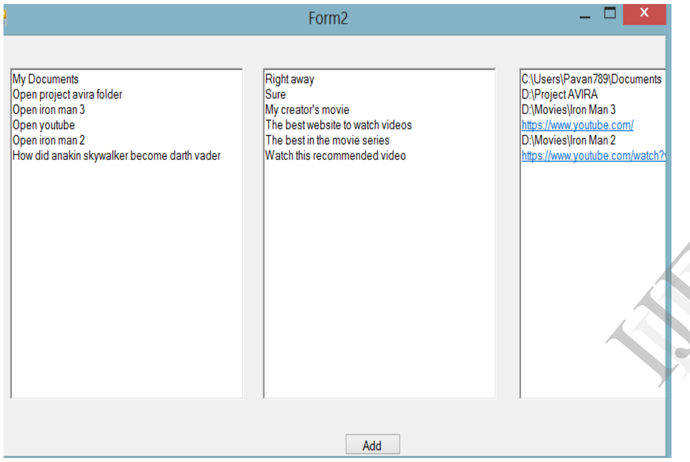


Nothing is impossible.
So wish for it

Nothing is impossible.
So wish for it.

Fig 8: Auto Read

A feature which can read out the contents of a file



My Documents	Right away	C:\Users\Pavan789\Documents
Open project avira folder	Sure	D:\Project AVIRA
Open iron man 3	My creator's movie	D:\Movies\Iron Man 3
Open youtube	The best website to watch videos	https://www.youtube.com/
Open iron man 2	The best in the movie series	D:\Movies\Iron Man 2
How did anakin skywalker become darth vader	Watch this recommended video	https://www.youtube.com/watch?

Fig 9: Custom Commands

A feature that will enable the users to define their own commands.

VI. CONCLUSION

In this paper, we have proposed a system which will enable the users to interact with the computer via speech. The speech synthesis and recognition process combined together forms an interactive system to search and control the computer via speech processing. The custom command feature is a highlight of the system as it enables the user to add their own commands.

Who can benefit from the proposed system?

- Persons with mobility impairments or injuries that prevent from keyboard access.
- Persons who have or who are seeking to prevent repetitive stress injuries.
- Persons with writing difficulties.
- Any person who wants a hands-free access to the computer.

The future enhancements include:

- Adding new synthesized voices.
- To avoid time delay for fetching grammar word.
- To implement Speech coding and Speech audio encoding/decoding.
- To make auto typing more accurate.
- To setup a Speech Authentication system.

Finally, to make the user interface as attractive and as simple as possible.

ACKNOWLEDGEMENT

We thank our guide Mrs. Snehalatha N in guiding us to finish this work on time. And also a hearty thanks to Mr. Abhilash C B for introducing us to technical platform where we can showcase our work. And last but not the least, our classmates in encouraging for us to go ahead with this work

REFERENCES

1. On the Integration of Speech Recognition and Statistical Machine Translation E. Matusov, S. Kanthak, and H. Ney Lehrstuhl f¨ur Informatik VI, Computer Science Department RWTH Aachen University 52056 Aachen, Germany {matusov,kanthak,ney}@informatik.rwth-aachen.de
2. Speech Recognition as Emerging Revolutionary Technology Parwinder pal Singh Er. Bhupinder singh Computer science &Engg. Computer science &Engg IGCE, PTU Kapurthala IGCE, PTU Kapurthala
3. Halimah B.Z.Dep. of Info. Science,UKM, Selangor, Malaysia.hbz@ftsm.ukm.my,Azlina A.Dep. of Indus. Comp.UKM, Selangor,Malaysia.aa@ftsm.ukm.my Behrang P. Dep. of Info. Science, UKM, Selangor, Mlaysia.hani_p114@yahoo.com Choo W.O.UTAR, Kampar,Perak, Malaysia.kenny@yahoo.com Voice Recognition System for the Visually Impaired: Virtual Cognitive Approach, IEEE2008
4. From AUDREY to Siri. Is speech recognition a solved problem? Roberto Pieraccini Director, ICSI the International Computer Science Institute at Berkeley
5. IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 21, NO. 5, MAY 2013 Machine Learning Paradigms for Speech Recognition: An Overview Li Deng, Fellow, IEEE, and Xiao Li, Member, IEEE
6. Toth,J. Inst. of Telecommun., Slovak Univ. of Technol., Bratislava, Slovakia Kondelova,A. Rozinaj,G. (2011) „Natural languag e processing of abbreviations? pp. 225 – 228.
7. Raitio,T., „ HMM-Based Speech Synthesis Utilizing Glottal Inverse Filtering" Vol. 19, pp. 153 – 165.
8. Kenneth Thomas Schutte "Parts-based Models and Local Features for Automatic Speech Recognition" B.S., University of Illinois at Urbana-Champaign (2001) S.M., Massachusetts Institute of Technology (2003).
9. Language development and everyday functioning of children with hearing loss assessedat 3 years of age, Teresa Y. C. Ching, Kathryn Crowe, Vivienne Martin, Julia Day, Nicole Mahler, Samantha Youn, Laura Street, Cassandra Cook, Julia Orsini,International Journal of Speech-Language Pathology Apr 2010, Vol. 12, No. 2: 124–131.