

## Speech Based Search Engine System Control and User Interaction

Prof.R.Vijayasarithi<sup>1</sup>, Prof.Dr.CManoharan<sup>2</sup>

<sup>1</sup>Asst.prof, Department of Computer Application, Annai Mathammal Sheela Engg College, Namakkal.

<sup>2</sup> principal, Annai Mathammal Sheela Engg College, Namakkal

---

**Abstract:** To make a solution to the users who cannot able to interact with the system directly without the keyboard and mouse. There is no such a need of typing by keyboard. This research will helps the users to type and work with the system on the basis of voice control using the speech synthesis and speech recognition. With this any user can access the computer without the help of mouse and keyboard. The user will have the ability to use the computer without touching it. This will be done by using the features of Artificial Intelligence and the Natural Language Processor.

Also this system will be very much useful to the differently abled persons and the visually challenged persons who are difficult to read and know the details in their education and general issues in their real world. Normally, the visually challenged persons acquire knowledge and exchange 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. The process of speech based search engine system control and user interaction will be helpful to the persons who are lacking in the computer knowledge and to interact with the computer without other person's help. This research 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 research we have introduced a new approach of using Speech recognition for educating visually challenged persons. A visually challenged person can easily interact with the computer systems without the help of other persons.

**Keywords:** speech recognition, visually challenged persons, 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)<sup>[2]</sup>. E.g. text-to-speech synthesis may use a syntactic parser on its input text and speech recognition's output may be used by e.g. information extraction techniques. Speech processing technology enables natural interaction with all kinds of computers, from cell phones, PDAs, and PCs<sup>[7]</sup>.

#### Speech processing can be divided into following categories:

- Speech Recognition - This deals with analysis of the linguistic content of a speech signal.
- Speaker Recognition - where the aim is to recognize the identity of the speaker.
- Speech Coding - a specialized form of data compression is important in the Telecommunication area.
- Voice Analysis for medical purposes, such as analysis of vocal loading and dysfunction of the vocal cords.
- Speech Synthesis - the artificial synthesis of speech, which usually means computer-generated speech.
- Speech Enhancement - Enhancing the intelligibility and/or perceptual quality of a speech signal, like audio noise reduction for audio signals.

#### Benefits of speech interaction include:

- Interact with disabilities people (e.g. Visually Challenged)
- Users can move around their homes or offices freely and still interact with computers without having to sit down and use a mouse or keyboard.
- Writing documents through speech.
- Hearing documents through speech.

This system will be very much useful for physically challenged persons and visually challenged persons who are difficult to read and know the details about their education and general issues in their real world<sup>[2]</sup>. Normally, the visually challenged persons acquire knowledge and exchange 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. The process of speech based search engine, system control and user interaction will be helpful to the persons who are lacking in the computer knowledge and to interact with the computer without other person's

help. This research is based on the fact that the computer will be able to interact with the user and fulfill their needs in the computer world<sup>[5]</sup>.

In this Research we have introduced a new approach of using Speech recognition for educating visually challenged persons. A visually challenged person can easily interact with the computer systems without any help of other persons. India is now home to the world's largest number of blind people. The 37 million people across the globe who are blind, over 15 million are from India and 15 million blind people in India, three million, 26% of whom are children, suffer due to corneal disorders with 34% of the illiterate population in the world, India has the largest number of illiterates by far.

### **1.1 OBJECTIVE OF THE RESEARCH**

The Objective of this Research to make an intelligent environment in the speech processing with the computer system by getting the Input from the user as speech input and it works for the request by the user and the process will be made easily. This is also interacts with the users who are working with the system. Also to make an innovative approach of establishing a communication with the visually challenged persons who can able to work with the system directly by typing and operating using keyboard and mouse.

### **1.2 PROPOSED SYSTEM**

The proposed system mainly focused on the process of making an interaction between the computer systems with their own speaking language. This system can directly access the Google API to recognize the human speaking sentence quickly. To the use of knowledge engineering to perform understand the human statement and perform the action. This system will be very useful to the persons who can directly speak with their computer<sup>[1]</sup>. It helps the visually challenged students to get details about their education and other details in the internet without typing in the keyboard or by other means. It will enable a person to get the basic details in any website that is displayed in the screen through voice. It will dictate the details on the website or speaks with the person who needs the details on any particular file or website or anywhere in the system. Our ultimate goal is making Robotic Intelligence operating System (RIOS) with the help of this library and Natural Language Processing for Visually challenged person. An innovative approach of developing a specific interface for the users.

### **1.3 BENEFITS IN THE PROPOSED SYSTEM**

- Users can work without the use of keyboard and mouse.
- Visually challenged persons will be able to access the computer system without other person's help.
- Users can access the system using their own voice itself.
- Also they can able to interact with the system. (i.e.) Any user can synthesize and recognize at a time.

### **1.4 FEASIBILITY STUDY**

When a request is made to develop a system, a study is undertaken to ascertain if the research is feasible whether it is worth the cost that the development of the system will involve, whether the necessary hardware and software are available if not whether it is possible to buy them and finally whether the system will be accepted and accommodated by those for whom the system is developed<sup>[3]</sup>.

After the problem is clearly understood and the solution is proposed the next step is to make the feasibility analysis. The existing system is more time consuming and inefficient. Typing all the data using speech is inconsistent. Hence to access each application using the keyboard and mouse which is time consuming and difficult to the visually challenged persons.

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<sup>[5][9]</sup>

During the feasibility analysis for this research, the following areas are considered preliminary investigations examine the research feasibility that is whether system is useful to the organization in many ways. Accurate information can be obtained at right time. The four tests of feasibility are

- Economic Feasibility
- Feasibility Analysis
- Technical Feasibility
- Operational Feasibility
- Social Feasibility

#### **1.4.1 ECONOMIC FEASIBILITY**

Economic justification includes a broad range of concerns that includes cost benefit analysis. This system 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 the information within minutes, hence the performance is improved.

Economic feasibility is checked by whether the financial benefits are exceeds the cost. This system uses only the Open source software which is economically feasible.

The proposed system will minimize the time and efforts involved in processing, hence it is economically feasible.

#### **1.4.2 FEASIBILITY ANALYSIS**

During system analysis the feasibility study of the proposed system was carried out to see, how far it would be beneficial to the organization<sup>[9]</sup>. A feasibility analysis is to test the system proposal according to its workability, impact on the organization, ability to meet user needs and effectively use of resources for the present day online messaging reliable resources of data transfer is not available and instant message is possible<sup>[8]</sup>.

#### **1.4.3 TECHNICAL FEASIBILITY**

The proposed system is said to be technically feasible if the necessary hardware and software are available for the system. The technical feasibility issues during the feasibility stage of the investigation include,

- The apt technology is adopted to develop the system
- The system can be expanded and the organization should provide the sufficient equipment to develop the system.

The proposed system is technically feasible if the following basic requirements are available,

- Windows 2000 and above
- New library RIOS.DSP will be used. It will integrates the Synthesis and Recognition process easily,
- Advanced domain such as Natural Language Processing will be used.

7

#### **1.4.4 OPERATIONAL FEASIBILITY**

Operational Feasibility is the ability and desire of the management, users and others to use and support the proposed system.

The proposed system offers greater levels of user friendliness combined with effective services to the users. 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<sup>[10]</sup>. Thus the proposed system increases the efficiency and provides more benefits to the organization.

#### **1.4.5 SOCIAL FEASIBILITY**

Social Feasibility is the ability to make changes in the society or the country that the research will be very helpful in its development of the education to the differently abled persons who cannot able to complete their basic education and their knowledge from the internet 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.

### **2.1 PROBLEM DEFINITION**

Speech recognition applications include voice user interfaces such as voice dialing (e.g., "Call home"), call routing (e.g., "I would like to make a collect call"), domotic appliance control, search (e.g., find a podcast where particular words were spoken), simple data entry (e.g., entering a credit card number), preparation of structured documents (e.g., a radiology report), speech-to-text processing (e.g., word processors or emails), and aircraft (usually termed Direct Voice Input).

Synthesized speech is created by concatenating pieces of recorded speech that are stored in a database. Systems differ in the size of the stored speech units; a system that stores phones or diaphones provides the largest output range, but may lack clarity. For specific usage of domains, the storage of entire words or sentences allows high-quality output. Alternatively, a synthesizer can incorporate a model of the vocal tract and other human voice characteristics to create a completely "synthetic" voice output. These recognition and synthesis process can be combined in a single program and this library provides a thin layer to combine these processes together to form both interaction and recognition process.

### **2.2 OVERVIEW OF THE RESEARCH**

As the Web transforms from a text only medium into a more multimedia rich medium the need arises to perform searches based on the multimedia content. In this paper we present an audio and video search engine to tackle this problem. The engine uses speech recognition technology to index spoken audio and video files from the World Wide Web, when no transcriptions are available. If transcriptions (even imperfect ones) are available we can also take advantage of them to improve the indexing process.

Our engine indexes several thousand talks and news radio shows covering a wide range of topics and speaking styles from a selection of public web sites with multimedia archives. Our Web site is similar in spirit to normal web search sites; it contains an index, not the actual multimedia content. The audio from these shows suffers in acoustic quality due to bandwidth limitations, coding, compression, and poor acoustic conditions. Our word-error rate results using appropriately trained acoustic models which shows remarkable resilience to the high compression, though many factors combine to increase the average word-error rates over standard broadcast news benchmarks. We show that, even if the transcription is inaccurate, we can still achieve good retrieval performance for typical user queries (77.5%).

Speech Recognition is technology that can translate spoken words into text. Some SR systems use "training" where an individual speaker reads sections of text into the SR 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 accurate transcription. Systems that do not use training are called "Speaker Independent" systems. Systems that use training are called "Speaker Dependent" systems<sup>[2][6]</sup>.

The term voice recognition refers to finding the identity of "who" is speaking, rather than what they are saying. Recognizing the speaker voice recognition can simplify the task of translating speech in systems that have been trained on specific person's voices or it can be used to authenticate or verify the identity of a speaker as part of a security process. "Voice recognition" means "recognizing by voice", something humans do all the time over the phone. As soon as someone familiar says "hello" the listener can identify them by the sound of their voice alone.

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. After those process completed the synthesis can speak the data to the user. The data flow diagram of this research is given in the figure 2.2 as data flow diagram.

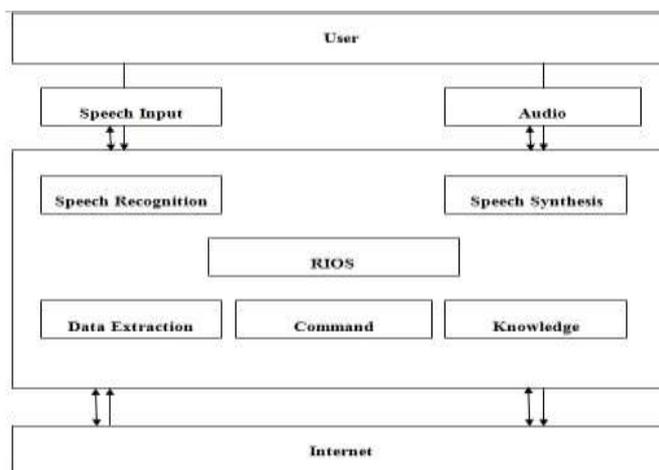


Fig 2.2 Data Flow Diagram

## 2.3 MODULE DESCRIPTION

### 2.3.1 Speech Synthesis

### 2.3.2 Speech Recognition

### 2.3.3 Data Extraction and Knowledge Understand

### 2.3.4 Command Processing

#### 2.3.1 SPEECH SYNTHESIS MODULE

- The text in any file or a source can then converted to the voice through synthesizer.
- It will convert the digital string to the respective voice through the US or UK Dictionary which we were used.
- Then it will dictate the related word in the dictionary.

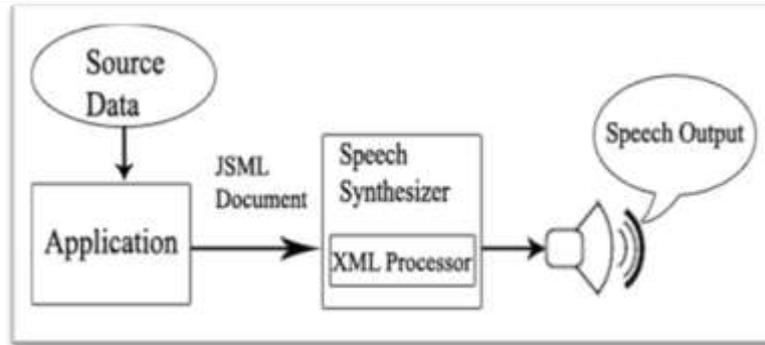


Fig .2.3.1 Speech Synthesis

A speech synthesizer converts written text into spoken language. Speech synthesis is also referred to as text-to-speech (TTS) conversion. The proposed form of Speech Synthesis is shown in the Fig 2.3.1

### 2.3.2 SPEECH RECOGNITION MODULE

- Speech recognition is the process of converting spoken language to written text or some similar form.
- To get the command or voice from the user through microphone and recognize the acoustics.
- Then it will check the clarity of voice and pass to the Extraction part.
- We use the globalized grammar database using “Google API” via internet to perform Speech Recognition.
- Speech recognition is the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words.
- The recognized words can be the final results for linguistic processing in order to reply the persons.

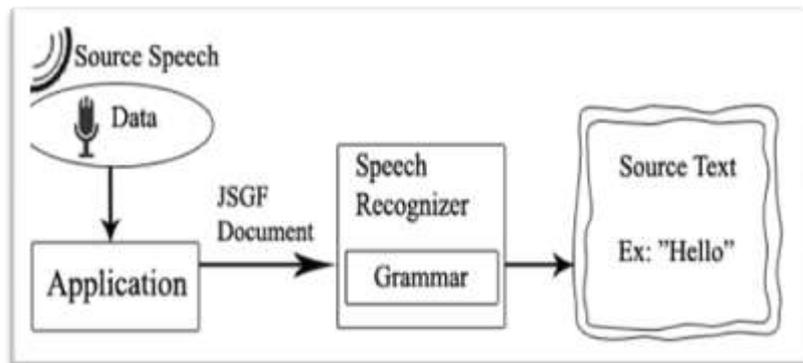


Fig .2.3.2 Speech Recognition

Speech recognition is the process of converting spoken language to written text or some similar form. Fig. 2.3.2 shows the process of Recognition.

The basic characteristics of a speech recognizer supporting the Java Speech API are:

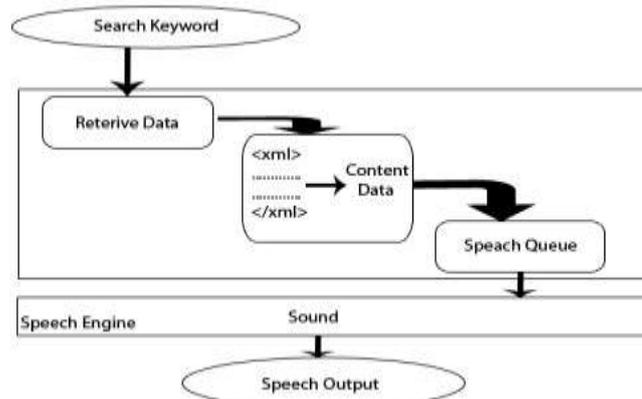
1. It is mono-lingual: it supports a single specified language.
2. It processes a single input audio stream.
3. It can optionally adapt to the voice of its users.
4. Its grammars can be dynamically updated.
5. It has a small, defined set of application-controllable properties.

The major steps of a typical speech recognizer are:

1. Grammar
2. Signal processing
3. Phoneme recognition
4. Word recognition.

### 2.3.3 DATA EXTRACTION AND KNOWLEDGE UNDERSTAND MODULE

- The voice signal can be then converted to the digital format based on the Parser.
- It will convert the digital string to word.
- Then it will process the command given by the user and sent to the synthesizer.
- It redirects to synthesizer.

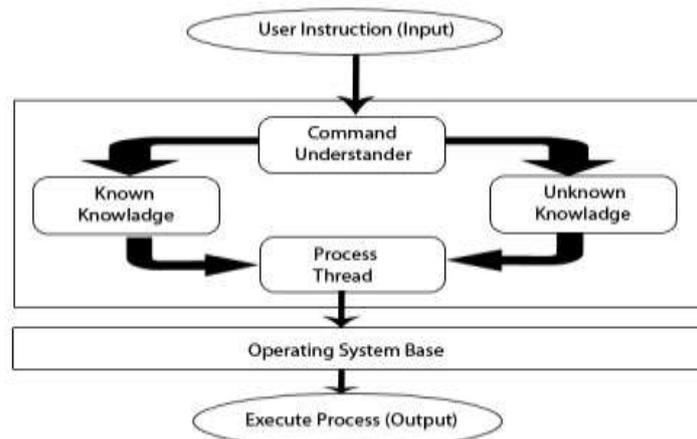


**Fig .2.3.3 Data Extraction and Knowledge Understand**

- Data Extraction performs the role of getting the raw factor of the user input from the internet. These raw factor (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 Understand module will understand the user’s keyword to perform both searching process as well as command processing methods.

**2.3.4 COMMAND PROCESSING**

- After understand the knowledge it can do the correct task, what the user need.
- Then the interaction processes simultaneously speak with user for effective communication and verify the tasks.



**Fig 2.3.4 Command Processing**

After understanding the knowledge, it does the correct task, what the user needs. Then the interaction processes simultaneously speak with user for effective communication and for verifying the tasks.

**II. Input Design**

The input design is the link that types the information system to the world of its user. It consists of developing specifications and procedures for data preparations and to input into the computer for processing. Simply we can Input Design is the process of converting the user-generated inputs into computer-based format.

**THE IMPORTANT FEATURES ARE AS FOLLOWS**

- The input screen is not over crowded, as the user can understand a piece of information from the screen. Instead more screens are provided to understand the data entry process.
- The input validation is being done at program level to check errors and help messages to be provided to effectively enter data into data entry screen in the software such as password checking, input checking etc.,
- As human is prone for errors and mistakes, confirmations for critical data entries are being provided. In the research scheduler confirm messages are provide at such a critical areas like deletion or existing without saving etc.

### 3.2 OUTPUT DESIGN

All the output screens are information and interactive in such a way that the user can fulfill his/her requirements. Output is information delivered to users through the information system. All the process of quoting of tender, acquiring business, packing of goods, scheduling the transportation and follow up till delivery etc., are shown in the output screen. Therefore an effective output design is an important feature of design specification. Since useful output is essential for gaining user acceptance of the system, the system analysis should try and follow the objectives, which are useful for designing acceptable output.

Most commonly outputs are printed on paper or displayed on screens. Outputs often have to be distributed to the user. The increase in online, screen displayed output, that is personally accessible, has cut down on the problem of distribution. Appropriate distribution is still an important objective for the system analyst.

### III. CONCLUSION

This research represents an interactive communication between human and computer. The speech synthesis and recognition process combines together and form an interactive system to search and control the computer via speech processing. This could be done by using the concept of Natural Language Processing in Artificial Intelligence. These processes can help the human quickly to access the content from the internet and makes to listen easily. So the human no need search and read the content from the internet. And he/she can easily control the system by using the system without using the keyboard and mouse.

#### 3.1 FUTURE ENHANCEMENTS

The future goal is make it up to simple compare to this current system. And additionally will be implement Voice analysis, Speech enhancement, Speech coding, Speech audio encoding/decoding, Speaker analysis. The following features will be under implementing our ultimate goal to make Robotic Intelligence operating System (RIOS) with the help of this library and Natural Language Processing for physically challenged person.

- Adding new synthesized voices.
- Creating new library for create this synthesise voice.
- Automatic grammar generator for a large JSGF.
- To avoid time delay for fetching grammar word.

Portably imported to any application for interaction environment.3

### Knowledgements

The author prof.R.vijayarathi, wish to thank our chairman Mr.M.Mani, principal prof.Dr.C.Manoharan, Hod Prof.M.Geetha and staff members of “Annai Mathammal Sheela Engineering College Namakkal, TamilNadu” for his continuous encouragement for carrying over my research works.

### References

- [1]. Alani, H. Southampton Univ., UK Sanghee Kim ; Millard, D.E. ; Weal, M.J. ; Hall, W. Lewis, P.H. ; Shadbolt, N.R. (2011), Automatic ontology-based knowledge extraction from Web documents’, Vol. 18, pp. 14 – 21.
- [2]. Dragon Naturally Speaking <http://www.scansoft.com/naturallyspeaking>
- [3]. Fetch Labs., Fetch Technol., El Segundo, CA, USA(2011), ‘Automatic Speech Recognition Based on Non-Uniform Error Criteria’ **Vol.** 20, pp. 780 – 793.
- [4]. FreeTTS <http://freetts.sourceforge.net>
- [5]. JavaTM Speech API Programmer’s Guide <http://java.sun.com/products/java-media/speech/forDevelopers/jsapi-guide/>
- [6]. Language development and everyday functioning of children with hearing loss assessed at 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.
- [7]. Roe, David B., and Jay G. Wilpon. Whither speech recognition: the next 25 years. *IEEE communications magazine*, v. 31, Nov. 1993: 54-62.
- [8]. Raitio, T., ‘HMM-Based Speech Synthesis Utilizing Glottal Inverse Filtering’ **Vol.** 19, pp. 153 – 165.
- [9]. Toth, J. Inst. of Telecommun., Slovak Univ. of Technol., Bratislava, Slovakia Kondelova, A. Rozinaj, G. (2011) ‘Natural language processing of abbreviations’ **pp.** 225 – 228.
- [10]. 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).