

Voice Interactive Browser

¹Ashwini M.D, ²Bhavya V, ³Divya K, ⁴Preethi S, ⁵Asha G.R
^{1,2,3,4} B.Tech, ⁵Assistant Professor

Computer Science Engineering Department
BMS College Of Engineering, Bengaluru, India.

Abstract-- Today, Web Browser is satisfying the need of an internet user to access information on the web; it is basically done through keypad action. But for the today's technology, this is not flexible for partially sighted people and also for those who wish to browse with hands free. However the advancement in technology enables us to ease the work of the user with speech interactions. A Speech Browser enables us to interact through voice with internet. Through the conversion of text-to-speech and speech-to-text, the voice browser renders the information from the World Wide Web in an audio format so that it facilitates to browse any information hands freely from anywhere, any place and at any time. However, voice browser is a Generic Application interface which is built using android application developing tools such as Android SDK, Android Manager, Speech Analyser and Speech Recognizer and editor like Eclipse. This facilitates the users to access an internet by interacting with their devices as if they are communicating with one another.

Index Terms--Internet, Browser, Voice, Speech Analyser, Speech synthesizer, Android Application.

I. INTRODUCTION

Many browsers that exploit voice as input and using speech synthesis it exploits pre-recorded sound for output. By using speech synthesis and pre-recorded voice material to present the contents of web pages with the variety of aural effects that is used to give different emphasis to list hyperlinks, headings etc. [1]. The speech technology has advanced so much so that it is promising to provide interaction between human and computer. We have to choose an application very carefully to engineer while considering various human factors instead of finding keys, speech provides a very interesting multi-mode interfaces. Speech is a very advantageous over other conventional media, for many easily imaginable applications (example interacting with the system while one's hand is occupied by some other important tasks or if one is physically challenged). However, it would be more natural and comfortable if speech interaction is provided instead of keyboard and mouse interaction [2]. The speech recognition for input and speech synthesis for output provides an easy to use interface that is ideal to access the internet [3].

Based on those issues on context based information the major cases is on retrieval in which the visual information is focused. The reason behind the focus on visual information is because of the availability of plenty of images and videos on internet which includes real world applications. However, the audio is a kind of important media from the human perception for carrying information. This is because human has not only visual sense, but also includes an auditory sense. They obtain lot of information from auditory media from day-to-day life through radio, television, music etc. [4]. The process where in the computer is made to speak is through the technology called TTS (text-to-speech) speech synthesizer which converts the audio input into text form. The system speaks out only particular words that answer to the user [5]. One of the oldest and natural way of communication is speech, the communication among are carried out through human and human interface that includes speaking, and listening to one another. According to this system, it can identify the command and give output through voice [6]. Voice recognition helps to map voice commands with its actions. This can be achieved by converting speech to text. The input command will try out to find a match with the trained commands and maps to the results [7]. It has been a dream of people to be able to communicate with all kinds of devices using speech as input and output. Command and control is simplest way to interact in terms of interaction model. The basic essential requirement is CFG (context free grammar) that allows particular rules that defines a set of allowable utterances [8]. One of the flexible platform for speech was mobiles, over other input modalities this is due to the fact that they can access the information easily through descriptions. And mobiles also act as ambient medium so that users can perform other task instead of performing complete attention.

A voice browser is an interface that provides interpretation of the markup language and produce output in the form of voice. It also provides interface to the PSTN (Public Switched Telephone Network). As the name suggests, the system delivers the voice dialogues with pages such as visual web pages which in turn deals with HTML pages. From users perspective the goal is to facilitate devices which do not have full browsers or even screens to provide them, like visual browser and related technology [9]. The existing system is of visual oriented browser since interactions requires full attention of users, the aged people who do not have good eye sight to read from the screen, people who have arthritis will not be able access internet efficiently. To overcome this telephone is considered as an interface for communication. The user and the browser using voice and telephone require very less attention from the user to operate. Hence this can be considered as ideal communication tool for environments where user is engaged in other important tasks [10].

Millions of people in their daily activities use the web browser that has become popular source of information. The browser that currently available is GUI browser. The people who are physically handicapped cannot use the normal browsers which are not capable of providing voice commands. One can access the complete browser using Vocal access network. Voice XML, JAVA speech

API, Microsoft speech software development tool kit is the different speech conversion methods which are available [11]. This system is based on Microsoft speech SDK platform which is used for the conversion of voice to text.

We got an overview of a browser that facilitates voice interactions and what are the techniques that it users to provide a voice features. Section 2 describes the related works that are held on the topic. Section 3 design it describes the structure of the proposed model.

II. RELATED WORKS

Usually a Web browser requires physical interaction with devices that are equipped with screen and GUI (Graphical User Interface). Our system allows browsing the internet through voice commands as inputs by developing VUI (Vocal User Interface). This system implements CE (Context Extraction) algorithm to analyse, to classify and to return specified part of the Web pages to users. Speech capabilities on the user's machine are accessed over the internet by the Microsoft Speech SDK which allow applet to be transmitted [12]. World Wide Websites and new ways of browsing will be enhanced by this. Control browser, fill out forms, Control Applets can be done by Speech recognition and WWW/Internet can also be enhanced. Informing the used the progress of applets and web pages can be done by speech synthesis. This improves the browsing time by minimizing the amount of audio which is sent across the internet.

I. Speech Recognition Systems

The motive of speech recognition for a device is to be able to "Hear", "understand" and "act upon" spoken information^[13]. The translation of spoken words into text is called Speech recognition. When an individual speaker reads a part of text into some SR systems use "Training". The specific voice of the person will be analyzed by these systems and fine tune the recognition of that voice which is more accurately [14].

Types of Speech Recognition system:

Speaker dependent:

Many number of voice recognitions systems are available in market. The thousands of words can be recognized by most powerful system, generally extended training is required. The system are easier to develop and cheaper to buy, but not flexible as speaker adaptive or speaker independent.

Speaker independent:

This system can be operated for any speaker of a particular type. These systems are more expensive and very difficult to develop. There is no training involved so, it can recognize any one system.

Speaker Adaptive:

This is currently emerging third variation. It usually begins with the speaker independent model and these models are adjusted more closely to each individual in the brief training period.

II. Topology of speech recognition system

The following are the topology of the speech recognition system; Speaker Dependent, Speaker Independent, Violated word recognizer, connected word system and Spontaneous recognition. Initially, a system should be trained according to the user voice, but the system doesn't require training the system by user, hence they are developed to support from any user or speaker. Even though only one word is accepted once at a time and their recognition system allowed to the user to speak continuously naturally [15]. With the planned speech and short paired system will allow speaker to speak slowly and each word distinctly and allows speaker to speak spontaneously.

III. Approaches to speech recognition.

- Acoustic Phonetic approach.
- Pattern Recognition approach.
- Artificial Intelligent approach.

Acoustic Phonetic approach: It is also called as rule-based approach. To search, it includes the knowledge of phonetic and linguistic approaches for the process of searching. To define or express anything or everything, the rules have been that helps in decoding the architecture of black board. Because of the difficulty in express rules, it has a poor performance for improving this system.

Pattern Recognition approach: Training a speech patterns and the recognition and the pattern, are the two methods of comparing a pattern. The test patterns are defines based on the sequence of measurement made by input signal in the parameter measurement phase. Some reference pattern are compared with the un-known test pattern and the test pattern versus reference pattern are measured for similarities, which matches the similarity scores of the unknown test pattern from phase of pattern classification.

Artificial Intelligent approach: This is a combined approach of acoustic phonetic and pattern recognition approach. To classify sounds in Artificial Intelligence (AI), neural network is implemented by expert system [16]. From the problems the knowledge are taken from variety of sources and are incorporated and compiled as a basic idea.

IV. Smart phone technology:

Today’s smart phones do much functionality, but natural ways of interacting with smart phone remain same. Some wants to develop next version of existing voice recognition is the next big thing of the IT industries. It means recognizing human speech and converting it to text. User can command their smart devices via speech. Speech recognition technologies have rapidly evolving because of the proliferation of portable computer terminals like smart phones [17]. So far, the information has primarily been used by the one who connects to web via a traditional computer accessing internet in which large amount of information on World Wide Web is available on fingertip. However advance in technology like wireless communication, speech recognition, speech synthesis technology and voice based browser made it possible to access internet hands freely anywhere, any place and at any time. The speech recognition for input and synthesis for output provides an ease access to internet that are of hands free operation. Web browsing operation which is in smart phone basically accepts voice commands as an input as per the advance technology [18].

An efficient speech recognition system use different techniques such as Mel Frequency Cepstrum Co-Efficiency (MFCC), Vector Quantization (VQ), and Hidden Markov Model (HMM). The idea behind the research on speech recognition is human interaction with machine. By means of algorithm which is implemented in android devices are used in the process and technology of automatic speech recognition for converting speech signal into sequence of words [19]. Presently speech understanding systems are capable of vocabularies of thousands of words in operational environment through speech as input.

III. DESIGN

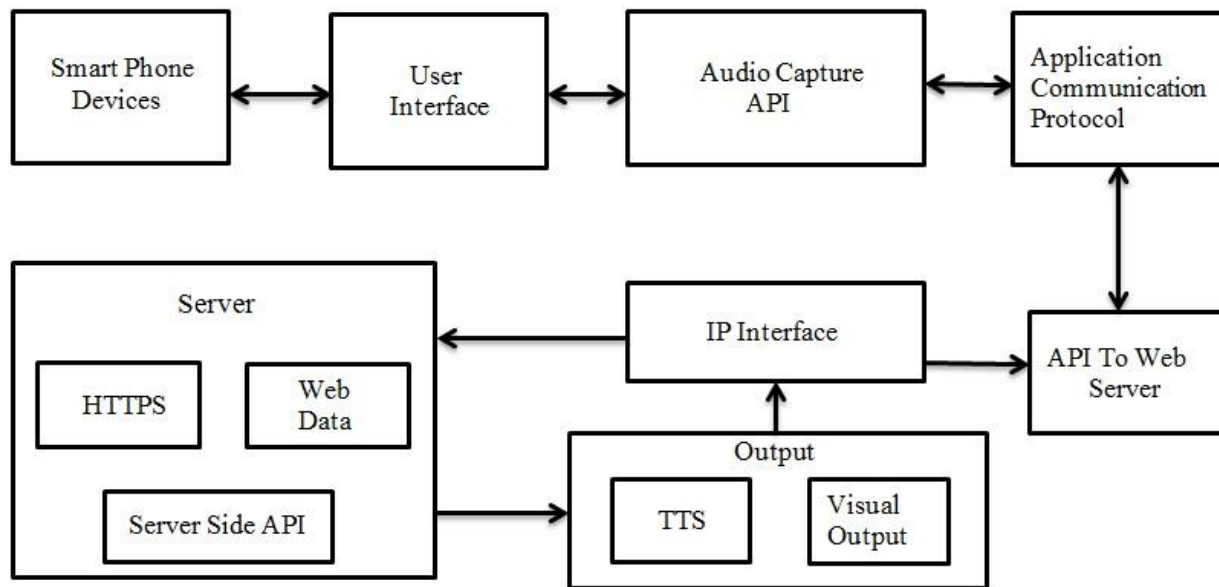


Fig. I

Figure 1 depicts the data processing flows as per the above design which describes that, the user sends his request from his smart phone device through user interface that is in turn called an application. The application accepts the input and this will be recognized by audio capturing API to analyse the audio. Later application communication protocol precedes the analyzed data to IP Interface which searches the information in the server for the input data given as a request. After finding the data it delivers it for converting text to speech via application control protocol it reaches user interface and been delivered to user in voice format.

IV. CONCLUSION

Using a web browser with voice interactions will be helpful for the partially sighted people and also for the people who desires to lead their life as per current trends in technology. It eases the work for the people who are busy in multiple tasks and easy to access. This can also involve in the minimization of time consumption for typing and reduces strain given to eyes by focusing on display screen.

We provided detailed information of the application and its features. The different knowledge and information about the technology required for the development of the application may be presented. This application user interface can be developed using Android SDKs.

V. ACKNOWLEDGMENT

We are grateful to BMS College of Engineering for having provided us with the facilities needed for the successful completion of this work. The work reported in this paper is supported by the college through the TECHNICAL EDUCATION QUALITY IMPROVEMENT PROGRAMME [TEQIP-II] of the MHRD, Government of India.

REFERENCES

- [1] Dave Raggett and Ben-Natin, Voice Browsers, W3C-Voice-19980128 28 January 1998.
- [2] Priyanka N. Waghela, Ankitha A. Shirsekar, Sandhyarani D. Narayankar and Vijaya Pinjarkar , Speech enabled browser using Microsoft speech API, Proceedings of National Conference on New Horizons in IT – NCNHIT 2013 ISBN 978-93-82338-79-6.
- [3] Raji Agarwal, Yeshwanth Muthuswamy and Vishu Viswanathan, Voice Browsing the Web for Information Access, Media Technologies lab TX 75265.
- [4] Li GuoHui, WU DeFeng and ZHANG Jum, Concept Framework for Audio Information Retrieval: ARF,J. Computer Science and Technology Vol. 18, No. 5, pp. 667-673 September 2003.
- [5] Deepa V.Jose, Alfateh Mustafa, Sharan R, A Novel Model for Speech to Text Conversion, International Refereed Journal of Engineering and Science (IRJES) Volume 3, Issue 1 (January 2014), PP. 39-41.
- [6] Preethi Saini and Praneeth Kaur, Automatic Speech Recognition, International Journal of Engineering Trends and Technology - Volume4 Issue2- 2013
- [7] Stuart Goose and Carsteen Moller, A 3D Audio Only Interactive Web Browser: Using Spatialization to Convey Hypermedia Document Structure, Siemens Corporate research Princeton, NJ08504.
- [8] Murad Al-Rajab and Haifaa Kattan, Toward a New Algorithm for Hands Free Browsing, International Journal Of Computer science and Security (IJCSS), Vol(1): Issue (3): 2011.
- [9] The New Era of Browsing-Voice Browsing by Khusbu, Manika Kapoor and Ayesha Tafsir at International Journal of Engineering and Computer science Vol. 1, Issue 2 November 2012 Page No. 59-62 ISSN:2319:7242.
- [10] Josiah Poon and Chris Nunn, Browsing the Web from A Speech-Based Interface , Proc. Of Human-Computer Interaction (INTERACT'01), Vol. 09, No. 13, pp. 302-309 July 2001, Tokyo, Japan.
- [11,12] Gobbur Uday, Kale Ketan, Umbarkar Dipak and Neharkar Swapnil , Voice Based Internet Browser , International Journal of Computer Application Vol. 66, No. 8 March 2013 ISSN 0975-8887.
- [13] Arnon Amir, Savitha Srinivasan and Alon Efrat , Search the Audio, Browse the Video- A Generic Paradigm for Video Collections , EURASIP Journal on Applied Signal Processing 2003:2, 209-222 Copyrights 2003 Hindawi publishing Corporation.
- [14] Ummuhansyifa U and Nizar Banu P.K., Voice Based Search Engine and Web Page Reader, International Journal of Computational Engineering Research (IJCER) May 2013.
- [15] Michael K. Brown, Stefen C. Glinski, Bernard P. Goldman and Brian C. Schmult, Phone Browser: A Web-Content-Programmable Speech processing Platform , Lucent Technologies Inc. Murray Hill, NJ 07974.
- [16] Pipe Beach AB, Position Paper-Standards for Voice Browsing ,Date 1998-10-09.
- [17] Renu Tarneja, Huma Khan, Prof. R. A. Agrawal & Prof. Dinesh. D., Voice Commands Control Recognition Android Apps, International Journal of Engineering Research and General Science Volume 3, Issue 2, March-April, 2015 ISSN 2091-2730
- [18] Suma Swamy1 and K.V Ramakrishnan, An Efficient Speech Recognition System , An International Journal (CSEIJ), Vol. 3, No. 4, August 2013
- [19] Muhirwe Jackson, Automatic Speech Recognition:Human Computer Interface For Kinyarwanda Language,Computer Science of Makerere University, August 2005
- [20] M.A.Anusuya, Speech Recognition by Machine: A Review ,(IJCSIS) International Journal of Computer Science and Information Security,Vol. 6, No. 3, 2009