International Journal of Future Generation Communication and Networking
Vol.14, No. 1, (2021), pp. 2510–2516

A Voice Controlled Web Application

Prof. Praful Barekar , Pankaj Shamnani , *Siddhesh Patki , Gurutva Patle , Saurabh Nitnaware , Prof. Bhagyashree Ambulkar

Department of the Computer Technology , YCCE, Nagpur
Department of the Computer Technology , YCCE, Nagpur
Department of the Computer Technology , YCCE, Nagpur
Department of the Computer Technology , YCCE, Nagpur
Department of the Computer Technology , YCCE, Nagpur
Department of Computer Science , G H Raisoni Institute of Engineering and Technology, Nagpur
praful.barekar20@gmail.com , pankajshamnani1000@gmail.com , siddhpp22@gmail.com , gurutavpatle@gmail.com , Snitnaware11@gmail.com , Bhagya.ambulkar@gmail.com

Abstract

The way human interacts with computers has been totally changed by the voice controlled systems. Voice-activated apps allow users to interact with the computer without using their hands. Voice-activated apps have become much more efficient and widely accepted by people. Modern web-based applications provides user friendly and user interactive interfaces. However, in few cases, for example: people with visual impairments cannot experience the functionality of such applications. The voice-controlled web application empowers users and provides voice as a means of communication with the system. In this article, we provide voice-activated web applications, how they work, benefits, and applications.

Keywords: *Speech recognition system, voice controlled system, speech to text conversion (STT), text to speech conversion (TTS).*

1. INTRODUCTION

The traditional method of interacting with computers using the keyboard and mouse does not meet the requirements to bear all users. Therefore, it becomes important to create websites with support for all types of users. Our work is focused on creating such a voice-activated web application that allows all users to interconnect with the system hands-free. This type of web application will allow users to access the entire website without using hands. You can perform various operations on a website, such as scrolling, page redirection, filling in data, and more.

Various web technologies such as screencasts, browsers, and magnification methods have been produced for visually impaired people over the last few years. These applications helping many users for reading, voice recognition, or providing ways for magnification of screen to understand the content of a website.

Voice controlled websites can give users the flexibility to choose how they interact with computers. They also improve the user experience by providing an additional method of interacting with web applications. It improves the internet experience and allows users to communicate their instructions using voice.

In this article, we introduce voice-activated web applications using Web Speech API. The Web Speech API provides two separate functionalities - speech recognition (TTS) and speech synthesis (STT) [1]. Our website takes an input as a voice command, converts it into a meaningful text, and then performs various

ISSN: 2233-7857IJFGCN
Copyright ⓒ2021SERSC

operations including searching, page redirection, scrolling, data filling and much more.

BACKGROUND

The idea of human interacting with a machine using speech has started a revolution on voice controlled systems. This has greatly improved human-computer interaction. Voice-activated apps can be used to solve difficulties that are being faced by people having disabilities.

Voice controlled systems are based on the principles of statistics. The working of ideal speech recognition system is shown in Fig. 1. The input data to the application is the speaker's speech, which is then transformed into a speech waveform and then it is passed to a feature extraction module which is used to remove noise and finally it is converted to a proper presentation format. The signal generated is then provided to a search engine that is related with sound patterns. The information about phonetics and environmental diversity, and gender and dialectal differences between speakers has been included by the sound model. The language model includes semantic knowledge. [5]

A voice controlled or speech recognition system is capable of handling many of the ambiguities corresponding with speaker quality, speed, and speech style; recognize main speech areas, possible and probable words, unspecified words and variations related to grammar ; [3] process noise interference and foreign accents and calculate estimates of the reliability of the results.

It is the primary function of the search engine component. The adaptation block is used to change the output of sound or language models to improve overall performance[7].

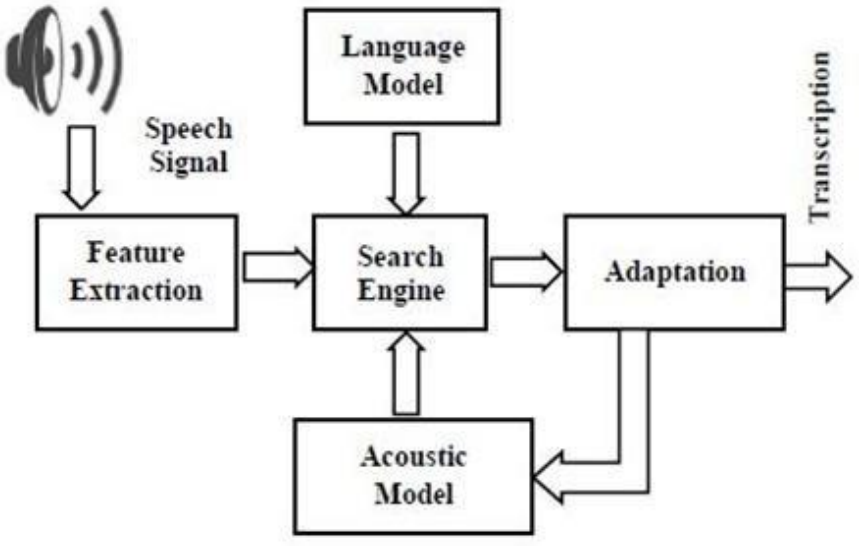

Figure 1. Architecture of speech recognition system

2. TAXONOMY OF WEB SPEECH

The capabilities of the speech recognition system has been identified by the various parameters. In Table 1, these parameters are grouped. The categorization given here are based on typical design

ISSN: 2233-7857IJFGCN
Copyright ⓒ2021SERSC

assumptions for a voice recognition system that might be related to a specific task. These parameters are somehow fixed in the system.

Table 1. Taxonomy Of Web Speech Table

Parameters	Easy Task	Difficult Task
Vocabulary Size	Small	Unlimited
Speech Type	Isolated Words	Continuous Speech
Speaker Dependency	Speaker Dependent	Speaker Independent
Grammar	Strict Syntax	Natural Language
Training Method	Multiple Training	Embedded Training

2.1. Vocabulary size

A vocabulary size is very important to the recognizer and its performance is directly related to the vocabulary size. The small vocabulary has an order of 100, medium has an order of 1000 and large has an order of 10000. A recognizer with small vocabulary can only recognize 10 digits.[2]
Small – 1 to 100 words or sentences
Medium – 101 to 1000 words or sentences
Large – 1001 to 10,000 words or sentences
Very-Large –10,000 words or sentences or more

2.2. Speaker dependency

The voice recognition job may or may not depend on the speaker. Regardless of the speaker, internal recognition is more difficult because speech representation should be global to surround all the probable nuances, and accurate to differentiate between different vocabulary words. [4]

For a system which is totally dependent on speaker, learning is usually done by the user, but for applications with a large amount of words, it takes more time for the single user. In those cases, an intermediate method is used, known as speaker adaptation. The speaker independent models have been filled with the system and then slowly interprets to specific aspects of the user. [2]

ISSN: 2233-7857IJFGCN
Copyright ⓒ2021SERSC

2.3. Grammar

Voice recognition systems are thoroughly equipped with a many language knowledge to minimize the number of words to choose from. This can range from strict syntactic rules, where in words that can follow each other, are determined by specific rules, to probabilistic linguistic models, which considers the probability of an output based on statistical knowledge of the language.

2.4. Training

There are different ways to train an automatic speech recognition system. When every single word of the vocabulary trained many times, then it might be able to build vigorous word model, and hence the better performance must be expected. Many systems can be trained with only one example of each word, or even without it. A training sessions per each word is called as the training passes.

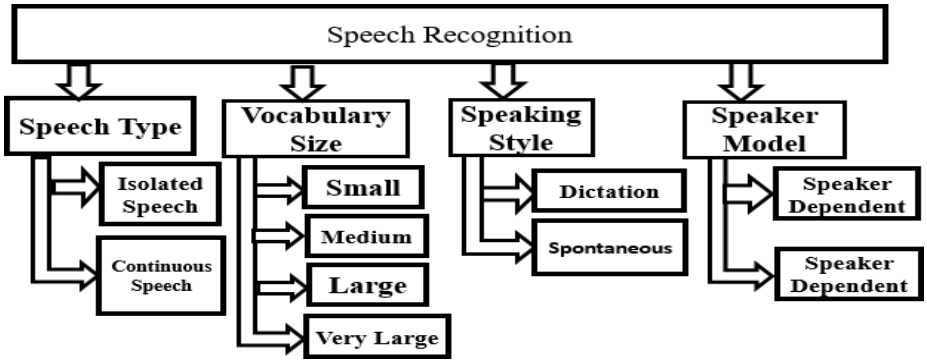

Figure 2. Classification of speech recognition system

3. ARCHITECTURE

The architecture of our web application uses the Web Speech API through the microphone of the user's device, which takes the user's speech as input and uses speech-to-text (STT) to translate it to an appropriate text.[8] As every web application keep a set of predefined tasks, the converted text will look for secret code or keyword present in the database and run the corresponding functions attached to that keyword if found. If the keyword is not found in the database, it will use Text-to-Speech (TTS) to convey the information / problem encountered. [6]

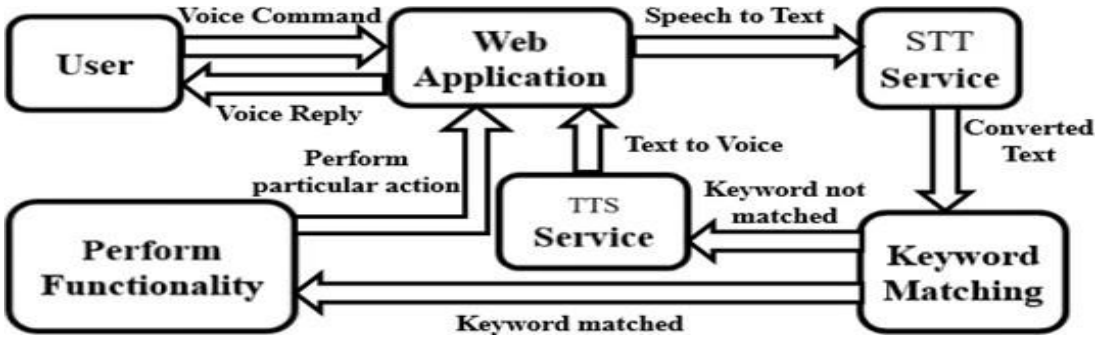

Figure 3. Architecture of speech recognition system

ISSN: 2233-7857IJFGCN
Copyright ⓒ2021SERSC

The architecture of our web application uses the Web Speech API through the microphone of the user' device, which takes the user's speech as input and uses speech-to-text (STT) to translate it to an appropriate text.[8] As every web application keep a set of predefined tasks, the converted text will look for secret code or keyword present in the database and run the corresponding functions attached to that keyword if found. If the keyword is not found in the database, it will use Text-to-Speech (TTS) to convey the information / problem encountered. [6]

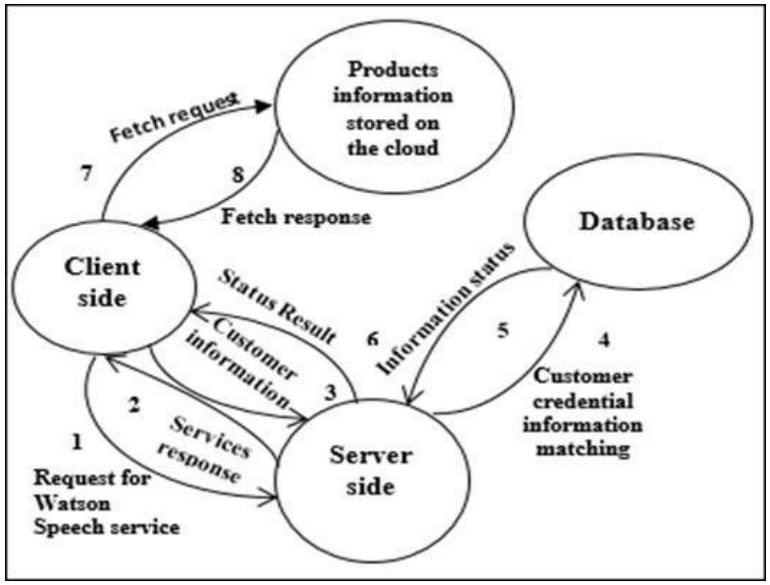

Figure 4. FlowChart of speech recognition system

The performance of the Web Speech API depends upon the dimensions of the terminology supporting at a particular level of precision, complication, and speed of processing.[9] We have implemented a continuous voice recognition type so that the user can continuously send a request to the web application. Another aspect of Web Speech API is that it does not depend on the speaker, meaning the web application supports speeches and pronunciation of words globally, making it useful for a larger audience.

4. IMPLEMENTATION

The architecture of our web application is shown in the figure above, now we will talk about the implementation of our web application.

In our web application, we used the speech-to-text (STT) and text to speech (TTS) functions of the Web Speech API. Our web application has a) a client side that will render the entire user interface (UI) where we used HTML, CSS and BootStrap, b) a server side or backend, where we used PHP and JavaScript, c) the database we used a MySql database for our web application, which will store all the information in our web application, including the keywords that will be used to map keywords to speech-to-text output. The domain of our site is grocery, this is a grocery site. [10]

Information flow: When the user says something, the speech-to-text (STT) module converts that speech to the appropriate text. Since each web application has its own predefined operations, the converted text will search the database for the keyword. If a keyword is found, it will perform the

ISSN: 2233-7857IJFGCN
Copyright ⓒ2021SERSC

corresponding JavaScript operations attached to that keyword, and text-to-speech (TTS) converts the text that matches that keyword into speech that will speak it to the user as output.

If the keyword is not found in the database, then text-to-speech (TTS) converts the common phrase present in the database to speech and speaks it to the user as output.

The Web Speech API will manage various functions of the website, including searching, scrolling, redirecting, populating data, and more.

5. RESULT

In this article, we introduced voice controlled web applications with the help of Web Speech API. The Web Speech API provides two functionalities - speech recognition and speech synthesis (also known as text to speech or tts).It opens up new possibilities for research and control Mechanisms [1]. Our web application takes speech commands as input, converts it into text and then performs a wide variety of operations including searching, scrolling, data filling, page redirection and much more.

6. CONCLUSION

We have implemented a voice-activated web application using the Web Speech API, which basically does two operations: speech recognition (STT) and speech synthesis (TTS). Our web application accepts voice command input and then converts it to appropriate text, performs keyword mapping from the database, and performs many tasks including searching, scrolling, redirecting, populating data, and more. The application might be expanded to recognize speaker found on the voice. Moreover, it provides modifications such that it allows users to provide commands to the web application in their own mother tongue. This would remove all the linguistic barriers from the web application. An interface can be drawn individually for each user which makes the web application more attractive which helps for increasing UI and UX. This web application can be implemented to other areas such as online education system, online emergency assistance and etc.

7. REFERENCES

7.1. Online Article

[1] “Using the Web Speech API,” MDN Web Docs. [Online].
Available:https://developer.mozilla.org/enUS/docs/Web/API/Web_Speech_API/Using_the_Web_SpeechAPI.[Accessed:03-Nov-2020].

[2] Classification of recognition systems. [Online].
Available: http://wwwhomes.unibielefeld.de/gibbon/Handbooks/gibbon_handbook_1997/node303.html. [Accessed: 03-Nov-2020].

[3] W3C, “Introduction to web accessibility”, available from: https://www.w3.org/TR/2018/REC-WCAG21-20180605/

7.2. Conference Proceedings

ISSN: 2233-7857IJFGCN
Copyright ⓒ2021SERSC

[4] K. Nahon, I. Benbasat and C. Grange, "The Missing Link: Intention to Produce Online Content Accessible to People with Disabilities by Nonprofessionals," 2012 45th Hawaii International Conference on System Sciences, Maui, HI, 2012, pp. 1747-1757.

[5] K. Christian, B. Kules, B. Shneiderman and A. Youssef, “A Comparison of Voice Controlled and Mouse Controlled Web Browsing,” Proceedings of the Fourth International ACM Conference on Assistive Technologies, no. ACM, pp. 72-79, 2000.

7.3. Journal Article

[6] M. Akram and R. Bt Sulaiman, "A Systematic Literature Review to Determine the Web Accessibility Issues in Saudi Arabian University and Government Websites for Disable People", International Journal of Advanced Computer Science and Applications, vol. 8, no. 6, 2017.

7.4. Book

[7] N. Indurkhya, F. J. Damerau, Handbook of natural language processing Vol. 2, CRC Press, 2010.

[8] B. Hashemian, "Analyzing web accessibility in Finnish higher education", ACM SIGACCESS Accessibility and Computing, no. 101, pp. 8-16, 2011

[9] M. A. Anusuya and S. K. Katti, "Speech Recognition by Machine," International Journal of Computer Science and Information Security, IJCSIS, vol. 6, no. 3, pp. 181-205, 2010.

[10] A. B. BAJPEI, M. S. SHAIKH and N. S. RATATE, “VOICE OPERATED WEB BROWSER,” International Journal of Soft Computing and Artificial Intelligence, vol. 3, no. 1, pp. 30-32, May-2015

ISSN: 2233-7857IJFGCN
Copyright ⓒ2021SERSC

ISSN: 2233-7857IJFGCN
Copyright ⓒ2021SERSC