

# Voice Controlled Cellular Communication (V3C) System for Special Citizens

Kashif Zafar  
National University of  
Computer and Emerging  
Sciences  
Islamabad

Abdul Rauf Baig  
National University of  
Computer and Emerging  
Sciences  
Islamabad

Ayesha Khan  
National University of  
Computer and Emerging  
Sciences  
Islamabad

## ABSTRACT

In this paper, we describe the formatting guidelines for IJCA Journal Submission. This paper presents a system for communication and control by disabled people based on automatic recognition of phonemes. This system allows users to navigate around an alphabet board by making phonemic utterances, thus enabling the user to spell out messages. Phoneme recognition provides an alternative to speech recognition technologies for people who have lost the ability to speak but remain capable of producing simple repeatable utterances. Voice Controlled Cellular Communication (V3C) aims at developing a voice-controlled tool for operating computer targeting physically handicapped and blind users having difficulties using a standard keyboard and mouse. It presents an interface that allows a user to activate any web page element through visual enumeration (Indexing) by an appropriate command.

## General Terms

Computational Linguistics, Speech Recognition, Machine Learning et. al.

## Keywords

Communication, cellular, speech recognition

## 1. INTRODUCTION

The Voice Controlled Cellular Communication System provides a complete description of all the functions and specifications of voice controlled cellular system for special citizens. V3C has been designed to operate in schools of disabled students and it allows them to control the computer through voice. All the applications are controlled through voice including activities such as opening, closing, typing, etc. Indexing technique is used to make communication with computer through voice easier. Through an implementation of a user centric incremental search algorithm, elements are matched according to user expectation as characters are entered into the interface. Using the user's browsing history; V3C improves the efficiency of the selection process and find potentially interesting page links for the user within the current web page. The foremost objective of this research is to give an ease to the special citizens in using computer. Secondly, to deploy in schools of special students so

that performance can be measured and improved along with the evaluation of its usability and user's response. V3C can operate in any environment and no conditions are needed to deploy the system. The external environment has no effect on the performance of the system. The user is required to have handset and microphone in order to communicate with the computer. The system has the capability of training the user incase the user has some problems in using the system.

## 2. LITERATURE REVIEW

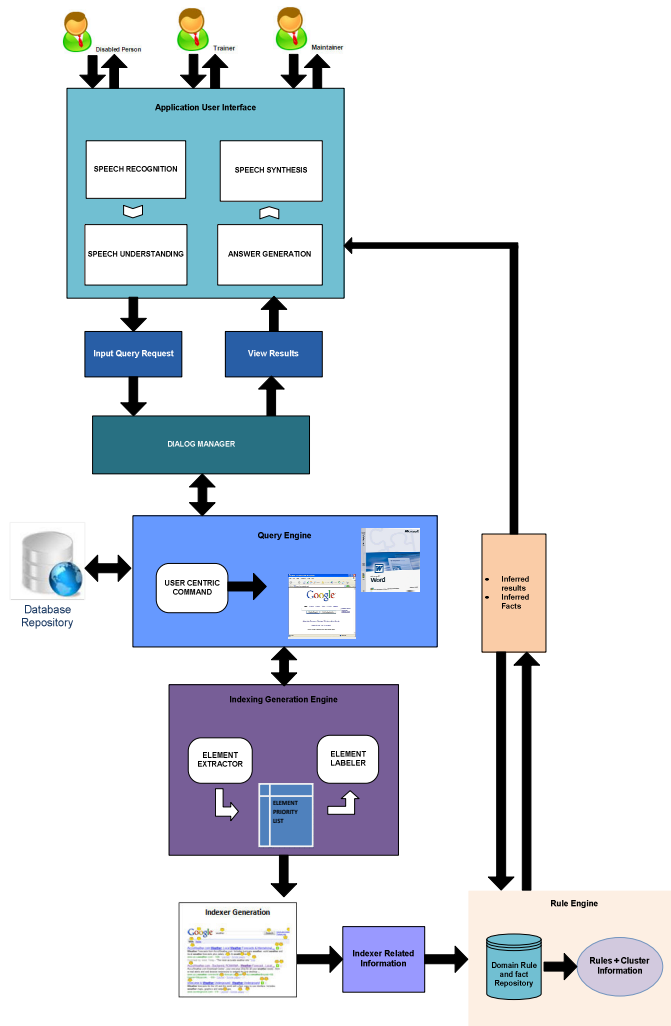
The voice controlled cellular system is a hybrid of two techniques. Chun et al, presents project "Indtal" ("Speak-it") that has succeeded in developing a Danish voice-controlled utility for internet browsing targeting motor-handicapped users having difficulties using a standard keyboard and/or a standard mouse. The system has been designed and implemented in collaboration with an advisory board of motor-handicapped (potential) end-users and underlies a number of a priori defined design criteria: learnability and memorability rather than naturalness, minimal need for maintenance after release, support for all web standards, conforming to certain recommendations, independency of the language on the websites being browsed, etc. These criteria have lead to a primarily message-driven system interacting with an existing browser on the end users' systems. In Brondsted et al, a voice-controlled human-computer interface has been presented that enables severely handicapped individuals to operate a computer. This system consists of six main blocks: the keyboard and the mouse layouts; the D6106 speech recognizer; the headset incorporated with a microphone and two mercury switches; the keyboard circuit and the mouse circuits; the mouse control circuit; and the microcontroller. A prototype system has been built and tested. In Brondsted et al, a voice controlled internet browsing for motor-handicap users is presented and they address the issues for such systems. In some papers [3, 4, 5, 9], they explored the working of voice controlled systems for web browsing for adult and blind users. J. Lin et al, presents an intelligent control system for wheel chair for elderly citizens to assist them in control and usability. In the work by L. M. Boulton et al, they present a computer hardware and software to assist visually impaired and blind users in their interaction with computers. M. C. Su et al, presents voice controlled system for disabled people with special focus on human computer interaction. In R. Iglesias et al, they present a computer graphics

access for blind users using hepatic and audio virtual environment.

### 3. SYSTEM ARCHITECTURE

#### 3.1 Voice Controlled Tool

V3C aims at developing a voice-controlled tool for operating computer. It is targeting physically handicapped and blind users having difficulties using a standard keyboard and mouse. Voice controlled cellular system addresses this situation by presenting an interface which allows a user to activate any web page element through visual enumeration (Indexing) by an appropriate command. By the implementation of a user centric incremental search algorithm, elements are matched according to user expectation as characters are entered into the interface. Using the user's browsing history; it can improve the efficiency of the selection process and find potentially interesting page links for the user within the current web page. The system architecture is shown in Figure 1.



**Figure 1: Voice Controlled Cellular Communication System**

#### 3.2 System Requirements

There is a DSP device which will be used in our system for speech recognition. The Ram of the system where V3C will be deployed should be 2.0 GB or high so that our system can work efficiently. Many different tools will be used for the implementation of the voice controlled cellular system. These tools include Microsoft Visual Studio.NET 2008 for C# programming, MATLAB 7.0, and SQL server 2008 for the storage of data. The behavior of user is stored in Database to monitor his frequency of use of different applications. The system will be made secure enough so that his patterns will just be monitored in order to make the system efficient for him.

#### 3.3 Assumptions and Dependencies

The system is implemented using the message-driven approach to reduce the future maintenance than systems build on own browser engines implemented either from scratch or as extensions to modifications of open-source browsers like the GTK Web browser Dillo or Mozilla. Some users want to use the voice control offered by V3C either occasionally to spare the muscles or in combination with their existing aid devices to avoid odd situations when moving from keyboard to mouse input or vice versa. The users operating the keyboard with sticks usually have to remove the sticks before operating their mouse device. The link selection mechanism relies on web page elements from the current page being extracted, assigned a relative priority for the current page, and unlabeled elements being given a label before selection can be commenced. Speech endpoint detection based on calculation of short-time signal energy, spectral energy, number of zero-crossing of signal, adaptive threshold values and information about duration of speech fragments.

#### 3.4 Interface Requirements

There are three logical user interfaces: disabled person, system trainer and maintainer. This system is specifically designed for the disabled persons specially blind and handicapped. They will use the system to interact with the external world using browsers. There will be a system trainer for the disabled users for guidance and provides tutorial for the system. The tutorial is different for people with different disabilities. A maintainer modifies a software product after delivery to detect and correct latent faults in the software product before they become effective faults.

#### 3.5 Voice Recognition Module

The analog signals are converted in to digital form with the help of microphone and the digital signal are processed by digital signal processor (DSP) that is used for the recognition of the voice pattern by mean of acoustic analysis that describes the frequency, pitch and intonation and the language model determines the phonetics. This device performs recognition process with the maximum accuracy.

#### 3.6 Stimulus-Response Sequences

The user action and system response is validated and checked before the implementation of the system in real world environment. The Table 1 below gives the detailed analysis.

**Table 1. User actions and system response**

<b>User Action</b>	<b>System Response</b>
<ul style="list-style-type: none"> <li>• <i>User validates his identity by entering user name and password.</i></li> </ul>	
	<ul style="list-style-type: none"> <li>• <i>System response to the authentication of the user.</i></li> </ul>
<ul style="list-style-type: none"> <li>• <i>After authentication the user proceeds to the voice recognition.</i></li> </ul>	
	<ul style="list-style-type: none"> <li>• <i>System then asks the user to input the analogue signals using a microphone.</i></li> </ul>
<ul style="list-style-type: none"> <li>• <i>User inputs his voice in the analog form that is to be converted into the digital form.</i></li> </ul>	
	<ul style="list-style-type: none"> <li>• <i>The voice is than digitized and is transferred for processing purpose.</i></li> </ul>
	<ul style="list-style-type: none"> <li>• <i>The acoustic analysis of the digital signals to determine frequency, pitch and intonation of the signals.</i></li> </ul>
	<ul style="list-style-type: none"> <li>• <i>Than analyzed voice is mould according to the language model.</i></li> </ul>

#### 4. SPEECH SYNTHESIS/RECOGNITION

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. Synthesized speech can be 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 diphones provides the largest output range, but may lack clarity. For specific usage domains, the storage of entire words or sentences allows for 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.

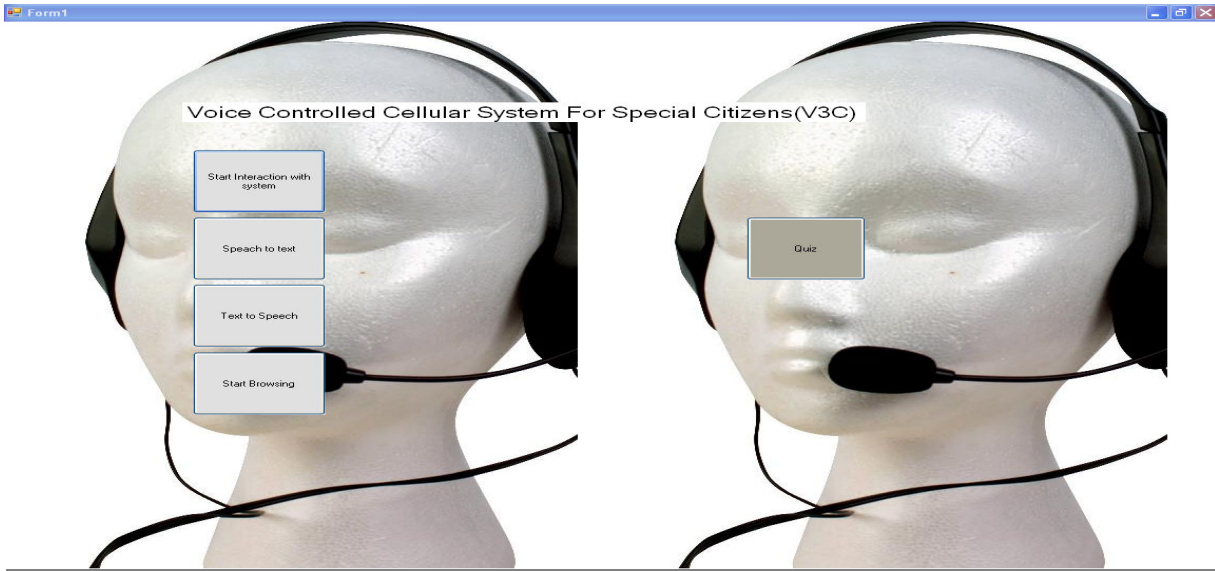
The speech is recognized by using the speech recognition module. After the recognition of speech, the next step is the understanding of the speech. This has two key components *semantic analysis* and *syntax analysis*. In many cases, the recognition procedure generates ambiguities (equally sounding words etc.) that result in word errors. If, for a recognized word, several alternatives exist, in some cases a grammar analysis can exclude some of them. The same holds for a *semantic analysis* that erases words that fails the syntax analysis and evaluates the speakers using their own stored reference patterns. The reference patterns of other speakers and reference patterns averaged over several speakers. For speakers using their own reference pattern the median word recognition error rate fell from 11.7% to 0.4% with the use of *syntax analysis*.

##### 4.1 Dialog Manager

In general, a speech-based user interface requires both, speech input (recognition) and speech output (speech synthesis). It is important to state that input and output have narrow links to each other and a dialog is by far more than the sum of both. The *dialog manager* plays a key role in the system. Its design, comfort and 'intelligence' is the main factor for the user's acceptance of the whole application. The user interface of the system is shown in Figure 2.

In complex applications, the dialog manager performs a user modeling with which, for example, trained and untrained users can be distinguished and the dialog is flexibly adapted to the individual dialog step, the user's reaction and the requirements of the task.

This system requires high performance and the voice recognition processing with the maximum accent accuracy and the recognition rate of over 95% of a noisy environment. Dynamic Type wrapping approach (DTW) and the Hidden Markov Model (HMM) are used for the development of the accurate voice recognition system.



**Figure 2. User Interface**

## **5. PERFORMANCE**

### **5.1 Software Quality**

Software is built to meet a certain functional goals and satisfy certain qualities. According to the software, the system qualities are classified into three categories that are the user, the producer and the manager of the project. The software has to be reliable and flexible as well as it must behave reasonably even in circumstances that were not anticipated in the requirement specifications. It must also meet up all the specified requirements and the system has to be maintainable in corrective, adaptive and perfective manner with respect to the cost.

### **5.2 Robustness**

The system should be robust i.e. the software has to be reliable and flexible as well as it must behave reasonably even in circumstances that were not anticipated in the requirement specification. Users should be notified about any type of error due to invalid input by the user or being generated by the system.

### **5.3 Quality of Design**

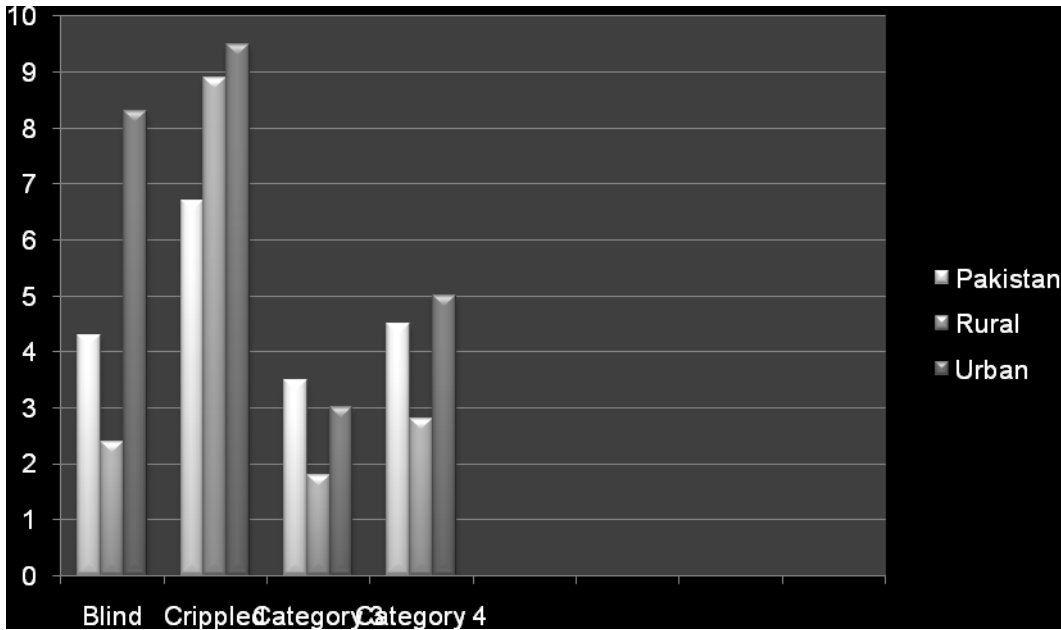
The system should be reliable enough so that it performs the required functionality and storage of the behavioral data because in case of any system crash, system can lose essential previously stored information. V3C would be affordable for the disabled institutes and normal people can utilize some contents of this software.

### **5.4 User friendliness**

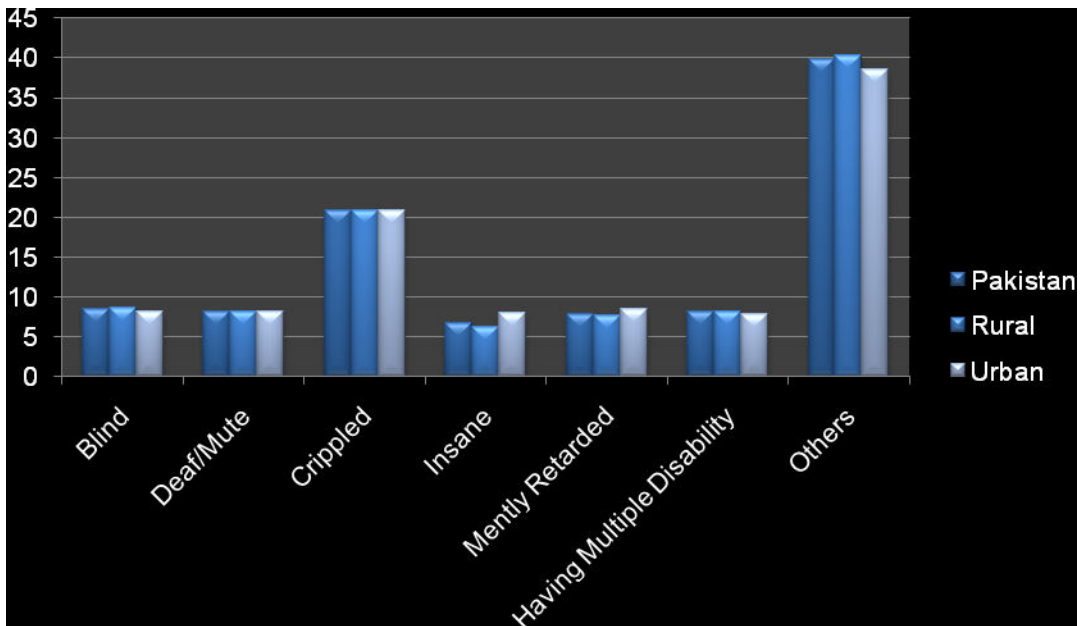
As this system is primarily designed for the disabled persons so it has to be comparatively easy and as the name of the system demands the ease of usability. The system should be user friendly and acts as a trainer for the new users and records the activities of the previous users in order to provide a friendly atmosphere to the authorized users.

## **6. IMPLEMENTATION**

In V3C, handicapped and blinds are considered as the main interactive actors in the system. They input in the form of voice and system reply back in the form of voice for blinds, although handicapped can see the screen and precede their functionality. As there is the conversion of analog input into digital input and then after processing, the digital output is converted into analog output. Handicapped and blinds will have ultimate access to the system i.e. they are able to change password of their assigned IDs and they can control every application of the windows through voice. But they have few limitations for i.e. they could not reboot the system; they could not delete important files of the windows. Maintenance and support for the system would be provided by the technical team which developed the system and teachers or trainers would also have full access to change the settings of the windows. System authorization is not given to the students that may cause failure and loss of data. Separate administrator (can be a teacher) can change the security settings and to decide accessibility constraint. The Figure [3-5] shows the statistical data for the categorization of disabled people.



**Figure 3. The detailed breakdown of disabled people in Pakistan**



**Figure 4. The detailed breakdown of disabled people in Punjab**

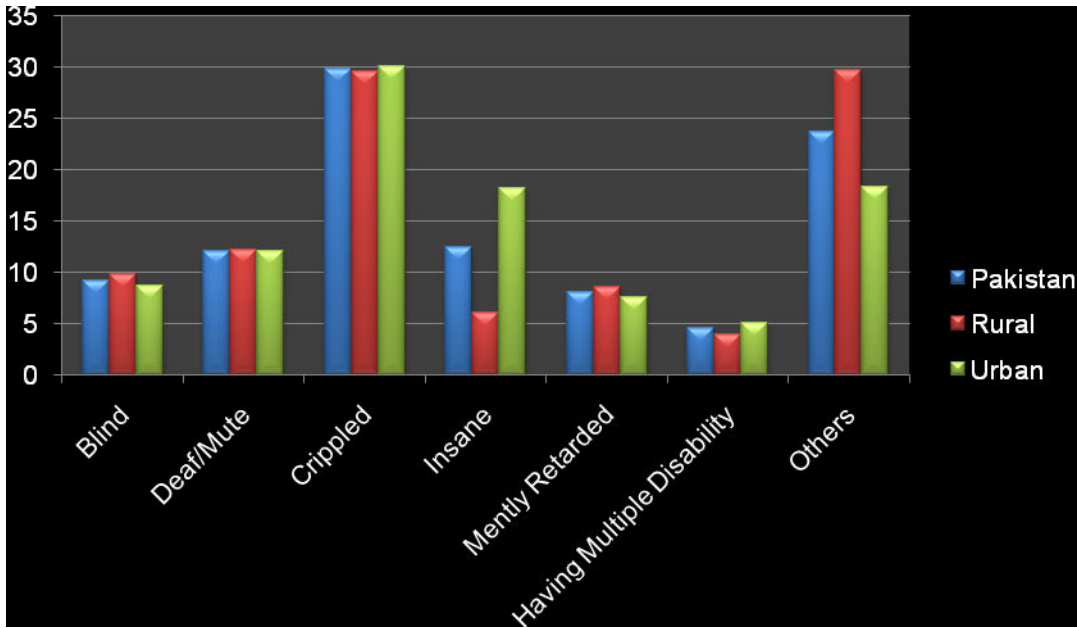


Figure 5. The detailed breakdown of disabled people in Capital

## 6.1 Screen Shots

The Figure 6 shows the screen shot for the speech to text converter using the indexing technique. The module for test is shown in Figure 7 for the quiz activity.

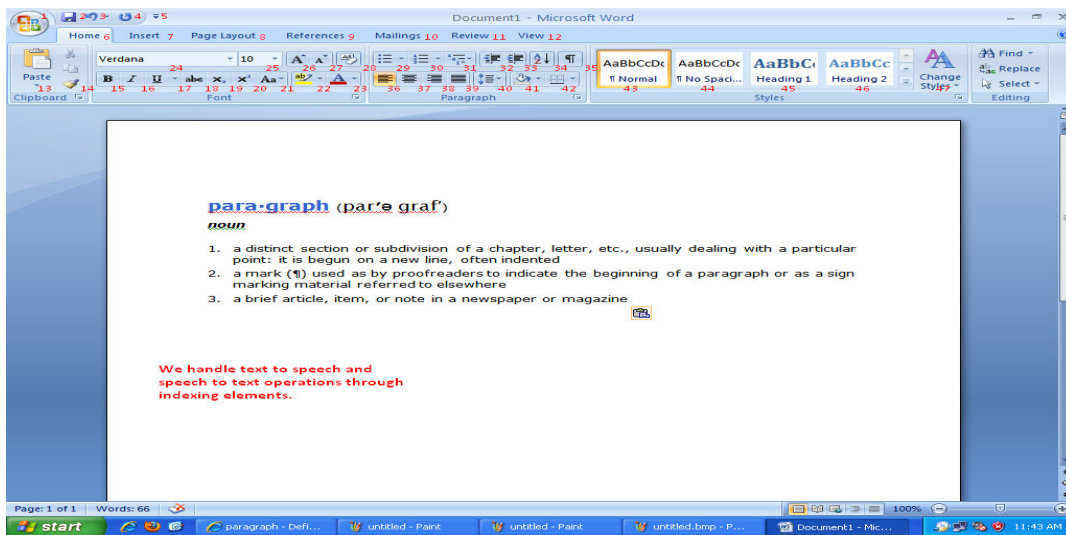
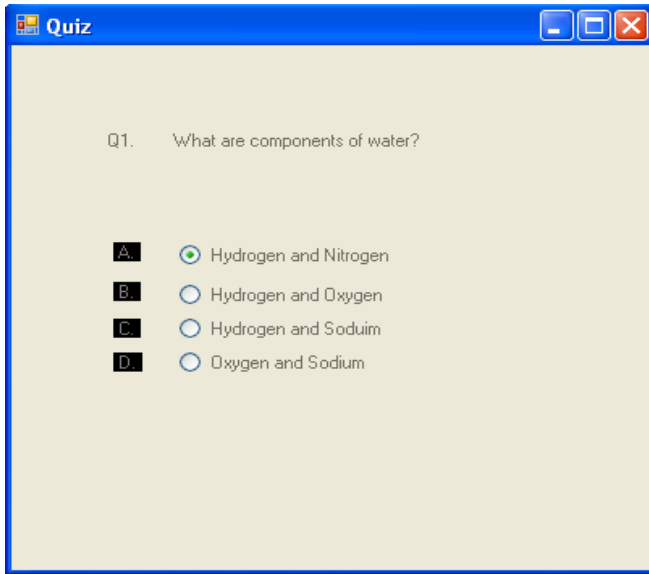


Figure 6. Speech to Text module



**Figure 7. Quiz Activity module**

## 7. CONCLUSION

V3C has been designed to operate in education sector of disabled citizens and it allows them to control the computer through voice. It can be used in special schools for the learning and training of special students. All the applications are controlled through voice including activities such as opening, closing, typing, etc. Indexing technique is used to make communication with computer through voice easier. Through an implementation of a user centric incremental search algorithm, elements are matched according to user expectation as characters are entered into the interface. Using the user's browsing history; V3C improves the efficiency of the selection process and find potentially interesting page links for the user within the current web page. The foremost objective of this research is to give an ease to the special citizens in using computer. Secondly, to deploy in schools of special students so that performance can be measured and improved along with the evaluation of its usability and user's response.

## 8. REFERENCES

[1] T. Brondsted, E. Aaskoven, "Voice-controlled internet browsing for motor-handicapped users: design and

implementation issues" In Proceedings of the speech communication and technology, EUROSPEECH, p-185-188, Portugal, 2005

- [2] M. C. Su, M. T. Chung, "Voice-controlled human computer interface for the disabled". Computing and Control Engineering Journal, p225-230, 2001
- [3] Hadjadj, Djamel & D. Burger, "BrailleSurf: An HTML Browser for visually handicapped people". In Proc. of 14th conference on "Technology and Persons with Disabilities", Los Angeles 1999
- [4] K. Christian, B. Kules, B. Shneiderman, A. Yousef, "A comparison of voice controlled and mouse controlled web browsing". In Proceedings of the 4<sup>th</sup> International Conference on Assistive Technologies, ACM, New York, USA, 2000
- [5] S. Morley, H. Petrie, A. Neill, P. McNally, "Auditory navigation in hyperspace: design and evaluation of non-visual hypermedia system for blind users". In proceedings of the 3<sup>rd</sup> International Conference on Assistive Technologies, ACM, New York, USA, 1998
- [6] L. M. Boulton, "Computer hardware and software to assist the visually impaired and blind". Australian and New Zealand Journal of Ophthalmology, Vol. 21, Issue 1, P. 7-14, WILEY, 1993.
- [7] J. Liu, H. Zhang, B. Fan, G. Wang, "A novel economical embedded multi-mode intelligent control system for power wheel chair". In Proceedings of the International Conference on Computing, Control, and Industrial Engineering, CCIE, June, 2010
- [8] R. Iglesias, S. Casado, T. Gutierrez, J. I. barber, C. A. Avizzano, S. Marcheschi, M. Bergamasco, "Computer graphics access for blind people through a hepatic and audio virtual environment". In Proceedings of the 3<sup>rd</sup> International Workshop on Hepatic, Audio and Visual Environments and Their Applications, IEEE, Derio, Spain, 2004
- [9] V. L. Hanson, J. T. Richards, C. C. Lee, "Web access for elder adults: voice browsing". In Proceedings of the 4<sup>th</sup> International Conference on Universal Access in Human Computer Interaction: Coping and Diversity, Springer, Berlin, 2007