A Voice Controlled Web based FTP Tool

Muhammed Maruf Öztürk
Comp. Engineering Dept.
Sakarya University,TURKEY

Ünal Çavuşoğlu
Comp. Engineering Dept.
Sakarya University, TURKEY

Ahmet Zengin
Comp. Engineering Dept.
Sakarya University, TURKEY

ABSTRACT

File Transfer Protocol (FTP) is a protocol which used for file transfer operation via port 21. Various tools are being used for file transfer management. These tools may be either client or web based. In this work a web based voice controlled ftp tool (Wb-FTP) is proposed. And have measured performance analysis of this tool and compared with other ftp tools. This tool has been designed with Asp.NET and C# language. The developed software is robust enough to cope with all file transfer operations. This study shows that voice control is reliable and preferable for essential ftp operations.

General Terms

Your Speech Recognition, Web Software, Voice Controlled Software.

Keywords

Speech recognition, Ftp, Web based ftp, Voice-controlled.

1. INTRODUCTION

In our world, web site count are getting increased day by day. Therefore file transfer management software is needed for web file transfer operations. The most common used protocol is File Transfer Protocol [1]. FTP's architecture was represented firstly in 1976 by Michel Gien [2]. FTP is used for data transfer between host to another host and this operation is performed over a TCP-based network. Ftp management is inaccessible for many people. Ftp management is that requires a keyboard and mouse to login, file download, file upload and other operations. Some people have trouble about using keyboard and mouse who lack movement ability [3]. This problem can be coped with with a speech recognition software. To develop a speech recognition software, we should achieve specific objectives given below [4].

- Speech recognition engine,
- Specific grammar,
- An algorithm,
- Voice control software

Voice recognition and speech recognition are confused by many researchers. Voice recognition is not concerned about speech pronouncing. Voice recognition system is composed of voice command device and voice recognition. However speech recognition is concerned about analyzing of voice. These systems analyze human voice and use it to fine tune the recognition of that person's speech, resulting in more accurate transcription.

The voice algorithms composed of two specific phases. The first one is training phase, the second one is testing phase as given in Figure 1 [5].
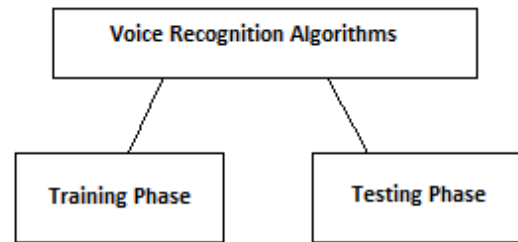

Fig 1: Voice recognition algorithms

In training phase speaker has to provide samples of their voice that base voice can be recognized. Voice recognition training can also include requests to repeat words that sound similar so the system can learn to distinguish between them. Testing phase sample voices match with stored words to ensure validation.

Voice control technology is not a new technology [6]. Various works had been done before. In this paper we have tried to show a simple steps to the necessity of adjusting speech recognition. Wb-FTP tool offers a solution to these people. Wb-FTP system listens words which have been entered before, performs ftp operations upon hearing these commands. This paper will focus on the ftp management via voice input and discuss performance of Wb-FTP.

2. LITERATURE REVIEW

Speech recognition applications can be defined into three categories; command controlled, dictation command and control applications. Control applications are related to providing the user to control items within environment with voice commands. Speech technology can also be used for authentication purposes using cryptology that we can obtain a robust system. Speech recognition eases normal life for physically disabled people [7]. In various studies, computer and electric devices have been controlled by voice. In a study [8], two voice controlled program was developed. The program for control of electric devices by voice was tested with Lithuanian and English speech recognition engines. The accuracy of recognition of commands which written as either Lithuanian or English was checked. In a speech-enabled windows application work [9] an e-commerce management system was developed using speech recognition. In this work C# and .NET was used for application design. For performance measurement a comparison work [3] was performed. This study shows that voice control adds approximately 50% to the performance time for certain types of tasks. Also speech recognition had been used robot technology before. In that work a VoiceBot was introduced and evaluated [10]. The results of the study showed that voice-controlled robotics are feasible, and portend extremely well for future research.

3. VOICE RECOGNITION

Voice recognition methods and techniques has been undertaken for an active area. Speech technology is being developed by many researchers to improve the accessibility of electronic services, software applications including games, robot applications and automatic control [11]. Speech recognition have been used in online shopping before [14]. Also speech recognition has been used in media file control [13]. Voice recognition and generation technologies offer a potential solution to ftp management problems by augmenting the capabilities of a ftp [3]. Voice controlled systems are generally based on logic or a special algorithm to perform voice operations. Wb-FTP is based on logic operations. Designed system recognizes human speech automatically. Voice controlled systems generally recognize words about %70 correctness. And system results correctness have reached approximately %72. In addition, this ratio varies according to the nature of pronunciation. Voice recognition engines can be divided into three categories. : discrete, continuous, and natural speech input [12]. We have used natural speech input. And proposed Wb-FTP is speaker independent.

4. SYSTEM DESIGN

Voice recognition is an alternative to traditional methods of interacting with a computer, such as textual input through a keyboard. A designed voice-controlled system can replace, or reduce the reliability on, standard keyboard and mouse input. The implemented Wb-FTP system is intended for file transfer control management and based on user given voice inputs. Wb-FTP is managed with voice commands given by the user. A voice user interface(VUI) is well designed for ftp control. Most users desired the usage given by the VUI clear and easy [15]. Also user can choose VUI or keyboard/mouse. Designed system is composed of a microphone that provides to user telling commands, a speech recognition engine to recognize the words and a VUI to voice input. The structure of the main system components relation is seen at Figure 2.
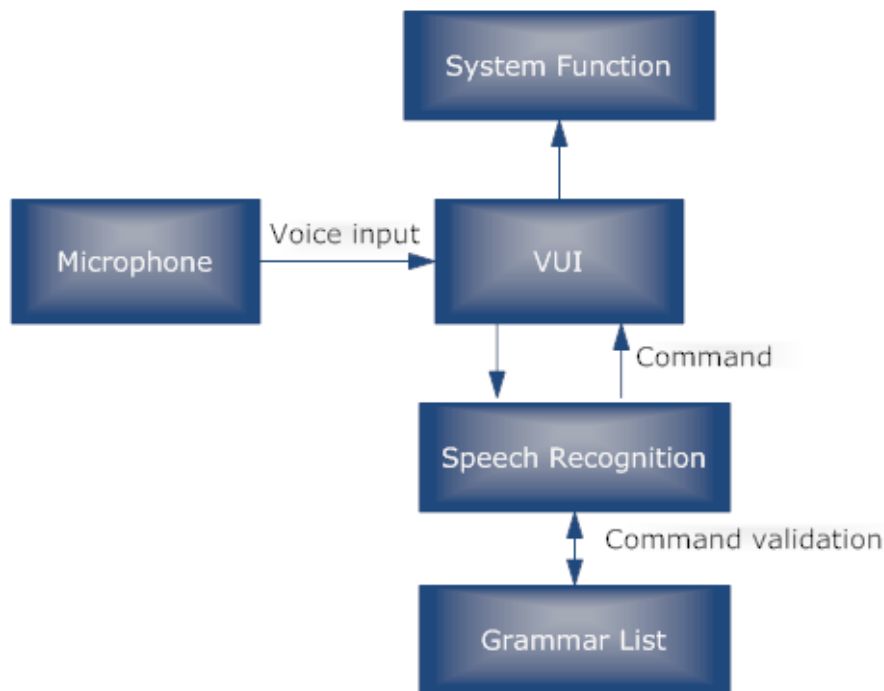

Fig 2: General voice controlled system architecture

4.1 System Architecture

A general overview of Wb-FTP is seen in Figure 3. The main Wb-FTP blocks are VUI, speech recognition and grammar list. Grammar list includes specific commands that system works optimum performance. Reporting is related to messages which given to the user about Wb-FTP process. Wb-FTP system architecture resembles a cycle. In this cycle all of unit should works correctly. If a unit of cycle works undesirable format, system operation stops.
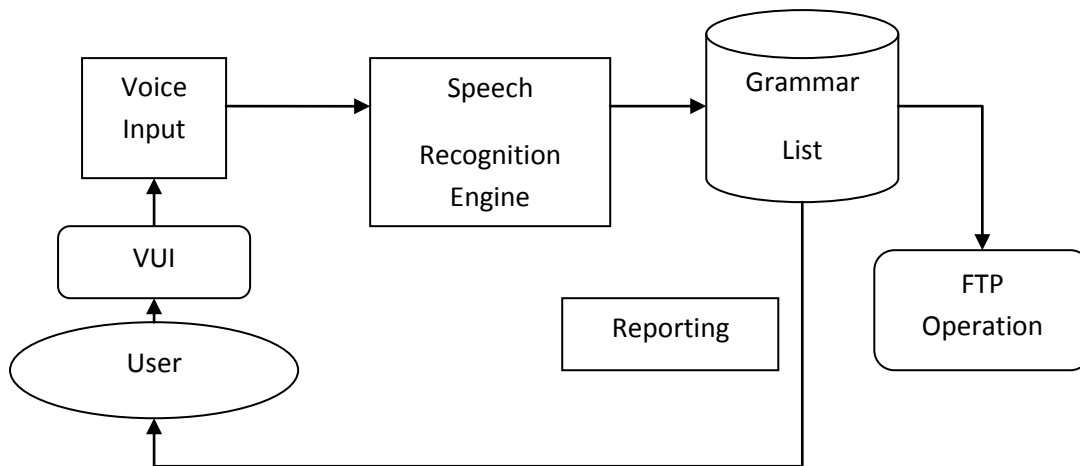
Fig 3: Wb-FTP System Architecture

As seen in Figure 3, after user voice entry, speech recognition engine works. Speech recognition engine consists of Speechlib, Windows 7 Operating System and .NET FrameWork 4.0. Most current speech recognition systems contains a model. Used Speechlib C# library has been designed on Hidden Markov Model (HMM) layer to represent acoustic models [16]. Speech recognition engine starts as soon as Wb-FTP web page loads. A grammar rule object has been created for grammar list database. If input voice verificated on grammar list database, user is informed about recognized word and Wb-FTP operation works. Long easily identifiable words have been used in grammar for the success of Wb-FTP [17]. To optimize memory and cpu usage an algorithm has been developed, it's pseudo-code is given below.

1. Start speech recognition engine and record time at T1
2. Save the frequent Wb-FTP operations on P1
3. If T1>10 seconds than disable speech recognition engine.
4. Save P1 operation list into database and select frequent list as follows:

        insert into L1(op1,op2,...) values(P1)
        select P1.op1,P2.op2.,.. Pk
        from L1

5. Save the frequent Wb-FTP operations in Lk.
6. Operate Lk and P1.

Used algorithm provides us that we can save frequent ftp commands and then use these commands rapidly.

4.2 Wb-FTP Interface Design

Wb-FTP have been designed with Visual Studio 2010 and .NET Framework 4.0. We can upload a file from user pc, list all of files in ftp server and pass between ftp directories. User can focus any component in Wb-FTP interface via focus() function. Wb-FTP is a tool which launches speech engine automatically aftermath the web page interface opening. Directory passing operations can be performed via Wb-FTP. In Figure 1 an ftp connection screen is seen. In considered solution two option are available. User can connect to web site file directory by entering required ftp information. Besides, a speech recognition engine has been created for this operation. Embedded ftp connection informations are held by

system. And if user command is "connect" an ftp connection is created. Connection flow schema is given in Figure 4.
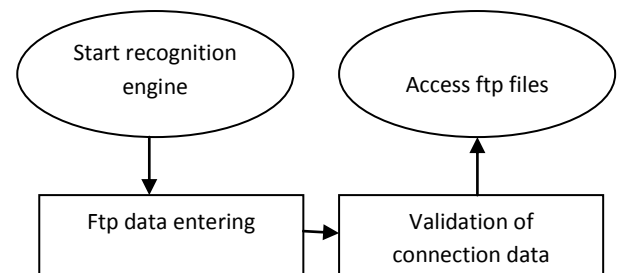
Fig 4: Wb-FTP Connection flow

Wb-FTP has been created with Visual Studio 2010. System codes are based on C#. For manual FTP connection information entering arbitrarily, Asp.NET TextBox components have been used. A file can be selected from client file system and uploaded to root ftp file directory. User can perform this operation with a microphone saying "upload" command. Designed Wb-FTP connection screen is seen in Figure 5. In this screen we can see ftp server, user name, password and file field. These fields can be filled by either manual or voice-controlled.

Accessing process of ftp files is important part of Wb-FTP connection flow. In designed system, pronounced words which used for ftp username and password are tested for validation. If system confirms user information, access permission is given to user for other operations. In this flow, recognition engine is essential for the other operations. While designing this system, similar works which include interface design have been analyzed. Because if a user friendly interface can't be generated, system fails. Components which used in system interface are designed by this way. To relieve recognition engine system is traced specific times. If any command is not taken from user, recognition engine is halted. This operation is performed by using an algorithm which is mentioned in part 4.1. In this system, user can upload whichever file is desired. Interface components are colored by different colors. Thereby ftp connection fields and label fields are distinguished easily.
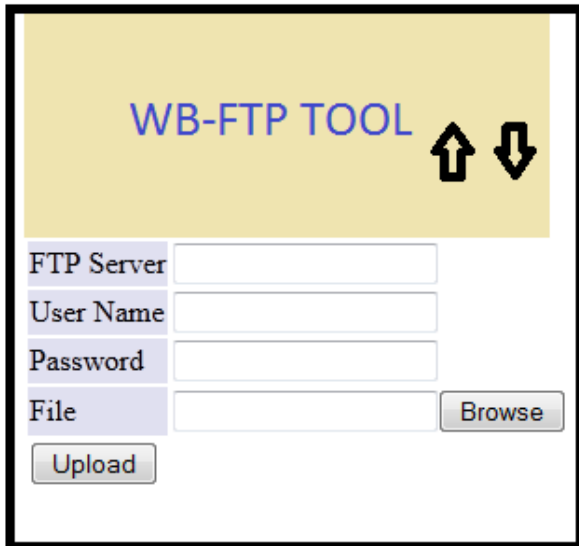
Fig 5: Ftp connection screen

User can connect desired ftp file server via entering ftp information. If user choose to enter via "connect" command these information are not necessary, because they are embedded in Wb-FTP. In Wb-FTP, speech recognition starts after Wb-FTP loading, waits until ten minutes if there is not any word said, speech recognition stops. This operation performed via "Asp.NET Cookies". Speaker independent recognition provides high accuracy [18] so we have developed speaker independent system. Before Speech Recognition some requirements have to be done [9], i.e. defining grammar. It is necessary to use an event handler as a callback function so that every time Wb-FTP recognizes a word from the grammar, it invokes the event handler. Of course, these words have to be added to the grammar before Wb-FTP can use them. Some example codes are given below:

```
objRecoContext = new SpeechLib.SpSharedRecoContext();
    objRecoContext.Hypothesis += new
_ISpeechRecoContextEvents_HypothesisEventHandler(Hypo
_Event);
```

```
objRecoContext.Recognition += new

_ISpeechRecoContextEvents_RecognitionEventHandler(Reco
_Event);
    grammar = objRecoContext.CreateGrammar(0);
    //Activate the Menu Commands.
    menuRule = grammar.Rules.Add("MenuCommands",
SpeechRuleAttributes.SRATopLevel |
SpeechRuleAttributes.SRADynamic, 1);
    object PropValue = "";
    menuRule.InitialState.AddWordTransition(null,
"Connect", " ", SpeechGrammarWordType.SGLexical,
"New", 1, ref PropValue, 1.0F);
    menuRule.InitialState.AddWordTransition(null,
"Upload", " ", SpeechGrammarWordType.SGLexical,
"Open", 2, ref PropValue, 1.0F);
    menuRule.InitialState.AddWordTransition(null,
"Download", " ", SpeechGrammarWordType.SGLexical,
"Close", 3, ref PropValue, 1.0F);
    menuRule.InitialState.AddWordTransition(null, "Choose
first", " ", SpeechGrammarWordType.SGLexical, "Exit", 4,
ref PropValue, 1.0F);
    menuRule.InitialState.AddWordTransition(null, "Choose
second", " ", SpeechGrammarWordType.SGLexical, "Cut", 5,
ref PropValue, 1.0F);
    menuRule.InitialState.AddWordTransition(null, "Choose
third", " ", SpeechGrammarWordType.SGLexical, "Copy", 6,
ref PropValue, 1.0F);
    menuRule.InitialState.AddWordTransition(null, "Choose
forth", " ", SpeechGrammarWordType.SGLexical, "Paste", 7,
ref PropValue, 1.0F);
    menuRule.InitialState.AddWordTransition(null, "Close
Connect", " ", SpeechGrammarWordType.SGLexical,
"Delete", 8, ref PropValue, 1.0F);
    grammar.Rules.Commit();
    grammar.CmdSetRuleState("MenuCommands",
SpeechRuleState.SGDSActive);
```

We can see voice commands and functions at Table 1.

Table 1. Wb-FTP voice command functions table

Command Name	Function
Connect	Connection of ftp server
List	Lists ftp files and local files
Choose First	Selects first local file
Choose Second	Selects second local file
Choose Third	Selects third local file
Choose fifth	Selects fifth local file
Upload First	Uploads first local file
Upload Second	Uploads second local file
Upload Third	Uploads third local file
Upload fifth	Uploads fifth local file
Download First	Downloads first local file
Download Second	Downloads second local file
Download Third	Downloads third local file
Download fifth	Downloads fifth local file
Close Connection	Closes ftp connection

Wb-FTP grammar list is composed of above commands. Firstly an ftp connection must be launched. Aftermath ftp file list is ordered. User can select a file from list then an upload or download operation can be performed. We have trained on

five file number. But on this grammar, various count of file operation can be added as much as wanted. Finally ftp connection should be closed.
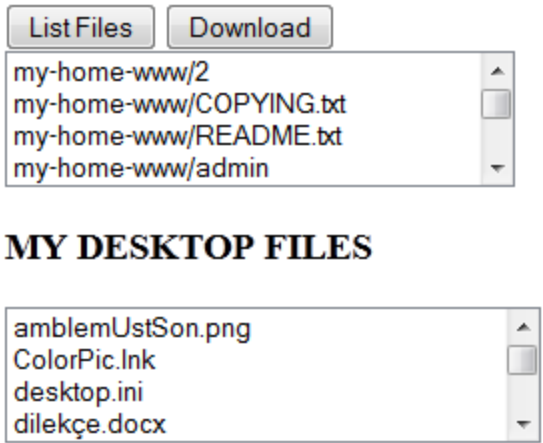
Fig 6: Ftp list download screen

In Figure 6, tested ftp server file list and pc desktop file list are shown. User can select desired file via voice command and upload to ftp server. When the file transmission is completed, a "successfully uploaded" message is appeared. Besides, file transmission completed message is informed via voice message.

5. SYSTEM PERFORMANCE

Wb-FTP is well designed. Web based components can be controlled via voice commands. The implemented Wb-FTP has been evaluated with ten users. The results has been collected with questionnaire and discussions. Voice controlled systems generally work slower than by hand applications. So designed system can't be hoped that proposed system works as fast as manual controlled system. Wb-FTP software has been 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, client and server. The software has to be reliable and flexible that we can add new module. Wb-FTP has been evaluated response with manual and voice. Performance result is seen in Figure 7.
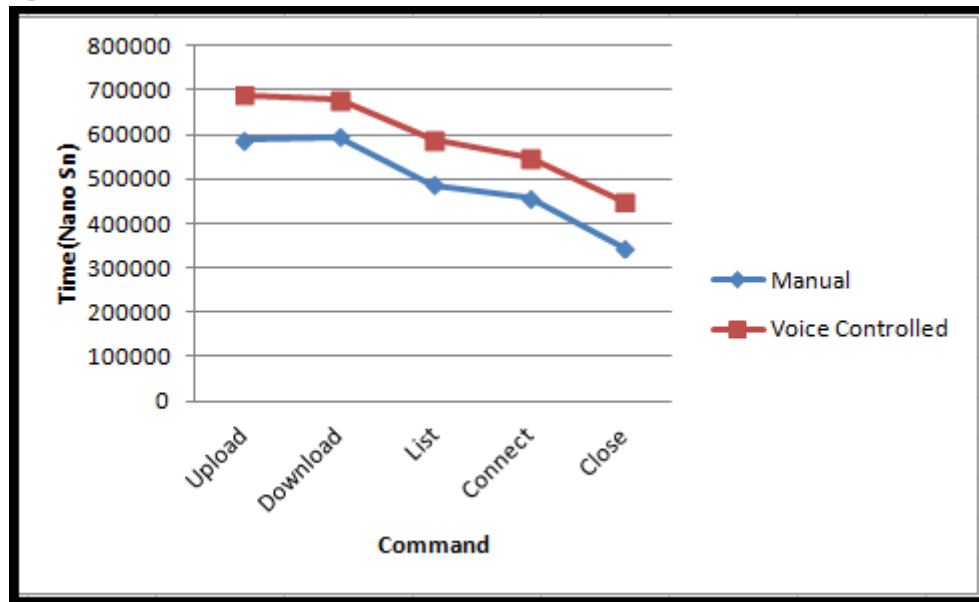

Fig 7: Performance comparison of Wb-FTP and manual

There are significant differences in operation performance times. For these operations, ftp closing took on average 3.1 (ms). This time is consistent with the other operations. Upload time is the maximum task of ftp operations. Voice-controlled operations have taken more time than manual operations. This is acceptable in theory. But if we consider that voice-controlled management provides to user a comfortable usage, voice-controlled task times are not so much. The comparison of Wb-FTP with the other tools is seen in Table 2.

Table 2. Comparison of Wb-FTP with other tools

	Ftp Tasks	Flexible	Modular	Manual/Voice Option
Wb-FTP	Sufficient	Yes	Yes	Yes
Ftp Clients	Sufficient	No	No	No
Web based	Sufficient	No	No	No

6. SUGGESTION AND FUTURE WORKS

In this work Wb-FTP has been designed. Designed ftp management tool is web based and need to be well designed. Therefore design innovations in ftp tool will ease controlling of ftp. If Wb-FTP is performed with cryptography via personal voice, robust and trustworthy tool can be obtained. One clear direction for future research is to explore more common website interface for ftp control. Speech fluency is another factor that effects the performance of speech recognition ratio [19]. Therefore with increasing of long words, voice entering should be clear. On the other hand, Simpson and Levine [20] have proved that voice controlled systems reduce physical requirements. Voice controlled software provides the user many degrees of freedom for control and input [8]. With respect to this work, a result can be concluded that physically disabled people can use Wb-FTP for file transfer operation to their web site. Besides these conclusions, a VoiceBot [21] has been designed for arm movement. In the near future we could also handle with web based operations via VoiceBot. FTP operations are driven generally sequential. These operations changes person by person. So if a prediction method can be developed, operation time will be reduced.

7. REFERENCES

[1] Sohail, S. and ElGindy, H. 2003., Parallelized File Transfer Protocol (P-FTP). In Proceedings of the 28th Annual IEEE International Conference on Local Computer Networks, USA.

[2] Gien, M., 1976, "A File Transfer Protocol", Computer Networks, Vol 2, No 4, pp. 312-319.

[3] Christian, K., 2000, "A Comparison of Voice Controlled and Mouse Controlled Web Browsing". In Proceedings of the fourth international ACM conference on Assistive Technologies, pp. 72-79, USA.

[4] R. Zhou, K. P. Ng, Y. S. Ng, 1994, "A Voice Controlled Robot Using Neural Network", In Proceedings of the Second Australian and New Zealand Conference, pp. 130-134.

[5] Muda L., Begam M. and Mumtaj E. I., 2010, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", Journal of Computing, Vol 2, No. 3, pp. 138-143.

[6] Derek F., Kevin M. D., 2008, "Voice Controlled Debugging of Spreadsheets", Proc. European Spreadsheet Risks Int. Grp., pp. 155-164.

[7] Hawley, M.S., Enderby, P., Green, P., Cunningham, S.P., Brownsell, S. Carmichael, J., Parker, M., Hatzis, A., O'Neill, P. and Palmer, R., 2007, "A speech-controlled environmental control system for people with severe dysathia, Medical Engineering and Physics, Vol 29, No 5, pp. 586-593.

[8] Rudžionis A., Ratkevičius K., Dumbliauskas, T., Rudžionis V., 2008. "Control of Computer and Electric Devices by Voice", Elektronika lr Elektrotechnika, No. 6, pp. 11-16.

[9] Shi H., Maier A. 2006, "Speech-enabled windows application using Microsoft SAPI", IJCSNS International Journal of Computer Science and Network Security, Vol. 6, No 9A, pp. 33-37.

[10] House B. , J. Malkin and J. A. Bilmes, 2009. "The VoiceBot: A voice controlled robot arm", In Proceedings of CHI 2008: ACM Conference on Human Factors in Computing Systems, (HFCS' 09), ACM, Boston, MA., pp: 1-10.

[11] Mustaquim, M. M., 2011, "Automatic speech recognition an approach for designing inclusive games", Multimedia Tools and Applications.

[12] Tan, C. L., Shankar, J. P., "Web Navigation in Small Communication/Computing Devices using Voice Recognition", Centre for Wireless Communications.

[13] Wang Y. F. H., Hamerich S. W., Hennecke M. E., Schubert V. M., 2005, "Speech-controlled Media File Selection on Embedded Systems", In SIGdial, pp. 217-221.

[14] Guoqiang D., Yaoyao L., Lingchao H., Jianping W. 2008, "Design and Implementation of Voice Web Pages for Online Shopping Based on .NET and Streaming Media", International Conference on Management of e-Commerce and e-Government, pp. 226-229.

[15] Sorsa T., Haloen K. 2002 ,"Mobile Melody Recognition System with Voice-Only User Interface", International Symposium/Conference on Music Information Retrieval ISMIR.

[16] Hemphill C., Thrift P., 1995, "Surfing Web By Voice", ACM Multimedia 95-Electronic Proceedings, pp. 215-222.

[17] Ferreiros J., Pardo J. M., de Cordoba R., Macias-Guarasa J., Montero M. J., Fernandes F., Sama V., d'Haro L. F., Gonzales G., 2012, "A speech interface for Air Traffic Control terminals", Aerospace Science and Technology, Vol 21, No 1, pp. 7-15.

[18] Ayres T., Nolan B. 2004, "Voice Activated Command and Control with Speech Recognition over Wireless Networks", ITB Journal, Vol 10, pp. 37-54, Indonesia.

[19] Oviatt S., 1994, "Interface Techniques for Minimizing Disfluent Input to Spoken Language Systems", In Proceedings of the SIGCHI Conference on Human Factors in Computins Systems, pp. 205-210.

[20] Simpson and Levine 2002, "Voice Control of a Powered Wheelchair", IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING, Vol. 10, No. 2, pp. 122-125.

[21] House B. , J. Malkin and J. A. Bilmes, 2009. "The VoiceBot: A voice controlled robot arm", Proceedings of CHI 2008: ACM Conference on Human Factors in Computing Systems, (HFCS' 09), ACM, Boston, MA., pp: 1-10.