ISSN 2047-3338

Speech Recognition and Synthesis Tool: Assistive Technology for Physically Disabled Persons

F. Reena Sharma and S. Geetanjali Wasson

Abstract— **In this paper, attempt has been made to develop a Speech Recognition and Synthesis Tool (SRST) as assistive technology to provide a solution for communication between two physically disabled persons; blind and deaf. An off-line chat room has made where two physically challenged persons can communicate to each other in US English accent via USB Serial Adaptor. SRST is designed in Microsoft.NET 3.5 framework using C# Programming in Microsoft Visual Studio 2008 Environment. Microsoft Windows Speech Application Programming Interface (SAPI) 5.3 and, system speech recognition and system speech synthesis namespaces are used for speech to text conversion and vice-versa. The blind and deaf students trained on the tool and window speech recognition system inbuilt in windows vista. Then this communication tool is implemented in real setting for the physically challenged persons. It has been observed that they communicated to each other effectively only in noise free environment and blind students taken more time than deaf students to familiar with tool commands and speech recognition commands.**

Index Terms— **Assistive Technology, C# Programming Microsoft SAPI, Speech Recognition and Speech Synthesis**

I. INTRODUCTION

ASSISTIVE Technology (AT) is defined in federal legislation as any item, piece of equipment, or product system, whether acquired commercially off the shelf, modified, or customized, that is used to increase, maintain, or improve functional capabilities of individuals with disabilities. Examples of AT include wheelchairs, electronic communication systems, computer adaptations, and thousands of other commercially available devices [1].

A tremendous variety of assistive technology is available today worldwide, providing the opportunity for nearly all people including physically and mentally challenged persons to access Information Technology (IT) – e.g., Braille display, Screen readers**,** voice recognition programs, speech synthesizer, screen magnifier, teletypewriters conversion modem etc. This technology can include hardware, software, and peripherals that assist people with disabilities. People who are blind cannot read instructions presented only in a visual format; people who are deaf cannot understand content that is presented only aurally. So they cannot access the technology for their educational needs and its communication in educational settings [2]. Microsoft Windows is compatible with a wide variety of assistive technology products such as screen readers, magnifiers, and specialty hardware that meet the needs of computer users with all types of physical impairments.

Speech is the most natural way of communication. It also provides an efficient means of man-machine communication. Generally, transfer of information between human and machine is accomplished via keyboard, mouse etc. But human can speak more quickly instead of typing. Speech interfacing provides the ways to these issues [3]. Speech interfacing involves speech synthesis and speech recognition.

Speech Recognition is a technology that allows the computer to identify and understand words spoken by a person using a microphone. Speech recognition allows a computer to interpret any sound input (through either a microphone or audio file) to be transcribed or used to interact with the computer. Whereas, Speech synthesizer takes the text as input and converts it into the speech output i.e. it act as text to speech converter. Speech recognizer converts the spoken word into text [4].

The paper is organized as follows. Second section described the need and significance of the work, third section illustrated the literature review of previous work, contribution of the paper is given in fourth section, fifth section showed the methodology of work, sixth section elaborated the basics of software used for making tool and tool features. Seventh section showed the hardware and software configuration needed and steps for implementation the tool in real setting. Findings and discussion are given in eighth section. Ninth section concluded the paper and scope of future of work is given.

II. NEED AND SIGNIFICANCE

Text-to-speech synthesis and continuous speech recognition are the two rapidly developing assistive technologies by which all able and disabled persons can access information communication technology for their educational needs. Many international organizations like Microsoft SAPI [14] and Dragon-Naturally-Speech [15] as well as research groups[6 are working on this field and people with disabilities is taking benefit from speech recognition programs. But these packages

F. Reena Sharma, Research Scholar, Department of Electronics & Communication Engineering, Bhaddal Institutes, India (Corresponding Author Email: reenashrama1723@gmail.com)

S. Geetanjali Wasson, Assistant Professor, Department of Electronics & Communication Engineering, Bhaddal Institutes, India

are expensive and available separately. There currently does not appear to be any assistive technology that directly addresses communication between blind and deaf persons. However, few seemed to incorporate both text-to-speech and speech-to-text softwares.

The goal of this research is to incorporate speech recognition (speech-to-text) and speech synthesis (text-to-speech) technology into a chat room, thus, providing a solution to communication between two physically disabled persons specially blind and deaf. This technology does not require any additional equipment besides computers, USB Serial Adaptor and Headphone with mike facility.

III. LITERATURE REVIEW

In the past decade, much works have been done in the field of speech recognition and speech synthesis for communication.

One promising study [5] at California State University at Northridge explored the performance of learning disabled college students using voice recognition technology to complete the university's written proficiency exam. With the use of this innovation, the learning disabled students achieved the same distribution of scores on the exam as their non disabled peers.

Another exploratory study [6] focused on a single subject-a sixth grade student with learning disabilities. Wetzel was interested in whether middle school students could learn to use a voice recognition system, in this case IBM VoiceType, and whether this system would enhance their communication skills. Wetzel found that the student was able to learn to use the software, but that difficulties with the system's recognition accuracy and the complexity of editing compromised this student's success.

The authors [7] describe four methods that persons who are blind or visually impaired use to access information: enhanced image, braille, synthetic speech and optical character recognition. These devices can be used separately or in combination to access consumer products, personal computers and printed information.

Microsoft also developed a standard by which applications could effectively communicate with assistive technologies. This standard, called Microsoft® Active Accessibility® (MSAA) has been available since Windows 95. Windows Vista has built on this to improve accessibility into two other areas: 1) an Ease of Access center to assist users in finding ways to make using their computer easier; and 2) new technologies, such as better speech recognition and magnification [8]. Apple has now built accessibility into its Carbon application programming interface (API), which allows Mac OS X applications to more effectively communicate with assistive technologies. the Linux developer community has produced a basic core set of accessibility features, as well as a combined screen reader/screen magnification application, Braille output software, and an on-screen keyboard. Each of these products was developed for the popular GNOME desktop, a graphic interface environment that runs on both Linux and Unix.

Microsoft commissioned Forrester Research to measure the market for accessible technology in the United States and to better understand how accessible technology is being used. Research advances can lead to a unique product coming to the market, such as the Xbox Kinect. The science behind Kinect came from multiple research areas, including depth sensing, speech recognition, gestural interface, computer vision, and sound processing [9]. Microsoft Research has a group in Redmond and another in Beijing working together to improve spoken language technologies. Our main goal is to build applications that make computers available everywhere, and work with Microsoft Tellme to make this vision a reality.

The research paper outlined how speech recognition is being used today to enhance the educational process for students and teachers alike. Dragon Naturally Speaking software offered speech output capabilities so that students can have text read aloud to them. The system worked with most popular screen readers, such as JAWS, helping visually impaired students to work more effectively [10].

Marshall H. Raskind, a learning disabilities researcher at the Frostig Center in Pasadena, Calif., found that voice recognition software could make a significant difference for many people with dyslexia. It is concluded that speech recognition not only allows dyslexics to communicate more efficiently, but may even help them overcome their condition [11].

J-Say is designed to make it possible for blind people to work with speech recognition software. It acted as an interface between Dragon Naturally Speaking and “JAWS” screen reading program. This combination of programs has made it possible for a totally blind person to dictate and compose documents hands-free. This solution requires technical aptitude, commitment and a significant amount of training [12].

The researchers developed a new, statistically trained, automatic text-to-speech (TTS) system having a relatively small training database, a small-footprint (less than 2 megabytes) hidden Markov model HMM flexible, easy modification of spectrum, gain, speaking rate, pitch range of synthesized speech. With its easy training and compact size, the new HMM is ideal for quick prototyping of a personalized TTS [13].

IV. CONTRIBUTION

The authors have developed a chat room where two physically challenged blind and deaf persons can communicate to each other in US English accent. To accomplish this, Speech Recognition and Synthesis Tool was designed on Microsoft.NET 3.5 framework using C# language in Microsoft Visual Studio 2008 environment. Then this communication tool was implemented in real setting for the testing of this tool as communication technology for blind and deaf persons and noted certain critical observations.

V. METHODOLOGY

The methodology for the current research is as follows in Fig. 1:

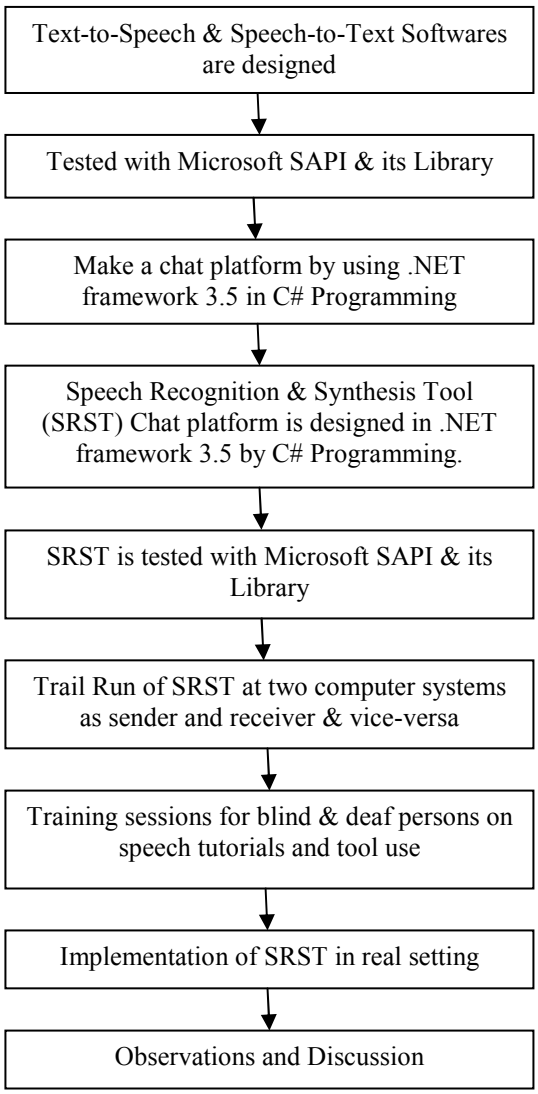

Fig. 1: Methodology of work

VI. SPEECH RECOGNITION & SYNTHESIS TOOL DESIGN

SRST is designed in Microsoft .NET Framework 3.5 using Microsoft Visual C# object oriented language in Microsoft Visual Studio 2008 Environment.

A. Fundamentals Basics

i) The *Microsoft .NET Framework* is a software technology that is available with several Microsoft Windows operating systems. It includes a large library of pre-coded solutions to common programming problems, a runtime or virtual machine that manages the execution of programs written specifically for the framework, and a set of tools for configuring and building applications. The .NET Framework is a key Microsoft offering and is intended to be used by most new applications created for the Windows platform. The .NET Framework has two main components: the common language runtime and the .NET Framework class library. The .NET Framework is an environment for building, deploying, and running Web Services and other applications. It consists of three main parts: The Common Language Runtime, The Framework classes, and The ASP.NET. .NET is a general-purpose software development platform, similar to Java [14].

ii) *Microsoft Visual C#* is Microsoft's implementation of the C# specification, included in the Microsoft Visual Studio suite of products. It is based on the ECMA/ISO specification of the C# language, which Microsoft also created. While multiple implementations of the specification exist, Visual C# is by far the one most commonly used. Visual C# is also heavily used by ASP.NET web sites and standalone applications based on the .NET Framework. C# is a new programming language, very similar to Java. An extensive class library is included, featuring all the functionality one might expect from a contempory development platform - windows GUI development (Windows Forms), database access (ADO.NET), web development (ASP.NET), web services, XML etc [15].

iii) *Microsoft Visual Studio* is an integrated development environment (IDE) from Microsoft. It is used to develop console and graphical user interface applications along with Windows Forms applications, web sites, web applications, and web services in both native code together with managed code for all platforms supported by Microsoft Windows, Windows Mobile Windows CE, .NET Framework, .NET Compact Framework and Microsoft Silverlight [16].

iv) *Microsoft Speech Application Programming Interface (SAPI)* is an API developed by Microsoft to allow the use of speech recognition and speech synthesis within Windows applications. It is possible for a 3rd-party company to produce their own Speech Recognition and Text-To-Speech engines or adapt existing engines to work with SAPI [17]. SAPI 5 however was a completely new interface, released in 2000.

v) The *System.Speech.Synthesis namespace* can be used to access the SAPI synthesizer engine to render text into speech using an installed voice, such as Microsoft Anna. The SAPI 5.3 synthesizer now supports the W3C standard Speech Synthesis Markup Language (SSML), a markup language that allows you to finely tune how the synthesizer will produce words, such as pronunciation, speed, volume and pitch, of the produced phrase [18].

vi) The *System.Speech.Recognition* engine is used to recognize a user's voice and convert it into text. The SAPI 5.3 recognition engine now supports the W3C standard; Speech Recognition Grammar Specification (SRGS), a markup language that defines how and what words are recognized. SAPI 5.3 also added support for Semantic Interpretation [18].

B. Properties & Features of SRST

1) Properties Details:

File Description	mtkvoiceresponse
Type	Application (.exe)
File Version	1.0.0.0
Product Name	mtkvoiceresponse
Product verson	1.0.0.0
Copyright	Copyright@2012
Original Filename	mtkvoiceresponse.exe

2) Features of SRST

Different Forms of SRST are as follows:

i) Application Form
ii) Login Form
iii) Container Form
iv) Speech To Text Form
v) Text To Speech Form

Features of each form as follows:

i) Application Form

It has 'mtkvoiceresponse' Icon at Desktop to run the Application (.exe) file by opening it.

ii) *Login Form*

It consists of two labels, two textboxes and two buttons. The labels are 'User Name' and 'Password' and its two boxes for writing it. After writing User Name and Password, click 'Login' Button to open Container Form or click 'Cancel' button to exit the program. The screenshot of Login Form is shown in Fig. 2.

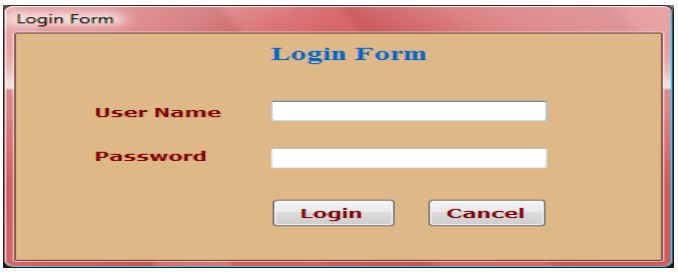

Fig. 2: Login Form of SRST

iii) Container Form

When both User Name and Password are correct, the next form labeled as 'frmcontainer' is opened as shown in Fig. 3.

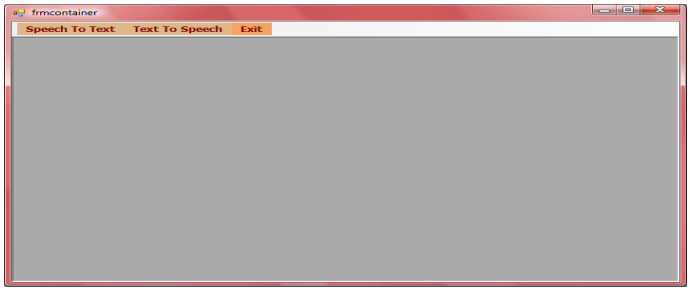

Fig. 3: Container Form of SRST

This Container Form consists two forms; *'Speech To Text'* and *'Text To Speech'* and 'Exit' button to exit all forms. This form can be maximized, minimized or close function.

iv) *Speech To Text Form*

When click on '*Speech To Text'* button in Container Form, it opens the form label as Send Text as shown in Fig. 4.

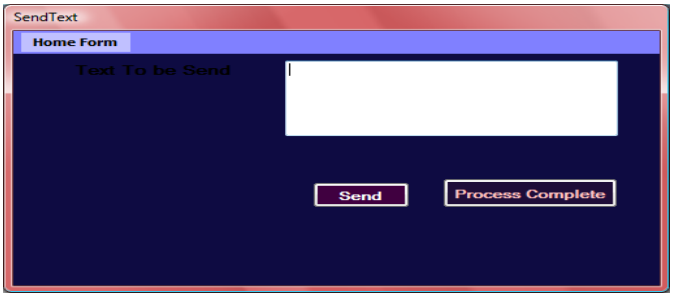

Fig. 4: Speech To Text Form of SRST

It consists *'Home Form'* button, *'Text to be Send'* label & its text box, *'Send'* Button and '*Process Complete'* Button. In 'Text To be Send' text box, the text is automatically convert from sender speech using headphone with mic and click 'Send' button to send the text to receiver end. Click 'Process Complete' button to close the Com Port. The Speech To Text used Windows speech library.

v) Text To Speech Form

When click on 'Text To Speech' button in Container Form the new form label as Convert Text to Speech Format opens as Fig. 5.

Fig. 5: Text To Speech Form of SRST

It consists Home Form button at top, Content To Read, Whole Conversation labels & its text boxes and 7 button. 'Home' button to return back to Container Form; '-Rate+', 'Dec', 'Inc' for increasing and decreasing rate of speed of voice conversion from text received; 'Ready To Read' button for opening the Com Port for receiving the text from sender end; 'Process Complete' button to close the Com Port link; 'Listen Whole Conversation' button to hear the whole text received and 'Clear Conversation' button to delete the text.

VII. IMPLEMENTATION

SRST is installed on desktop or laptop having the following specifications:

Computer System Configuration: Microsoft Window Vista and above, 512 MB RAM, 40 GB HD, Realtek High Definition Audio Sound Card.

Softwares Required: Microsoft Visual Studio 2008, Microsoft .NET 3.5 Framework.

External Aids Required: Headphones with mike facility, USB Serial Adaptors, COM PORT (DB9)

A. Training

- Sessions were conducted to familiar the sample students with Microsoft Speech Recognition and Speech Synthesis engines tutorials and with SRST for effective and efficient implementation of SRST in noise free environment.
- Samples of five blind and five deaf students of 10^{th} grade were selected randomly from Deaf & Dumb School and Blind School of English Medium.
- All students were familiar with US Accent English.
- Blind students got training in two parts; Windows Speech Recognition technology with Microsoft Narrator assistive technology, in built in Vista operating systems and SRST working and how to use it, whereas deaf students got training in SRST working and how to use it.
- Intially, five sentences were given to blind and deaf students to practice for thrice and observed the accuracy of the SRST.
- Each sentence 'Hello', How are you?', 'I am fine', 'Where are you going?' and 'I am going to school' were sent and received from one computer to other and vice versa by both blind and deaf students mutually.
- Then blind students were given extensive training for speech recognition, half an hour for five days for window SR engine recognize the words of speaker more accurately.
- Feedback was taken from each student verbally to modify the system i.e., add error message when does not communicate etc.

B. Implementation of SRST

- Both the computer systems were ready with necessary hardware, softwares and external aids for the communication between blind and student.
- The noise free environment was provided for the effective communication.
- The trained blind and deaf students were given one computer system for communication.
- One educational topic was given to discuss for five minutes to each pair.
- The both deaf and blind students acted as sender and receiver at many times for effective communication.
- The time taken by blind students' more than deaf students, as they were communicating and operating tool with their voice.

Below the diagram shows one example of SRST to show its working. One blind person has own system and deaf student its own and they conversed to each other as follows in Fig. 6.

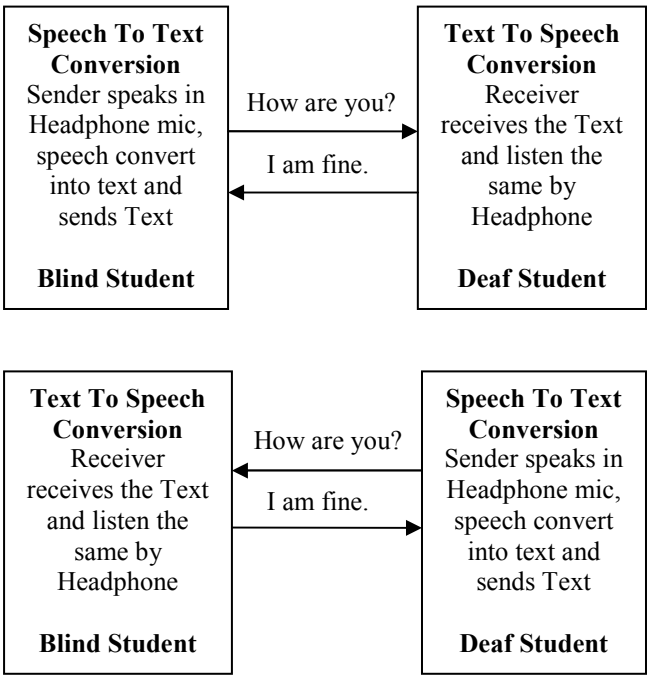

Fig. 6: Blind and Deaf student communicate to each other

VIII. FINDINGS AND DISCUSSION

There were certain findings during implementation and after implementation. These are as follows:

- The blind students have taken more time to use SRST and SR system than deaf persons, as they can use manually the system.
- The noise free environment room made tool effective to work.
- One pair did not communicate properly because of blind student could not follow the voice instructions to use commands.
- One pair was performed well and felt satisfied with the tool.
- Other three pair was satisfactory satisfied with the tool.

Discussions for the findings are as follows:

- The room or the location where SRST has to used should be noise free.
- The blind students should require longer period of time for speech recognition system and SRST with more simplified instructions
- More interactive functions should be added to use pdf files and text files to be sent through SRST for effective communication.
- Many serial adaptors can be avoided by making one to many and many to one communication possible.

IX. CONCLUSIONS AND FUTURE SCOPE

Many companies are producing a variety of applications or packages which incorporated speech synthesis and recognition separately as assistive technology for physically disabled persons to increase their access to the current technology. However, few seemed to incorporate both.

In this Speech Recognition & Synthesis Tool, the both Text to Speech and Speech to Text software were combined effectively and worked as communication medium between two physically disabled persons. Now technology can be used for everyone. With certain improvement in the future, the communication between two different physical disability persons can communicate with ease like normal persons as SRST can be designed for the communication through Ethernet/LAN/Internet/Wireless in future for large distance. It can be designed to one to many communications for educational purposes. It can also design for other languages like Hindi, Punjabi etc.

REFERENCES

[1] Ronald J. Seiler (2007), Assistive Technology for Individuals with Cognitive Impairments, The Idaho Assistive Technology Project, Center on Disabilities and Human Development, University of Idaho, Moscow, December 2007

[2] Access Technology Training-What is accessed technology? Royal National Institute of Blinds, Retrieved on 10th January, 2012 from http://www.rnib.org.uk/professionals/solutionsforbusiness/trainingandconferences/technology/Pages/access_technology_training.aspx

[3] 'Accessibility technology', Wikipedia online encyclopedia, Retrieved on 10th January, 2012 from http://en.wikipedia.org/wiki/Accessibility

[4] Kumar,Kuldeep & Aggarwal,R.K. (2011) "Hindi Speech Recognition System Using HTK", International Journal of Computing and Business Research, ISSN (online):2229-6166, Volume 2 Issue 2 May 2011.

[5] Higgins, E.L., & Zvi, J.C. (1995). Assistive technology for postsecondary students with learning disabilities: From research to practice. Annals of Dyslexia, 45: 123-143.

[6] Wetzel, K. (1996). Speech-recognizing computers: A written-communication tool for students with learning disabilities? Journal of Learning Disabilities, 29(4): 371-380.

[7] Schreier, E.M.*, Levanthal, D.D., Uslan, M.M. (1991). Access technology for blind and visually impaired persons. Technology and Disability,1 (1), 19-23.

[8] What is Accessible Technology Retrieved on 12th Januray, 2012 from http://accessibletech.org/ access_articles/ general/whatIsAccessibleEIT.php

[9] 'The Microsoft Projects on Speech Processing' Retrieved on 18th February, 2012 from http://research.microsoft.com/en-us/groups/srg

[10] Dragon Naturally Speaking,'Helping All Students Reach Their Full Potential, A White Paper for the Education Industry from Nuance Communications, March 2009

[11] Raskind, Marshall H. and Higgin, Eleanor L. "Speaking to read: The effects of speech recognition technology on the reading and spelling performance of children with learning disabilities"Annals of Dyslexia, Volume. 49, Number 1/December 1999.

[12] Voice Recognition for Blind Computer Users, Retrieved on 25th February, 2012 from http://www.abilitynet.org.uk/content/factsheets/pdfs/ Voice%20Recognition%20for%20Blind%20 Computer % 20Users.pdf

[13] Alias, Frank Soong Microsoft Speech Research Group, Retrieved on March 12, 2012 from http://research.microsoft.com/en-us/groups/speech/ default.aspx #people

[14] .NET Framework Conceptual Overview, Microsoft Developer Network Platform, Retrieved on March 10, 2012 from http://msdn.microsoft.com /enus/ library/w0x726c2%28v =vs.90%29.aspx

[15] 'Getting Started with C#', Microsoft Developer Network Platform, Retrieved on March 10, 2012 from http://msdn.microsoft.com/enus/ library/z1zx9t92.aspx

[16] Introducing Visual Studio, Microsoft Developer Network Platform, Retrieved on March 12, 2012 from http://msdn.microsoft.com/enus/ library/fx6bk1f4.aspx

[17] Dunn,Michael (2007) Give Applications a Voice, Speech synthesis and recognition in .NET, Tech Brief Articles, Retrived on 14th March, 2012 http://reddevnews.com/articles/2007/02/15/give-applications-a-voice.aspx?sc_lang=en

[18] Brown, Robert. Exploring New Speech Recognition And Synthesis APIs In Windows Vista, Talking Windows, MSDN Magazine, http://msdn.microsoft .com/hi-in/magazine/ cc163663 (en-us).aspx