MULTIMEDIA COURSEWARE FOR VISUALLY IMPAIRED (MCVI)

MOHAN A/L ARJUNAN

A report submitted in partial fulfilment
of the requirements for the award
of the degree of
Bachelor of Computer Science (Software Engineering)

Faculty of Computer Systems & Software Engineering
Universiti Malaysia Pahang

30th MAY, 2011

ABSTRACT

Multimedia Courseware for Visually Impaired (MCVI) is a speech recognition system that is aimed for visually impaired people who has low vision but have the speaking and listening ability. The objective of this system are to develop a simple and easy prototype of Speech Recognition learning application that reacts to speech commands and to help the visually impaired people or students in learning process through hearing, speaking and touching since hard learning materials are infective and required man power to guide them. The technique that had been used to develop this system was Linear Predictive Coding (LPC). The methodology that had been used was Rapid Application Development (RAD). This system enables visually impaired people or students to get accessed to computers and learn about things, lives an environment surroundings which to make them familiar and sense living and non-living things exist around them by hearing.This thesis discussed on the preparation and analysis that had been done in order to develop this system. A complete reference and research details had been inserted in this thesis. This thesis will be a quick reference to refer on the specification and requirements of the system.

Keywords: Multimedia Courseware for Visually Impaired, Linear Predictive Coding (LPC).

ABSTRAK

Sistem pembelajaran multimedia untuk orang yang bermasalah penglihatan ini ialah satu sistem pembelajaran yang berasaskan pengenalan suara. Ia fokus kepada pengguna yang mampu bertutur dan mendengar. Objektif sistem ini dibangunkan ialah untuk menghasilkan satu system pembelajaran prototaip yang mudah dimana ia dapat bereaksi berdasarkan kepada suara input untuk membantu pelajar atau orang dewasa yang menpunyai masalah penglihatan atau buta kerana media cetak tidak sangat berkesan sebagai bahan pembelajaran untuk mereka dan memerlukan tenaga pengajar. Sistem ini diharap dapat membantu golongan bermasalah penglihatan untuk mengakses komputer dan belajar tentang benda-benda hidup dan bukan hidup disekitar mereka bermediumkan bunyi yang dihasilkan daripadanya. Jadi, mereka boleh membiasakan diri mereka pada masa hadapan dengan objek-objek di sekitar mereka dengan menggunakan naluri pendengaran. Theses ini membincangkan persediaan dan analisis yang telah dibuat untuk menbangunkan system ini. Rujukan dan kajian yang lengkap diselitkan dalam thesis ini. Thesis ini dapat menjadi bahan rujukan yang mudah dan cepat untuk rujukan spesifikasi dan keperluan system.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

According to the World Health Organization (WHO), range between 20/200 to 20/400 is considered severe visual impairment, or severe low vision, Secondly, 20/500 to 20/1,000 is considered profound visual impairment, or profound low vision, Thirdly, more than 20/1,000 is considered near-total visual impairment, or near total blindness, Finally, no light perception is considered total visual impairment, or total blindness. [1]

Visually impaired people use other sense organs such as touching, smelling and hearing to stimulate the brain to understand the surroundings Object recognition is a problem for them. In learning aspect, hearing is very important method to input the information for them. According to the National Network for Child Care (NNCC) and The Americans With Disabilities Act, to help a blind kid, the guide can provide activities with sensory experiences. Children with visual disabilities learn through hearing and touch. Sand and water play, collages, play dough, and finger painting are good learning activities. In addition, guidance can read aloud stories that have a

predictable story line. Stories that offer interesting descriptions of actions or objects are suitable them [2]. According to the Eyeway Organisation, if a person became blind at the age of nine or later, he or she has certain visual memory. They would perceive ideas and concepts differently from someone who became blind at birth. In addition, a blind person can similarly benefit if somebody reads to them [3]. Academician should not modify academic standards for visually impaired students. All students must meet the required level of understanding and performance competencies for the course, although visually impaired students may need to be modifications in the evaluation or testing method. Students with visual impairments are constantly challenged by classroom instructional strategies. Although they can easily hear lectures and discussions, it can be difficult for them to access class syllabi, textbooks, overhead projector transparencies, maps, videos, written exams, demonstrations, and films. A large part of traditional learning is visual, fortunately, many students with visual disabilities have developed strategies to learn. [4]

There are ranges of visually impaired person. Partially sighted students between 70 and 80 percent of all legally blind persons have measurable vision. Partially sighted students may use readers, audio taped texts, raised line drawings, etc. In addition, the partially sighted student may use large print books, closed circuit TV magnifier, or a large print typewriter. There are two basic difficulties that the partially sighted student is confronted with that the blind student is not. First, the partially sighted student is sometimes viewed by instructors and classmates as "faking it." Since this student does not use a white cane or guide dog, people have difficulty believing that the student needs to use adaptive methods when using print materials [6]. Another difficulty is more subtle, the psychological response that large print evokes in the sighted reader. Large handwriting may give the sighted reader the idea that "a child has written this." Instructors sometimes assume the partially sighted student is only trying to make an assignment longer. For the blind students, most blind student use a combination of methods including readers, audio taped books and lectures. Students may use raised drawings or diagrams, charts, and illustrations, relief maps, three-dimensional models of

physical organs, shapes, etc. Talking calculators and reading machines are beneficial to students, although not often readily available [5].

There are many multimedia courseware in the market for various age of user. Instead of considering user's age, user's physical conditions and disabilities factor also have to be taken under consideration where disable people also can use the system. In this learning system, hearing and speaking method is used to communicate with the user. This Multimedia courseware for visually impaired (MCVI) is focusing on hearing where speech technology (speech to speech) which had been developed by using Microsoft Speech Application Programming Interface (SAPI) since user is unable to see the computer screen clearly.

1.2 Problem Statement

Blind or visually impaired students face difficulties in learning process. Books or other hardcopy materials do not help much since it is in printed form. In computer based task or learning, blind people using Braille keyboard in order to communicate with the computer. Unfortunately, not all users are knowledgeable in using Braille keys to type. In addition, manually man power are needed to guide the visually impaired students in learning process. Therefore, to help the blind students in learning and problem solving, hearing and speaking method is choose for this multimedia learning process. This method is easy communication since user does not need to navigate the computer by using normal keyboard or mouse.

i. Visually impaired does not gain much benefit from the hardcopy or printed materials.
ii. Not all students are knowledgeable in Braille.
iii. Man power needed all the time to guide the visually impaired.

1.3 Objective

i. To develop a simple and easy prototype of Speech Recognition learning application that reacts to speech commands.

ii. To help the visually impaired people or students in learning process through hearing and speaking and touching.

1.4 Scope

System:

i. The tool to develop this learning system is Microsoft Visual Studio 2010 (VB.Net 2010).

ii. Windows Speech Application Programming Interface (SAPI) as the programming engine.

iii. Sounds produced by animals, musical instruments, nature, transport, etc were included as the learning subject and for project presentation purpose (prototype). Any learning can be done.

iv. Theory and exercise are the sub modules in this learning system. For example in Animal theory module, system teaches the user about the animal information, to spell and pronounce the animal's name. System is also produces animal's sound.

v. For example, in Animal exercise module, user needs to talk by command in order to answer the system based on what the user has listened in theory module. User will be asked about the basic features about the

animal, to spell and pronounce the animal name, and identify the particular animal by sound produced.

vi. Module that involving touch screen also included to test the primary user on Directions and Counting. This method stimulates user touching and hearing sense.

User:

i. Visually impaired (partially sighted and blind) people as the primary user.

ii. Normal user (tutor/assistant) as the secondary user to guide the new primary user on how to use the courseware.

Iii User cannot be mute or deaf or both since speech communication are used.

1.5 Organization of Thesis

This thesis consists of six (6) chapters. Chapter 1: Introduction- This chapter is the introduction about the project that had been be developed. It consists of introduction, problem statement, objective, scope, and thesis organization. Chapter 2: Literature Review- This chapter explained the case study of the project. There are 2 general structures of this study, the technique that has been used and the former system that are already created. Chapter 3: Methodology- This chapter discussed more close on the overall work flow in the development of the project. It justified the technique along with the equipment and the software. Chapter 4 - Implementation - This chapter discuss on how MCVI system had been developed in development environment structurally and logically. Chapter 5 – Result and Discussion- This chapter discussed on the results or output produced as expected and the result is further discussed. Chapter 6- Conclusion. This chapter concludes about the entire system.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter explained the case study of the project. There are two (2) general structures of this study, the technique that has been used and the former system that are already created.

Speech recognition is the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words. The recognized words can be the final results, as for applications such as commands & control, data entry, and document preparation. They can also serve as the input to further linguistic processing in order to achieve speech understanding. Speech recognition systems can be characterized by many parameters. An isolated-word speech recognition system requires that the speaker pause briefly between words, whereas a continuous speech recognition system does not. Spontaneous, or extemporaneously generated, speech contains disfluencies, and is much more difficult to recognize than speech read from script. Some systems require speaker enrollment where a user must provide samples of his or her speech before using them,

whereas other systems are said to be speaker-independent, in that no enrollment is necessary. Some of the other parameters depend on the specific task. Recognition is generally more difficult when vocabularies are large or have many similar-sounding words. When speech is produced in a sequence of words, language models or artificial grammars are used to restrict the combination of words. The simplest language model can be specified as a finite-state network, where the permissible words following each word are given explicitly. One popular measure of the difficulty of the task, combining the vocabulary size and the language model, is *perplexity*, loosely defined as the geometric mean of the number of words that can follow a word after the language model has been applied. Finally, there are some external parameters that can affect speech recognition system performance, including the characteristics of the environmental noise and the type and the placement of the microphone. [7]

Figure 2.0 shows the process of how speech is converted into text in a particular speech recognition system.

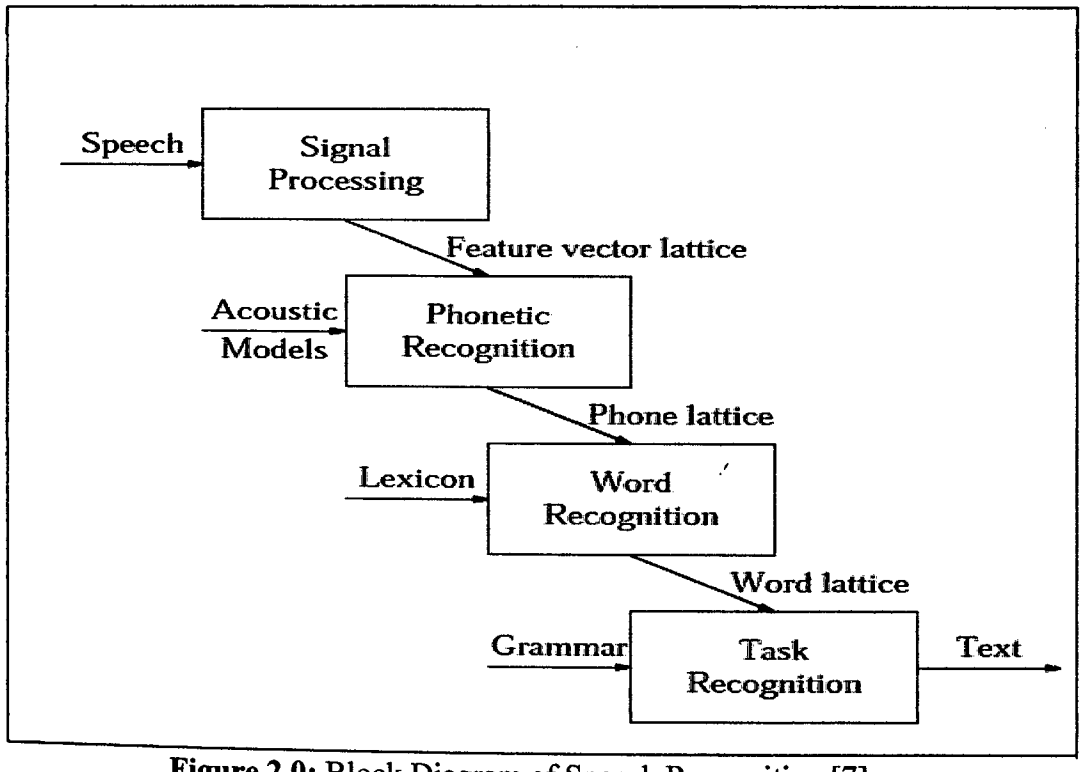

Figure 2.0: Block Diagram of Speech Recognition [7]

2.2 Existing system

Below are the few existing systems that use speech recognition technique:

i. Thai Automatic Speech recognition

ii. Alphabet Generator for Kids Using Speech Recognition

iii. Teaching Orientation and Mobility Skills to Blind Children Using Computer Generated 3-D Sound Environments

iv. Speech Recognition in Windows

2.2.1 Thai Automatic Speech Recognition

This research was performed as part of the DARPA-Babylon program aimed at rapidly developing multilingual speech-to-speech translation capability in several languages. It is built on extensive background in ASR (Automatic Speech Recognition), language portability, and speech translation, the group has built Arabic-English and Thai-English Speech-to-Speech translation systems in less than 9 months per language. This system has been used in an external DARPA evaluation involving medical scenarios between an American Doctor and a naive monolingual Thai patient. *Hidden Markov Model* has been used in this system. The objective of this system is to develop a robust and flexible Thai Speech Recognizer that can be integrated to Thai-English speech translation. [8]

2.2.2 Alphabet Generator for Kids Using Speech Recognition

Alphabet Generator for Kids Using Speech Recognition is a system that enables children to generate alphabet through their voice tone. It is a learning program for kids where alphabet will be generated according to the utterance of the particular kid. It is useful to train the kids to utter alphabets correctly. *Autocorrelation method in Linear Predictive Coding* technique has been used in this system. The objectives of this system are to develop a prototype of Speech Recognition application that generates selected capital alphabets and to implement Linear Predictive Coding approach to recognize specific alphabet through human voice. [9]

Figure 2.1 and 2.2 show the interfaces of the Alphabet Generator for Kids Using Speech Recognition.

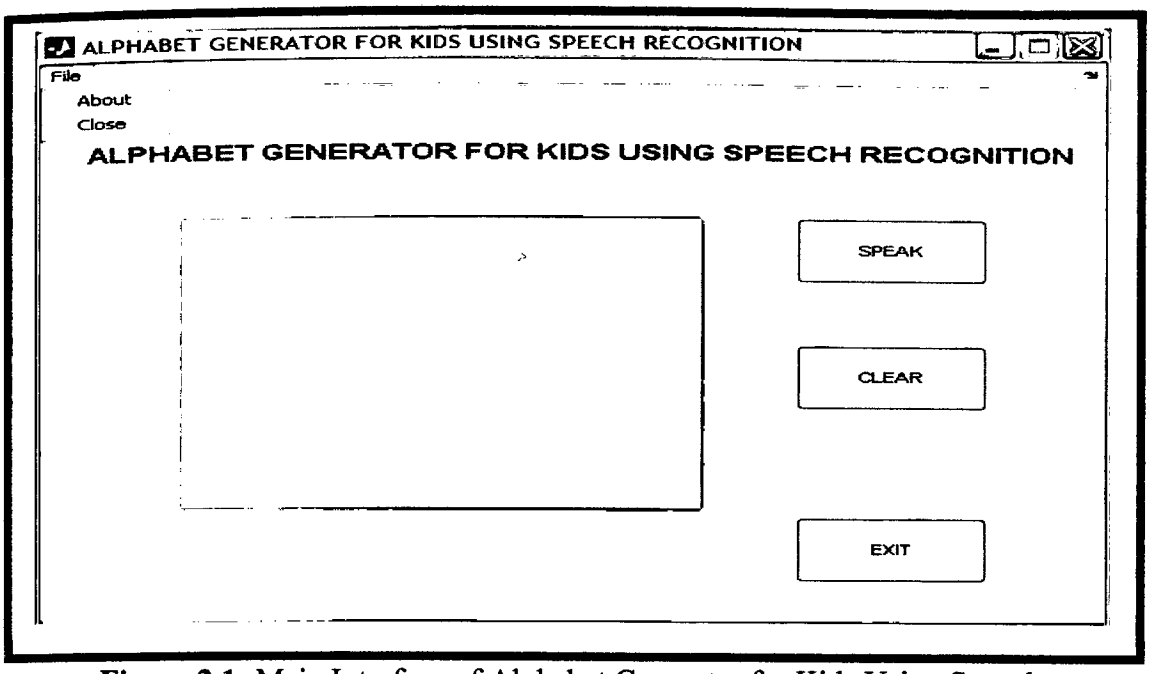

Figure 2.1: Main Interface of Alphabet Generator for Kids Using Speech Recogniton

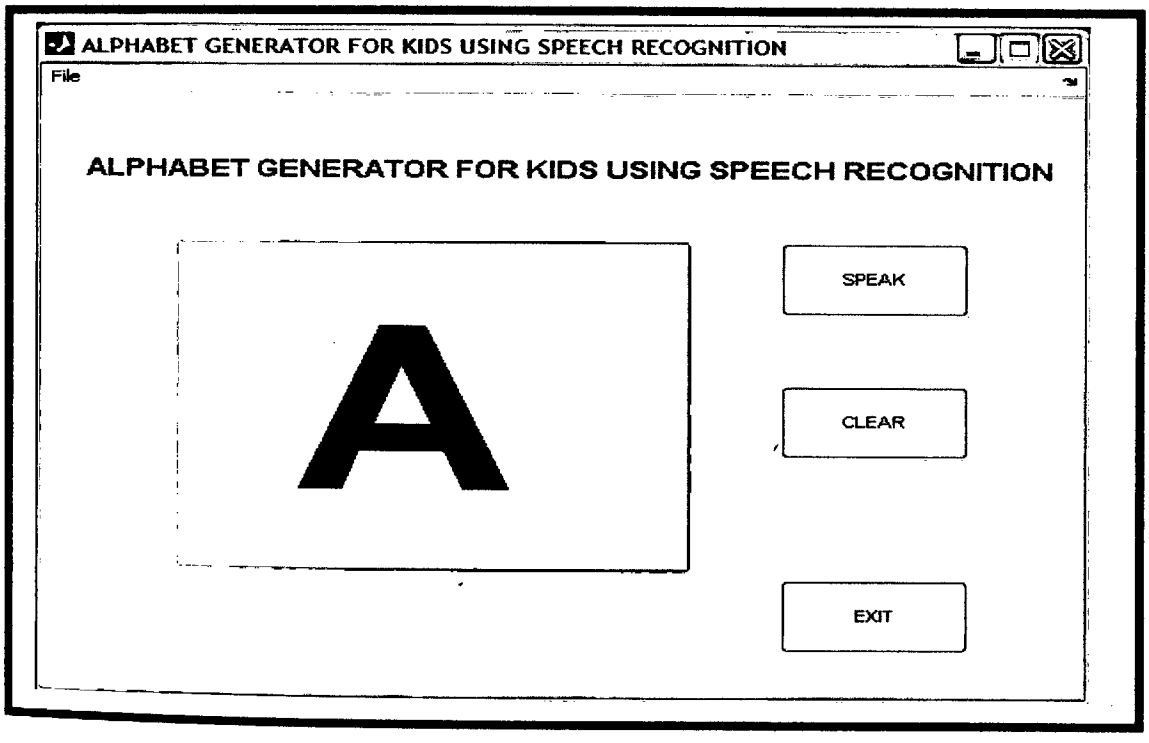

Figure 2.2: Result generated for 'A' utterance

2.2.3 Teaching Orientation and Mobility Skills to Blind Children Using Computer Generated 3-D Sound Environments

This system describes a computer program designed to teach orientation and mobility skills to visually impaired persons. The system utilizes off-the-shelf computer hardware and a proprietary virtual reality authoring library to create 3-D spatial audio environments. Simulation environments currently being developed emphasize sound identification, localization, and tracking skills which are requisite to effective orientation and mobility in real world settings. Data describing how well a blind individual is able to identify, localize, and track various sound sources within each virtual training environment will eventually be collected and correlated with actual orientation and mobility performance data. The technique used is Simulation Performance Platform (SPP). The objective of this system are to teach orientation and mobility skills to visually impaired persons and to create a series of three-dimensional sound environments with special features built-in which are designed to teach orientation and mobility skills to children who are blind. [10]

Figure 2.3 shows a screen shot of a prototype orientation and mobility acoustic training environment. In this example, three sounds are available for playback. The arrow over the "Horse" icon indicates that the OM training specialist has selected this sound for activation. The icon in the center of the screen represents the orientation of the user's head, and the line extending from the head icon is the actual direction the head is pointing. Each of the sound sources can be activated or moved manually or automatically. Student performance data are automatically recorded so they can be plotted or correlated with the actual position and orientation of any sound generating source within the environment. Each square of the grid represents 2 meters to scale.

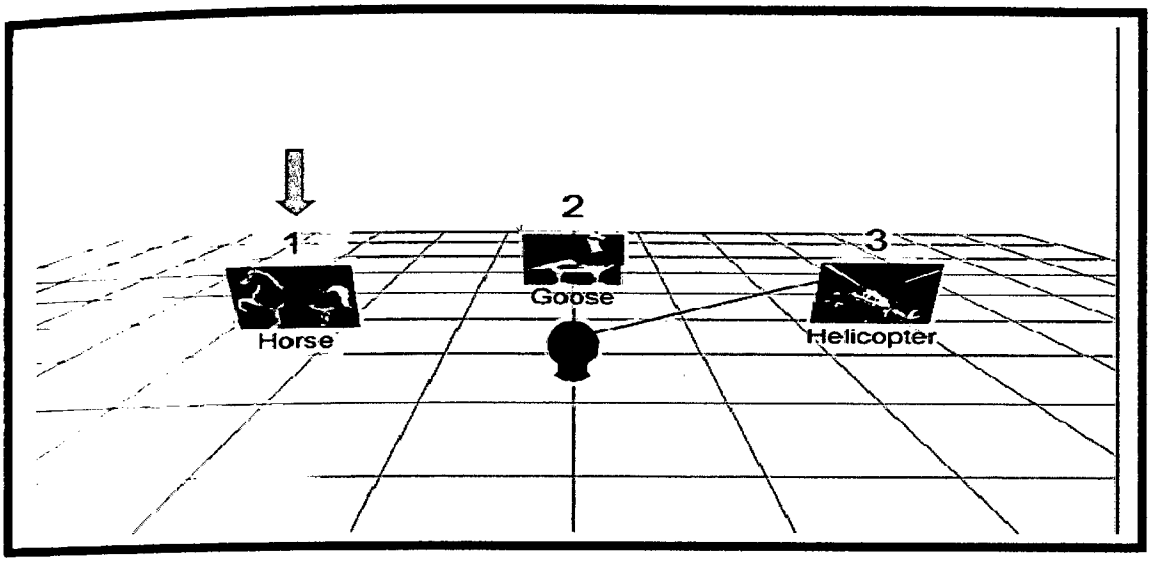

Figure 2.3 Mobility acoustic training environment [10]

2.2.4 Microsoft Speech API in Windows

The *Microsoft Speech Application Programming Interface* is developed by Microsoft to allow the use of speech recognition and speech synthesis within Windows applications. Applications that use SAPI include Microsoft Office, Microsoft Agent and Microsoft Speech Server. There have been two main families of the Microsoft Speech API. The first is the SAPI versions 1 to 4 which are all similar to each other with extra features in each newer version. The second main is SAPI 5 which has a new interface. It was released in 2000. Since then several sub-versions of this API have been released. All versions of the API have been designed in order that a software developer can write an application to perform speech recognition and synthesis by using a standard set of interfaces, accessible from a variety of programming languages. In addition, 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. [11]