Voice Recognition with Image Producer

by

Rubyatul Adawiah Binti Rosly

Dissertation submitted in partial fulfillment of
the requirement for the
Bachelor of Technology (Hons)
(Information Technology)

DECEMBER 2004

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

t
Tk
7882
·R 8a4
2004

1. Speech processing systems
2. Automatic speech recognition
3. IT/IS--Thesis

CERTIFICATION OF APPROVAL

Voice Recognition with Image Producer

By

Rubyatul Adawiah Binti Rosly

A project dissertation submitted to the

Information Technology Programme

Universiti Teknologi PETRONAS

in partial fulfillment of the requirement for the

BACHELOR OF TECHNOLOGY (Hons)
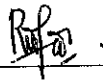
(INFORMATION TECHNOLOGY)

Approved by,

_____

(Mr. Jale b. Ahmad)

UNIVERSITI TEKNOLOGI PETRONAS

TRONOH, PERAK

December 2004

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or person.

RUBYATUL ADAWIAH BINTI ROSLY

ABSTRACT

Voice recognition is a process of taking the spoken word as an input to a computer program. The development of this project is to delve the usage of voice recognition in the field of education and examine how voice recognition is accomplished. The method used to dictate user voice is by using discrete voice recognition. The application has the ability to record voice and produce a text and image according to the corresponding spoken word. This application is to provide new method of learning English word and as an alternative to busy parents or teachers to teach English pronunciation to primary students when they are not around. The problem that started this study is incorrect English pronunciation and parents do not have much time to spend with their children to teach English pronunciation. Primary school students, teachers and parents are the targeted users for this application.

ACKNOWLEDGEGEMENT

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF GRAPH

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND OF STUDY

Voice recognition allows user to use his/her voice as an input device. Voice recognition used to dictate text into the computer or to give commands to the computer such as opening application programs, pulling down menus, or saving work.

Older voice recognition applications require each word to be separated by a distinct space. It allows the machine to detect and determine which part to begin or stop. This style of dictation is called discrete speech. The first software-only dictation product for PC's, Dragon Systems' DragonDictate for Windows 1.0, using discrete speech recognition technology, was released in 1994. Discrete speech is slow, unnatural means of dictation, requiring a pause after each and every word [9].

Continuous speech voice recognition applications allow users to dictate text fluently into the computer. These new applications can recognize speech at up to 160 words per minute. While this voice application takes place, user still can do other tasks because they are hands-free.

Voice recognition uses a neural net to "learn" to recognize user voice. The software remembers the way users speak even though different person has different accents and inflection. Voice recognition ability has been improved for the past few years. However, some users still got the problem with its accuracy maybe because of the nature of their voice and the way they pronounce the words itself.

The style of dictation that is going to be used in this project is discrete speech. This application produces text and image. The idea of this project started because primary students are expected to be able to listen and speak in English during their primary school time. Primary students especially who come from rural area cannot speak English very well and they have pronunciation problem in English word.

1.2 PROBLEM STATEMENT

The development of this project is to settle few problems. The first problem is that primary student cannot pronounce English word correctly. How are they going to improve their pronunciation in English when they come from non-English speaking family and their English teacher can guide them only eight hours per week? It is hope that they can pronounce English word correctly and if they do, it encourages them to speak English confidently. According to the questionnaires that have been conducted to 35 respondents, incorrect English pronunciation is the second biggest problem after lack of English vocabulary.

The second problem is that parents do not have much time to spend with their children to teach English and if they are willing, do they can pronounce the English word very well? This problem experienced by most of primary students when they learn English lesson. If they are lack of vocabulary, they always can refer to the dictionary but how they are going to settle pronunciation problem when their parents and teachers are not around.

1.2.1 Theoretical Framework

Theoretical framework is a concept model that makes the logical sense of the relationship among the several factors that have been identified as important to primary students' problem in learning English lesson. To describe it in more simple word, it was drawn in schematic diagram (figure 1.1) as in below. Incorrect English pronunciation and busy parent has been identified as dependant variables, which are the main focus of this research. The independent variables are short English class hour, lack of speaking practices and parent leisure time that influences all dependent variables in negative way. Parent accent is the moderating variable which has a strong contingent effect on the independent variable which is parent leisure time and all the dependent variables. The intervening variable is no environment encouragement which helps to conceptualized and explained the influent of the independent variable (lack of speaking practices) on the dependent variables. It affected relationship between dependent and independent variables between time 1 and 3 that is at time 2.

3

1.2.2 Schematic Diagram

Time 1 ──────────────→ Time 2 ──────────────→ Time 3

```
┌─────────────────┐                                    ╭──────────────╮
│ Short English    │                                    │  Incorrect   │
│   class hour     │                                    │   English    │
│                  │                                    │ pronunciation│
└─────────────────┘                                    ╰──────────────╯

┌─────────────────┐      ╱─────────────────╱
│ Lack of speaking │ →   ╱ No environment  ╱
│    practices     │    ╱  encouragement  ╱
└─────────────────┘    ╱─────────────────╱

                        Intervening
                         Variable                        ╭──────────────╮
                                                         │ Busy parent  │
┌─────────────────┐                                      ╰──────────────╯
│ Parent leisure   │
│      time        │                                      Dependent
└─────────────────┘                                        Variable

  Independent              ◇
  Variables              ╱   ╲
                       ╱ Parent's ╲
                       ╲pronunciation╱
                         ╲        ╱
                           ◇

                     Moderating Variable
```
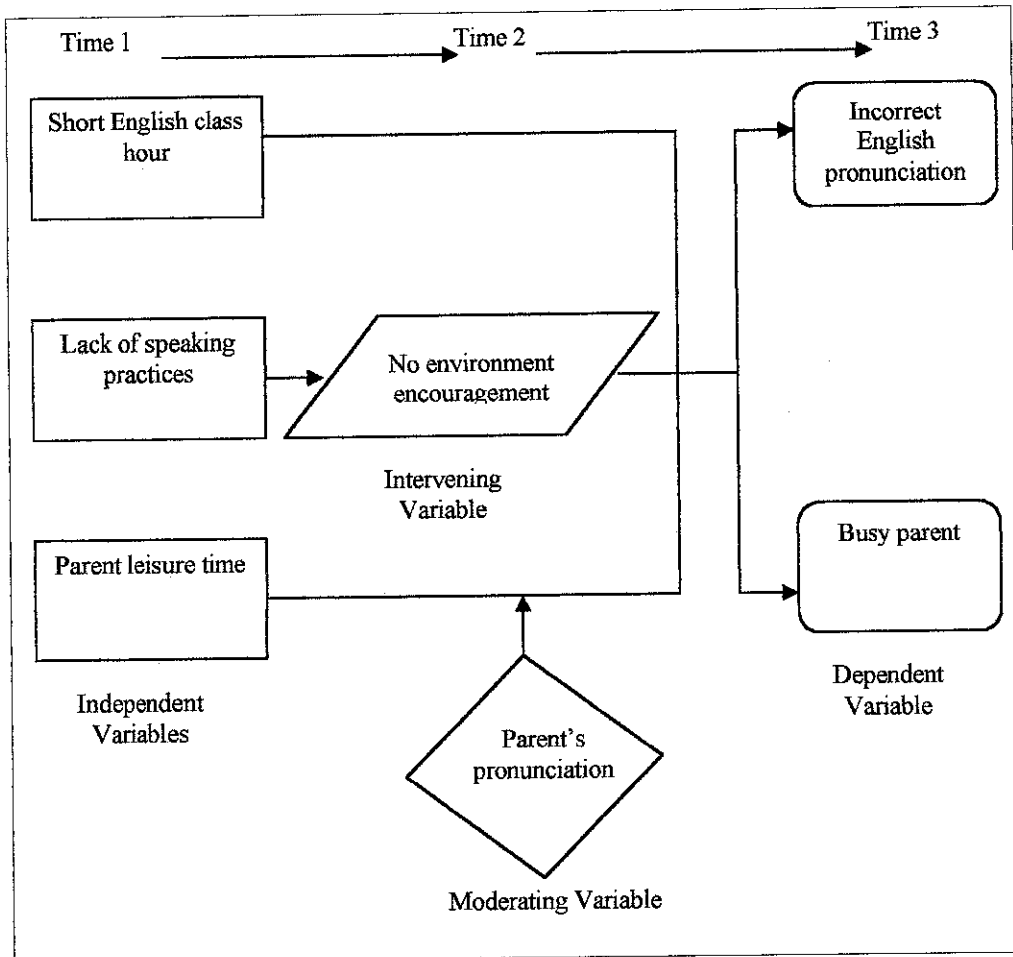
Figure 1.1: Schematic Diagram

1.3 OBJECTIVES

The first objective is to provide a new method of learning English words. It gives an alternative to users besides just by using traditional learning method. The image that was produced helps to attract users to learn English words. Parents also find that this application is easy to use and even a primary students can use it by themselves without guidance from the adult.

The second objective is as an alternative to the busy parents or teachers when they are not around to teach English pronunciation. The interactivity used in user interface is simple and easy to understand, so that students do not need parents or teachers to be around them when using this application. From research that has been done, it is proven that voice recognition help to minimize parent's time to teach student rather by using traditional learning method.

1.4 SCOPE

The targeted users for this application are students in primary schools with aged between 7 to 9 years old, parents and English teachers. That is why words used in this system are easy to understand and simple. This application is to be used at primary school and home, where the lesson takes place. However, the environment where this application is placed should be as silent as possible. The main hardware equipments used are microphone and sound card. The words used are limited due to the time constraint. This is because the application need to be trained to recognized every word and each training takes a few days time.

CHAPTER 2

LITERATURE REVIEW AND/OR THEORY

Voice recognition is a technology by which sounds, words or phrases spoken by humans are converted into electrical signals, and these signals are transformed into coding patterns to which meaning has been assigned. The power of voice was described as a below.

According to Schacter (1999) in his analysis of newer educational technologies, the fourth and eight-grade students that have been exposed to the computer assisted instructions or integrated learning system technology or simulations and software that teaches higher order thinking or collaborative networked technologies or design and programming technologies show positive achievements on researcher constructed test, standardized tests and national tests [1]. Schacter comes to this conclusion after his study in over 700 empirical research studies, in the entire state of West Virginia [1].

Kulik (1994) drew several conclusions from his work on research technique called meta-analysis [2]. From the positives finding, students who used computer based instruction score higher compare to the students without computers in his tests of achievement, student learn more in a less time if they helped by computer-based instruction and student even showed a positive attitude in their class if that class provides computer-based instruction.

Jay Sivin-Kachala (1998) draw the same conclusion as Kulik except for their inclusive findings which report that the level of effectiveness of educational technology are effected by specific student population, the software design, the educator's role and the level of the student access to the technology [3].

The above researches are not too specific on what computer technologies that they were focused on. Let's consider below findings on the effect of computer-based voice recognition software.

Many businesses that involve in education, believe that voice recognition is futuristic. The reality is they are actually the one who responsible of integrating this technology in nowadays class room. In fact, the likelihood of voice recognition technology courses replacing keyboarding courses is very probable by the year 2000. While this statement may seem unbelievable to many business educators, according to Wallace (1996) Pentium computer equipped with voice recognition technology software now permits users to achieve voice input recognition speeds of 40-60 words per minute in far less time than a person can develop comparable keyboarding skills [10].

According to Joyner (1993), postsecondary students at the end of one semester of keyboarding instruction using a microcomputer during the input process will keyboard at an average of 36 gross words per minute [11]. For a long time, keyboarding or typewriting has been a focal point of the business education curriculum. Based on the Joyner keyboarding speed facts, it is clear that voice recognition user able to input data more efficiently compare to typewriting or keyboarding.

Some students have difficulty to record their thought by doing the traditional use of paper and pencil [4]. If integration of voice recognition technologies included as part of their writing curriculum, it does benefits the writing skill of at-risk ninth grade students (Gary, 2001).

D. Roberts (2002) points out from his research result that one of her participant loves the voice recognition software because it cut minimize his work into half [5]. The used of voice recognition had increased attraction in the private sector. It changed the normal way of physically keying in the information of the data entry [6]. Speed command held an advantage over the keyed data entry (Dampier, 1996).

By referring to Kletke (1996), the continual improvement in both hardware and software of voice recognition, were describe as removing the gap between man and machine [7]. A research of eight to thirteen years old students in Ireland by McTear and O'Hare (1999) indicate that voice or speech recognition software might helpful for the physically challenge and it does merit additional study [8].

Not much data has been found about English learning ability among primary students in Malaysia. However, an interview has been done to several teachers of Sekolah Rendah Kebangsaan Hashim Awang, Kepala Batas (2004). According to them, the most common problem that students experience when studying English are lack of English vocabulary, incorrect English word pronunciation and no full time guidance from teacher and parent and short English class hour that is only eight hours per week [15]. They added that student can always improved their English skill by doing exercises regularly and speak English a lot.

CHAPTER 3

METHODOLOGY / PROJECT WORK

3.1 PROCEDURE IDENTIFICATION

In this project development, Waterfall Model (figure 3.1) had been chosen. However, the Waterfall Model that was learned in theory has been modified so that it can be adapted with real project. This type of model is chosen because this project is not a simple linear model but involves a sequence of iterations of project development activities. This model always allows referring back the previous stages if mistake has occurred.
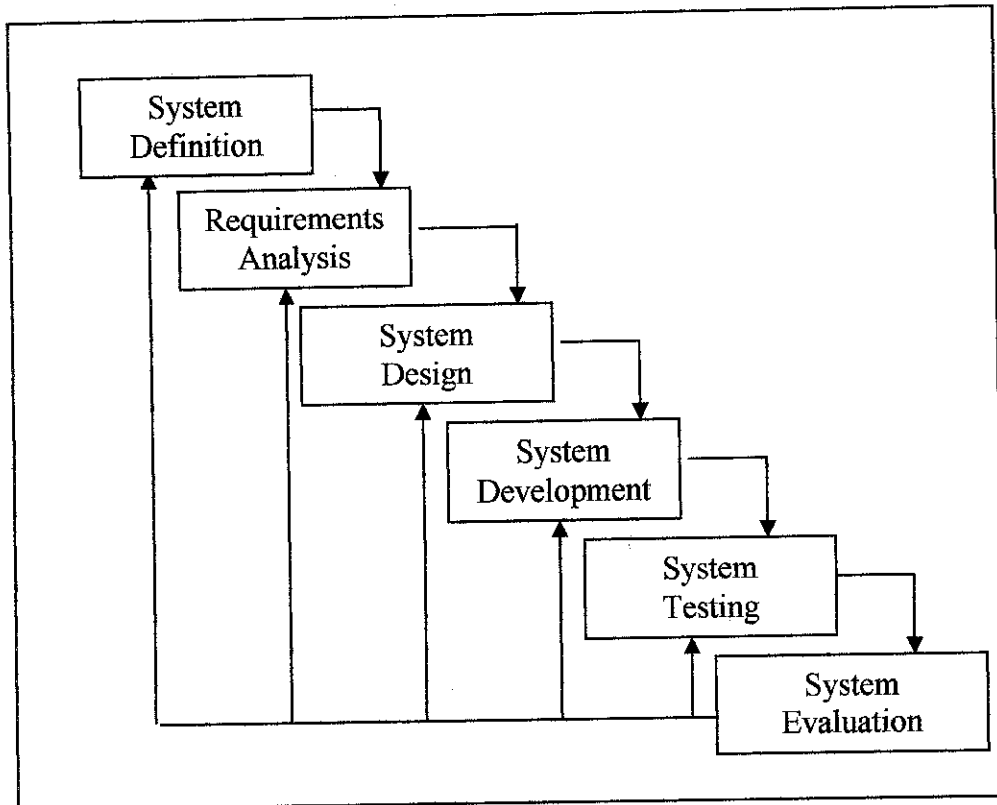
Figure 3.1: Waterfall Model

3.1.1 System Definition

In this stage, it is concerned with identifying problems, opportunities and objectives of the project. These three important components were recognized from the scenario and case study that has been done. Problem is the first in line before we can recognize opportunities to improve the situation of student education method and its objectives. The problems were identified by doing the theoretical framework according to the problem statements. In the theoretical framework, the focus of research is the dependent variables, followed by independent variables, intervening variables and moderating variables.

The problems that started this study are incorrect English pronunciation and busy parent. The opportunities are situations that this problem can be improved through the use of computerized system where it can provide an environment to encourage student speak English regularly. The objectives are to provide a new method of learning English word and as alternative to busy parents and teachers to teach English pronunciation to students.

3.1.2 Requirements Analysis

Among the tools used to define information requirements in this project are investigating data and information from Internet, interviewing, distribute questionnaires, observing and gather information by doing research on the existing similar applications. Most of requirement analysis data is taken from the existing researches and work papers that related to the topic. Interview also has been done to a few English teachers of Sekolah Rendah Kebangsaan Hashim Awang, Kepala Batas. Questionnaires also have been distributed to the 35 respondents in order to gather related data.

3.1.3 System Design

Information collected earlier in the requirements analysis is used to accomplish the logical design of application. The data flow of system will be identified in data flow diagram. Part of them is by devising the user interface. A few diagrams have been drawn in this phase, which are data flow diagram, flow chart and use case diagram.

3.1.4 System Development

This phase is where the approved design need to be developed on the original platform and software, code and removed syntactical errors from the computer programs. Local personal computer has been identified as the main platform of system development.

3.1.5 System Testing

The purpose is to catch problems before the system is signed over to users. Any errors that occurred is debug in this phase. The application is tested to the several computers to recognize any problem that occurred. One of the problem that has been identified is that some sound card does not support Microsoft Speech SDK 4.0 which leading to Microsoft Speech SDK installation error. Other problems are when Microsoft Visual Basic 6.0 environment come to the front and Microsoft Speech SDK engine cannot be load. When this occurs, voice cannot be recognized by the application.

3.1.6 System Evaluation

A key criterion must be satisfied users and evaluators. A few checklist of system functionality was done to ensure that system really provided what user expected from it. In this last phase, the application that has been developed is presented to the evaluator.

3.2 TOOLS REQUIRED

Tools are divided into two parts which are software used and hardware used.

3.2.1 Software Used

Before this, the software that had been chosen was Matlab 6.0. However, because of the developer does not familiar with its coding, the main software is changed to Visual Basic 6.0. The other software used is Microsoft Speech SDK. Microsoft Visual Basic 6.0 is needed to write applications incorporating SAPI automation in Microsoft Speech SDK, or for compiling the Visual Basic sample code. The supported operating systems for this application are:

- Windows XP Professional or Home Edition; all language versions.

- Microsoft Windows 2000 all versions; all language versions.

- Microsoft Windows Millennium Edition; all language versions.

- Microsoft Windows 98 all versions; all language versions.

- Microsoft Windows NT 4.0 Workstation or Server, Service Pack 6a, English, Japanese, or Simplified Chinese versions.

- Windows 95 or earlier is not supported.

3.2.2 Hardware Used

The hardware requirements are:

- A Pentium II\Pentium II-equivalent or later processor at 233 MHz with 128 megabytes (MB) of RAM is recommended.

- A microphone or some other sound input device to receive the sound is required for speech recognition. In general, the microphone should be a high quality device with noise filters built in. The speech recognition rate is directly related to the quality of the input. The recognition rate will be significantly lower or perhaps even unacceptable with a poor microphone.

- Not all sound cards or sound devices are supported by SAPI 4 or 5, even if the operating system supports them otherwise.

- Recommended RAM is 300MB

3.3 DEVELOPMENT DURATION

In general the development duration took 14 weeks. Please refer to the Gantt chart in Appendix Section for further information about the system development duration.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 FINDINGS

The finding focuses on the existing similar application. Two applications have been identified and analyze which are Dragon NaturallySpeaking and ReadPlease 2003.

4.1.1 Dragon NaturallySpeaking

Dragon NaturallySpeaking has abilities to dictate memos, reports, and other documents, enter data, fill-in forms, send e-mail, and work on the Web-all by voice. It performs complex tasks on the user computer with a spoken word or phrase, customize their vocabulary and commands to match user's workflow, listen to incoming e-mail and documents read aloud, reduce the need for manual transcription, and much more. It is integrated with the full Microsoft Office suite, and Corel WordPerfect, Dragon NaturallySpeaking software works with virtually all Windows-based applications, including enterprise information management systems. The key benefits of Dragon NaturallySpeaking:

- Accurate

  Dragon NaturallySpeaking users can easily achieve accuracy rates of up to 99%. Dragon NaturallySpeaking actually gets better the more user use it, and has won over 160 awards for accuracy and ease-of-use. Includes a comprehensive general English vocabulary of 250,000 words, and it's easy to create custom vocabularies for different specialties or interests.

- Fast

- Instantly convert what users say into text three times faster than typing.

- Easy

  In just about a few minutes, user will be dictating directly into virtually any Windows-based application, including the Microsoft Office suite.

- Powerful

  Perform complex tasks on computer instantly with a spoken word or phrase. Create users own specialized commands-from simple dictation shortcuts to scripting commands that can automate virtually any process.

- Mobile

  Productive on-the-go. Import recordings from any ScanSoft-certified Pocket PC or digital handheld recorder for automatic transcription when user synch with PC. Users can also dictate using array microphones and wireless headsets, eliminating the need to be wired to the PC.

- Accessible

  Dragon NaturallySpeaking 7.3 Professional Solutions can help people with disabilities achieve maximum productivity on PCs, as well as help prevent carpal tunnel syndrome and other musculoskeletal disorders. Dragon NaturallySpeaking 7.3 Professional meets Section 508 standards, established by the US government to make software accessible for people with disabilities.

- Enterprise Ready

  Install and manage multiple seats of Dragon NaturallySpeaking Professional from a central network location.

The Key Features of Dragon NaturallySpeaking:

- Most Accurate Ever
- Fastest Ever
- Expanded input device support
- Support for array microphones and wireless headsets
- Most Accessible Ever
- Most Application Support Ever
- Enterprise Ready
- New Vocabulary Optimizer
- Improved RealSpeak Proofing
- Expanded Custom Voice Shortcuts

4.1.2 ReadPlease 2003 User Interface

The user interface of ReadPlease 2003 has been analyzed to get the clear idea of interface to be designed. Below are the findings about the ReadPlease 2003 User Interface (figure 4.1) and a description on how to use ReadPlease 2003.
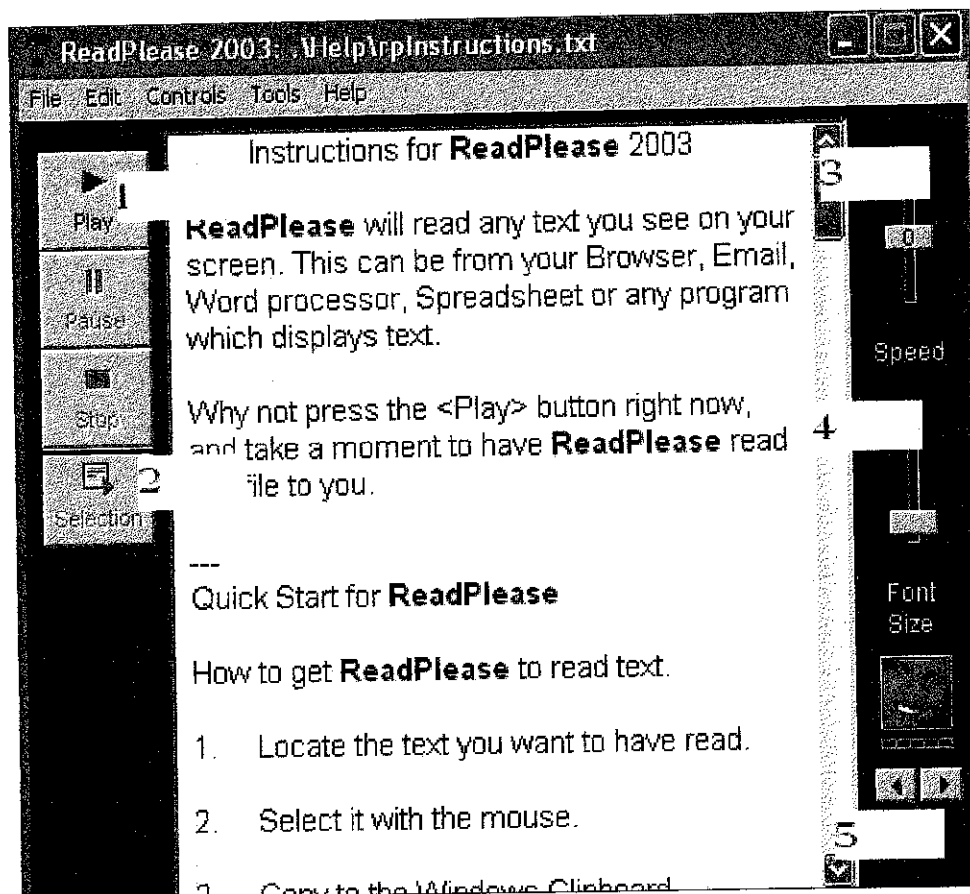
Figure 4.1: ReadPlease 2003 main menu

Functions Buttons in ReadPlease 2003 main interface:

1. VCR like control – Play, Pause, Stop
2. Play the selected text
3. Narrators voice speed controller
4. Font size controller
5. Narrators gender voice - Male or Female

ReadPlease reads any text you see on your screen. ReadPlease 2003 functionality is as below:

- User locates the text that they want to have read.
- User selects the text with the mouse.
- User copies to the Windows Clipboard.
  (Use Edit + Copy menu, or the hotkey - Control + C).
- User launches ReadPlease, and Paste it into the ReadPlease window.
  (Use Edit + Paste menu, or the hotkey - Control + V).
- User presses the green ReadPlease Play button and ReadPlaese start reads the text to the user.

4.1.3 Questionnaires result

Questionnaires have been conducted to the 35 respondents in order to find related data about voice recognition usage among primary school students. The result of the questionnaires is as the following:

Question 1

How do you rank four main problem of primary student when learning English subject?

1	2	3	4
First Rank Problem	Second Rank Problem	Third Rank Problem	Forth Rank Problem

Problem	Ranking of importance
Incorrect English word pronunciation	2
Lack of English vocabulary	1
Short English class hour (8 hours per week)	4
No full time guidance either from parent or teacher	3

19

Discussion 1

The main problem is lack of English vocabulary, followed by incorrect English word pronunciation, no full time guidance from parent or teacher and the last problem is Short English class hour that is only eight hours per week. The first problem can be solved by using a dictionary. That is why the focus of this project is the other three problems.

Question 2

During your primary schools time, how many hours your parent taught English lesson to you in a week? Circle the answer

      A. None or less than 1 hour per week

      B. 1-3 hour per week

      C. 3-5 hours per week

      D. More than 5 hours per week

Discussion 2
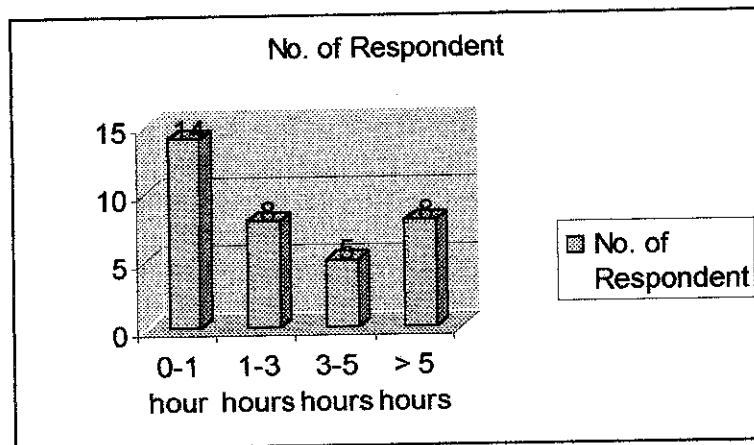
Graph 4.1: Total hour that parent spend to teach English to their children every week

Most of the respondent's parent does not spend enough time to teach them English during their primary school time which is the time used only from 0 to 1 hour in every week.

Question 3

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree

Do you agree that voice recognition application that can produce text based on the spoken word can settle English pronunciation problem?

Discussion 3
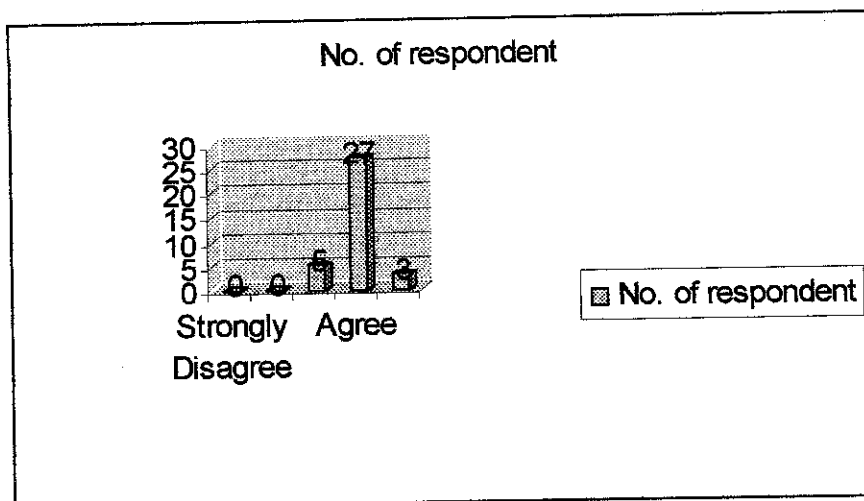
Graph 4.2: Response to whether voice recognition application can settle English pronunciation or not.

27 respondents agree that voice recognition that can produce text based on the spoken word can settle English pronunciation problem, 5 response to neither agree nor disagree and 3 response to the strongly agree. This shows that such application can settle word pronunciation problem.

Question 4

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree

Do you agree that a busy parent should use this application?

Discussion 4
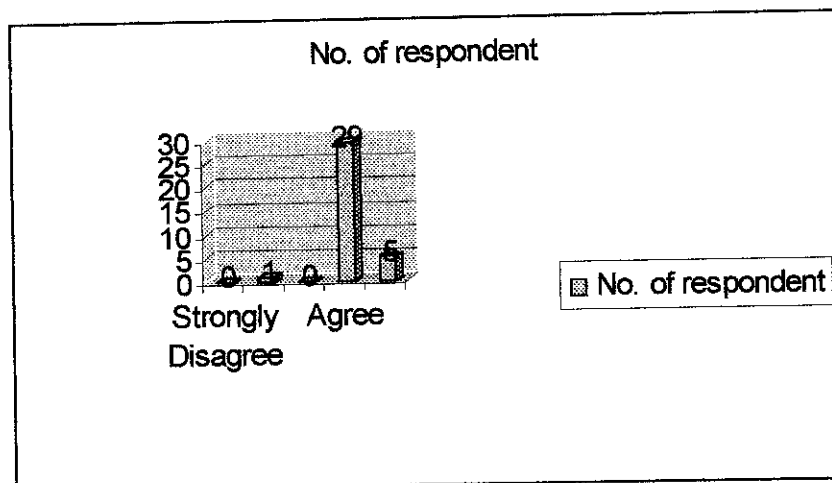
Graph 4.3: Response to whether busy parent should use this application or not

29 respondents agree that busy parent should use such application. This help to minimize the time used to teach English pronunciation to their children.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree

Do you agree that parent's time used to teach student pronounce English word by using voice recognition cut to half rather than by using traditional method?

Discussion 5
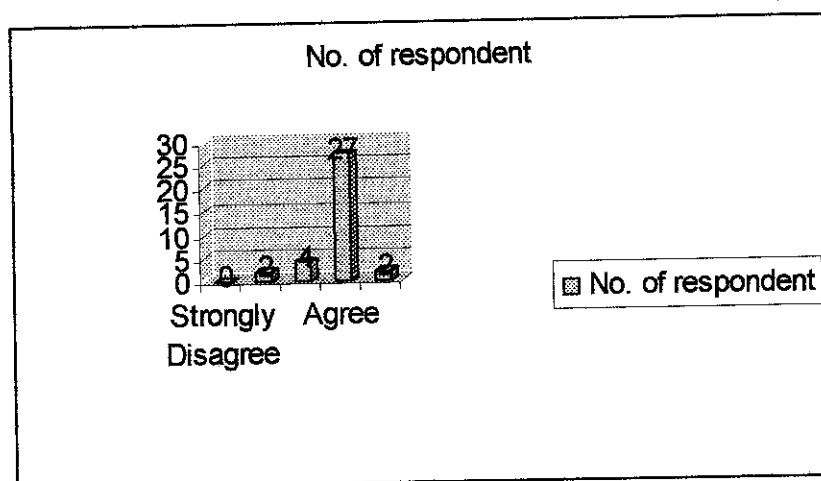
Graph 4.4: Response to whether voice recognition application can cut parents time to half or not rather by using traditional method

27 agree that such application can cut half of parent's time rather than by using traditional method.

Question 6

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree

Do you agree that this application should be used in English class at primary school?

Discussion 6
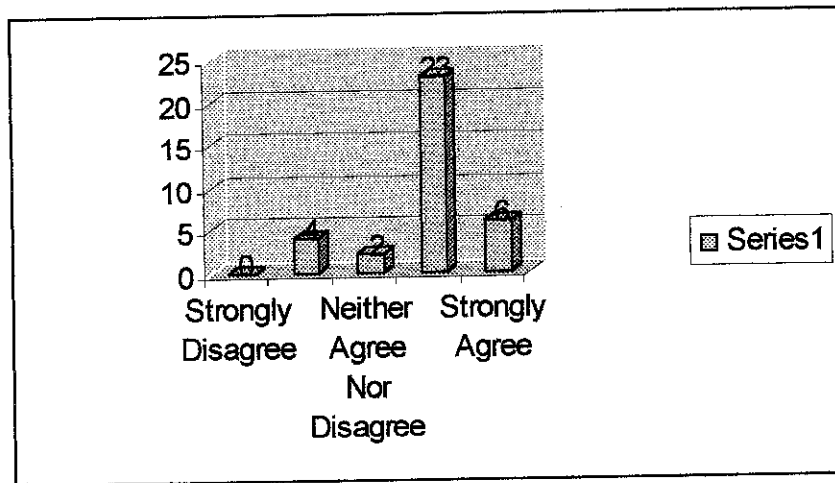
Graph 4.5: Response to whether the voice recognition application should be used at primary schools or not

29 respondents agree that this should be used in primary schools. That is what has been discussed in scope.

Question 7

What is student age that you think should be produced to this application? Circle the answer

A.   7-8 years old

B.   8-9 years old

C.   9-10 years old

D.   10 years old and above

Discussion 7
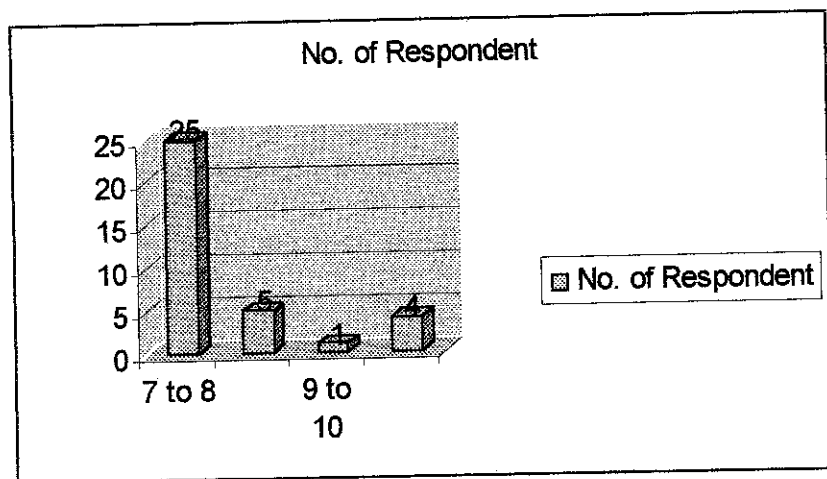
Graph 4.6: Response to which is most suitable target user of this application

25 respondents select student at age 7 to 8 years old as primary user for this application. The main targeted user for this application is primary student at age 7 to 9 years old as specify in scope section.

4.2 DISCUSSION

After analyzing the existing similar applications, the flow of the process that includes in the system has been determined. The brief explanations about this are described as below.

4.2.1 "V- Class" Use Case Diagram

The use case diagram provides a bird's-eye view of the actors which are students, teachers and parents and it provides an overall description of the external functionality of the system. Teachers and parents teach students how to use or launch the application. Students, teachers or parents then choose "V-Class" application by clicking the picture or label at main menu. If "V-Class" button is clicked, it launches the application. A virtual voice asks student what he/she would like to report the missing item in that class. Student must pronounce the list of word that is displayed. Any pronounced word appears at list box. If student say "Delete", it deletes the listed order in list box and if student say "Exit", the virtual voice added the total sum and informs the price of total loss and end the session. The image box displays the corresponding image of item.
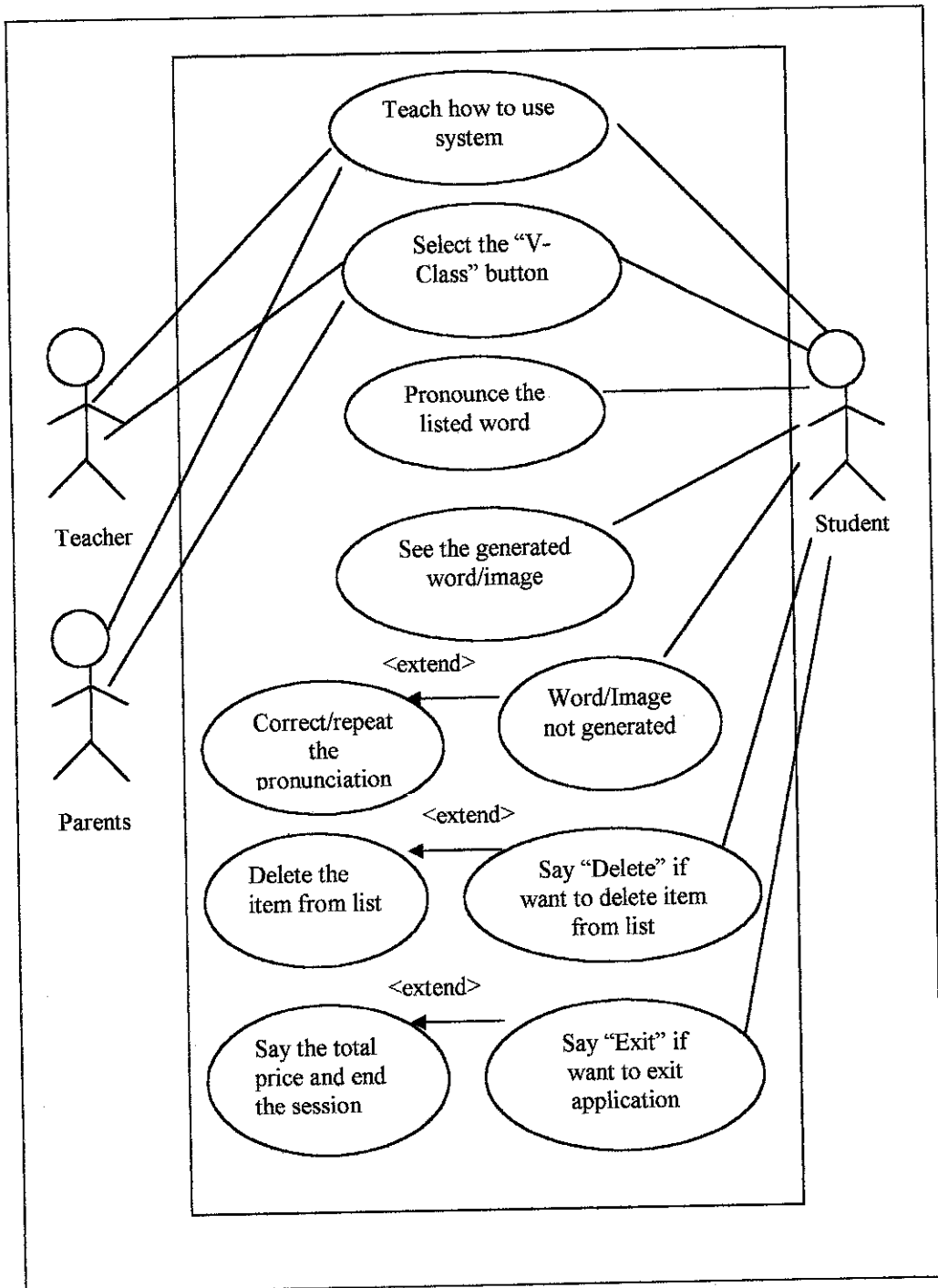
Figure 4.2: "V-Class" Use Case Diagram
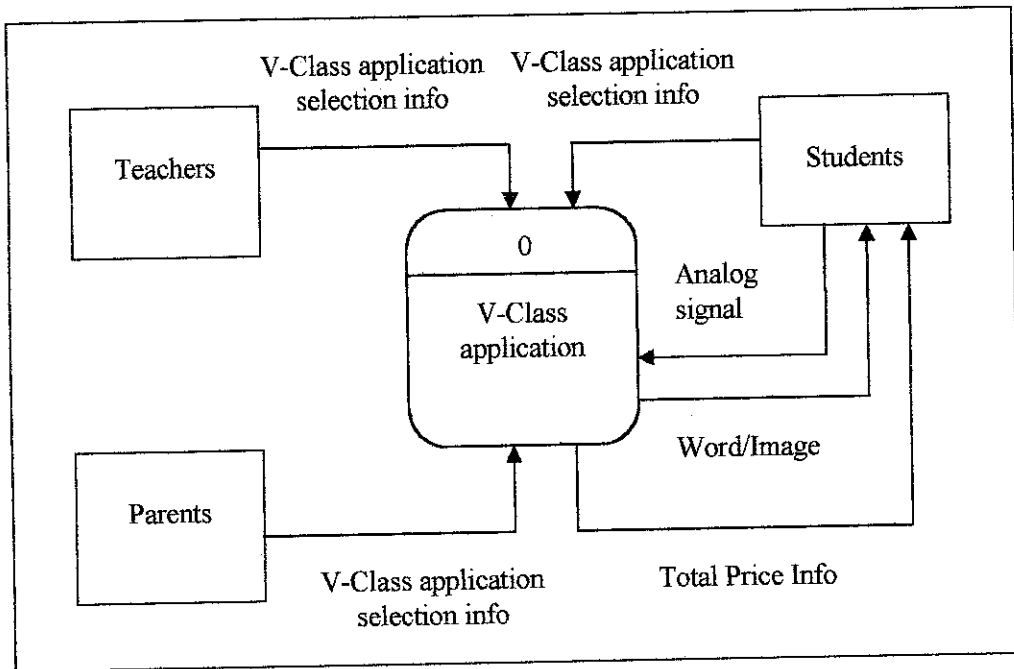
Figure 4.3: "V-Class" Context-level Data Flow Diagram

4.2.2 "V-Class" Context-level Data Flow Diagram

Data Flow Diagram focuses on the data flows in and out of application and the processing of data. In this application, teachers, parents and students input "V-Class" application selection info. Students then send analog signal data via the microphone. The application responses by produced the word or image according to the input analog signal. It also gives total price info to the user.

28

4.2.3 "V-Class" Flow Chart

After "V-Class" button is selected from the main menu, "V-Class" starts by saying: "Hello! I'm your voice teacher. A few items are missing in my class. Please pronounce the missing items, and let's count the loss together". User needs to input their voice by saying the listed items from menu. If user does not do that, "V-Class" asks again by saying: "I'm sorry. What are you saying?" or "I'm sorry. I cannot detect your pronunciation." "V-Class" keeps saying that until recognized phrase is detected. If the phrase is detected equal to words in grammar file, "V-Class" repeats the phrase, together with its price. It also produces text in list box and image in image box. "V-Class" counts the total items price. If the listed word is wrong and "V-Class" detected phrase "Delete" from user, "V-Class" delete the last word in list box and if not, "V-Class" keep asking the items. If user input "Exit" phrase, "V-Class" asks "Would you like to report another item?", if no, "V-Class" gives the total items price and if yes, "V-Class" ask items that user want to report again. After the price is produced, "V-Class" ends the session and starts the new session.
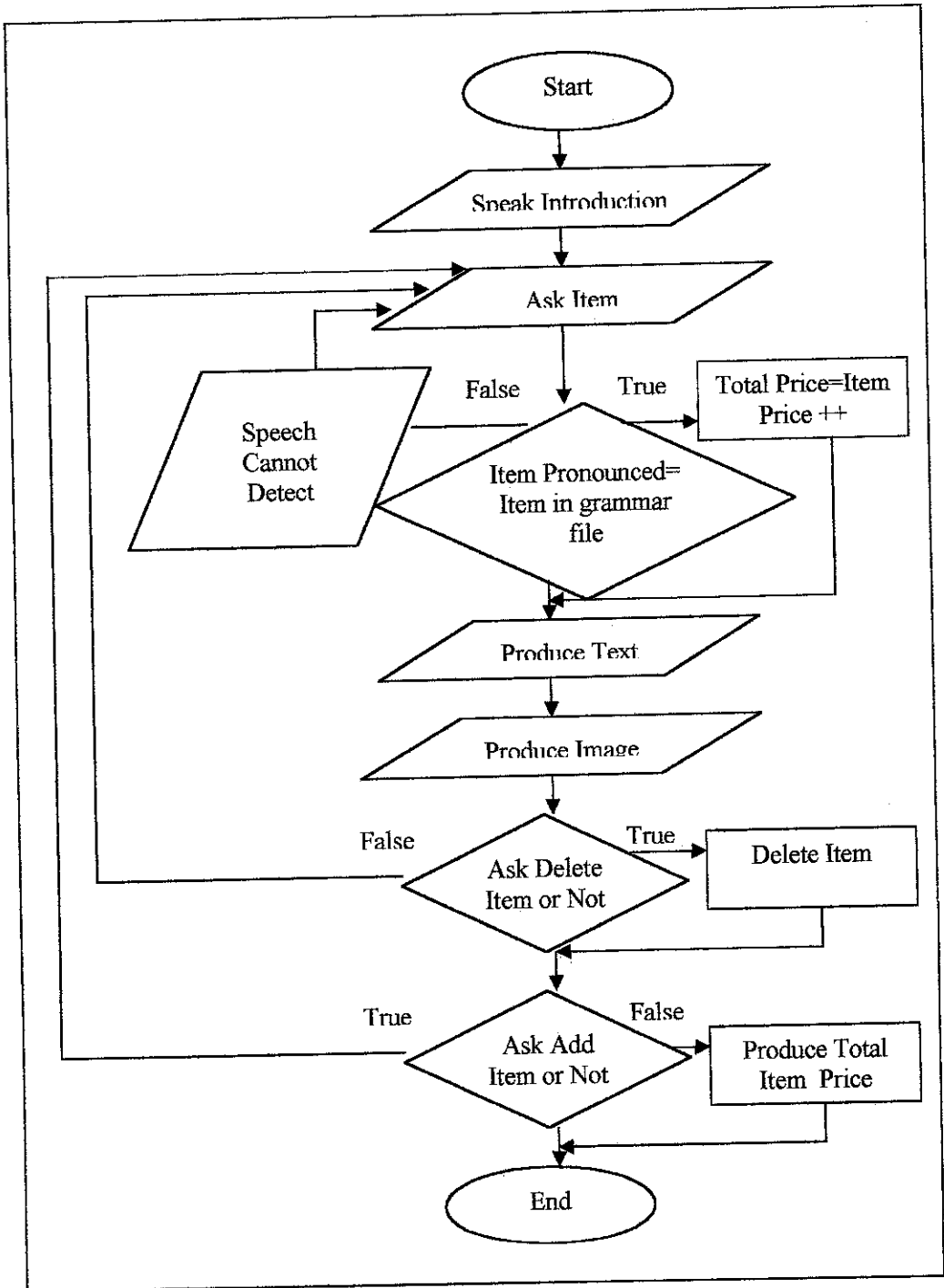
Figure 4.4: "V-Class" Flow Chart

4.2.4 "V-Class" Screen Shot
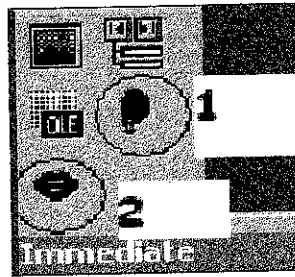
Figure 4.5: DirectSS and DirectSR tool

After Microsoft Speech SDK 4.0 is installed, two new tools are added in Visual Basic 6.0 toolbox. The tools are labeled as in above figure:

1- DirectSR

2- DirectSS

DirectSR records user phrase and DirectSS speaks the parse in double quote in the coding. Some of users sound card does not support multiplex recording and playing, and to ensure that no error is occurred, each time DirectSR is activated, DirectSS is deactivated and vice versa. That is why, while "V-Class" is speaking, it cannot detect any input voice from user.
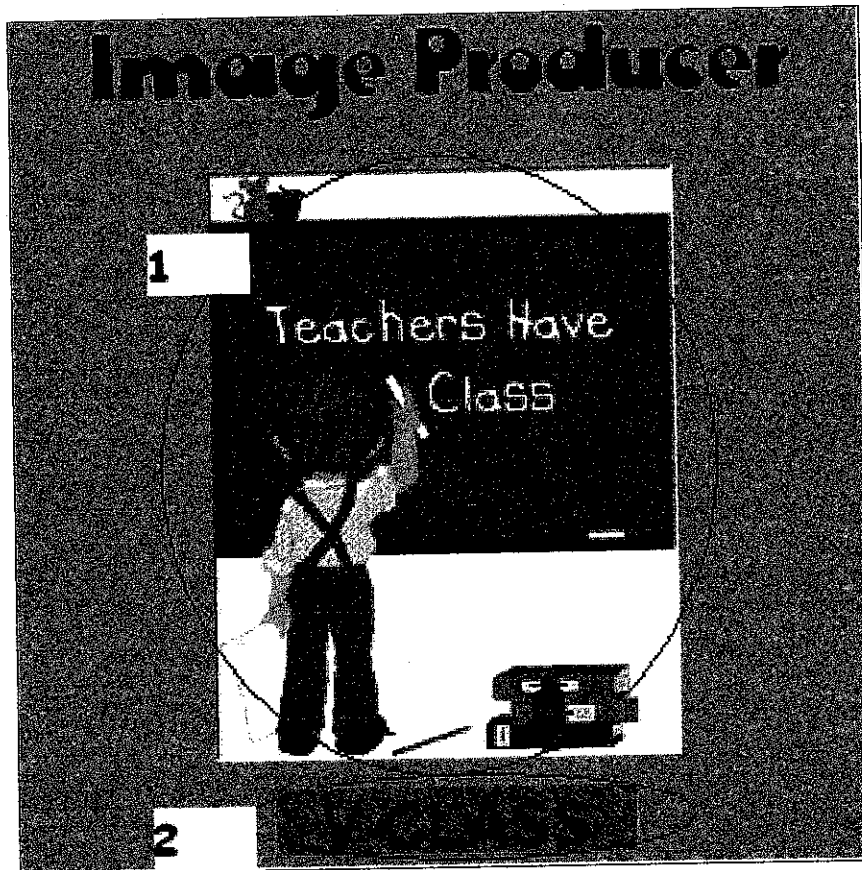
Figure 4.6: "V-Class" Main Menu Screen Shot

In "V-Class" main menu, user needs to click on picture or "V-Class" label to enter the "V-Class" application.

1- V-Class picture

2- V-Class label
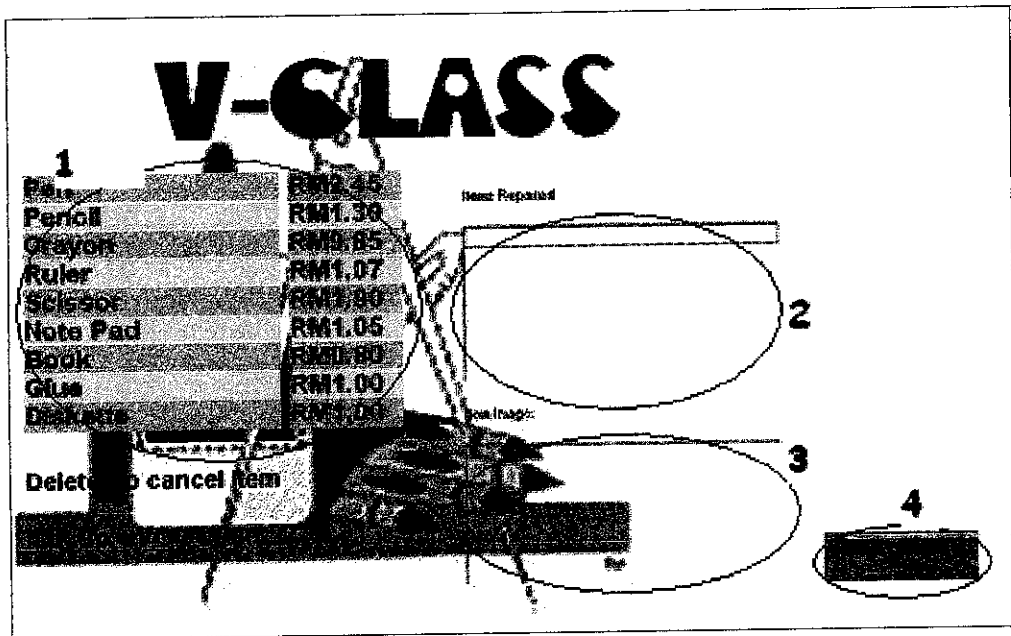
Figure 4.7: "V-Class" Application Screen Shot

User needs to pronounce the listed word in item label only. Item reported list box is to produce listed text, item image box is to display item image and to enter main menu item user must click on main menu label.

1- Item label

2- Item reported list box

3- Item image box

4- Main menu label

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 RELEVANCY TO THE OBJECTIVES

Voice recognition is a process of taking the spoken word as an input to a computer program. This process is important to virtual reality because it provides a fairly natural and intuitive way of controlling the simulation while allowing the user's hands to remain free. The development of this project is to delve into the uses of voice recognition in the field of virtual reality, examine how voice recognition is accomplished, and help to the advancement of voice recognition technology. Since that many higher education centre using voice recognition to increase the effectiveness of their learning process, it is important to young students to get familiar with this technology.

In order to give a commitment toward the carrier enhancement, many of the parents experienced the difficulty to spend time with their children. This means that the ignorance of teaching their children with English lesson also occurred. They give a hundred percent trust to teacher to teach their children without realizing that teacher also have to divide their commitment to many students. This application is an alternative to busy parents to teach their children to pronounce English words.

This application not only benefits busy parents, it also helps students to pronounce the English word correctly. It plays role as one of the interesting medium for teachers or parents to teach English word to students. The new thing especially computerized learning equipment like this will get the student attentions compare to common learning equipment.

5.2 SUGGESTED FUTURE WORK

In the future, it is suggested that this application just not focus on English lesson only. It can be expand to other subjects such as mathematic, science and Melayu Language. Although part of this application used mathematics, it is just a simple addition algorithm operation. The continuity can be made so that other mathematic operations can be added such as subtraction, division and multiplication.

The limitation of this application is that sometimes it cannot detect our pronunciation correctly although the word has been repeated a few times. It is happen because of the noise around it which is result from radio wave, hand phone reception wave, talking people and others. To solve this, a better noise filtering technology and coding should be used and added. Maybe a more high-tech microphone should be used which will filter all of these noises. Users must choose a silence environment in order for this application to function properly.

Due to the time constraint, the developer cannot add other attraction element in the prototype. If it is possible, the animation can be added in the application so that student will find it more interesting. This can increase their attention and focus to use this application in a longer time.

CHAPTER 6

REFERENCES

1. Schacter, John, 1999, "Does Technology Improve Student Learning and Achievement? How, When and Under what Conditions?," *Journal of Educational Computing Research* (20) : 9

2. Kulik A, James 1994, *"Meta-Analytic Studies of Finding on Computer-Based Instruction"*

3. Sivin-Kachala, J. 1998. Report on the *"Effectiveness of technology in schools, 1990-1997"*. Software Publisher's Association

4. James Abernethy, Gary 2001. *"The Effect of Computer-Based Voice Recognition Software on the Composition Writing of At-Risk Ninth Graders"* Department of Instructional Technology, College of Education, University of South Florida

5. D. Robert, Kelly, 2002, "Voice Recognition Software as a Compensatory Strategy for Postsecondary Students with Learning Disabilities," *Journal of International Conference on Computers in Education* 2002 : 1506

6. Dampier, R.I, 1996, "Speech Versus Keying in Command and Control: Effect of Concurrent Tasking," *Journal of International of Human Computer Studies* 45: 337-348

7. Kletke, Marilyn, 1996, "The Impact on User Performance and Satisfaction of a Voice-based Front-end Interface for a Standard Software Tool," *Journal of International of Human Computer Studies* 45: 287-303

8. Mc Tear, M & O'Hare, 1997, "Speech Recognition in the Secondary Schools Classroom: an Exploratory Study". *Journal of Computer and Education* 33, 27-45

9. Munro, Jay, 23 October 1998, "Watch What You Say." *PC Magazine Online*. 23 October 1998 <http://www.zdnet.com/pcmag/features/speech/intro1.html>

10. Wallace, I., 1996 "You can make it happen with voice recognition technologies", *Proceedings of the 13th Annual Atlantic Coast Business and Marketing Education Conference, Raleigh, North Carolina* (7): 119-124.

11. Joyner, R. L., 1993 "Postsecondary keyboarding students' speed and accuracy outcomes when using electronic equipment", *Journal of the Delta Pi Epsilon* (35): 175-189.

12. Schwalbe, Kathy, 2000, *"Information technology Project Management"* Third Edition, Boston, Course Technology.

13. E. Kendall, Kenneth, 2002, *"Systems Analysis and Design"* Fifth Edition, New Jersey, Pearson Education Inc.

14. Microsoft Corporation, 31 July 2001, "Microsoft Speech Release Notes". 31 July 2001 <http://www.microsoft.com.my speechsdk51\readme.htm>

15. English teachers of Sekolah Rendah Kebangsaan Hashim Awang, Kepala Batas, Pulau Pinang (2004).

CHAPTER 7
APPENDICES

7.1 GANTT CHART

7.2 QUESTIONNAIRES

7.3 "V-CLASS" VISUAL BASIC CODING

7.1 GANTT CHART

ID	Task Name	Duration	Start	Finish	Predecessors
3	Form 03 submission	10 days	Mon 7/19/04	Fri 7/30/04	
4	**System definition**	**8 days**	**Wed 7/21/04**	**Fri 7/30/04**	**2**
5	Define system goal	3 days	Wed 7/21/04	Fri 7/23/04	
6	Requirement analysis	5 days	Mon 7/26/04	Fri 7/30/04	
7	System specification and requirement	5 days	Mon 7/26/04	Fri 7/30/04	
8	**System design**	**15 days**	**Mon 8/2/04**	**Fri 8/20/04**	**7**
9	analyse prototype functionality	5 days	Mon 8/2/04	Fri 8/6/04	
10	Develop outline prototype design	5 days	Mon 8/2/04	Fri 8/6/04	
11	Obtain approval for construction	1 day	Mon 8/9/04	Mon 8/9/04	
12	Progress report submission	5 days	Mon 8/16/04	Fri 8/20/04	
13	**System development**	**30 days**	**Tue 8/10/04**	**Mon 9/20/04**	**11**
14	Construct the system	25 days	Tue 8/10/04	Mon 9/13/04	
15	Prepare for transition	5 days	Tue 9/14/04	Mon 9/20/04	
16	**System testing**	**60 days**	**Tue 9/21/04**	**Fri 12/10/04**	**15**
17	Test generate data	4 days	Tue 9/21/04	Fri 9/24/04	
18	User feedback	2 days	Fri 9/24/04	Mon 9/27/04	
19	Prototype revision	5 days	Tue 9/28/04	Mon 10/4/04	18
20	Prototype debugging	5 days	Tue 10/5/04	Mon 10/11/04	19
21	Final draft submission	1 day	Fri 10/15/04	Fri 10/15/04	
22	**System evaluation**	**2 days**	**Thu 12/9/04**	**Fri 12/10/04**	
23	Oral presentation	4 days	Sat 10/30/04	Wed 11/3/04	
24	Project dissertation submission	1 day	Fri 12/10/04	Fri 12/10/04	

This is a legend/key from a Microsoft Project Gantt chart, rotated 90 degrees.

Task	▭	Milestone ◆
Split	⋯⋯⋯	Summary ▬▬▬
Progress	▬▬	Project Summary ▬▬▬

External Tasks	▭
External Milestone	◈
Deadline	⇨

Project: GarYP
Date: Tue 1?

41

7/21

Task		Milestone ◆	External Tasks
Split		Summary	External Milestone
Progress		Project Summary	Deadline

Project: Gantt Chart_FYP
Date: Tue 12/14/04

8/2

8/10

9/21

Task		Milestone ◆
Split		Summary
Progress		Project Summary 53

External Tasks	
External Milestone ◆	
Deadline ⇩	

Project: Gantt Chart_FYP
Date: Tue 12/14/04

Task

Split

Progress

Milestone

Summary

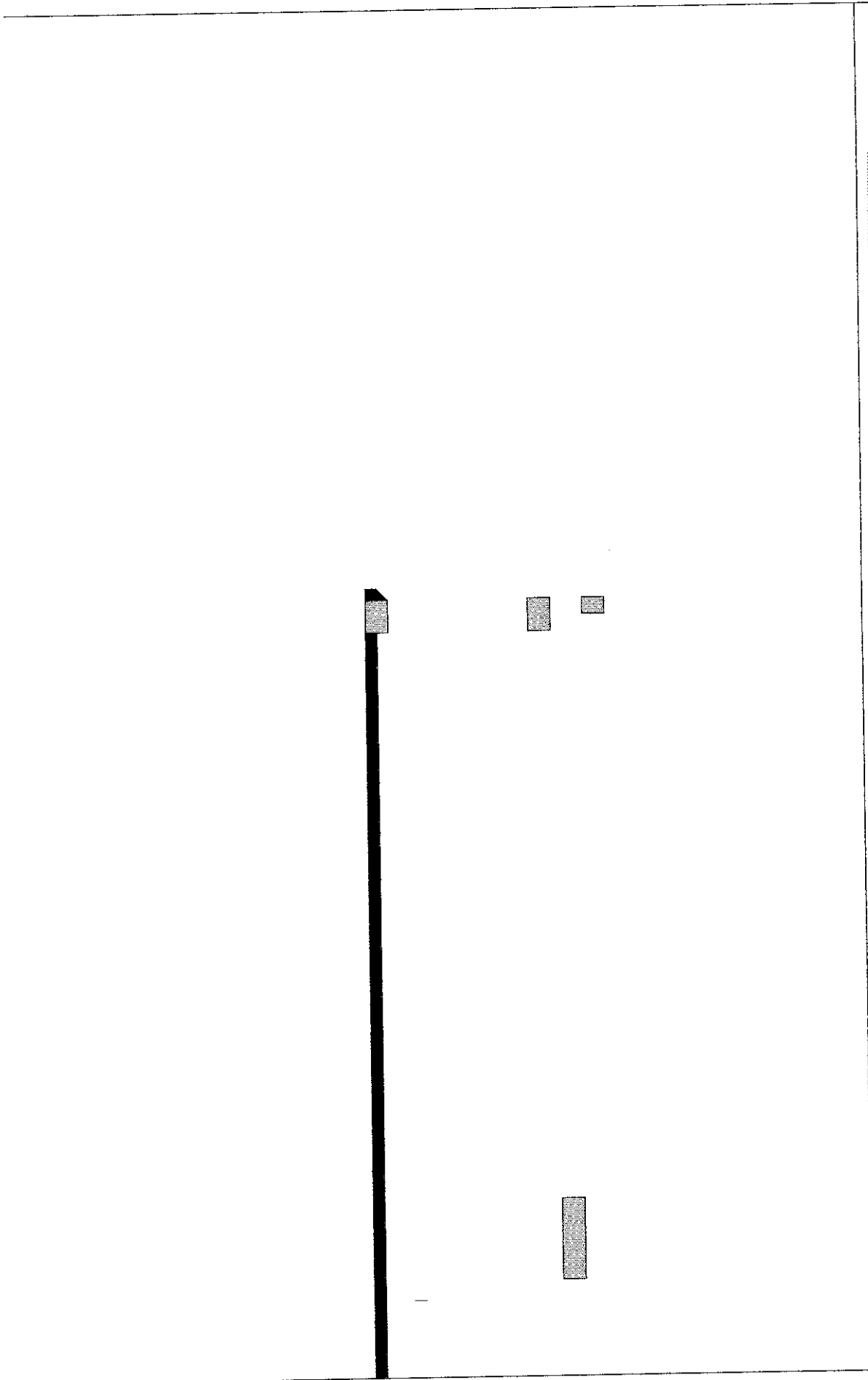Project Summary

External Tasks

External Milestone

Deadline

7.2 QUESTIONNAIRES

This questionnaire is conducted to gather information about voice recognition usage among primary school student.

1. How do you rank four main problem of primary student when learning English subject?

1	2	3	4
First Rank Problem	Second Rank Problem	Third Rank Problem	Forth Rank Problem

Problem   **Ranking of importance**

English word pronunciation   _____

Lack of English vocabulary   _____

Short English class hour (8 hours per week)   _____

No full time guidance either from parent or teacher   _____

2. During your primary schools time, how many hours your parent taught English lesson to you in a week? Circle the answer

A. None or less than 1 hour per week

B. 1-3 hour per week

C. 3-5 hours per week

D. More than 5 hours per week

For questions between 3 to 6 please refer to the below response number

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree

Do you agree that voice recognition application that can produce text based on the spoken word can settle English pronunciation problem? _____

4. Do you agree that a busy parent should use this application? _____

5. Do you agree that parent's time used to teach student pronounce English word by using voice recognition cut to half rather than by using traditional method? _____

6. Do you agree that this application should be used in English class at primary school? _____

7. What is student age that you think should be produced to this application? Circle the answer
   A. 7-8 years old
   B. 8-9 years old
   C. 9-10 years old
   D. 10 years old and above

7.3 "V-CLASS" VISUAL BASIC CODING

Public gMode As Integer - - - - - - 'keep track of state we are

Public gAdd As Boolean   'is set to true if user want to add item

Public gRestoreGrammar As Boolean   'Set to true to reload vclass item

---

'This function is called back whenever a phrase is recognized.

'Phrase which is the verbatim phrase heard/to repeat what user said

'Parsed separate the word in a sentence/to say word in double quote

Private Sub DirectSR1_PhraseFinish(ByVal flags As Long, ByVal beginhi As Long, ByVal beginlo As Long, ByVal endhi As Long, ByVal endlo As Long, ByVal Phrase As String, ByVal parsed As String, ByVal results As Long)

If (parsed = "") Then   'This for mismatch

    'Most sound cards cannot multi-plex recording and playing, so it is

    ' imperative. Then the listening state of recognizer needs to be disable

    DirectSR1.Deactivate

    If (Rnd > 0.5) Then 'Order user to do input again

    DirectSS1.Speak "I'm sorry. What are you saying?"

    Else

        DirectSS1.Speak "I'm sorry. I cannot detect your pronunciation."

    End If

Else

47

```vb
'Recognizer is activated
If (parsed = "exit") Then
        DirectSR1.Deactivate


        If (gAdd = 0) Then
            gMode = 2
            DirectSS1.Speak "Would you like to report another item?"


        Else
            gMode = 3
            DirectSS1.Speak "Is that all loss item?"
        End If


ElseIf (parsed = "delete") Then
    'This to delete last item in list
    Item = List1.ListCount - 1
    If (Item >= 0) Then
        List1.RemoveItem (List1.ListCount - 1
        DirectSR1.Deactivate 'Recognizer deactivated, system will speak
        DirectSS1.Speak "Oh! Sorry"
    End If


ElseIf (parsed = "no") Then
    'Asked question which the customer said no to what we do depends on
'context gMode
    If (gMode = 2) Then
        gMode = 3
        gAdd = True
```

48

```
'If no item reported, stop listening and speak again
DirectSR1.Deactivate
DirectSS1.Speak "Is that all we miss?"


ElseIf (gMode = 3) Then
    gMode = 0
    gRestoreGrammar = True 'to load v class grammer
    DirectSR1.Deactivate
    DirectSS1.Speak "Please report another item?"


Else
    gMode = 0
    gRestoreGrammar = True
End If


ElseIf (parsed = "yes") Then
    'This is to deactivate recognizer
    If (gMode = 2) Then
        List1.AddItem ""
        gMode = 3
        DirectSR1.Deactivate
        DirectSS1.Speak "Do you want to add again?"


    ElseIf (gMode = 3) Then
        'This to total up the loss
        'This to tell user the total loss
        total = 0


        For i = 0 To List1.ListCount - 1
            total = total + Val(List1.List(i))
```

```
        Next i
        List1.AddItem "---The sum is RM" + Format(total, "###0.00")
        gMode = 1
        gRestoreGrammar = True
        DirectSR1.Deactivate
        DirectSS1.Speak "The total of losses are RM" + Format(total,
"###0.00") + " at the next window please."
        End If


    Else
        DirectSR1.Deactivate
        List1.AddItem parsed


        If (Phrase = "pencil") Or (Phrase = "pen") Or (Phrase = "crayon") Or
(Phrase = "eraser") Or (Phrase = "ruler") Or (Phrase = "scissor") Or (Phrase =
"note pad") Or (Phrase = "book") Or (Phrase = "glue") Or (Phrase =
"diskette") Then
            gAdd = True


        End If
        'Use use Phrase, not Parsed, to repeat what user said
        DirectSS1.Speak Phrase
    End If
End If


    'This to highlight the last item in the listbox
    List1.ListIndex = List1.ListCount - 1
    End Sub
```

'This to repeat again the session

Private Sub DirectSS1_AudioStop(ByVal hi As Long, ByVal lo As Long)

If (gMode = 1) Then

    List1.Clear

    gAdd = 0

    gMode = 0

    DirectSS1.Speak "Hello! I'm your voice teacher. A few items are missing in my class. Please pronounce the missing items and let's count the loss together"

Else

    If (gMode = 2) Or (gMode = 3) Then

      HighLightOrder (False)

      DirectSR1.GrammarFromFile App.Path + "\yesno.txt"

    ElseIf (gRestoreGrammar) Then

        HighLightOrder (True)

        DirectSR1.GrammarFromFile App.Path + "\vshop.txt"

        gRestoreGrammar = False

    End If

    'Recognizer is activated

    DirectSR1.Activate

End If

'This to produce image

If List1.ListCount <> 0 Then

N = List1.ListCount - 1

```
If (List1.List(N) = "2.45 Pen") Then
Image1.Picture = LoadPicture(App.Path + "\pen.bmp")
End If


If (List1.List(N) = "1.30 Pencil") Then
Image1.Picture = LoadPicture(App.Path + "\pencil.bmp")
End If


If (List1.List(N) = "9.85 Crayon") Then
Image1.Picture = LoadPicture(App.Path + "\crayon.bmp")
End If


If (List1.List(N) = "1.07 Ruler") Then
Image1.Picture = LoadPicture(App.Path + "\ruler.bmp")
End If


If (List1.List(N) = "1.90 Scissor") Then
Image1.Picture = LoadPicture(App.Path + "\scissor.bmp")
End If


If (List1.List(N) = "1.05 Note Pad") Then
Image1.Picture = LoadPicture(App.Path + "\notepad.bmp")
End If


If (List1.List(N) = "0.80 Book") Then
Image1.Picture = LoadPicture(App.Path + "\book.bmp")
End If
```

```
If (List1.List(N) = "1.00 Glue") Then
Image1.Picture = LoadPicture(App.Path + "\glue.bmp")
End If


If (List1.List(N) = "1.00 Diskette") Then
Image1.Picture = LoadPicture(App.Path + "\diskette.bmp")
End If
End If
End Sub
```

```
'This is to load background picture
Private Sub Form_Load()
gMode = 0
gRestoreGrammar = False
Form1.Picture = LoadPicture(App.Path + "\bground.bmp")
End Sub
```

```
Private Sub Form_Terminate()
'This is to terminate the form, recognizer is unnecessary
DirectSR1.Deactivate
End Sub
```

```
'This is to show main menu
Private Sub lblmenu_Click()
frmmain_menu.Show
Form1.Hide
End Sub
```

```
Private Sub Timer1_Timer()
On Error GoTo GeneralFail
'This needed on the first callback, and then the timer isn't needed anymore
Timer1.Enabled = False


'This is to initialize the recognizer. ofcommand and control engine
DirectSR1.Initialized = 1


engine = DirectSR1.Find("MfgName=Microsoft;Grammars=1")
DirectSR1.Select engine


'To compile and load vshop grammer now.
'The entire listed item is highlighted
HighLightOrder (True)
DirectSR1.GrammarFromFile App.Path + "\vshop.txt"


'This is to load the synthesizer, selects male voice from microsoft
'If male voice=0, if female voice=1
voice = DirectSS1.Find("MfgName=Microsoft;Gender=0")
DirectSS1.Select voice
```

```
'Start the parse
DirectSS1.Speak "Hello! I'm your voice teacher. A few items are missing in
my class. Please pronounce the missing items, and lets count the loss
together"


'If VB environment come to front
'Prompt message box speech recognition cannot be loaded
Form1.Show
GoTo done
GeneralFail:
MsgBox "Speech recognition cannot be loaded."
done:
End Sub
```

```
Sub HighLightOrder(burgers As Boolean)
    If (burgers) Then
        fColor = &H0
        Label2(29).ForeColor = &HFFFFFF
    Else
        fColor = &HC0C0C0
        Label2(29).ForeColor = &H0
    End If
    For i = 0 To 10
        Label2(i).ForeColor = fColor
    Next i
    Label2(23).ForeColor = fColor 'delete
    Label2(26).ForeColor = fColor 'exit
End Sub
```