Oral Dictionary

by

Ashylla Bt. Md. Radzi

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

JANUARY 2006

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Oral Dictionary

by

Ashylla Bt. Md. Radzi (3671)

A project dissertation submitted to the
Information Communication Technology Programme
Universiti Teknologi PETRONAS
in partial fulfillment of the requirement for the
BACHELOR OF TECHNOLOGY (Hons)
(INFORMATION COMMUNICATION TECHNOLOGY)

Approved by,

_____

(Mr. Ahmad Izuddin B. Zainal Abidin)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK
January 2006

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 persons.

ASHYLLA BT. MD. RADZI

3671

ABSTRACT

Oral Dictionary application is developed with aim to enhance the existing dictionary, therefore promote flexibility in English dictionary usage through implementation of speech recognition technology. Speech recognition addresses an effective and faster way for word definition searching as word input is retrieved through human's voice, regardless the knowledge on the word spelling where current dictionary relies to. Capability of providing pronunciation playback based on phonetically British English standard overcome the major limitation of current English dictionary. The application is using the text-to-speech (TTS) technology which is different from other commercial dictionary software who use to record each sample of word. In dealing with homophone words (words having similar pronunciation but different spelling), the system successfully provide an approach to reduce the misrecognition. The approach is by providing possible matches, relevant to the phonetic matching of the word. Oral dictionary also offers hassle-free solution to users who only want to hear the pronunciation of the word without having to search throughout the dictionary. Currently, the scope of Oral Dictionary only cover 100 words, within the noun context and using British English (Br.E) standard, as it is English language correctly spoken or written. The application is developed using the combination of vocabulary collection process and incremental delivery model. The major processes involve in vocabulary collection process are words collection, syllabification and ASCII-phonetic transcription. Process activities in Oral Dictionary application will then follow the incremental model which allows system to be reworked in response to change request. Since the output speed is critical, binary search tree algorithm is chosen as best data structure to be implemented in this application to improve the searching time performance. Procedures to obtain the result of the system effectiveness are discussed in this paper, where product evaluation / testing is conducted to two groups of users and the result on searching time comparison, misrecognition rate, verification on British-English standard, system functionality and overall system interface design are discussed in detail throughout this paper.

ACKNOWLEDGEMENTS

First of all author would like to praise the Almighty God for giving author the opportunity to complete this Oral Dictionary project. Author would like to express the most gratitude to FYP supervisor, Mr. Ahmad Izuddin B. Zainal Abidin, Mr. Nordin, the FYP Coordinator and Ms. Nor Shuhani who are willing to give their full support, advice, and knowledge from the beginning of the project to the end. With their wisdom, inspirations and experience, this project is then successfully completed. A special thank you to Ms. Ena Bhattacharya for her effort in helping the author to understand the complexity that lies in English language.

Sincere thanks also go to all colleagues and course mates who have been helping and supporting the author through all the difficulties that author has encountered in making this Final Year Project a reality.

Most of all, author would like to thank her beloved parents and family for their blessing and continuous support. Without them, it would be impossible for author to complete this project on scheduled. Their understanding and endless support are also one of the most important factor in the success of this project.

Last but not least, a thousand thanks and gratitude to those who are involved directly or indirectly in making this project a success. May ALLAH bless them. Thank you very much.

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Over recent years, numerous researches on speech recognition technology were performed to signify the factors to successful implementation of this technology with regards to natural language processing of human. Speech recognition technology which has the capability to produce a typed manuscript in response to human's voice, is continued to grow with its implementation in various areas including medical, customer inquiries-response, security as well as education. The speech recognition involved study on human production of sound as well as the sound perception by machine. As for education, the speech recognition is likely to contribute most for language learning, especially for pronunciation which is often lacked of attention in the language classroom. General acceptance of English language by many countries provides a direction in speech recognition study, especially in determining which standards to be applied for text-to-speech feature. Text-to-speech or TTS synthesizer is defined as a computer based system that should be able to read any text aloud [1].

The study of this project will focus on implementation of speech recognition in English dictionary application. In general, dictionary is defined as a collection of words in alphabetical order; providing the definition or concept of the word, along with phonetic information for correct pronunciation. Dictionaries can be divided into different type as they deliver different purpose, such as bilingual, thesaurus and specialized dictionary. Traditionally, a dictionary came in the form of hard-covered book and normally is owned individually, or a family by least, mostly for word definition searching. With the emergence of technology, various dictionary softwares were developed and later become more accessible when it is available online. However, regular usage of dictionary has exposed author with some limitation that current dictionary has, and as a result, it influenced author to the development of Oral Dictionary project which aims to enhance current dictionary by allowing voice input for word definition searching. The TTS feature

will be used to provide user with correct word pronunciation based on phonetics applied in British English standard.

1.2 Problem Statement

The major problem actually derived from the English words itself. English words are not pronounced the way it is spelled, thus create problem for users, especially for non-native English speaker who tends to pronounce based on its spelling. Besides, there's pronunciation confusion in English word, as one word may have different ways of pronunciation such as for word *read*, the pronunciation might sound as *rid* or *red*. In the case of homophones, two or more words are likely to have similar sound of pronunciation but have different meanings, such as *see* and *sea*. Apart from that, the problems which influenced the idea of Oral Dictionary are derived from limitation feature of any basic dictionary itself. The limitation result is based on comparison made to existing dictionaries which are traditional dictionary, dictionary software and online dictionary.

1.2.1 Problem Identification

- Even though the words in dictionary are arranged alphabetically, it will remain useless if the spelling of the English word is unknown or unsure, especially when it involves lengthy word.
- Regularly flipping through pages in traditional software is a tiring work especially whenever large vocabularies need to be defined.
- Existing dictionary software depends directly on word spelling. Null result or incorrect definition is obtained if the user mistakenly spelled the word (e.g. *dye* is spelled *die*)
- Online dictionary relies heavily to internet connection availability. User will not be able to use the dictionary if they are not connected to internet. Another problem raised is lagging of output display due to slow internet connection.
- Traditional dictionary is large in size and some are quite heavy to be carried along for regular reference.

- Existing dictionary is lack of ability to provide user on correct English word pronunciation.
- Phonetic information for word pronunciation provided in dictionary is not understood by user who has no knowledge of phonetic transcription.

1.2.2 Significance of the Project

- With implementation of speech recognition, Oral Dictionary is able to retrieve input from user's voice
- Voice input is easy to perform as it doesn't require any specialized skills
- Speech recognition provides faster time for word definition searching using Oral Dictionary application
- Text-to-speech (TTS) feature enables Oral Dictionary to provide correct phonetically pronunciation of English word based on British English standard
- Implementation of speech recognition technology provides alternative and accessibility to visual impairment users, thus perform as attempt to bridge the gap between the disabilities and normal users

1.3 Objectives and Scope of Study

The primary objective of this project is to fulfill the University Technology of PETRONAS (UTP) requirement upon completion of Information and Communication Technology (ICT) course. As for the Oral Dictionary application itself, the objectives are as the following :

- To enhance existing dictionary application through implementation of speech recognition
- To help user in finding definition of English word in more efficient way
- To provide user with correct pronunciation of English word effectively

1.3.1 Relevance of the Project

There are few characteristics which determine the relevance of Oral Dictionary project development. As discussed earlier in the problem statement, English word is mostly not pronounced the way it is spelled. Therefore, by implementing the speech recognition in word definition searching, the accuracy of the output will be determined from the way user pronounced the word, regardless their knowledge of the word spelling. Since the process of recognizing homophone word is prone to ambiguous result, the system will be designed to provide with possible matches for user to choose instead of an exact output which might not be the intended word. The relevance of Oral Dictionary is also measured by its capability to provide user with British English standard pronunciation which is most lacked by existing dictionaries. Apart from that, the project is economical as the voice input can be performed as long as the user has a microphone. With the implementation of text-to-speech (TTS) engine, the system can read out loud the word pronunciation together with the definition, thus provide an attempt to bridge the gap between the normal and disabled users. The read-out-loud mechanism is essential for visual impairment to obtain the output of word definition.

1.3.2 Feasibility of Project within Scope and Time Frame

Since the speech recognition is a new knowledge to author, much time will be consumed in order to understand the underlying concept of this technology as well as ensuring its successful implementation in Oral Dictionary. Due to the time constraint faced by author, scope of the study was narrowed down to make it possible to be completed within the time frame with available skills while remained inline with the FYP standard. English dictionary is chosen instead of other languages due to the fact that it is mostly demand and widely used in worldwide communication. British English is a standard used to provide guideline in pronunciation and definition of the words for Oral Dictionary. In fact, British English or Queen's English is English language correctly spoken or written [4]. The application is purposely developed for normal users as the intended users while visual impairment still can use it with the assistance of TTS engine to read out loud the definition for them. The prototype will limit the words or vocabulary to be covered in Oral Dictionary only to maximum of 100 words. In order to meet the FYP standard, the selection of some words should allow the demonstration of how challenges in speech recognition are overcame (e.g. homophone words). Supposedly, an actual dictionary consists of words in all different contexts (verb, noun, adjective etc.) but for the purpose of prototype, the context of word coverage is limited to noun only. The above decision is made to ensure the project at least meets the objectives through a prototype rather than trying to achieve an *impossible* ready-for-market product.

CHAPTER 2

LITERATURE REVIEW

2.1 Speech Recognition Technology

As information technology continues to make more impact on many aspects of our daily lives, the problems of communications between human beings and information-processing machine becomes increasingly important. Up to now, man-machine communication is still dominated by means of keyboard. The situation is not because of the strong desire to produce words by means of fingers, but inability of machine to understand speech [3]. Speech is at first sight an obvious substitute because it is most widely used and natural means of communication between people. However, this deceptively simple means of exchanging information is, in fact, extremely complicated. Therefore, continuous research was done on how human speech can be interpreted by machine and finally, a technology called speech recognition was introduced. Speech recognition is the process of automatically extracting and determining linguistic information conveyed by a speech wave using computers or electronic circuits [2].

The technology of speech recognition has progressed greatly over the past few years. Ever since research of this technology began in 1936, the largest barrier to the speed and accuracy of speech recognition was the speed and power of the computer itself. With currently CPU average at or above a Pentium III and RAM levels of 500 MB and up, accuracy levels have reached 95% and better with transcription speed at over 160 words per minute [24]. The study on speech recognition initially started in 1936 at AT&T's Bell Labs. Until 1980's, most of the research was performed by universities which was funded primarily by U.S. Military and DARPA - Defense Advanced Research Project Agency. When the technology reached the commercial market, several independent research "camps" began competing to develop the speech recognition. Covox was the first company to launch a commercial product in 1982. Covox has brought the digital sound to the Commodore 64, Atari 400/800 and finally to the IBM PC in mid 80's. Along with this introduction of sound to computers came *'speech recognition'*. [24]

A toy dog called "Radio Rex" was the first success story in the field of speech recognition, where the dog will pop out from its kennel when "Rex" is called [25]. The dog was held within the kennel by an electromagnet, and the electromagnet is energized as the current flowed through a circuit bridge. The bridge is sensitive to 500 cps of acoustic energy. The energy of the vowel sound of the word "Rex" caused the bridge to vibrate, breaking the electrical circuit, and allowing a spring to push Rex out of his kennel. However in late 1940s, the U.S. Department of Defense sponsored the first academic pursuits in speech recognition [25]. In attempt to intercept and decode Russian messages, the U.S. sought the development of an automatic language translator. The first, and most difficult step was to solve the problem in creating a program that could recognize speech. The project was a dismal failure. There was misrecognition, where phrases typically mistranslated such as :

"The spirit is willing but the flesh is weak"

to

"The vodka is strong but the meat is disgusting"

Despite the dismal failure, appreciation and interest for the field began to grow. As a result, the government funded the Speech Understanding Research (SUR) program in Carnegie Mellon University, MIT, and some other commercial institution [25]. The agency that funded the research is known as Defense Advanced Research Project agency (DARPA).

Among early key advances in speech recognition technology were [25] :

- In 1952, as government-funding research began to gain momentum, Bell Laboratories developed an automatic speech recognition system that successfully identified the digits 0-9 spoken over telephone.
- In 1959, MIT developed a system that successfully identified vowel sound with 93% accuracy.
- In 1966,a system with 50 vocabulary words was successfully tested.

- In early 1970's, the SUR program began to produce results in the form of the HARPY system. This system could recognize complete sentences consist of a limited range of grammar structures. The program required massive amounts of computing power to work, 50 state-of-the-art computers.
- In 1980s, Hidden Markov Model (HMM) became the standard statistical approach for computation.

2.2 Theories in Speech Recognition

The major characteristic of a speech recognition system is the number of words it can recognize correctly. However, the performance declines with the increasing number of words. The increasing number of words or vocabulary in a speech recognition system will increase the complexity, and therefore decrease the performance [26]. The time performance decrease since more time is used to search a word from large vocabulary, causing the system to be slower. Based on the best-match case, the system become less effective for its inability to differentiate words such as "two" and "to". However, the use of grammar making the problem is possible to be resolved since the grammar is capable to speed up speech recognition system by narrowing the range of words to be search. It also increases the performance by eliminating inappropriate word sequencing. But yet, grammar does not allow random dictation which is a problem for some application.

Using continuous speech approach is more desirable as it is natural way of human speaking. However, it is difficult to use continuous speech in speech recognition system. The pause between words in discrete speech is more reliable because the silent gap is used to determine the boundary of the word, whereas an algorithm is needed to separate the word in continuous speech which still not 100% accurate. It is still difficult to achieve high performance of using continuous speech for large vocabularies as it requires huge computational power, otherwise the system will be very slow. In fact, processing a speech sample take about three to ten times longer than required for a person to say it.

2.2.1 Speech Recognition Approaches

Acoustic-Phonetic was the earliest approach used in developing fundamental principle of speech recognition system [26]. The theory is based on assumption that the word is divided into phonetic units that are finite and particular. The phonetic units are distinguished by the properties apparent in the speech signal. The process of recognizing the speech begins by dividing the speech into segments where appropriate phonetic unit is attached to it based on acoustic properties of this segment. The sequence of units obtained is later used to formulate a valid word. Later than was the Statistical Pattern Recognition approach [26], where the speech pattern are directly inputted into the system as the comparison model to speech patterns inputted in the system during training. This approach has the performance advantage compared to acoustic-phonetic approach as it depends on how much patterns are matched with the previously stored patterns. In general, this approach is more preferable because of its simplicity, invariant to speech vocabularies and higher accuracy which lead to better performance, compared to acoustic-phonetic approach.

2.2.2 Hidden Markov Model (HMM)

HMM tool is a mathematical based approach to fundamental principle if speech recognition. The tool was proven to provide better efficiency than earlier methods as the speech recognition is now capable to recognize more words with high accuracy rate. Markov model consist no. of states linked together with each state corresponding to a unique output [27]. Each link between two states is characterized by transitional probability. Moving from one state to another or remaining in the same state is function of corresponding transitional probability. A classical example illustrating Markov model as following: consider a three-state weather system with state one being rainy, state two for cloudy and state three for sunny. The probability of tomorrow being cloudy is 0.1, being rainy 0.1 and being sunny 0.8, where the combination of each probabilities must be equal to 1. HMM is used since the speech fragment is not observable by the speech recognition system. In HMM, a state can

represent many outputs; therefore a probability distribution of all possible outputs is associated with each state. In speech recognition system, each word is represented by a sequence of state, therefore it is essential to find this sequence of for any sequence of outputs. The sequence of state is determined according to the probability. However, checking all probabilities of all possible sequence could be a time consuming task especially for more complex HMM which is more then the 3-states example given previously. The problem is solved using an algorithm that utilizes the fact that probability of being in a certain state relies from the previous state. As mentioned earlier, major components of HMM is the probability between the states and probability distribution of each state. These probabilities must change to factors like language, no. of speakers etc. Determining these probabilities is part of what is known as training the speech recognition system. The training process depends on category used, speaker-dependent or speaker-independent. The speaker-dependent require user's speech sample where the probabilities is based on. However, for speaker-independence speech samples are accumulated from many speakers in addition to text which shows that the training process is more complicated since the spectrogram (measure of frequency vs. time) of the same word depends on the speaker.

2.3 Speech Recognition in Learning Field

While navigation of normal users is dominated largely by visual sensory, unfortunately for disabled user, especially for those with visual impairment, they depend entirely on their audio sensory. Speech recognition technology has provided an opportunity for the disabled to become productive members of the society by using it as an adaptive technology to accomplish multitude of tasks. However, a national survey on technology abandonment concluded that almost one-third of assistive devices are abandoned [8]. Study conducted by Tanya Goette (2001) demonstrated the importance of matching the task to the type of voice recognition technology (VRT) system to be used. This is because some tasks can only be easily accomplished by certain types of VRT [9]. From the

journal, author believed that part of this problem is due to the lack of technological product knowledge of developer. This finding will be used as a guideline for Oral Dictionary to be reviewed from its critical tasks to be accomplished and later used to determine what features should the speech engine support in dealing with large vocabulary size and appropriate techniques for faster output retrieval.

Speech recognition can be used in the field of education for a variety of applications, closely linked to the speech synthesis applications. Current uses of automatic speech recognition (ASR) generally involve assessing the accuracy of pronunciation of specified words. *'Talking & Listening Book'* of Speech Training Aid Research (STAR) project by Russell and his team from Speech Research Unit, DRA Malvern (1996) purposely to distinguish 'good' and 'poor' word pronunciation of children. The speech signal is compared with a word level Hidden Markov Model (HMM) and 'general speech' HMM, and the child is said to have produced 'good' pronunciation if it best match with word model. However, the reliability is questioned since according to Hereford and Worcester County Council Education Department (HWCC), 'good' pronunciation occurs within the context of variety of regional accents, and clearly not the same as Received Pronunciation (RP/ BBC English) since factors such as a child's confidence in speaking are also relevant [19]. Motivated by poor assumption that all speakers pronouncing words as described in Pronunciation Dictionary (PD), Humphries, Woodland and Pearce [20] proved that addition of accent-specific pronunciation has reduced error rate by almost 20% for cross accent recognition. Even though they fulfill the objective, but it is still difficult to get full coverage of all accents.

Rebecca Hincks (2001) however, suggested that extra pronunciation training using ASR based language learning is beneficial for students who began the course with an 'intrusive' foreign accent [7]. The result was based on her study on existing commercial English language learning product, *Talk To Me* to groups of students who were non-native English speakers. *Talk To Me* uses speech recognition to provide conversational practice, visual feedback on prosody and scoring of pronunciation. The result presents the major limitation of current commercial programs; its inability to diagnose specific

articulatory problems and give corrective rather than evaluative feedback. The result also indicates that mimicry does not necessarily improve pronunciation and such practice only applicable for beginning students of a language. However, Bryna Siegel [28] argued that imitation is the gateway to early learning, based on her study that children learn to speak and pronounce the word by imitating how their parents speak For Oral Dictionary, imitation practice is necessary since the primary purpose of a dictionary is to provide word definition. Playback of pronunciation is to give the idea to user on how the word is pronounced, rather than exhaustive effort to educate user on 'perfect' pronunciation like what a language learning program should have.

In determining the successful implementation of speech recognition in Oral Dictionary, focus should also be given to common problem associated with this technology. Significant challenges in speech recognition technology are recognition errors and lag separating the spoken words and its transcription [10]. There are many sources of error in queries, including the vocalization of the query itself, the estimation of fundamental frequencies in the sound and the transcription of this frequency into a discrete representation [11]. Recognition error or misrecognition occurs when the system recognize different word from what is intended by user as best match by the system. In continuous speech recognition, this type of error is reduced by speech engine through 'understanding' the context of word used in the sentence. Since Oral Dictionary only involve discrete word rather than a sentence, approach to minimize this error is very important to determine the accuracy of the output as well as its reliability. Karat J. in his studies suggested that a dialog should appear with list of possible alternative words based on probability metrics from recognition engine [10] to reduce the misrecognition error. Apart from that, lag in producing result of word definition however, can be reduced by introducing binary searching algorithm which is known to be much faster than sequential approach.

The findings discussed above are essential to ensure that speech recognition implemented in Oral Dictionary is able to address its function successfully to intended users as well as achieving its objectives. At the very least, studying the trends, challenges and issues

related to this technology is helpful to avoid the Oral Dictionary from reinventing the wheel in specified area.

2.4 Speech Interface Design and Human Factors

One of the many reasons for using speech recognition for interacting with machines is that speech is the natural mode of communication between humans [3]. In such a situation great care must be taken to ensure that the computer's reaction matches the person's expectations of it. Understanding of how human-computer interaction is different from human-human interaction is helpful in achieving realistic goal of speech-based HCI. Human factors should be taken into consideration before the interface is designed to ensure the product delivery will allow user to perform the task effectively. The cognitive aspect of human factors is given more attention, which include the mode of man-machine interaction, user's mental model of machine, user's short term memory and interference of competing tasks [3].

Three modes of man-machine interaction are command mode, question-and-answer mode and finally, menu mode. Which mode is employed depends on task involved and the experience of the user [13]. Difficulty in recalling the whole repertoire of available commands should be supported with help facility which can be accessed at any time. Question-and-answer mode is more appropriate for inexperienced or occasional users of a system. This mode is implemented to let user knows what to say next by providing explicit speech prompt [14]. This mode works best for naïve user which can be fairly error-free and lead to shorter transaction time. However, Ainsworth recommends that the list to be presented to user using question-and-answer mode should not exceed the span of human short-term memory [3].

Speech recognition application is prone to error, such as misrecognition. Misrecognition is serious in dictionary application as it changes the meaning of the utterance [14]. A few solutions to reduce this problem include providing list of possible matches, as phonetically relevant to the word recognized by the system. The early solution identified

in this problem area was the Lawrence Phillips' Double Metaphone phonetic matching algorithm [29]. The phonetic algorithm was successfully implemented by Adam Nelson in his project for name searching using the double metaphone [29]. However, the algorithm is currently limited to American-English pronunciation standard only. But yet, the work is appreciated for the attempt to homophonic problem as well as error handling for unavailable words in vocabulary.

One of the challenges faced in designing interface is the techniques deployed to keep user's attention. Although there are various techniques available, Oral Dictionary will implement structuring and alerting techniques [15]. The structuring is strengthened by providing easy navigation for the application. Meanwhile, alerting technique will be implemented for possible matches of result. The best-match word from the possible matches list will be highlighted or simply called word flashing.

2.5 Elements of System Usability

System usability is important for any application as it addresses the relationship between tools and their users [16]. Usability is defined as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in specified context of use. The goal of system usability is broken down into effectiveness, efficiency, learnability, memorability, errors and satisfaction.

Usability depends on a number of factors including how well the functionality fits user needs, how well the flow through the application fits user tasks and how well the response of the application fits user expectations [16]. Learning the design principle and design guideline is helpful in producing better user interface. But even the most insightful designer can only create a highly-usable system through a process that involves getting information from people who actually use the system. The importance of system usability is viewed from the perspective of user and the developer [16]. From the user's perspective usability is important because it can make the difference between performing a task

accurately and completely or not, and enjoying the process or being frustrated. From the developer's perspective usability is important because it can mean the difference between the success and failure of a system.

Six key elements mentioned previously are used to determine the system usability. The effectiveness of the system is measured on how good the system in performing the task which it is supposed to do. In term of learnability, how easy a system be learnt to use which includes to get started performing core task and later learning to accomplish wider set of tasks. Once learned, how easy a system to be remembered how to use is accounted to the goal of memorability. Whenever the user already familiar with the system, the efficiency is then measured on how quick user can perform the task and access the information [17]. A system is considered usable if it produces low error rates. Error defined here is any action which does not accomplish desired goals. This is measured on how many errors is the user likely to make, how easy for user to recover from the errors and the seriousness of each error [17]. Finally, the satisfaction is subjective and varies to each user. How pleasant the system to be used by the user should be obtained from user's opinion and questionnaire.

The key principle for maximizing usability is to employ iterative design, which progressively refines the design through evaluation from the early stages of design [16]. The evaluation steps enable the designers and developers to incorporate user and client feedback until the system reaches an acceptable level of usability. The preferred method for ensuring usability is to test actual users on a working system [16]. Achieving a high level of usability requires focusing design efforts on the intended end-user of the system. There are many ways to determine who the primary users are, how they work, and what tasks they must accomplish. However, clients' schedules and budgets can sometimes prevent this ideal approach. Some alternative methods include user testing on system prototypes, a usability inspection conducted by experts, and cognitive modeling.

2.6 British-English (BrE) and American-English (AmE)

English language is a West Germanic which was firstly introduced in England. It has been the primary medium for communication around the world since the last centuries. It was believed that the language was originated from the Germanic language with a significant amount of vocabulary from French, Latin, Greek and many other languages [21]. It has settled down at the stage now known as Modern English. The first English dictionary was published in 1755 by Samuel Johnson. As for the literature studies, author only covers 2 major standards in English language, which is British-English and American-English.

According to Jeremy Smith, when an American and British person meet, the obvious differences is the accent (pronunciation of words) and vocabulary (occasional different word for something, like foreign language) [30]. However, more subtle difference becomes apparent in syntax or grammar. The differences are based on the reason of inheritance, innovation and isolation. The inheritance factor is based on which society they grown up where American is brought up in America while British from Britain. Innovation through new activities, which by nature give rise to new terms, especially to those directly involved. In some cases, for instance technology, this exponentially affects the rest of population such as the word *morph.* However, isolation is described by dialectologist that some dialects are separated by geographical features that naturally separate people, such as hills, rivers or bogs. So dialect arises when a group is isolated long enough.

American English is the dialect of the English language used mostly in United States of America [21]. British English however, is a term used when describing formal written English and forms of spoken English used in United Kingdom [22]. The dialects and accents vary not only between regions in the UK, for example in Scotland, Northern Ireland and Wales but also within England. Although spoken American and British English are generally mutually intelligible, there are enough differences to occasionally cause awkward misunderstandings or even a complete failure in communication. George

Bernard Shaw once said that the United States and United Kingdom are "two countries divided by a common language". There are few categories can be used to differentiate the British English and American English such as pronunciation, word derivation and compounds, lexis, writing and etc. However, as part of the literature studies, only 2 categories are covered as relevant to Oral Dictionary project which is pronunciation and spelling differences.

2.6.1 Differences in Pronunciation

General American is the name given to American accent that is relatively free of noticeable regional influences [21]. It enjoys high prestige among Americans, but is not a standard accent in the way that Received Pronunciation (RP) as spoken by British newscaster which is known as BBC English. RP is the accent of English-English most often taught to non-native speakers and it is represented in pronunciation scheme of British dictionaries. Both Figure 2.1 and 2.2 below show the cardinal vowel system of American and English (RP), where the difference of tongue position results in different sound / pronunciation of the word. . In many areas the American ætÆ, when not the initial consonant in a word, is pronounced closer to a ædÆ, and in some cases can disappear altogether. Thus *latter* and *butter* sounds more like *ladder* and *budder*, and words like *twenty* and *dentist* can sound like *twenny* and *Dennis*. Most Americans are Rhotic [30], where r is clearly pronounce in *barn, cart, park* whereas non-Rhotic accent in Britain will make no distinction between *barn* and *bahn.* Besides, there are some words that are spelled similarly but pronounced different way. As for the word *fillet,* Americans pronounce it as *filay* while British called it *filit.* This described why choosing the English standard is important because the speech recognition process in Oral Dictionary rely solely on user's pronunciation to determine its output definition.

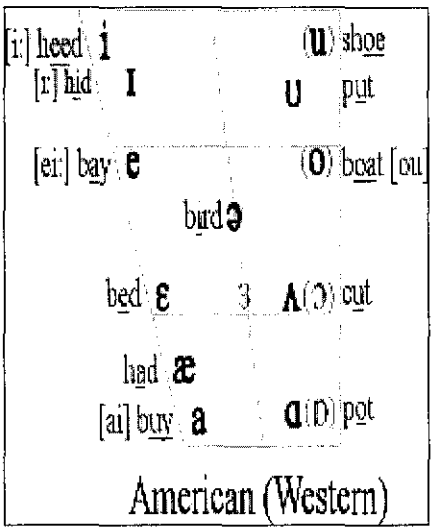


Figure 2.1: American English vowel system

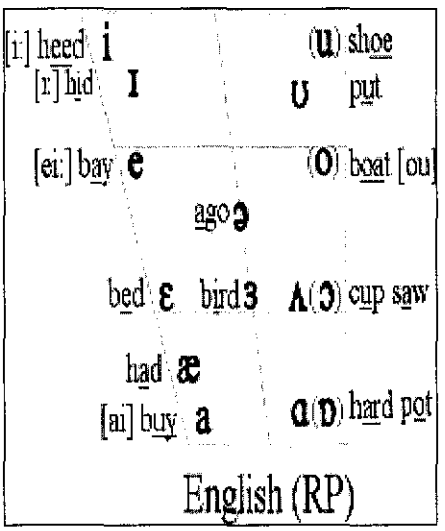


Figure 2. 2: British English vowel system

Source : http://www.uta.fi/FAST/US1/REF/usgbintr.html

2.6.2 Differences in Spelling

American English has many spelling differences from English as used elsewhere, some of which were made as part of an attempt to rationalize the spelling used in Britain at the time [30]. British English uses *grey,* while American English uses both *grey* and *gray*, but *gray* is far more common in American English. American authors tend to assign wistful, positive connotations to *grey*, as in "a grey fog hung over the skyline", whereas *gray* often carries connotations of drabness, "a gray, gloomy day." [31]. The American spelling is the international standard in science, although many British scientists used British spelling such as *sulphur* compared to *sulfur* used by American [31].

2.7 Phonetic Transcription

OUR STRANGE LANGUAGE

When the English tongue we speak
Why is "break" not rhyme with "freak"
Will you tell me why it's true?
We say "sew" but likewise "few";
And the maker of the verse
Cannot cap his "horse" with "worse";
"Beard" sounds not the same as "heard";
"Cord" is different from "word".
Cow is "cow" but low is "low";
"Shoe" is never rhymed with "foe";
Think of "hose" and "dose" and "lose";
And think of "goose" and not of "choose";
Think of "comb" and "tomb" and "bomb";
"Doll" and "roll", "home" and "some";
And since "pay" is rhymed with "say";
Why not "paid" with "said", I pray;
We have "blood" and "food" and "good";
"Mould" is not pronounced as "could";
Wherefore "done" but "gone" and "lone"?
Is there any reason known?
And short it seems to me
Sounds and letters disagree.

Author : E.L. Sabin

The poem illustrates the difficulties of English spelling and sounds [23]. English language itself is blatantly irregular; whose written form generally does not give any

direct information about pronunciation [32]. Therefore, in dictionary, phonetic transcription is used to tell user how the word is pronounced. Phonetic transcription or phonetic notation is the visual system of the symbolization of the sounds occurring in spoken human language [33]. Most phonetic is based on the assumption that linguistic sounds are segmented into discrete units that can be represented by symbols. There are 3 types of transcription; iconic, analphabetic but common type used by dictionaries is alphabetic which follows the International Phonetic Alphabet (IPA) standard as shown in Figure 2.3 :

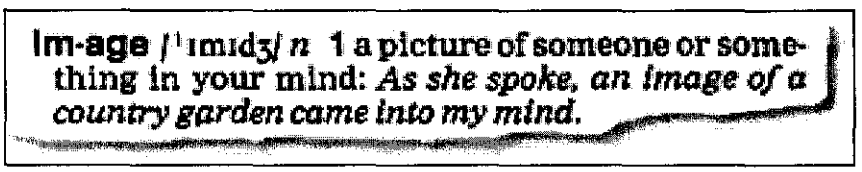


Figure 2.3 : Example of word in dictionary having phonetic transcription

Source : http://www.antimoon/phonetic_transc/

Generally, there are 2 types of alphabetic phonetic transcription: broad transcription and narrow transcription Broad transcription is usually just a transcription of the phonemes of an utterance, whereas narrow transcription encodes information about the phonetic variations of the specific allophones in the utterance.

Phonetic transcription contributes in language learning process since it is able to educate user on how the word is correctly pronounced, especially in dictionary. However, the advantage of this transcription is argued since it is only beneficial for those with phonetic knowledge. When the speech recognition technology arrived in, it seems to overcome this limitation with the text-to-speech (TTS) feature. The text-to-speech (TTS) model is developed based on the phonetic transcription concept, where the word must be pronounced with correct phonetic standard. Difficulties of representing the IPA symbols in computer system allow the Kirshenbaum [34] in early 2001 to be the major reference for his pioneer work on ASCII representation for phonetics.

CHAPTER 3

METHODOLOGY

3.1 Procedure Identification

The vocabulary collection described the processes involved in developing the vocabulary database for Oral Dictionary. As referred to Figure 3.1, the processes start with collection of words, performing word syllabification, transcribing into ASCII representation and finally deciding which action to be performed next.

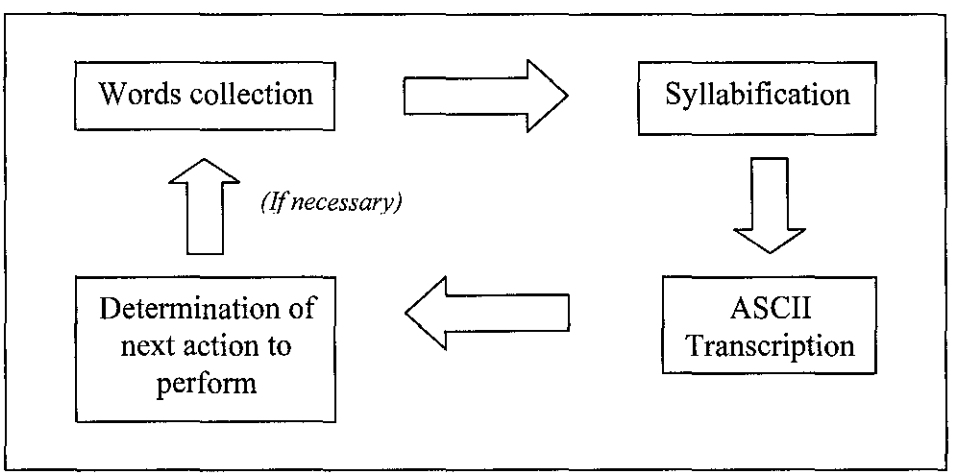


Figure 3.1 : Vocabulary Collection Process

Set of words are gathered from the dictionary. The words then will be divided into its corresponding syllables. Once the syllabification process is completed, the phonetic notation of each word is converted into its ASCII representation. The process is repeated until necessary amount of words is obtained and will proceed to the design stage. If the developer wishes to add more words for further enhancement, similar vocabulary collection process will be used.

Systems Development Life Cycle (SDLC) on the other hand, is agreed to be the systematic approach in any software development [13]. In Oral Dictionary, process activities will be regularly repeated as the system is reworked in response to change request. In addition to this, incremental delivery model is used as an approach to the

iteration process involved. The model which combines the advantages of both waterfall and evolutionary model allows system specification, design and implementation be broken down into series of increment that are each developed in turn [6]. This advantage leads to lower risk of project failure, because although problems may be encountered in some increments, it is likely that some will successfully be delivered. Based on Figure 3.2, the stages involved are requirement analysis, system design, coding development and finally testing and evaluation.

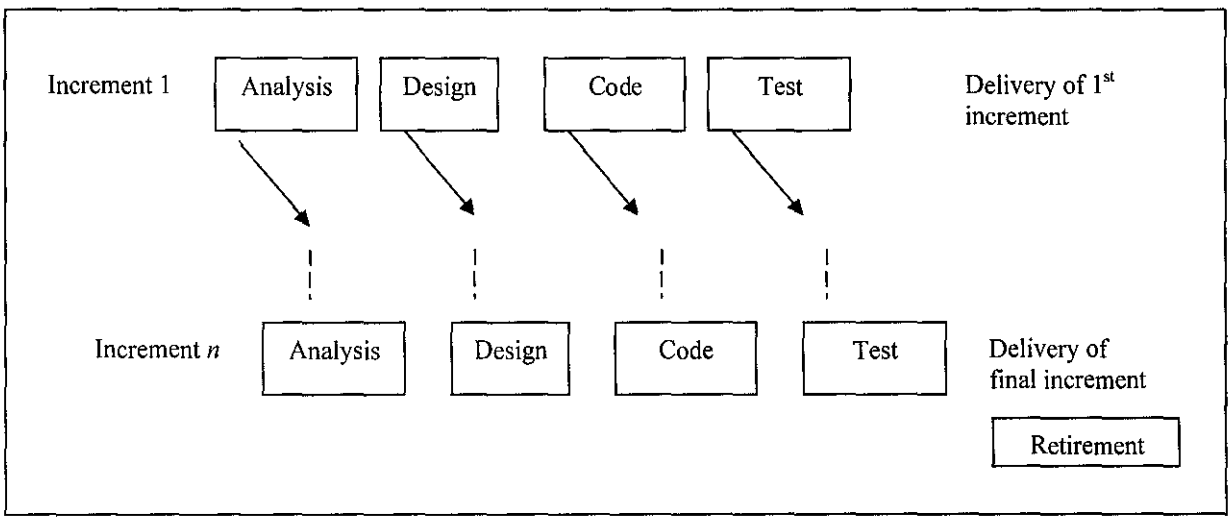


Figure 3.2 : Incremental Model

Based on the advantages discussed, combination of vocabulary collection process and incremental model is used for completion of Oral Dictionary project. The combination of methods is also be used as a methodology outline for this project. The process flow of the steps involved is shown in Figure 3.3 :

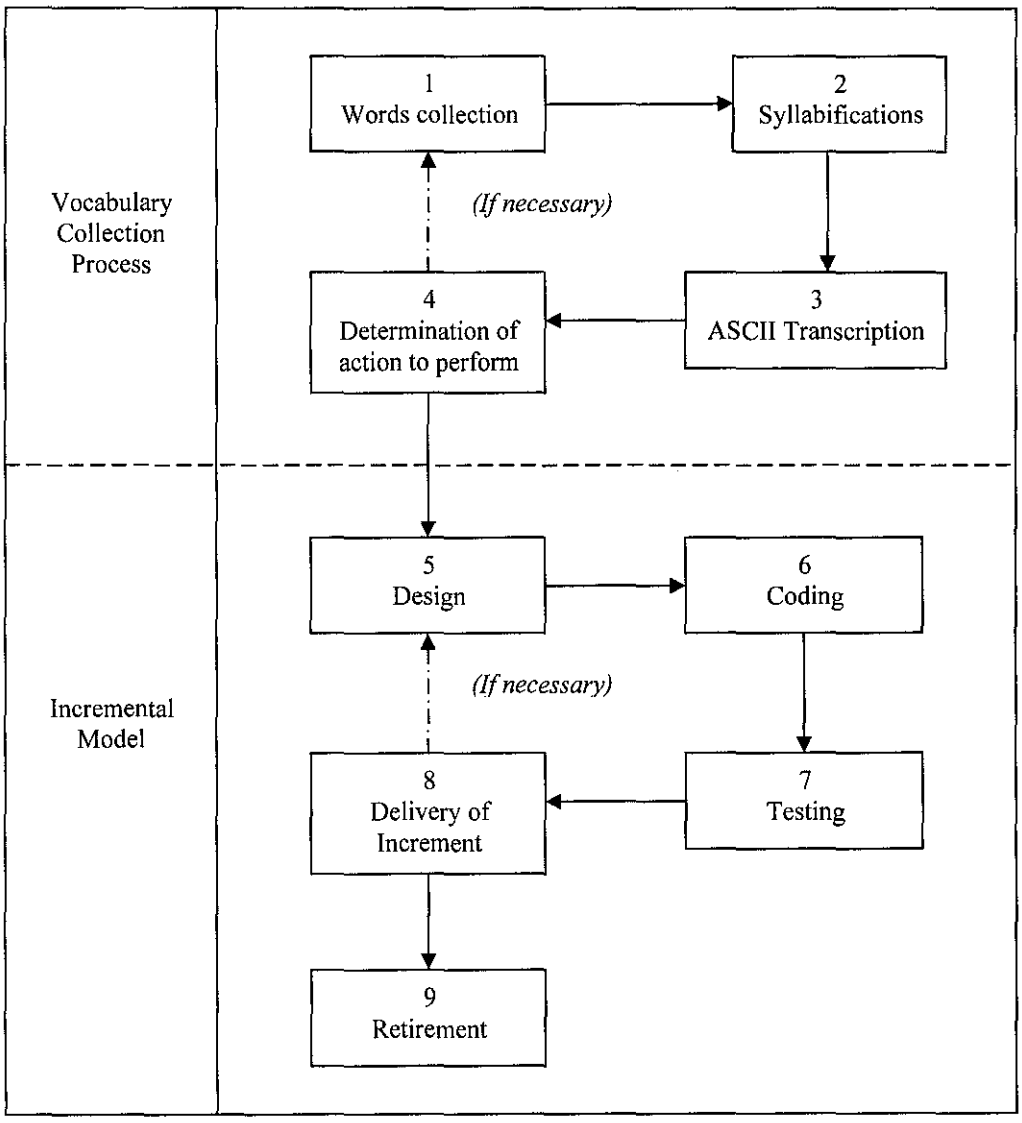


Figure 3. 3 : Combination of vocabulary collection process and incremental model

The combination methods consist of eight steps. The first step, *analysis* from incremental model is eliminated since the process of gathering and analysis of data is done in steps 1 to 4 of the combination model. The processes begin with words collection, word syllabification, ASCII transcription, system design, coding development, testing and delivery of increment. Each process is described in details on the next page.

3.1.1 Phase 1 : Words Collection

Pocket Oxford Dictionary 7th Edition is used as the major reference for words to be included in Oral Dictionary. This dictionary is used as reference since it has the reputation in the field of English dictionary. Apart from that, this dictionary is very comprehensive and used British-English as the standard for spelling, pronunciation, word usage etc.

For the prototype purpose, only 100 words are selected to be included in this project. No specific guideline to indicate how many words from each alphabet. However, the criteria of the words should finally be able to address the successful functionality of the system during the testing phase later. The criteria include simple to lengthy words, homophone, words with more than one definition etc. The words to be covered for this prototype are limited to noun context. Another important information is the definition of the word itself. The definition should be exact with the definition given by Pocket Oxford Dictionary to ensure the integrity of the information provided by the system. Although Oral Dictionary will no longer provide the phonetic transcription to user, it is still important to be considered as part of data collection. The phonetic transcription is necessary to provide information on the pronunciation of each word to users. Oral Dictionary will use the information from phonetic transcription to provide user better way of learning the pronunciation by introducing a new approach in later stage.

Example of the word expected from the collection is as the following :

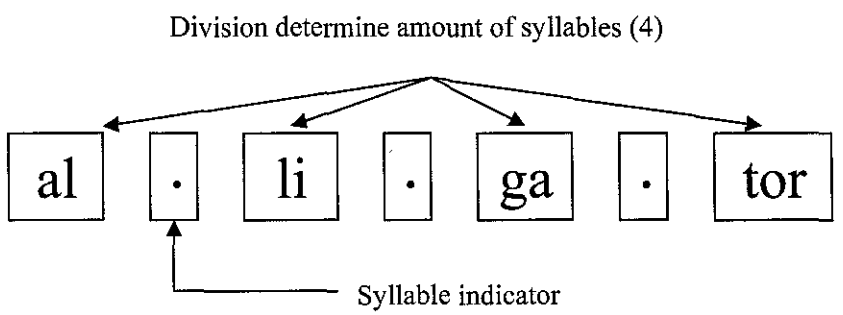
Table 3. 1 : Necessary word information

Word	Phonetic Transcription	Definition
zoo	**zu:**	Zoological garden [abbr.]; Public garden or park with collection of animals for exhibition and study

3.1.2 Phase 2 : Syllabification

Syllabification is the process of forming or dividing words into syllables. Syllable is the most basic element in any language and fortunately, it is countable. But according to the linguists of MacQuerie University in Australia, there is no exact definition that phonologist agreed upon what a syllable is. They believe that the variation in defining the syllable depends on the speaker awareness. From the research, they conclude that finding the syllable boundaries is much difficult than counting the number of syllables especially for those who have not exposed to alphabetic writing system.

Though, for Oral Dictionary, it is strongly recommended to use the dictionary reference rather than the speaker awareness to determine the syllable of the word. As for this stage, Oxford Advanced Learner's Dictionary software is used since Pocket Oxford Dictionary did not provide the word syllabification. The syllabification is one of the important parts because the variant in syllabification affect the effectiveness of the speech recognition system. The syllabification also will be helpful for determining the approach in dealing with homophone words later. The "." is used as symbol to indicate the syllable division and amount of syllable is measured by how many division are there separated by the small dot. Example of syllabification is illustrated by using word *alligator* :



Example of the syllabification result expected are :

Table 3.2 : Syllabification

Word	Syllabification	Amount of Syllables
bangle	**ban.gle**	2
bungalow	**bun.ga.low**	3

3.1.3 Phase 3 : Phonetic to ASCII Transcription

Phonetic transcritption primarily used in common dictionary to provide information on the word pronunciation. However, in Oral Dictionary this information will no longer be displayed to user. It is not even used for the pronunciation purpose in Oral Dictionary but it performs as important characteristic for evaluating the homophone words. The words are categorized as homophone to each other if they share similar phonetic transcription which indicates the pronunciation.

Almost existing dictionaries used International Phonetic Alphabet (IPA) standard to represent the phonetic transcription to user. The strange symbols however, are difficult to be understood by the computer system. Therefore, it is important to convert the original phonetic transcription into its corresponding ASCII symbol. The representation provided by Kirshenbaum [34] in Appendix A was used as major reference to convert the IPA symbol into readable ASCII symbol. Each ASCII representation will also include the syllabification of each word.

Example of the conversion is :

Table 3.3 : IPA Phonetic to ASCII transcription

Word	Phonetic Transcription	ASCII Transcription
zoo	**zu:**	**zUW**
tuna	**tju:na**	**tyUW:nAX**
eagle	**i:gl**	**IY:gl**

3.1.4 Phase 4 : Determine Subsequent Action

Once the phonetic transcription of each word is successfully converted into the ASCII symbols, the process continue to determine the next stage it will undergo. As mentioned previously, the above steps will be repeated if the developer needs to add any new vocabulary to the system. Otherwise, the developer will decide to proceed to the next stage since the words information was already complete to be used in the design process.

3.1.5 Phase 5 : System Design

Once the vocabulary collection process is complete, the system design phase will take place as part of the incremental model. The system will consider as well the background study, problem statement and literature studies presented in the early discussion in designing Oral Dictionary.

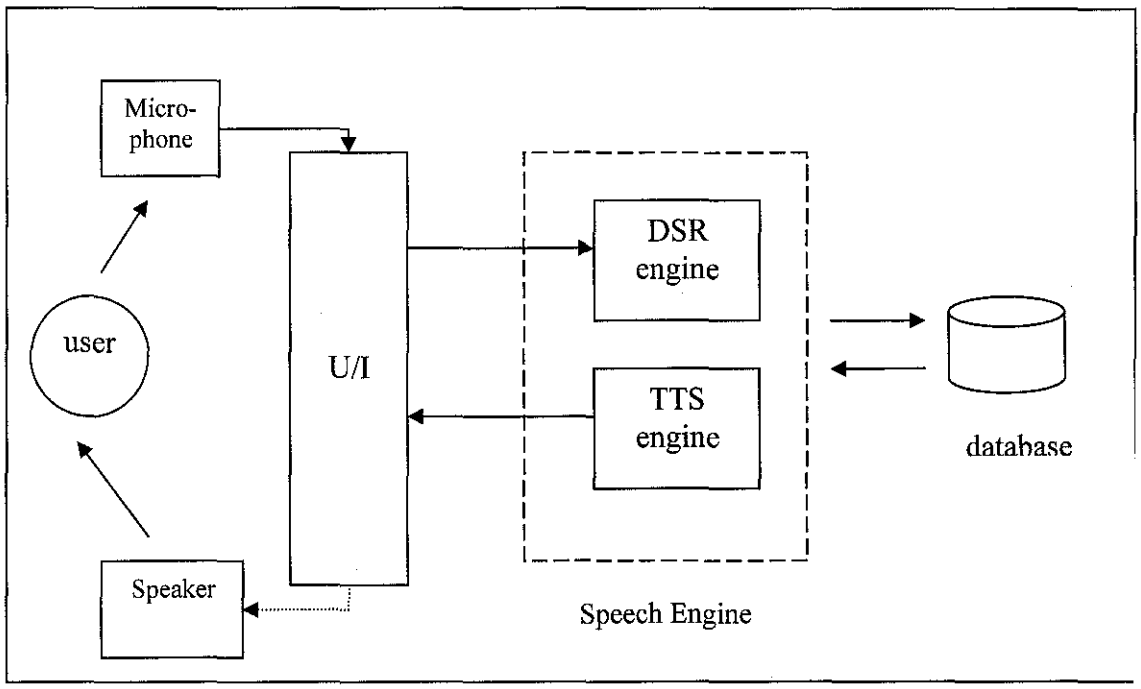


Figure 3.4 : System architecture for Oral Dictionary

The proposed system architecture for Oral Dictionary is shown in the Figure 3.4 above, involving 4 components which are user, user interface, speech engine and database. Generally, user interface enables the communication between user the system where it provides control for user and display of the system. User will perform voice input using microphone and Discrete-Speech-Recognition (DSR) engine will translate the voice into text (word), which is based on its availability in the text file. Once the user wants to hear the pronunciation of the word, the text-to-speech (TTS) engine will be converting the text (word) into phoneme-based audio, as stored in the database. Finally, the output of pronunciation sound will be heard by user through a speaker.

3.1.5.1 Functional Requirement

The background study performed previously will be used to prepare a functional requirement for Oral Dictionary. The functional requirement specifies functions that a system or component must be able to perform and later it will be used as reference during the testing process, purposely in preparing the test case and test plan. The functional requirement of Oral Dictionary list out that the system should be able to :

- The system should allow user to input the word using voice.
- The system should be able to give appropriate response when the word spoken is not recognized by the system.
- The system should be able to display the result from the word spoken using the best match case.
- The system should allow user to see possible matches word if the result is not the word intended by user. The possible matches should have exact phonetic matches.
- The system should be able to display the definition(s) of the word once the word is recognized.
- The system should able to read out loud the definition if requested, especially for user with visual impairment.
- The system should be able to playback the British-English phonetically correct pronunciation of the words.
- The system should allow user to type the word if result failed to be obtained using voice input.

Once the functional requirement has been specified, the flowchart is prepared to represent the sequence of activities, steps and decision points involved in Oral Dictionary.

3.1.5.2 Flowchart of Oral Dictionary

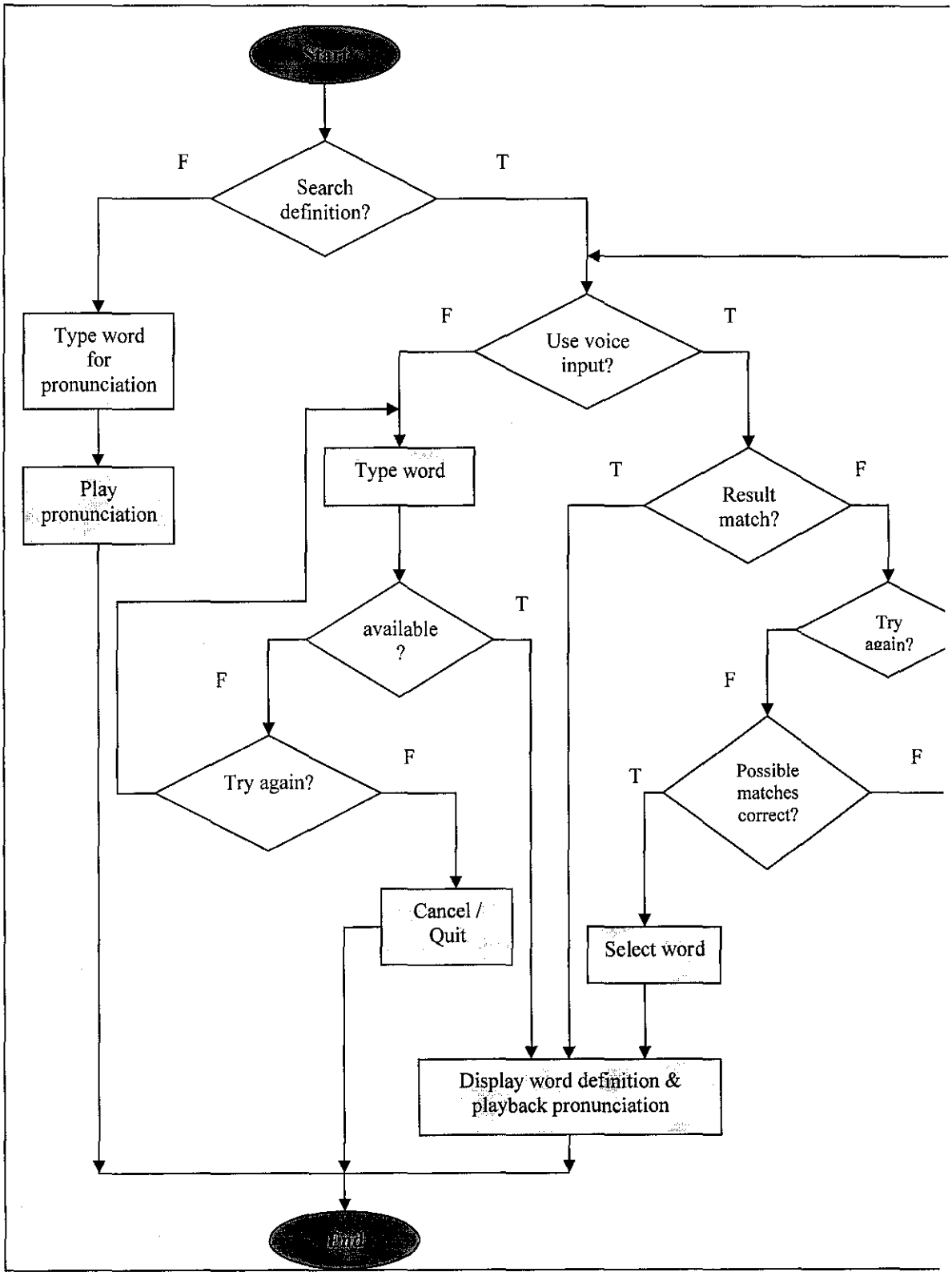


Figure 3.5 : System flowchart for Oral Dictionary

3.1.5.3 Storage of Words Information

The words information gathered earlier in the vocabulary collection process is then stored in the database. Microsoft Access is used as database to store those 100 words with the name `dictionary.mdb`. The table consists of 5 fields, which are ID, Alphabet, Word, Definition and ASCII. The ID is the auto-generated number to determine the primary key as well as the index of the word in the table. Although it is only one table involve, it is still necessary to provide the primary key for easier further enhancement work, especially when it involves additional table. The item of each field is as following:

Table 3.4 : Item field with data type and field description

Item	Data Type	Description
ID	Auto Number	Auto-generated field represented as index as well as the primary key for this table
Word	Text	Field to store all the words.
Alphabet	Text	Field which is determined by the first alphabet of the word, to allow alphabetical ordering
Definition	Text	Field which stores the definition of each word
ASCII	Text	Field which is resulted by conversion from phonetic transcription to ASCII symbols.

However, the words to be recognized by the speech recognition engine must be stored separately in a text file. The text file is named `Words.txt` and only consist the words. The language must be set earlier, where `langid=1033`, which is British-English. The engine will only recognize the words defined after the `<start>` tag as the following :

```
[Grammar]
langid=1033
type=cfg

[<start>]
<start>=airplane "airplane"
```

3.1.5.4 Interface Design

In order to accomplish this phase, 2 softwares are required which is Adobe Photoshop CS 9.0 for graphic editing and Microsoft Visual Basic 6.0 for designing the user interface. Since the user interface is the method of communication between the user and system, it is important to design the system with maximum usability. Misrepresentation will easily occur when the system perceived by user, is different from what is intended by developer. Therefore, a human-computer interaction (HCI) element is important to be incorporated into the design.

Basically, the menu section of Oral Dictionary will display 3 available options which is :

- Search word using voice (using voice input)
- Search word using keyboard (using keyboard input)
- Pronunciation only (convenient for user who only want to hear the pronunciation)

As the first menu is selected, user will perform the word input using their voice. Designing the speech interface however, is not exactly the same as designing the usual interface. Speech input is known for its prone to error. Therefore, it is important to identify and analyze the problems related to speech recognition and try to minimize as much as possible those errors associated with such problems using appropriate design.

Misrecognition occurs when the speech engine recognized word which is different from what the user spoke / intended. Misrecognition is likely to happen for homophone words; few words which sound alike to another. As for Oral Dictionary, recognition error is serious because it will change the meaning of an utterance. To overcome this problem, the system is designed to provide possible matches of words as spoken by user previously, as shown in the diagram below. The cut-off of the recognition is set to 80%, which means the system will display all the words which is minimum 80% phonetically matches with word recognized by the system earlier.

The possible words are in descending ranked, where the highest recognition confidence level will be at the top most and is highlighted. If the word is to be the intended word, the system will display the definition of the word as well as allowing the user to hear the pronunciation of the word as presented in the diagram. Otherwise, user can choose from the options available if the highlighted word is not what he or she looking for.

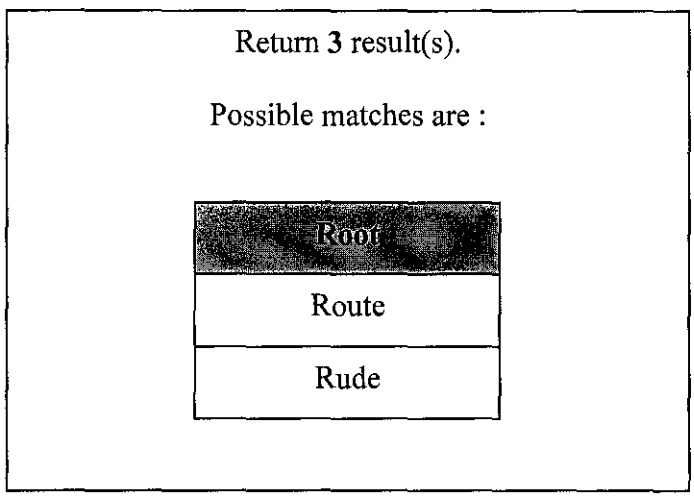


Figure 3.6 : Interface design for possible matches result

The system will tell the user that no result is found if the word is not available in the Oral Dictionary's vocabulary. There is also situation whereby the system will return the result as similar as above if it is not able to recognize the word spoken by the user. This is due to user over-emphasis on word or uneven voice of user who suffers from cold. In this case, the system will allow the user to input the word using keyboard by simply typing it in the text box provided. It will serve as an alternative to voice input means and hence, providing more flexibility for users in using Oral Dictionary application. Finally, for pronunciation only, a textbox is provided for user to enter any word which they want to hear the pronunciation. This option is convenient for user who is looking for the sound of the word without having to obtained unnecessary output such as the definition. The pronunciation must be ensured to provide the user with correct British-English phonetic pronunciation, which defeats the existing dictionary software who still depends on the sample recording o each word.

3.1.6 Phase 6 : Coding Development

Using human's voice is agreed to be the fastest mean for input compared to writing or typing. However, the system will remain useless if the time taken for output retrieval is longer than typing or other means. Even though the prototype of Oral Dictionary will not cover all words in dictionary, the approach has to be implemented as if in 'actual' dictionary which involves large size of vocabulary.

Since speech recognition is prone to error, it is important for the application to be designed in ways to minimize error as much as possible. Some English words are pronounced similarly but have different spelling (homophone words). The biggest limitation of Direct-Speech-Recognition (DSR) engine is, it only capable of providing one result for each word spoken by user, based on the best matched case. This limitation is crucial because some English have similar sound but different in spelling as well as meaning (homophone words). For homophone cases, the system is prone to misrecognition as the result (word) given by the system might not be the word that user spoke, and to the worse extent, wrong word definition is given to user. In order to reduce the misrecognition associated to homophone cases, Oral Dictionary is designed to provide possible matches result based on exact ASCII phonetic matches. The words are homophone if it has similar phonetic transcription. Based on this fact, string comparison of ASCII phonetic information is used to determine the possible matches result.

Example :

Word recognized by system : **root**

Phonetic transcription : **ru:t**

ASCII phonetics : **rUW:t** ------→ comparison model

Alternatives for the word *root* is determine as below :

Table 3. 5 : Phonetic string comparison

Word	Phonetic Transcription	ASCII Phonetics	Exact Phonetic Match?	Display?
route	ru:t	rUW:t	Yes	Yes
rod	rɔd	rOHd	No	No

As for above case, only word *route* will be displayed in list box as possible alternatives (matches) of the word *root.*

Finally, it is a good practice for Oral Dictionary to follow the standard naming convention and provide comments in the programming code. This practice will make the programming code more understandable and it is useful if future enhancement need to be done by other programmers. Last but not least, it is also helpful for programmer in tracking errors in Oral Dictionary application.

3.1.7 Phase 7 : Testing and Evaluation

The system will go through the testing process as soon as it is completed. The testing will be conducted primarily by author as the project developer, as well as group of real users. Generally, the purpose of testing is mainly to :

- To identify any problem or system defect
- To assess the system functionality
- To evaluate the system usability
- To obtain user's feedback pertaining to the use of voice recognition system

There are 2 groups of real users involved in the testing phase. The scope of users for testing is limited to University of Technology PETRONAS students and staffs only because of the time constraint faced. Group 1 consists of 10 UTP students, with equal representative from Technology and Engineering course. Group 2 consists of 10 UTP staffs and they represented as the user group from professional workers. Both are few focus groups of dictionary user. Although both groups used dictionary primarily for searching word definition, but yet they are different in term of frequency of usage, lifestyle, perception on dictionary etc.

Group 1 represents the students group whose frequency usage is higher for the educational and learning process. These group deals with tones of assignment in English, and they are likely to make frequent reference to dictionary. Therefore, they need a dictionary system which contains a lot of words and capable to provide quick result for word definition. Audio word pronunciation allows them to learn the sound quickly in order to suit their fast paced lifestyle.

Group 2 represents the professional worker, where dictionary is less often used. The members of this group mostly do not have strong acquaintance with softwares, especially dictionary software as well as speech recognition system. Therefore, the system should be tested whether it is easy to use as well as easy to be remembered.

3.1.7.1 Evaluation Procedure

The steps involved in the evaluation process includes :

1. Testing will take place in 2 different environments; silent as well as noisy surrounding :
    - Noise : Room with music from radio, fan sounds and 5 people talking
    - Silent : Similar room but no sound from talking human nor music, fan
2. Perform 2 areas for testing: Functionality and overall system interface design.
3. Both groups are tested on individual member, using similar set of task series.
4. The task should begin the most critical to least critical task as related to system functional requirement.
5. Average time taken and misrecognition rate from respondents is calculated.
6. Questionnaire filled by respondents are collected and analyzed. (Appendix B)

3.1.7.2 Set of Task Series

1. Search word definition for 5 given words using normal dictionary and repeat it using Oral Dictionary. Time used to complete each given word is recorded.
2. Search word definition for 5 given words in silent room. Record any misrecognition as well as how many repetition to obtain the correct output.
3. Repeat step 2 in noisy room for both groups.
4. Search word having homophone using voice input. Observe the output
5. Search unavailable word using voice input. Observe the output
6. Search word using keyboard input. Observe the output
7. Search unavailable word using keyboard. Observe the output.
8. Click button to hear pronunciation and definition. Listen to output.
9. Type any word for pronunciation only and click the Speak button. Listen to output.
10. Browse the menu independently. Observe the layout design and the function.
11. Answer the post-test questionnaire provided.

3.2 Tools Required

Tools required for Oral Dictionary project development are divided into software and hardware as the followings :

3.2.1. Software

- *Visual Basic 6.0*

  Visual Basic 6.0 will be the application interface for users since it is the development platform of Oral Dictionary

- *Microsoft Office Access 2003*

  Microsoft Access is used as database storage for information of 100 words collected in vocabulary collection process.

- *Microsoft Speech SDK 5.0 or higher*

  Microsoft Speech SDK should feature at least 2 basic engines, which are Discrete Speech Recognition (DSR) engine to convert sound into strings and Text-to-Speech (TTS) engine to process text input into digital sound.

- *Windows Notepad*

  The text file is used to store and limit the vocabulary to be recognized by the speech engine.

- *Adobe Photoshop CS 9.0*

  The graphic editor is used for editing images which to be included in Oral Dictionary interface, mainly for designing the splash screen.

3.2.2. Hardware

- *16-bit Sound Card*

  Sound card is compulsory to capture the user's voice and perform conversion from analog to digital and vice versa

- *Microphone*

  Microphone is required as voice input device. Headset is preferred since this type of microphone picks up less background noise compared to usual microphone. [5]

- *Speaker (Optional)*

  Speaker will perform as the output device for Oral Dictionary application. It is only an optional hardware if the headset is not available.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Result

4.1.1 Searching Time Comparison between Normal & Oral Dictionary

The testing is conducted to compare the time taken by user in searching word definition using normal dictionary and Oral Dictionary. Pocket Oxford Dictionary 7th Edition is used to represent the normal dictionary. 5 words involved are *apple, lava, route, xylophone and zoo.*

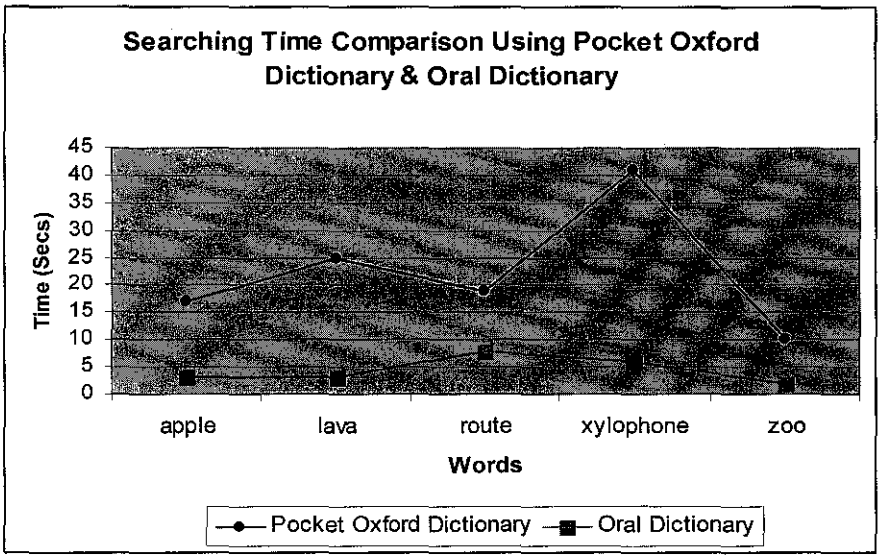


Figure 4.1 : Result for searching time comparison between Pocket Oxford Dictionary and Oral Dictionary

Figure 4.1 shows the result from comparison made for average time taken to search word definition using pocket Oxford Dictionary (POD) and Oral Dictionary (OD) by 20 respondents. For word *apple*, respondents took averagely 17 secs using POD and 3 secs by OD. Average of 25 secs is taken to search word *lava* using POD while only 3 secs is needed for OD. Word *route* took average of 19 secs for POD and 8 Secs for OD. POD needs 41 secs to search *xylophone* while OD only need 6 secs. Finally, word *zoo* is located averagely after 10 secs using POD and only 2 secs using OD. Overall, the difference is much significant as the Oral Dictionary provide faster result for each search.

4.1.2 Error & Misrecognition Rate

Table 4.1 : Result of recognition error by developer in silent and noisy environment

Speaker : Project Developer (Author)			
Word	**Syllable**	**How many attempt to correct result?**	
		Silent	**Noise**
zoo	1	1	3
route	1	2	3
eagle	2	1	5
caravan	3	2	5
xylophone	4	2	7

Result from Table 4.1 shows that in silent environment, only an attempt is needed to search *zoo* and 3 attempts performed before correct output is produced in noisy room. The word *route* required 2 attempts in silent room and 3 attempts in noisy surrounding. Only one attempt for system to produce the correct output for *eagle* in silent room, while 5 attempts in noisy room. Word *caravan* is attempted 2 times in silent room compared to 5 times in noisy surrounding. Finally, *xylophone* is produced after 2 attempts in silent room while noisy environment is produced by 7 attempts.

Table 4. 2 : Result of average recognition errors from respondents in silent and noisy environment

Word	**Syllable**	**How many attempt to correct result?**	
		Silent	**Noise**
zoo	1	1	3
route	1	3	4
eagle	2	2	5
caravan	3	2	5
xylophone	4	2	8

Table 4.2 shows that in silent environment, respondents perform averagely one attempt to search *zoo* and 3 attempts performed before correct output is produced in noisy room. The word *route* required 3 attempts in silent room and 4 attempts in noisy surrounding. 2 attempts are performed for system to produce the correct output for *eagle* in silent room, while 5 attempts in noisy room. Word *caravan* is attempted 2 times in silent room compared to 5 times in noisy surrounding. Finally, *xylophone* is produced after 2 attempts in silent room while noisy environment is produced by 8 attempts.

4.1.3 Verification of British-English standard in Oral Dictionary

The test is primarily tested to one of Oral Dictionary's major function where it is able to playback the pronunciation of the word if requested by user. This function is performed using the text-to-speech engine to synthesize the word and produce the audio output of the word. Since Oral Dictionary application is using British-English standard, it is important to test this function to verify that the words pronounced by the system is British-English phonetically correct pronunciation. However, not all English words are tested as some of them are pronounced similarly in any English standards. 10 English words are used for test case, where the criteria of these words must be pronounced differently from American-English standard. The original British-English and American-English audio output of each word pronunciation is obtained from Oxford Advanced Learner's Dictionary software. The original output is used as comparison model to verify this standard. The result obtained as the followings :

Table 4. 3 : Validation table for English pronunciation standard

Words	American-English	British-English	Oral Dictionary	Result
aluminium	aluminum	aluminium	*aluminium*	valid
apricot	a-pricot	ay-pricot	*ay-pricot*	valid
β	bayda	beeta	*beeta*	valid
cordial	corjul	cordee-al	*cordee-al*	valid
fillet	filay	filit	*filit*	valid
privacy	pry-vacy	priv-acy	*priv-acy*	valid
route	rout	root	*root*	valid
schedule	skedule	shedule	*shedule*	valid
tomato	tom-ay-do	tom-ah-to	*tom-ah-to*	valid
vase	vayz	vahz	*vahz*	valid

The result indicates that each word follows the same audio output of British-English standard in Oxford Advanced Learner's Dictionary software. Therefore, Oral Dictionary application is proven to apply British-English as the standard.

4.1.4 Result from Post-Test Questionnaires – System Functionality (Appendix B)

4.1.4.1 Group 1 – Students

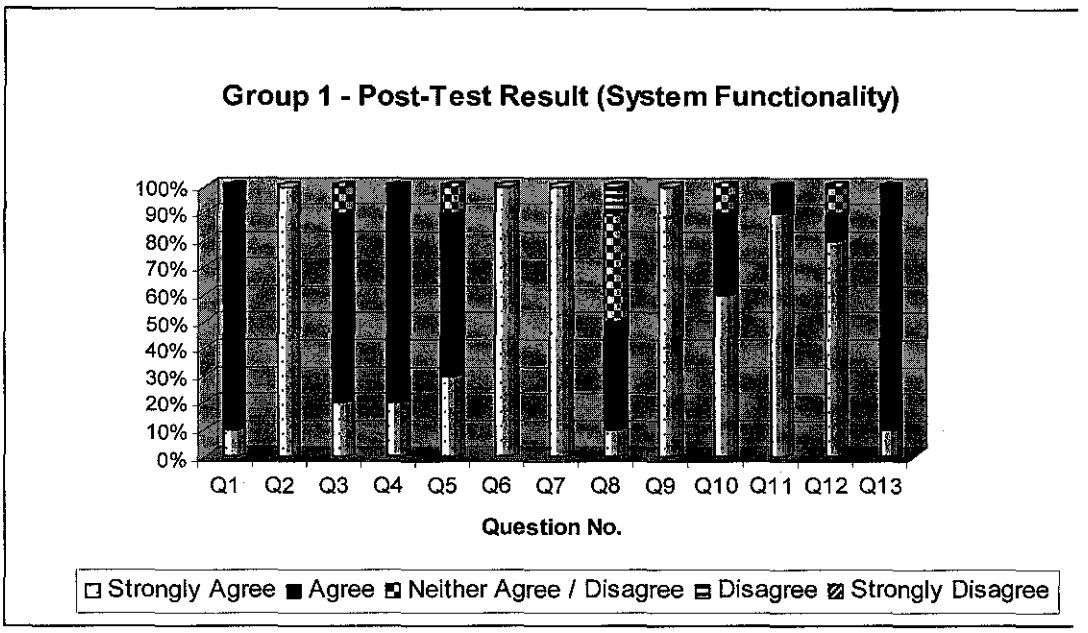


Figure 4.2 : Post-Test result of Group 1 on system functionality

From the result of post-test, the students group agreed the voice recognition function is easy to use. They agreed that using voice input is much simpler as they still can perform the task regardless of their knowledge on the word spelling. 90% respondents agreed the system produces output faster than using the normal dictionary. 90% agreed that the system produce output similar to what is spoken and this indicate that few misrecognition is still tolerable as the performance might be interfere with external factor such as noise. The students strongly agreed that the possible matches return relevant result to the word spoken and they found this function is very useful. They strongly agreed that keyboard function is useful since there are situations that speech recognition is not effective. Based on the list of words given, half of them disagreed that Oral Dictionary has large vocabulary. 100% strongly agreed that audio pronunciation is clear. 90% of respondents also agreed the read-out-loud definition is clear and useful.90% agreed that rephrase or retyping the word is easy to perform if mistakes occur and finally all of them agreed that Oral Dictionary clearly serve its purpose.

4.1.4.2 Group 2 – Professional Workers

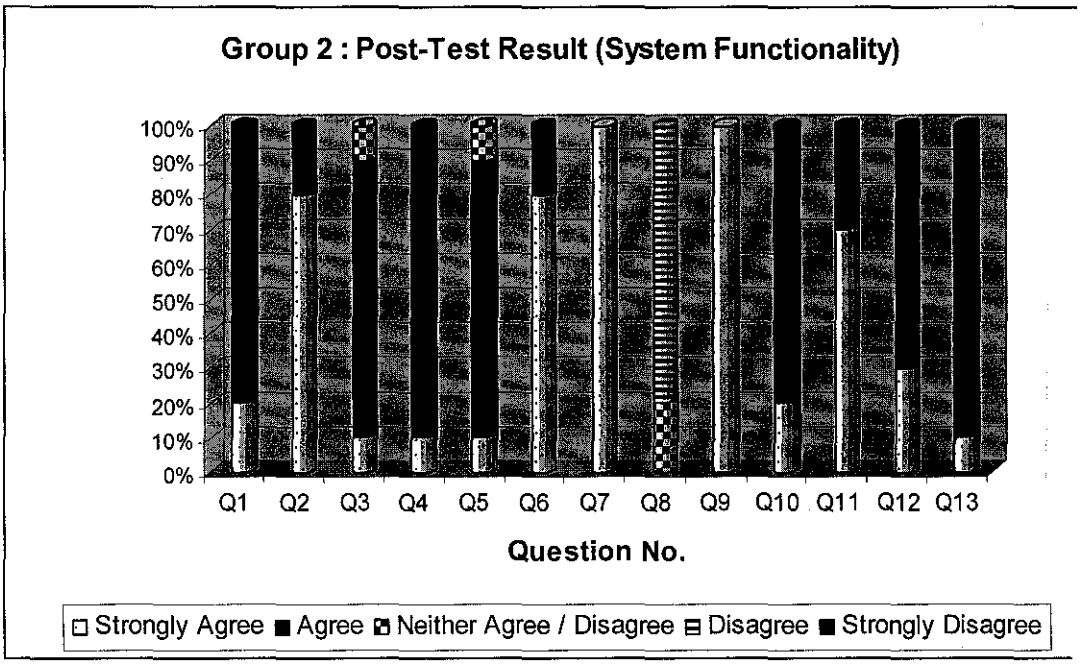


Figure 4.3 : Post-Test result of Group 2 on system functionality

From the questionnaire result, the professional group also agreed that the voice recognition function is easy to use as well as reduced the need to memorize the word spelling. 10% is neither agrees nor disagrees to Question 3 that the system produces output faster than expected. All of them agreed that the searching process is much simpler compared to normal dictionary. 90% agreed that the system produce result similar to what is spoken. The respondents agreed that the system return relevant possible matches and keyboard is a useful alternative to voice input. Based on the word choice, 80% disagreed that the system offers large vocabulary. Similar to student group, the workers also agreed that audio pronunciation is very useful as well as the read-out-loud definition. The respondents agreed that the buttons performed the expected action. The respondents also agreed that the mistakes are easy to be corrected and Oral Dictionary deliver its purpose clearly.

4.1.5 Result from Post-Test Questionnaires – Overall System Interface Design (Post-Test Questionnaire from Appendix B)

4.1.5.1 Group 1 – Students

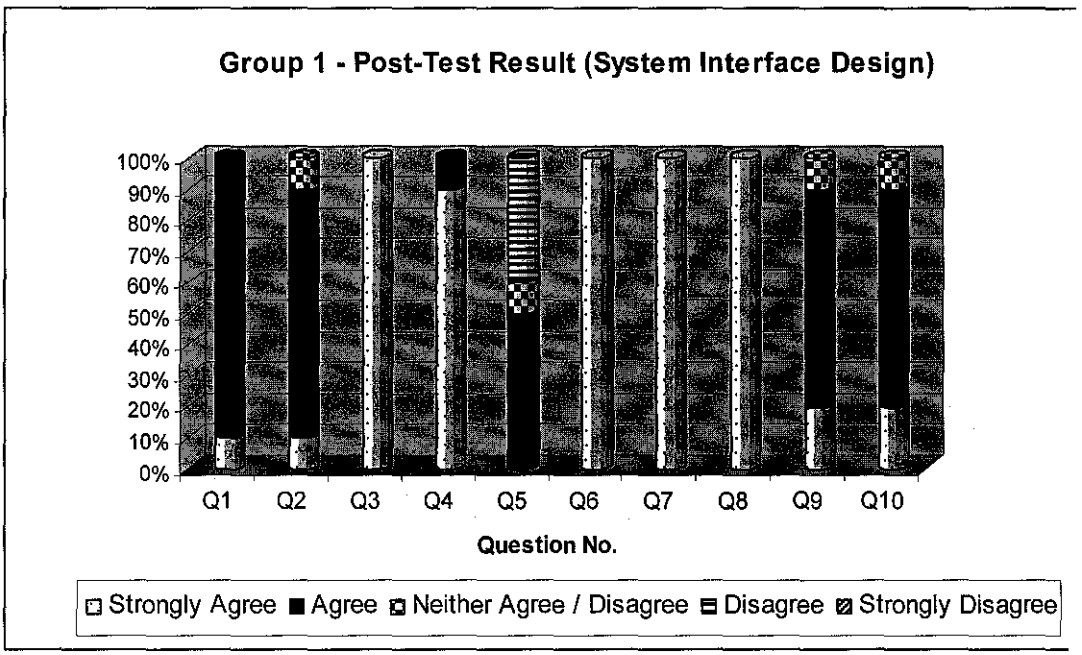


Figure 4. 4 : Post-Test result of Group 1 on overall system interface design

Respondent from students group agreed that the system overall is easy to use. 90% agreed the interface of the system is user-friendly while 10% neither agree nor disagree on the statement. The respondents strongly agreed to easily understand and follow the instruction. They also agreed the content definition is placed properly. Only 50% agreed the system used attractive colors. 100% respondents found the functions are easy to be remembered. 100% respondents also agreed the information is readable and be layered effectively. 90% agreed the presentation of content suit their preference. Finally, 90% found the alert technique using "word highlight" very useful.

4.1.5.2 Group 2 – Professional Workers

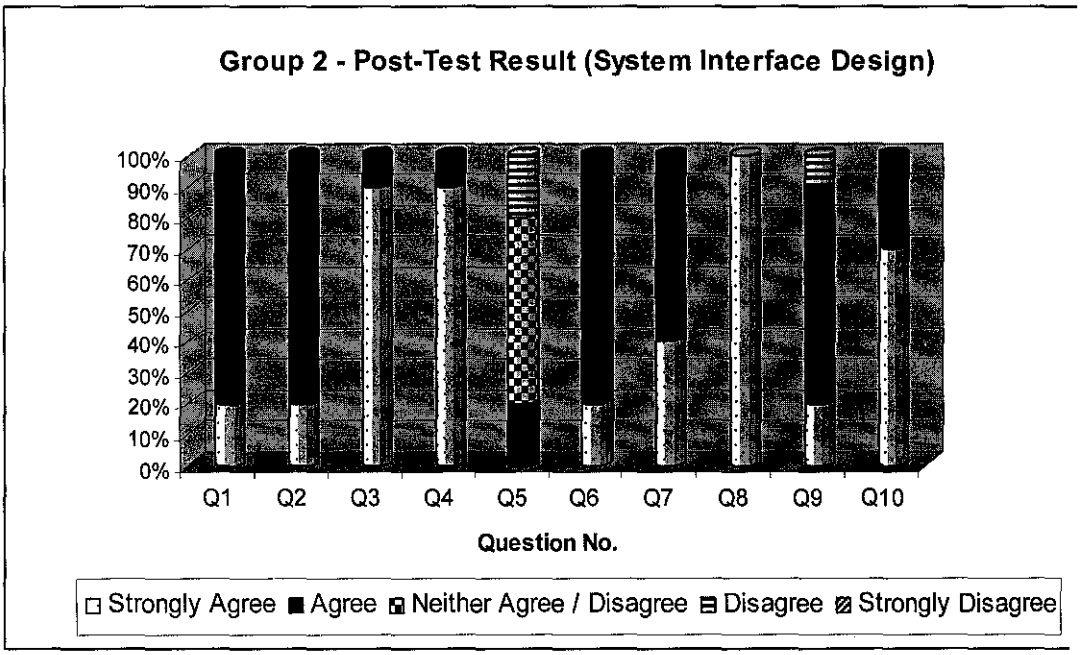


Figure 4. 5 : Post-Test result of Group 2 on overall system interface design

From the post-test questionnaire, respondents agreed the system is easy to use and the system interface is user-friendly. 90% of the respondents strongly agreed the instruction is easy to understand and the word definition is arranged properly. Only 20% agreed the system use attractive colors. The professional group also agreed the system has high degree of memorability. The respondents also agreed information is readable as well as be layered effectively on screens. Only 90% agreed the content is presented as what they prefer. Finally, the professional group also agreed that alerting technique used is very useful for highlighting words in possible matches.

4.2 Discussion

Oral Dictionary application is aimed to enhance the existing dictionary with the strategic implementation of speech recognition technology. The technology which allows user to search the word definition using the voice input is still be considered useless if it fails to deliver the purpose correctly as perceived by users. Therefore, the result from testing is used as main discussion in determining whether the project really address its functionality in a way expected by the real users, especially for the focus group. There are 5 relevant areas to be discussed which are vital for the success of Oral Dictionary project.

4.2.1 Searching time comparison using normal dictionary and Oral Dictionary

Based on result from testing, there is significant difference for time taken to search a word using normal dictionary and Oral Dictionary. Oral Dictionary successfully provides faster result for each definition searching. However, it is far more important to identify factors to such faster or lagging in time for both dictionaries. As for normal dictionary, the factors include position of front letter in alphabetical arrangement and length of word itself. However, for Oral Dictionary, the factors include the length of word and its syllable. The relationship between these factors and time performance is not exactly justified as some might interfere with factors that are very subjective to discuss such as human logical thinking and the user's experience also affect his efficiency (speed) in using dictionary.

Factor 1 : Position of front letter in alphabetical order arrangement

The word 'apple' has the advantage of position in POD because the front letter is 'a', allowing the word to be arranged in early pages of the dictionary. Therefore, the user does not have to flip a lot of pages compared to the word 'lava' which is located in the middle of the dictionary. As a result, the time taken for word apple is lesser than the word 'lava'. However, this rule does not apply for the word 'zoo' which took the shortest time of 10secs which simply because human logical thinking already perceived that the word

'zoo' will be located at the behind part of the pages as according to its position in alphabetical order. Therefore, logically user will no longer waste their time to search from the front pages but directly browse to the end pages to locate the word 'zoo'. A conclusion that can be drawn from this factor is the position of the front letter affect the searching time, where the searching time increase as order of the front letter is increase alphabetically without interference with external factors such as human logical thinking and user's experience of using dictionary.

Factor 2 : Length of words

From the result of searching time using POD, the word 'xylophone' took the longest time to be located. Even though human logical thinking already know that the word is arranged nearly the ending part of dictionary, the problem arise when the user need to compare the letters in the word alphabetically before it can be located. Comparing the letters in the word 'xylophone' require more time compared to other words as the word itself is lengthy and more time is used to compare between the letters. As for example, user started with "xy" and then moving to "xyl" then "xylo" and so on until they complete the word "xylophone". However, for the word "zoo", it only contains 3 characters which made much easier to locate the word in the dictionary once the page containing the words starting from letter "Z" is obtained. Therefore, it will require shorter time than to locate the word "xylophone" which is much lengthy. It can be concluded that the more lengthy the word, there is an increasing searching time using Pocket Oxford Dictionary, especially when the adjacent words have the letters closer to the intended word. As for example, locating "cordon" between the adjacent words of "cord", "cordial" and "corduroy" is more difficult than locating the word "velocity" in between of adjacent words of "veld", "vellum" and "velour".

Factor 3 : Number of syllables in a word

Similar to POD, "xylophone" is also the word which took the longest time for definition searching compared to word "zoo". This is particularly because of the word lengthiness

and its number of syllables. As for the word "zoo", it is much simpler and only has one syllable compared to word "xylophone" which is much lengthy and has 4 syllables. The speech engine need to listen to the syllables and combine all the syllables to determine it is the correct word as stored in the text file. Since there is only one syllable for the word "zoo", the speech recognition engine only need to listen to one syllable and simply can give the result from the text file without having to combine with other syllables. As for the word "xylophone", the speech recognition engine needs to combine the 4 syllables together and make the comparison with the list of words available in the text file. Therefore, more time is needed to listen as the syllables are increasing. As for the word with more syllables, the speech recognition engine might not recognize it at the first time and need few attempts before the correct word is recognized. This explains why lengthy and increasing number of syllables will also increase the searching time performance.

4.2.2 Error and misrecognition rate

Effort toward minimizing errors associated with speech recognition system is important. One of the common associated errors is misrecognition. The system will produce an error message if the speech recognition engine is not able to recognize the word spoken by user. The message is said to function successfully if the word spoken by user is not available in the text file which can be retrieved for Oral Dictionary. However, it is said fail to perform successfully if the system cannot recognize the word spoken by user which has been stored in the text file. The system is also said to perform in error if there is misrecognition, which is the system recognize the wrong word and produce the wrong result to user. Although the misrecognition cannot be totally eliminated, the system still has to keep the error level within an acceptable rate. From the testing, there are few factors that lead to misrecognition and other potential error.

Factor 1 : Effect of using different voice with error rate

The error seems to increase when the real user with no speech training profile is tested. The speech training profile of user is needed to allow the speech recognition engine to

'familiar' with the user's voice, the speaking rate, the pronunciation of letters and etc. This is important for the speech recognition engine to produce the spoken word based on the information of user's voice that it already stored. However, the error rate resulted from the test shows that differences are not significant, therefore indicates that Oral Dictionary is a speaker-dependent system. The system is not designed to be speaker-dependent because the voice input is not a sensitive data as used for authentication in a security application. Therefore, any user can use Oral Dictionary regardless of whose speech profile is trained in the computer system. However, it is still recommended that the speech recognition engine is trained by the user who is going to use the system to reduce the error rate as well as to provide effectiveness of Oral Dictionary.

Factor 2 : Effect of using different environment with error rate

The result showed that the error increased whenever different environment is used to conduct the testing. Both author and real user have the test conducted in silent as well as noisy environment. Same room is used to conduct the test but different environment has been created. The test is conducted in a noisy room surrounded with the sound of fan, music from the radio as well as 5 persons talking to each other in the room. However, no fan or radio is turned on and nobody is chatting when the room is tested for silent environment. From the test, it is obvious that noisy environment introduced more errors to the system compared to silent environment where author and real user have error of 8 and 10 for 5 words tested which is "zoo', "route", "eagle', "caravan" and "xylophone". However, the error shows significant increase in noisy environment where 23 errors produced by author and 25 resulted from the real user. When the error occurs, the user need to provide another few attempts to see the correct result produced by Oral Dictionary. More attempts indicate that the errors were introduced during their previous attempt to request for the words. In a noisy environment, the speech recognition engine is exposed to various sounds and the noise will interfere with the word spoken by user, causing the 'confusion' to the engine on which sound need to be entertained. The combination with the unknown sound of noise generates the new 'vocabulary' which is

not defined in the text file, therefore causing the system to produce the error message that the result is not found. Misrecognition also occurs in this test based on the assumption that the engine has combined the noise with the word spoken by user and unfortunately the word is available in the text file. This explains why error rate is higher when Oral dictionary is used in a noisy environment.

Factor 3 : Effect of homophone word with error rate

As discussed previously, homophone word is English word which has similar pronunciation but different in spelling as well as its definition. As for this test, the word "route" is used as one of the test case to see whether the misrecognition occur and how many attempts need to get the actual result because it is homophone with "root" and relatively close to the sound of "rude" and "rod". The misrecognition is mostly associated with homophone word and from the test; the system frequently produced the word "root" although the intended word is "route".

Although the word "eagle" is not homophone with other word, the system has once produced the word "bangle" as the result. It indicates that the system is able to hear the last syllables clearly compared to the early syllables which has similar sounds or homophone.

As for the other words which is "zoo", "caravan" and "xylophone"; the misrecognition did not occur during the test, probably currently in the text file there is no such word closely sound as these three words. Therefore it can be concluded that the error rate is higher if the word is known to have similar sound (homophone) with other words stored in the text file.

Factor 4 : Effect of no. of syllables and error rate

From the result, no. of syllables also has some effect on error rate produced by Oral Dictionary. The word "zoo" has slightly no error in silent environment compared to the

word "caravan" and "xylophone". The word "zoo" only has 1 syllable but the word "caravan" has 3 syllables and 4 syllables for "xylophone". As discussed earlier, the number of syllables affected the speech recognition engine because the engine needs to listen to more than a syllable and combine the syllables together before making the comparison with the words available in text file in order to produce the result. As the syllable increases, there is higher probability for the speech engine to overlook the earlier or middle syllables compared to the last syllables. Such as the word "caravan", the speech engine might overlook the syllable "ra" in the middle of the word because the pronunciation of that syllable is descended, causing the system to produce the error message indicating that the result is not found because the speech engine 'believes' the word is not stored in the text file. Therefore, it can be concluded that the error rate is higher when the no of syllables in the word increased.

4.2.3 System functionality

User satisfaction on the system functionality can be measured based on the questionnaire evaluation result from both student and professionals group. The voice recognition system is agreed as easy to use because user only needs to input the word using their voice to microphone. Therefore, it reduces the dependency on word spelling in searching word definition as well as being much simpler than normal dictionary. As expected, the system allows the word definition result to be obtained much faster than using normal dictionary. However, 10% of respondents from the students group expect faster than the current performance, which indicate that user's current fast-paced lifestyle required the output to be faster or similar to what they expect. The system is mostly agreed to produce the result as spoken by the user while 10% respondents from both group neither agree nor disagree to the statement. This indicates that the misrecognition rate is low and within the level tolerable by users since external factor such as noise should be considered. The function that provides possible phonetic matches is really useful as it is one of solution to reduce the misrecognition problem. The possible matches result return relevant results, typically words that have sound similar or closer to the word that is recognized by the system earlier. The keyboard input providing more flexibility for user in word definition

searching based on the fact that speech recognition may not be effective means in some situation, especially in noisy surroundings. This also gives options for those having sore throat to use the system by simply typing the word. It is more effective rather than using voice input which might difficult to be recognized due to significant change of voice. The respondents do not agree that the system has large vocabulary as responded to list of 100 words given. This indicate that, user preferred system which can offer them as much word as possible. The audio pronunciation works really well, as it quickly gives the user the idea of pronunciation. From the result, the students group highly prefers the audio pronunciation compared to phonetic transcription which is difficult to understand. The function that read-out-loud of definition is agreed to be useful as the output can be obtained by hearing rather than reading the sentence line by line. However, 10% of the respondents from the students group did not like this idea very well as the reading from text-to-speech (TTS) engine sound unnatural to normal human speaking. The idea at the first place was to give the visual impairment users with maximum accessibility. Users agreed that the buttons used in the system perform the actions as expected, which indirectly means that author has successfully use understandable icon and button as perceived by users. Any mistakes done by users are easy to correct, such as by rephrasing the word, retyping the word, clicking reset button etc. Therefore, it allows the user to explore the system without fear of making mistakes, and thus creating enjoyable experience in using dictionary application. However 10% neither agree nor disagree to the statement. The reason might be re-correcting the mistake is distracting to some extent. Finally, from user's point of view, Oral Dictionary has successfully delivered its purpose as a dictionary based on the functionalities that it offers to users.

4.2.4 Overall system interface design

For student group, the system is perceived as easy to use based on their experience working with several of softwares in their education days. Therefore, the responds from the professionals provide more values and they also agreed that system is easy to use. This result indicates the system is easy to use even by those who do not have close acquaintance with computer and dictionary application. The design of the interface is

user-friendly with the use of relevant icon button, readable font etc. The instruction used in the system is clear and easy to understand. This explains why using simple English sentence is important in order to guide user in following the instruction. Apart from that, users also agreed that the definition content is neat and properly arranged, where the Oral dictionary make it steps by step; user need to click some button to display the definition and other button for system to read-out-loud the word as well as the definition. From the questionnaire, some respondents did not find the color choice of pink as attractive. This measurement is very abstract to evaluate because some judge it based on their favorite colors. In term of system usability, Oral Dictionary has high-degree of memorability as users agreed all the functions are easily to be remembered as well as how to perform the task. The information is layered effectively on different screen, avoiding constraint to user's eye to see on crowded screen. Proper arrangement and appropriate selection for font and font size making the information very readable to users. Most of the users agreed that the information is written in a style that suits their preference. However, 10% of respondents from professional group disagreed to the statement, where the reason might be they prefer more professional looks and design. The use of alerting technique is very useful as applied in the system where the word with high percentage of phonetic matching will be highlighted for possible matches result.

CHAPTER 5

CONCLUSION

5.1 Conclusion

Oral Dictionary is aimed to provide enhancement to what existing English dictionary is lacked of. It still serves the same purpose of other English dictionary which is to provide user with list of words along with their definition as well as the pronunciation information of the word. However, different approach is used in Oral Dictionary in order to give user more useful and enjoyable experience in looking up for word definition rather than depending entirely on the word spelling as currently used in existing dictionary. Implementation of speech recognition technology allows the user to input the word they are looking for simply by using the voice. The advantage of using voice is it can be performed regardless of the user's knowledge on word spelling and thus, reduce the dependency on the word spelling to search word definition from dictionary. The ability to provide audio pronunciation of the word became another major advantage compared to traditional dictionary that is still using phonetic transcription, which is not understood by most users. Unlike other mobile dictionary which has to record each word for pronunciation playback, text-to-speech (TTS) technology implementation allows Oral Dictionary to stand out for its capability to provide audio pronunciation with correct British-English phonetic and yet reduce the consumption of recording memory. With the merging of speech recognition technology, Oral Dictionary will provide a new era in dictionary usage. With essential new features introduced by the use of speech recognition technology, users will gain as much benefits which is currently lacked from existing dictionaries. Generally, it can be concluded that the objectives of this project were achieved. The project is not just contributing to the growth of speech recognition application but also promoting how the invention of technology is useful in improving any field of study as well as human's life in general.

5.2 Recommendation

From the evaluation, respondents have listed some suggestion and comment regarding the Oral Dictionary application. Some of the recommendations are also from developer's observation throughout the development of this project. The additional enhancement includes :

1. *Develop mobile version of Oral Dictionary*

   Mobile dictionaries in various languages are currently available in the commercial market. Therefore, developing a mobile version provide more convenience for user as it can be carried along wherever they go, suitable with current mobility lifestyle.

2. *Include American-English standard*

   Providing American-English will give more flexibility to users instead of using British-English standard alone.

3. *Provide voice commands*

   Voice commands for menu selection and system navigation as well will allow maximum accessibility to visual impairment users.

4. *Include pictures and diagram*

   This will be helpful for better understanding of the word concept, as well as making the application more attractive to users.

5. *Provide more functionalities*

   Related functionalities as extension to dictionary function should be provided such as finding synonym, antonym words etc. These extra functionalities will provide more value to users in dictionary usage.

REFERENCES

[1] Dutoit, Thierry, *Introduction to Text-to-Speech Synthesis* available at http://tcts.fpms.ac.be/synthesis/introtts.html

[2] Furui, Sadaoki, *Digital Speech Processing, Synthesis and Recognition.* Marcel Decker, New York, 2001

[3] Ainsworth, W.A., *Speech Recognition by Machine,* Peter Peregrinus Ltd, London, 1998

[4] Fowler, F.G., H.W. *The Pocket Oxford Dictionary 7th Edition,* Oxford University Press, New York, 1990

[5] Goette, Tanya, *Factors Leading to Successful Use of Voice Recognition Technology* (2001)

[6] Sommerville, Ian, *Software Engineering 7th Edition,* Addison-Wesley, England, 2004

[7] Hincks, Rebecca, *Speech Recognition for Language Teaching and Evaluating : A Study of Existing Commercial Products* (2001)

[8] Phillips, B., & Broadnax, D.D. (1992, August) *National Survey on Abandonment of Technology : Final Report,* Washington, DC : Request Rehabilitation Center, National Rehabilitation Hospital.

[9] Goette, Tanya, *Factors Leading to the Successful Use of Voice Recognition Technology* (2001)

[10] Karat, J., Lai J., Danis C. & Wolf, C. *Speech User Interface Solution,* IBM T.J Watson Research Center.

[11] Dannenberg, R.B. & Hu, Ning, *A Comparison of Melodic Database Retrieval Techniques Using Sung Queries* (2002)

[12] Sprankle, Maureen, *Problem Solving and Programming Concepts 5th Edition,* Prentice Hall, New Jersey (2001)

[13] Peckham, J., *Human Factors in Speech Recognition,* Collins London (1986), pp. 172-187

[14] Yankelovich, Lai, *Speech in User Interface*

[15] *Attention and Memory Constraints* (2004)

[16] *Introduction to Usability* available at http://www.usabilityfirst.com/intro/index.txl

[17] *Introduction to Usability of Websites* available at http://66.102.7.104/search?q=cache:voDZfrTbuVEJ:www.becta.org.uk/page_documents/industry/advice/usability.doc+element+of+usability&hl=en

[18] Lafore, Robert, *Sams Teach Yourself : Data Structures and Algorithms in 24 Hours,* Sams Publishing, USA (1999)

[19] Russell, Brown, Skilling, Series, Wallace, Bonham & Barker, *Applications of Automatic Speech Recognition to Speech and Language Development in Young Children* (1996)

[20] J.J. Humphries, P.C. Woodland & D. Pearce, Cambridge University Engineering Department, *Using Accent-Specific Modelling for Robust Speech Recognition* (2002)

[21] *American-English* available at http://en.wikipedia.org/wiki/American_English

[22] *British-English* available at http://en.wikipedia.org/wiki/British_English

[23] *Our Strange Language* by E.L. Sabin available at http://home.planet.nl/~blade068/languagefun/pronunciation.htm

[24] Garfinkel, *History of Speech and Voice Recognition and Transcription Software-Dragon System* (1998) available at http://www.dragon-medical-transcription.com/history_speech_recognition.html

[25] Barber J, *History of Speech Recognition*, St. Norbert College Education available at http://www.snc.edu/compsci/newpages/cs2252003/projects/Voice%20Recognition/history.html

[26] Dynamic Living, Inc, *Speak Up for More Independence - Use Voice Recognition* available at http://www.dynamic-living.com/news-voice_activation.htm

[27] *Automatic Speech Recognition Shopping List Generator* available at http://wwww.halfbakery.com/idea/Automatic_20Speech_20Recognition_20Shopping List_20Generator

[28] *Helping Children with Autism Learn - Imitation as the Gateway to Early Learning*, Oxford University Press (June 2003) available at http://www.parent-wise.org/articles/autism.htm

[29] Nelson, Adam, *Implement Phonetic ("Sounds-like") Name Searches with Double Metaphone Part I: Introduction & C++ Implementation* (July 2003) http://www.codeproject.com/string/dmetaphone1.asp

[30] Smith, Jeremy, *Differences* available at http://www.peak.org/~jeremy/dictionary/chapters/differences.php

[31] *American vs. British English - Basic Differences and Influences of Change* available at http://www.uta.fi/FAST/US1/REF/usgbintr.html

[32] Malsori, *Importance of Phonetic Transcription,* Journal of the Phonetic Society of Korea, No. 31-32:239-242 (December 1996)

[33] *Phonetic Transcription* available at http://en.wikipedia.org/wiki/Phonetic_transcription

[34] Kirshenbaum, Evan, *Representing IPA Phonetics in ASCII,* Hewlett-Packard Laboratories (February 2001) available at www.kirshenbaum.net/IPA/ascii-ipa.pdf

APPENDIX A : List of IPA symbols and their corresponding ASCII symbols

IPA	*SEE Examples*	*ASCII*	*Partial Feature Set*
[i]	heel, me	IY	{vowel,voiced}
[ɪ]	hit	IH	{vowel,voiced}
[e]	*SAE bait*	EY	{vowel,voiced}
[ɛ]	met, head	EH	{vowel,voiced}
[æ]	hat	AE	{vowel,voiced}
[ɑ]	*SAE father, pot*	AA	{vowel,voiced}
[ə]	about, after, fern	AX	{vowel,voiced}
[ʌ]	up, fun	UX	{vowel,voiced}
[u]	soon	UW	{vowel,voiced}
[ʊ]	put, foot	UH	{vowel,voiced}
[o]	*SAE boat*	OW	{vowel,voiced}
[ɔ]	fork, taut	AO	{vowel,voiced}
[ɒ]	hot	OH	{vowel,voiced}
[ɑ]	bath, bar	AH	{vowel,voiced}
[eɪ]	wait, cake	EI	{vowel,voiced}
[aɪ]	kite, buy	AY	{vowel,voiced}
[ɔɪ]	coin, toy	OY	{vowel,voiced}
[oʊ]	bone, open	OU	{vowel,voiced}
[aʊ]	cow, out	AW	{vowel,voiced}
[ɪə]	ear, sheer	IA	{vowel,voiced}
[ɛə]	air, share	EA	{vowel,voiced}
[oə]	tour	UA	{vowel,voiced}
[p]	spin	P	{stop,bilabial,voiceless}
[b]	boo	B	{stop,bilabial,voiced}
[t]	stop	T	{stop,alveolar,voiceless}
[d]	dog	D	{stop,alveolar,voiced}
[k]	scan	K	{stop,velar,voiceless}
[g]	gate	G	{stop,velar,voiced}
[m]	mat	M	{nasal,bilabial,voiced}
[n]	not	N	{nasal,alveolar,voiced}
[ŋ]	king	NG	{nasal,velar,voiced}
[f]	fat	F	{fricative,labiodental,voiceless}
[v]	vat	V	{fricative,labiodental,voiced}
[θ]	thumb	TH	{fricative,dental,voiceless}
[ð]	that	DH	{fricative,dental,voiced}
[s]	sat	S	{fricative,alveolar,voiceless}
[z]	zip	Z	{fricative,alveolar,voiced}
[ʃ]	mesh	SH	{fricative,palatal,voiceless}
[ʒ]	measure	ZH	{fricative,palatal,voiced}
[h]	hot	H	{fricative,glottal}
[tʃ]	chair	CH	{affricative,palatal,voiceless}
[dʒ]	edge, jam	JH	{affricative,palatal,voiced}
[l]	lot	L	{approximant,voiced}
[r]	rot	R	{approximant,voiced}
[j]	yawn	Y	{approximant,voiced}
[w]	win	W	{approximant,voiced}

APPENDIX B : Example Questionnaire for Post-Test

ORAL DICTIONARY
POST-TEST QUESTIONNAIRE

Thank you for taking time to do the Oral Dictionary Questionnaire. Please fill out the questions provided below. Be rest assured that all information will be kept private and confidential.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree
1	2	3	4	5

The questionnaire will be divided into two sections; Functionality and Overall System Interface Design & Layout. With reference to the Likert Scale above, please circle the extent to which you agree on the following statements.

Section A : Functionality

1.	The voice recognition function is easy to use.	1	2	3	4	5
2.	The voice input reduces the need to memorize the word spelling.	1	2	3	4	5
3.	The system produce output faster as expected than using normal dictionary.	1	2	3	4	5
4.	Searching word definition is much simpler than using normal dictionary.	1	2	3	4	5
5.	The system produces result similar to the word spoken.	1	2	3	4	5
6.	Possible matches result is relevant and useful.	1	2	3	4	5
7.	Using keyboard is useful as alternative to voice input.	1	2	3	4	5
8.	There are many words available for searching.	1	2	3	4	5

9.	The word pronunciation is clear and works very well.	1	2	3	4	5
10.	The read-out-loud definition is clear and useful.	1	2	3	4	5
11.	The buttons performs action as expected.	1	2	3	4	5
12.	Mistakes are easy to correct (e.g. retype word)	1	2	3	4	5
13.	Oral Dictionary clearly addresses the system purpose as English dictionary.	1	2	3	4	5

Section B : Overall System Interface Design & Layout

1.	System is easy to use.	1	2	3	4	5
2.	Interface is user friendly.	1	2	3	4	5
3.	Instruction is clear and easy to understand.	1	2	3	4	5
4.	Definition content is neat and properly arranged.	1	2	3	4	5
5.	The colors chosen for this system is attractive.	1	2	3	4	5
6.	It is easy to remember where to find the functions.	1	2	3	4	5
7.	Information is layered effectively on different screen.	1	2	3	4	5
8.	Information is easy to read.	1	2	3	4	5
9.	Information is written in a style that suits me.	1	2	3	4	5
10.	The word highlighting for possible matches is useful	1	2	3	4	5

Do you have any comments or suggestion to improve this system? If yes, please state them down :

_____

_____

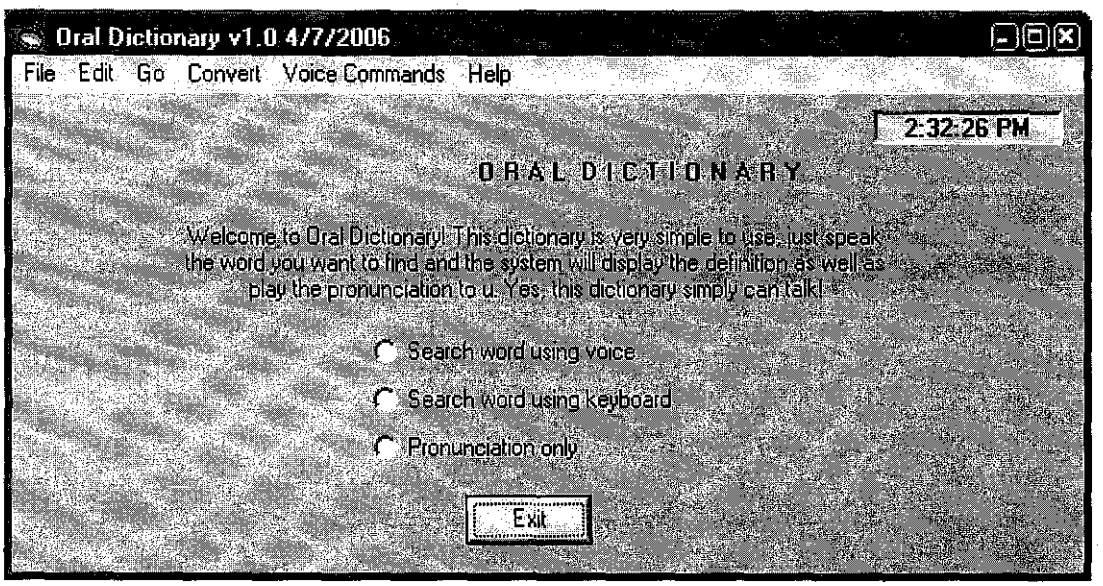
_____

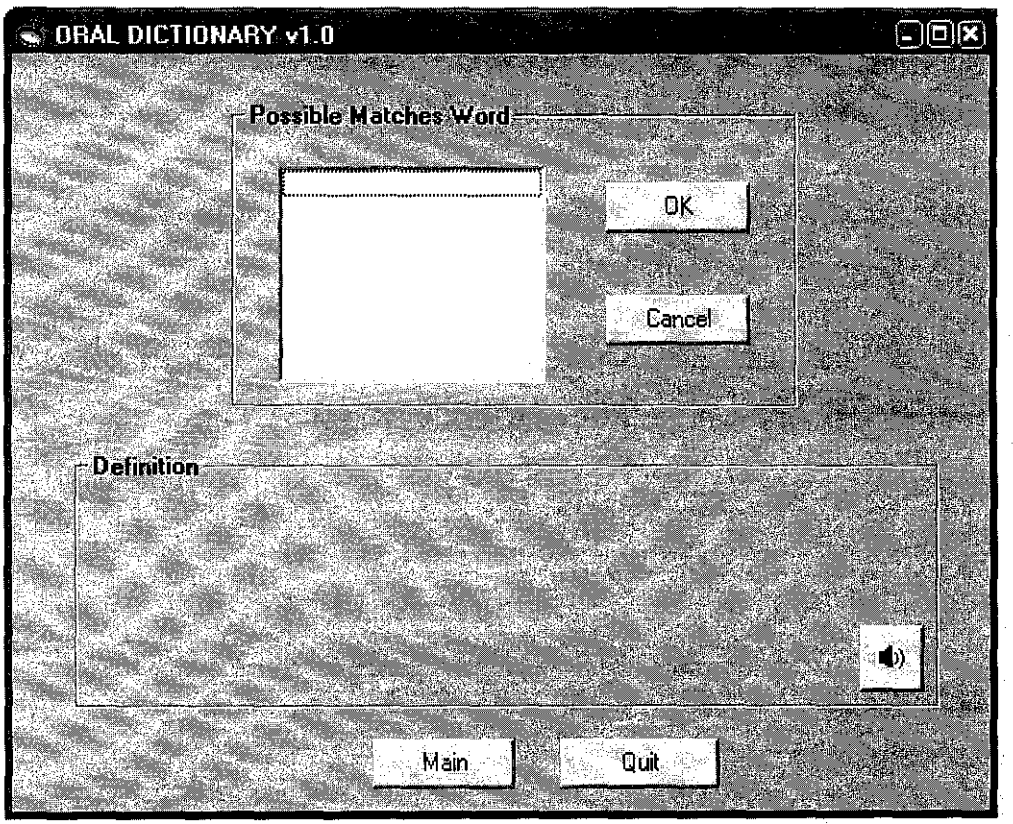
_____

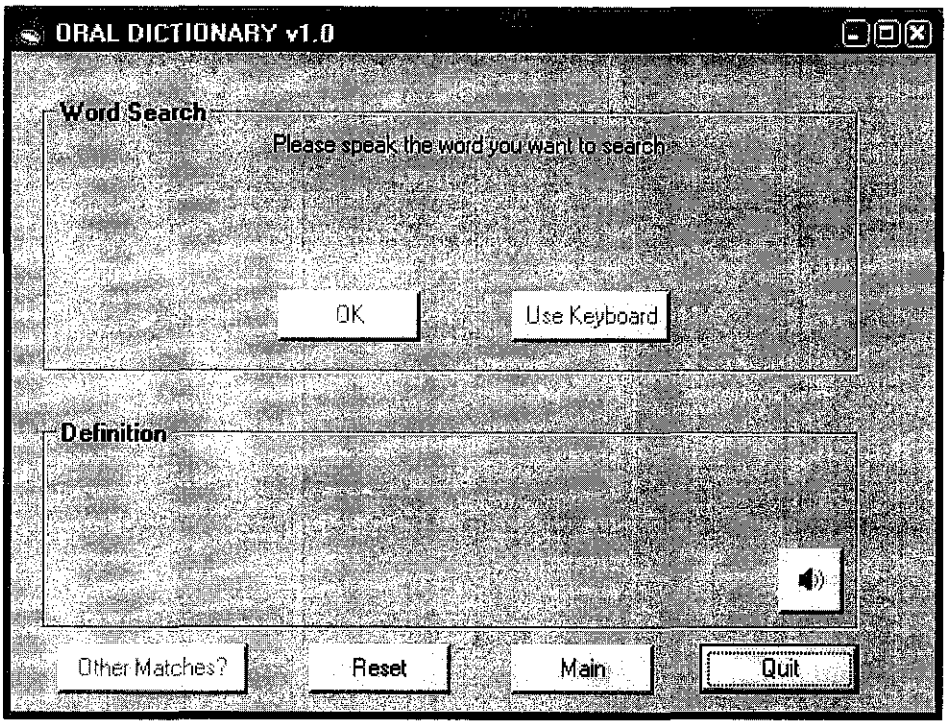
THE END

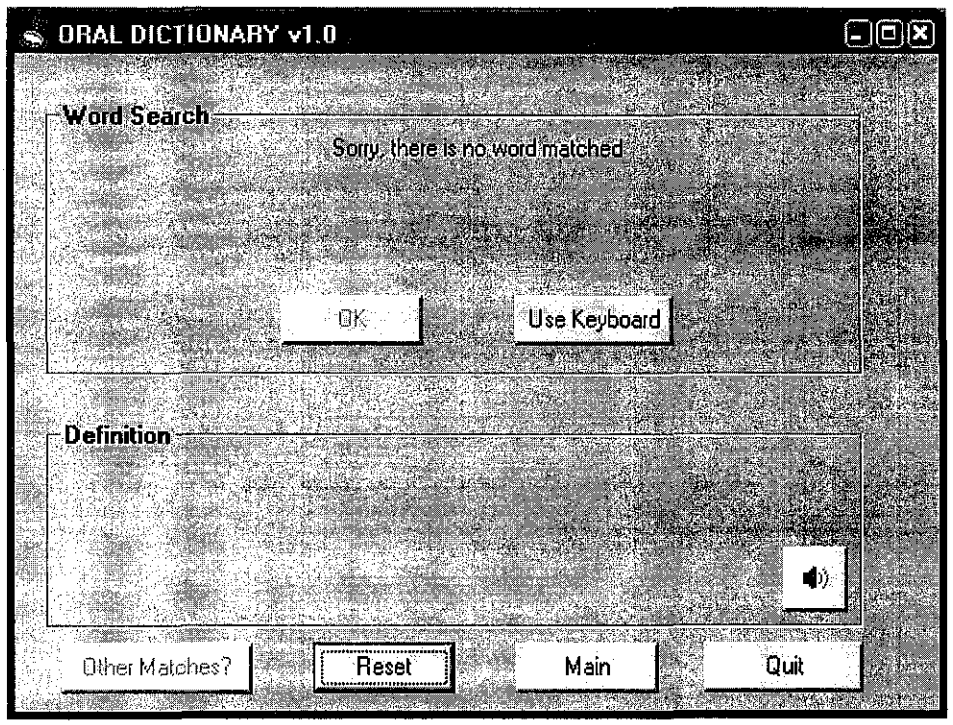
THANK YOU FOR YOUR TIME

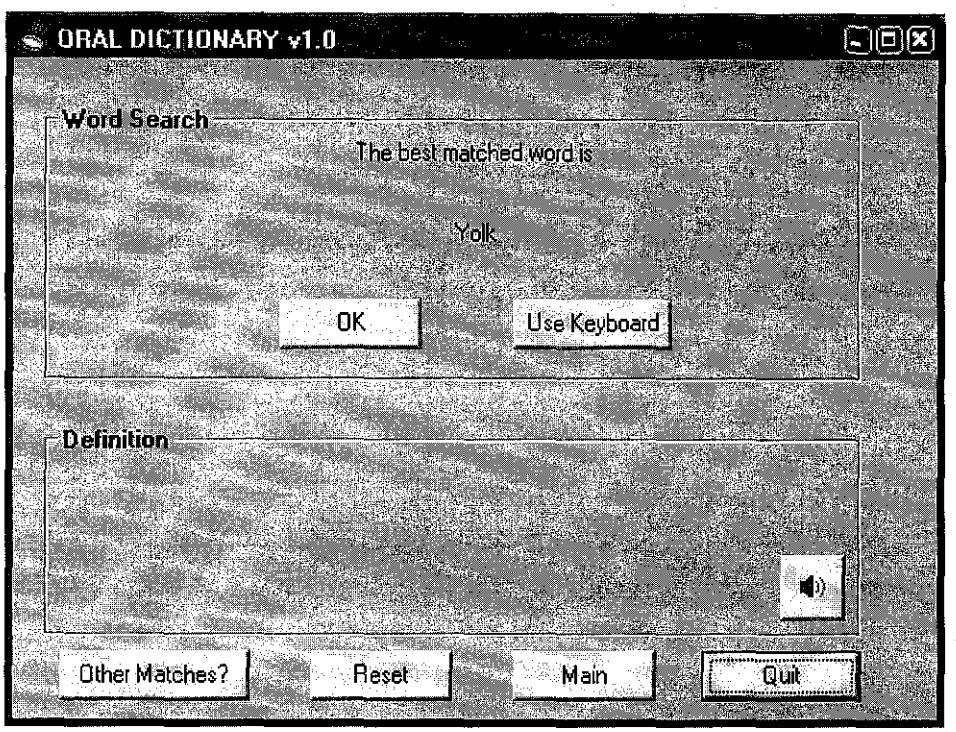
APPENDIX C : User Interface Design

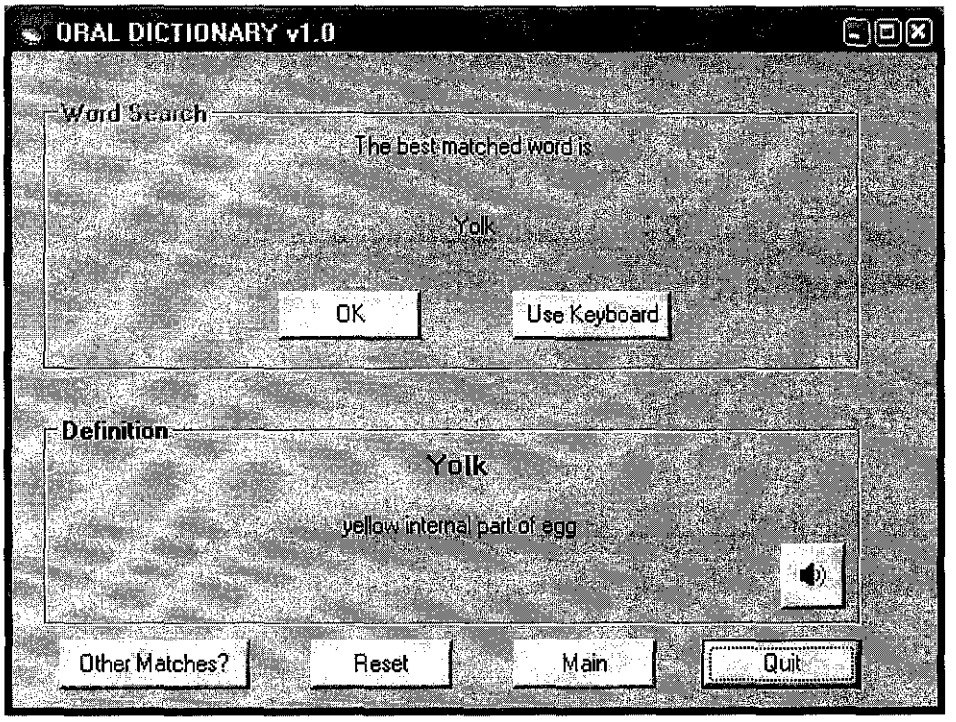





ORAL DICTIONARY v1.0
Word Search
Please speak the word you want to search
OK
Use Keyboard
Definition
Other Matches?
Reset
Main
Quit


ORAL DICTIONARY v1.0
Word Search
Sorry, there is no word matched
OK
Use Keyboard
Definition
Other Matches?
Reset
Main
Quit


ORAL DICTIONARY v1.0
Word Search
The best matched word is
Yolk
OK
Use Keyboard
Definition
Other Matches?
Reset
Main
Quit


ORAL DICTIONARY v1.0
Word Search
The best matched word is
Yolk
OK
Use Keyboard
Definition
Yolk
yellow internal part of egg
Other Matches?
Reset
Main
Quit