INFORMATION VISUALIZATION WITH VOCABULARY TRACKING FOR WORD RECOGNITION IN QUR'ANIC VERSES

RAJA JAMILAH BT RAJA YUSOF

THESIS SUBMITTED IN FULFILMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTER SCIENCE AND
INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2011

ORIGINAL LITERACY WORK DECLARATION

Name of Candidate : RAJA JAMILAH BT RAJA YUSOF

IC No : 740718-08-6442

Matric No : WHA040003

Name of Degree : Doctor of Philosophy

Title of Dissertation : Information Visualization with Vocabulary Tracking for Word Recognition in Qur’anic Verses

Do solemnly and sincerely declare that:

1. I am the sole author/writer of this work;

2. This work is original;

3. Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the work and its authorship have been acknowledged in this work;

4. I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;

5. I hereby assign all and every rights in the copyright to this work to the University of Malaya (UM), who henceforth shall be owner of the copyright in this work and that any reproduction or use in any form or by any means whatsoever in prohibited without the written consent of UM having been first had and obtained;

6. I am fully aware that is in the course of making this work I have infringed any copyright whether intentionally or otherwise, I may be subjected to legal action or any other action as may be determined by UM.

Candidate’s Signature : Date:

Subscribed and solemnly declared before,

Witness’s Signature :

Name :
Designation :

ABSTRACT

Muslims read or recite the Qur'an which is a source of guidance in all aspects of life. However, their recitation is sometimes without comprehension especially to those who do not understand Arabic language. To help reduce this comprehension problem, visualization is proposed based on the theory that the learning process involves the perception and the processing stages. In particular, the visualization helps to support the perception stage and is used as a strategy to solve the word recognition with meaning retrieval problem. The occurrences of the Arabic words are visualized since the *Qur'anic* verses contain many repeated words. This is achieved through Parallel Coordinate and word segmentation visualization technique. The next stage is to process the information and this is supported through a vocabulary tracking system. The system is able to track a user's personal vocabulary with the presentation of the percentage and position in context of the *Qur'an*. In direct consequence the users know their ability of recognizing the Arabic words in relation to the whole *Qur'an* to achieve reading comprehension. The hypothesis is that visualization with vocabulary tracking mechanism supports Arabic word recognition in the Qur'anic verses. The research methodology of this thesis starts with a preliminary study of identifying problems of *Qur'an* reciters and in solving the problem a study was carried out on how visual elements can help in reading comprehension. Then, an interface system called Parallel Coordinate of the *Qur'an* (PaCQ) was developed with appropriate functions and finally tested for its effectiveness. Particularly, PaCQ was tested for any improvement in the word recognition percentage scores and the time taken to complete word recognition exercise. An experimental study was set up starting with a pre-test of Arabic Word Recognition Test (AWRT) to all the targeted participants. Then stratified sampling was

used to divide the participants into the control and the experimental group for the AWRT post-test. It was found that the AWRT score percentage rank for participants after using PaCQ interface was significantly higher compared to before using PaCQ. The average AWRT percentage scores of set 4, 5 and 6 (the post-test) for the PaCQ group, show significant difference in the rank score compared to the control group. The average rank time to complete the post-test for the PaCQ group is significantly faster compared to the control group. The average time to complete AWRT for age group >= 35 had been significantly improved in the experimental group compared to the control group. It can be concluded that visualization and vocabulary tracking mechanism through the PaCQ interface improve word recognition performance both in the score and the time taken to complete the test.

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious, the Most Merciful. He, who has given me the strength to continue doing this research. Alhamdullillah, All Praise to Allah that he has answered my prayers to actually submit this thesis with the hope to pass with excellence. It feels like it has been too long. I thank all people involved, from the three supervisors that had helped me with this research (Prof. Dr Roziati, Zulkifli and Sapiyan), my husband (Mohd Bahrulnizam), my mothers (Siti Halimah Saad, Eishah Awang and Siti Halimah Awang), my father (Musa Mohamed), my children (all seven of them: Salsabiela, Tasniem, Hamidun, Muhsinin, Muhammad, Hafidz, Roshada), my siblings, friends, other researchers and participants. All that had helped in this research. I pray that this research is beneficial to at least some people. I also hope that my late father (Raja Yusof) gets his share out of this research.

Raja Jamilah Raja Yusof

Part-time PhD student

University of Malaya

2004 – 2011

TABLE OF CONTENTS

LIST OF FIGURES:

LIST OF TABLES

LIST OF STATISTICAL SYMBOLS, RELATIONAL SYMBOLS AND ABBREVIATION:

Abbreviation:

1. 10AL: 10 affixed alphabets that can be attached to a root words which are ا, ه, ي, ن, و, م, ت, ل, ء, س *(A, h, Y, n, w, m, t, l, ', s)*
2. AWRT: Arabic Word Recognition Test
3. AWV: Arabic Word Visualizer component
4. CG: Control group
5. GOMS: Goals, Operators, Methods and Selection Rules
6. II: Intermediate Interface component
7. In: Infix
8. IP: Information processing
9. LEG: Less than 35 years age group
10. LTM: Long Term Memory
11. MOG: 35 years and above age group
12. CG: Control Group
13. UIG:User Interface Group / the experimental group
14. PaCQ: Parallel Coordinate for the Qur'an
15. PCG: Parallel Coordinate Graph component
16. Pre: Prefix
17. PPI: Pixel per Inch
18. PUI: PaCQ User Interface component
19. QC: Qur'an Corpus software
20. QL: Qur'anic Lesson website
21. RSE: Rule-based stemming engine
22. SM: Sensory memory
23. STM: Short term memory
24. Suf: Suffix
25. TI: Tanzil.info website
26. UIG: User interface group or the experimental group
27. WFE: Word frequency effect
28. WSE: Word superiority effect
29. ZQ: Zekr Qur'an software

Relational symbols:

1. ∪ - union of	6. ≡ - congruent/similar to
2. ∈ - element of	7. ∀ - for all
3. ⊆ - subset of	
4. ℕ - set of natural numbers	
5. $\mathbb{N}^{+}$ - set of positive natural numbers	

Statistical symbols:

1. *df*: degrees of freedom	2. *SD*: standard deviation
3. *Sig*.: p-value to tell readers if the result is significant	4. *M*: Mean
5. *p* : the probability	6. *Mdn*:Median
7. Mann Whitney *U* : sum of the ranks	8. *t*: the ratio of the mean difference divided by the standard error of the sample mean.
9. *z*: the critical value at which the null hypothesis is rejected	

1.0 INTRODUCTION

Reading is a process of encoding a written text to its meaning. Although, it is almost an automatic task, the cognitive process involved is complex. In normal situation, a text is captured by our eye sight and then a slow serial process of encoding the text occurs. In contrast the parallel process of our visual perception is much faster in encoding images.

The main issue discussed in this thesis is comprehension problem faced by the Muslim community in reading the *Qur'an*. The approach adopted for integrating three main fields are of cognitive science, information processing and visualization, while the application domain is the *Qur'anic* verses.

This chapter starts with the motivation and background, followed by statement of problem, objectives, scopes, research questions, hypothesis, research methodology, novelty, thesis structure and finally the terms and terminology used in this thesis.

1.1 Motivation and background

The *Qur'an* is the word of Allah, God of the heavens and the earth as guidance to the Muslims revealed 1400 years ago (Denffer, 1983). It contains lessons drawn from daily experience, anecdotes from prophets of different era and places, stories which cover the beginning of time to the day of judgement and the life beyond, scientific facts which human knew not until revealed, strict but flexible rulings for daily life and other important messages (Husain, 1974). All these messages are written in the Arabic language and the purpose of Its revelation is to inspire human the living sense of Allah, and that He is real, not vague and near, not distant. The ability to understand the language of the *Qur'an* is of more primary importance to those who intend to specialize in the Islamic studies. Unfortunately, currently there is the problem of many Muslims reading this document without comprehension needing the involvement of many people to be overcome. This phenomenon is applicable to many non-Arabic speakers of the Muslim community that spreads in India, Indonesia, China and others, totalling to approximately 1.65 billion (Kettani, 2010).

One of the main reason why reading without comprehension occurs is because the nature of the Arabic language itself. Arabic language is considered as a difficult language to learn (Anonymous, 2006). There are many factors such as the writing system, the vocabulary, the morphological structure and the grammar. The writing system is different from Malay and English which uses the Latin script. In addition, the Arabic script is a cursive one. One who does not know the language would see the writing as a continuous character making it difficult to differentiate one word to another. This is especially true for the writing system used to write the Qur'an, the classical Arabic language. The vocabulary in the Arabic language is vast in number. The lack of knowledge of the root word and pattern system of the language adds to the difficulty. This is the reason for the complex morphological structure of an Arabic word. Even with an Arabic dictionary at hand, without this knowledge, one would struggle to search for the meaning of the word. The grammar is another issue that adds to this complexity list. Not only that there are singular and plural nouns, there is also the dual form and they can be masculine or feminine.

Consequently, the written guidance as a source of basic and essential knowledge to many Muslims is not attained from reading or recitation, instead heavily dependent upon classes conducted in schools, mosques or other places, lectures in radio or television, personal reading of books of well understood languages. Example of lessons and books aimed to understand the Qur'an are of translations, *tafsirs* (interpretation), *tanzils* (revelation), *asbabul nuzul* (historical background) and lexical intricacies.

Although understanding the words of the *Qur'an* in depth needs special knowledge and expertise, one can still generally comprehend the clear and simple literal messages of the *Qur'an*. Several solutions for comprehension of Qur'anic word literal meaning had been proposed. First, Abdelbaki (1983) proposed pictorial sketches for certain *ayat* (verses) of the Qur'an to aid in the comprehension of those verses. This is suggested for those in the intermediate level of non-Arabic speakers. Secondly, teaching the meaning of Qur'anic words based on word frequency is another approach popularly introduced in the web by Abdulazeez (2004). The strategy is to give priority to a more frequently occuring words since the Qur'an contain many repeated words. Thirdly, there are already word by word translations books of the Qur'an to help reciters learn the meaning of words instead of just giving meaning verse (sentence) by verse. There are also lessons conducted specifically to learn Qur'anic Arabic words meaning such as the *Harfiah* Qur'anic lessons by Hasim (2012). The focus of this lesson is to teach how to segment Qur'anic words to be able to identify word meaning and hidden pronouns in verbs to achieve comprehension during recitation of the Qur'an.

These solutions suggest relations to several theoretical studies related to visualization, reading comprehension in generally and word recognition specifically. In theory, to be able to achieve reading comprehension, word recognition is an important element (Perfetti, Landi, & Oakhill , 2007). This is where word superiority effect (WSE); (Reicher, 1969; Wheeler, 1970; McClelland & Johnson, 1977) and the word frequency effect (WFE) (Forster & Chambers, 1973) comes in. The word superiority effect shows that a letter is easier to be recognized in known words compared to non-words; while the word frequency effect shows that more frequent words are responded to more rapidly. It is envisaged that the study which relate visualization and word recognition will support comprehension to those who face the problem of comprehending Qur'anic verses.

1.2 Statement of problem

The main issue involved in this research study is a reading comprehension problem in context of Qur'anic verses. It is applicable to the non-Arabic speakers in the Muslim community whereby many recitation and memorization of Quranic verses occur without comprehension. This is mainly due to the problem of understanding the language. Reading theories suggest that reading is a process that involves information processing, accessing or retrieving information from memory that is transforming print to speech or print to meaning and as such our memory plays an important role in reading comprehension (Snowling & Hulme, 2007; ColHeart, 2007) . Reading is comprehension and involves two major components; the word recognition and the assembling of words into messages.

"One *must readily identify words and encode their relevant meaning into mental representation… Comprehension cannot be successful without the identification of* *words and the retrieval of their meaning" (*Perfetti, Landi, & Oakhill , 2007, *pg. 229)*.

However, the issue of reading comprehension is too complex to be solved. It is the knowledge level in the learning taxonomy selected as the focus of this research which attempts to solve some important aspects of reading comprehension, the word recognition or identification problem (which can be used interchangeably between each other). Some authors refer to it as word recognition and some as word identification (Pikulski, 2010). Also, the term word recognition is highly associated to the concept of word superiority and word frequency effect used in this research (see § 1.6). Therefore, from this point onwards word recognition is used.

Decomposing the word recognition problem, the question on how the learning process of word recognition should be supported arises. At the same time, Arabic words have complex morphological structure that needs to be dealt with. Therefore learning to recognize the Arabic words is different from a language such as English. This is where visualization and information processing theories comes into the role as part of the solution.

1.3 Main aim and objectives

The main aim of this research is to solve the problem of reading comprehension for non-Arabic speakers when reading the *Qur'an*. However, to achieve reading comprehension many processes and factors are involved. Therefore more specifically the objectives of this research are as follows:

- To design a word recognition system supported by visualization based on word superiority effect and word frequency effect
- To evaluate whether the above approach can improve word recognition in Qur'anic verses

1.4 Research Questions

Based on the objectives stated above, several related research questions are:

i) What are the underlying problems of comprehension amongst *Qur'an* reciters?

ii) How does visualization support word recognition based on word superiority effect and word frequency effect?

iii) What are the suitable techniques to visualize word information to support word superiority effect and word frequency effect?

iv) How should Arabic words be counted since the word can exist in many forms?

v) What are the platforms or architecture to implement the solution found based on the above questions?

vi) What are the performance measures used for system evaluation?

vii) Is there any improvement on the word recognition performance supported by the system above?

1.5 Scope and constraints

The 30th part called *Juz Amma* is used in this research due to three reasons, one is because *Juz Amma* is one of the most frequently read verses of the Qur'an. The second reason is that, *Juz Amma* contains 37 *suras* or chapters in which contain both *Madaniah* and *Makkiah* typed *suras* reflecting the whole Qur'an (114 *suras of either Madaniah or Makkiah*). Lastly, the amount of the Arabic data from the Qur'anic verses in *Juz Amma* is fewer to handle compared to the whole Qur'an so that the related data can be organized within a reasonable time. Targeted participants are selected from non-Arabic speaking community, that is the Malays who are mainly non-Arabic speakers facing the specified problem.

Word recognition is an important source of knowledge towards reading comprehension. The assessment of word recognition performance is based upon the ability to retrieve the meaning of the Arabic words in Malay or English language. However, the constraint is that a word can have several meanings depending on the context of the text. It is therefore, practical to associate a word to one possible meaning even though within the actual context certain meaning may not be suitable. This is easier for the assessment of participant to be design and evaluated.

1.6 Approaching the problem and hypothesis

A theory outlined by Hunt, Ellis, & Ellis (2004) states that the process of comprehension involves summarization, association to a central theme and relates to some inferring information. A fast reader may do this by looking at the title of the passage, the lead sentence of the paragraph, the recursive words and the dominant idea of the text being read. This can be related as an intra communication activity within the cognitive thinking of the readers.

Looking from the perspective of cognitive learning style, learning involves the restructuring and the integration of the pre-existing knowledge with the new acquired knowledge in the mind. The sensory (SM), short-term (STM) and long-term memories (LTM) are all involved throughout the process which is referred as the information processing stage. To be able to acquire new knowledge into the long-term memory, is to firstly attract the attention of the sensory memory and something of interest or any recognizable pattern will then lead the new knowledge into the short-term memory. It is then passed on from the short-term to the long-term memory through repeatable action or rehearsal. It is therefore hypothesized that information visualization attracts the sensory memory so that the information can be subsequently passed to the short-term memory. With the act of exploring, interpreting, filtering, visualization supports the rehearsal process so that the information from short term memory can then be passed to the LTM.

In relation toward word recognition process there are two well known theories which are the word superiority effect (WSE) (Reicher, 1969; Wheeler, 1970; McClelland & Johnson, 1977) and the word frequency effect (WFE) (Forster & Chambers, 1973). The word superiority effect shows that a letter is easier to be recognized in known words compared to non-words; while the word frequency effect shows that more frequent words are responded to more rapidly. Semantic word visualization through vocabulary, cognate words and Arabic word segmentations is to support WSE, while the parallel plot on word count and word position is to support the WFE leading to word recognition and subsequently reading comprehension through reinforced learning. This is expected to occur when the information has flow to the LTM. Figure 1.1 illustrates the model described above in solving the reading comprehension problem. So, the overall hypothesis is that visualization with vocabulary tracking mechanism supports Arabic word recognition in the reading process.

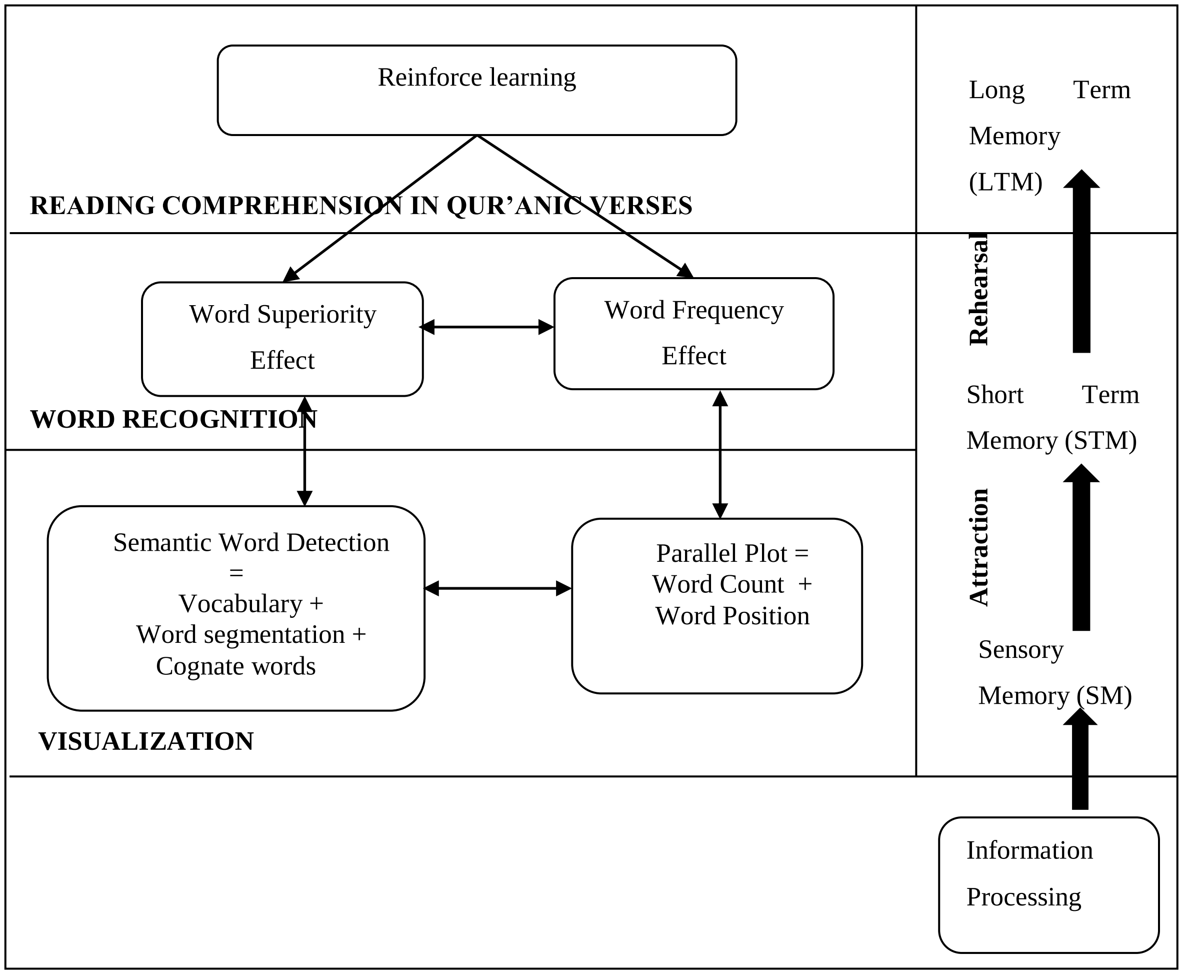


Figure 1.1: Integrated model for quick recognition in Qur'anic verses

1.7 Research methodology

The research consisted of 3 phases. It started with identification phase with literature investigation followed with investigating theories of reading comprehension and information theory through preliminary studies. Next, based on a hypothesized solution, the PaCQ system was designed and implemented with the embedded theoretical concepts in the second phase.

Finally the evaluations were conducted. First, the experimental was to test whether the PaCQ system has any significant effect in the word recognition performance. 90 participants are recruited from around the campus of University of Malaya. Equal numbers of female and male in the age category of below and at least 35 years are selected. This experimental study starts with a pre-test to obtain stratified participants which were then divided into the control and experimental groups. This is to make sure that there is almost equal number of participants with good and poor scores in the control and experimental group. Then the post-test was conducted to evaluate the word recognition performance via the Arabic Word Recognition Test (AWRT). The data obtained are mostly not normally distributed. Therefore, the significant tests (testing whether there is any difference in the scores obtained between the control and experimental group) were conducted using mostly the Mann- Whitney U test and the Wilcoxon test. Other tests used were the independent t-test, the Wilcoxon test and the Pearson's correlation test.

Next, the user satisfaction test was also conducted to find out the usability issues of PaCQ interface followed by the test of parallel plot interpretability. Finally interview sessions were conducted after modification of the PaCQ interface. Then, results are analysed and triangulated. Detailed descriptions are further discussed in chapter 4 and 7. Figure 1.2 shows the summary of the research methodology.

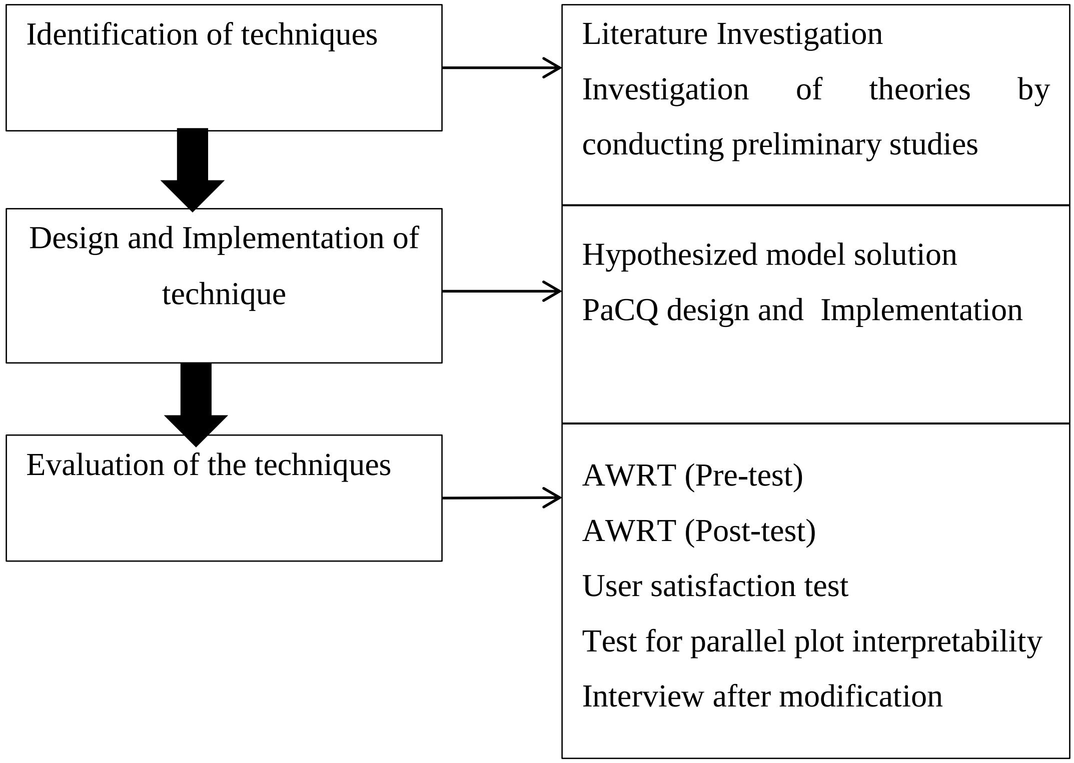


Figure 1.2:Research Methodology

1.8 Novelty and research investigation

The novelty of this research is combined theoretical concept and practical implementation resulting in development of computer software (PaCQ interface) that supports Arabic word recognition for reading comprehension in *Qur'anic* verses. While the research contributions based on cognitive science, information processing and visualization are as follows:

- Cognitive science:
  - Reading comprehension based on word recognition using visualization
  - Implementation of vocabulary tracking system to support *Qur'anic* Arabic word recognition

- Information processing:
    - Arabic word form pattern analysis and Arabic word recognition experimental study
- Visualization:
    - Application of parallel coordinate visualization technique in the PaCQ interface based on Goals, Operators, Methods and Selection Rules (GOMS) analysis

Currently, no such software as the PaCQ interface is available in the market. Furthermore, the research contribution based on theories of reading comprehension and visualization applied in the *Qur'anic* recitation domain can indeed be claimed as unique.

Several findings of this research are as follows.

- Parallel coordinate visualization technique is suitable to be used by novice and expert users with higher education background.
- In order to process the Qur'anic text for the purpose of counting words, stemming process generates inaccurate root word counts without using any word lexicon
- Integration of visualization and the vocabulary tracking system in the PaCQ interface helps in increasing the word recognition scores and the time taken to complete the word recognition process.

In general, the contribution of this research is to the Malay Muslim community specifically and the other non-Arabic speaking Muslims community in general. The complexity of the learning process in learning Arabic language with respect to comprehending the Qur'an can be reduced, hence contributing to the better understanding of daily ritual acts. Research findings, contributions, and outcome are further described in the last chapter.

Three main publications from this research are as in Appendix A

1.9 Thesis structure

Reading the Qur’an for comprehension is the main issue addressed in this thesis and chapter 2 covers the related topics on Qur’anic lessons, theories of information processing and reading comprehension. In chapter 3, information visualization for comprehension in *Qur’anic* verses is discussed. The four main issues addressed are word visualization, visualization based on aspects of quantity, visualization based on aspects of connectivity and visualization of multidimensional variables. Chapter 4 is the research methodology. Chapter 5 covers discussion on three preliminary studies. The first one is on problems related to recitation; the second is on an experimentation of using images in *Qur’anic* verses while the third is the contextual comparison of parallel and perpendicular plots.

Chapter 6 covers issues related to integrated learning for quick recognition in Qur’anic verses. It discusses comprehension problems on Qur’anic verses by decomposing issues to achieve word recognition and lastly outlining the system requirements of the PaCQ interface. Chapter 7 is the design and implementation showing modules of PaCQ interface through interface screen shots, component diagrams and algorithm. Chapter 8 is the experimentation descriptions and the results obtained and finally the conclusion is highlighted in chapter 9.

1.10 Terms and terminology in Qu'ran, language and visualization fields

There are several important terms and terminologies used repeatedly which carry certain meaning adopted by this thesis. These are divided into two categories, the Arabic terms pertaining to Qur'an and non-Arabic terms in relation to language. The Arabic terms are listed below:

- *Qur'an:* a sacred and unaltered book (original) to the Muslims believed as the words of God to have been revealed to the Prophet Muhammad (peace be upon him) by the angel Gabriel around 1400 years ago in *Mekah*.
- *Juz':* parts or portions of the *Qur'an*. The *Qur'an* is divided into 30 portions of approximately equal length.
- *Sura:* a chapter in the *Qur'an*. There are 114 chapters in various lengths, some are very lengthy and some are short.
- *Ayah:* a verse or a sentence in the *Qu'ran* in various lengths. The longest verse can occupy up to a page of the *Qur'an*. A page of the standard *Qur'an* (known as *Uthmani Qur'an*) is always written within the same number of pages even though the sizes of the *Qur'an* can be different.
- *Ayat:* many *ayah*, the plural of the word *ayah*
- Diacratics: marked symbols in Qur'anic verses that are non- alphabets such as in Table 2.2.

Non-Arabic and visualization terms are as follows:

- Word recognition: the cognitive process of identifying a particular word by a human while reading takes place. Word form: the morphological structure of a word.

- Morphology: Description of the structure of a word. A word may contain subwords or other unit of syntax related to the grammar of the language. For example the word ‘balls’, the unit syntax ‘s’ is a suffix that carries the concept of plurality.
- Phonology: The set of rules used to encode the sound to the spoken word.
- Orthography: Specifies, describes and defines the rules of using one particular writing system (written in a particular symbol) to write the language.
- Phonemes: Smallest unit of sound.
- Cognate words: words that originate from another language
- Base words: this term is related to the words that can be found one or more times in the *Qur’an*. To count words, the term base word refers to all words that can be written on the word axis versus the count axis if a word-count graph is drawn
- Root word: the smallest form of a word, the word form when all affixes are removed
- Occlusion is the number of identifiable points in relation to other visible points

2.0 READING THE QUR'AN FOR COMPREHENSION

Muslims around the world believe the *Qur'an* as the word of Allah. Coming from all kind of background and native language, Muslims have the traditions of memorizing the Qur'an which is in Arabic language. It is part of the Muslim ritual act, every ritual act; is related to the *Qur'an* from reading or reciting parts of the *Qur'an* (such as in the five daily prayer) to implementing the teachings of the *Qur'an*. However, since Qur'an is in the Arabic language, many Muslims read them without comprehension. It is ironic that all other reading habits is indeed related to comprehending the written message, nobody in actual fact would read anything which is not understood.

In this chapter, discussions start with human processing stages and followed with reading comprehension theory and then focusing on word recognition issues. Subsequently, the problems and solution of reading comprehension are discussed followed by the applicability of those theories on the Qur'anic verses comprehension which involves discussion on the Arabic and Malay language.

2.1 Information processing in human brain for reading comprehension

Based on the theory of Information Processing (IP) (Wortham, 2003), information is acquired, stored, and retrieved from the human brain. For reading comprehension the process involves the human sensory system acquiring the information and then storing them to the sensory memory (SM), short-term memory (STM) and long-term memory (LTM) (Reed, 2000; Huitt, 2003).

During SM stage, sensory information is stored briefly and human are more likely to pay attention to this stimulus if there are some interesting features attained through the sensory system i.e. light, smell, sound, cold, hot. Our sensory memory constantly receive huge amount of information and through attention, these information are filtered out for the brain to process and the criteria for the filtration can be based on endogenous control (i.e., selection based on voluntary attention, current aims, and knowledge) and stimulus-driven control (i.e., selection based on the intrinsic features of the sensory input (Macaluso, 2012). For example, human can decide to focus on one human face in a crowded room or focus to a conversation in a noisy room. This memory lasts for a short period of about ½ second for vision and 3 seconds for hearing. In order to transfer this information, apart from possessing an interesting feature, a recognizable pattern is more likely to go through the next stage, the STM. Perception is the interpretation of what we acquire from the sensory memory (Solso, MacLin, & MacLin, 2008 and Sobel, 2001). Macaluso (2012)

The information from the SM goes to the STM which is also commonly known as active memory. Here information is held for a short period; up to 20 seconds. It also has a capacity of holding 7± 2 items. As an example people tend to remember 7 digits number compared to 10 digits. This memory is easily forgotten if not rehearsed and not organized in a meaningful manner with the overall pre-existing mental structure. In the context of learning, any action or movement, the act of relating, explaining, and discussing leads to active learning and reinforced learning (Felder & Silverman, 1988). Consequently, this results in a better recall of the learning material. This is the basis of the theory on students retaining 90% of what they say as they do something.

The LTM holds information without limitation or capacity. Once the information is stored here, it is not easily forgotten. There are also two types of memory in this category, the preconscious and unconscious memory (Huitt, 2003). The preconscious memory is relatively easy to retrieve and recall while the unconscious memory cannot be assessed during normal consciousness.

Apart from the above model of memory, the dual coding theory of Bagozzi & Silk (1983) states that people represent information either verbally or as images. Verbal codes can be concrete or abstract while the image is concrete. For example episodic memory based upon one's person experience can be considered as concrete while semantic memory dealing with meaning can be considered as abstract. Recognition in this model is referred to as a piece of stimulus being compared to content in memory and recall contains this process plus another step. In recalling, the stimulus is first in form of uncertain cue which will be compared to establish mental concept to be further recognized. However, there are also other views on recognition and recall, but the majority related them as being at least overlapped in certain aspect.

According to a multidimensional model information is processed at varying depths: the physical and visual information is processed at the shallowest level, for example the process involving the word and letter shape. The auditory information is processed in the intermediate level, for example, the process involving word rhymes. Finally, semantic information is processed at the deepest level, for example the process involving how to figure out the meaning of a word or sentence (Bagozzi & Silk, 1983).

2.2 The reading processes

Kintch and Rawson (2007) regarded reading comprehension as consisting of several processing levels involving the process of decoding the text (i.e. from letters to the word or phrase) to its meaning and then combining them to form ideas in units or proposition. At the early stage of learning to read, focus is on decoding and identifying words as in the case of children and this process limits comprehension.

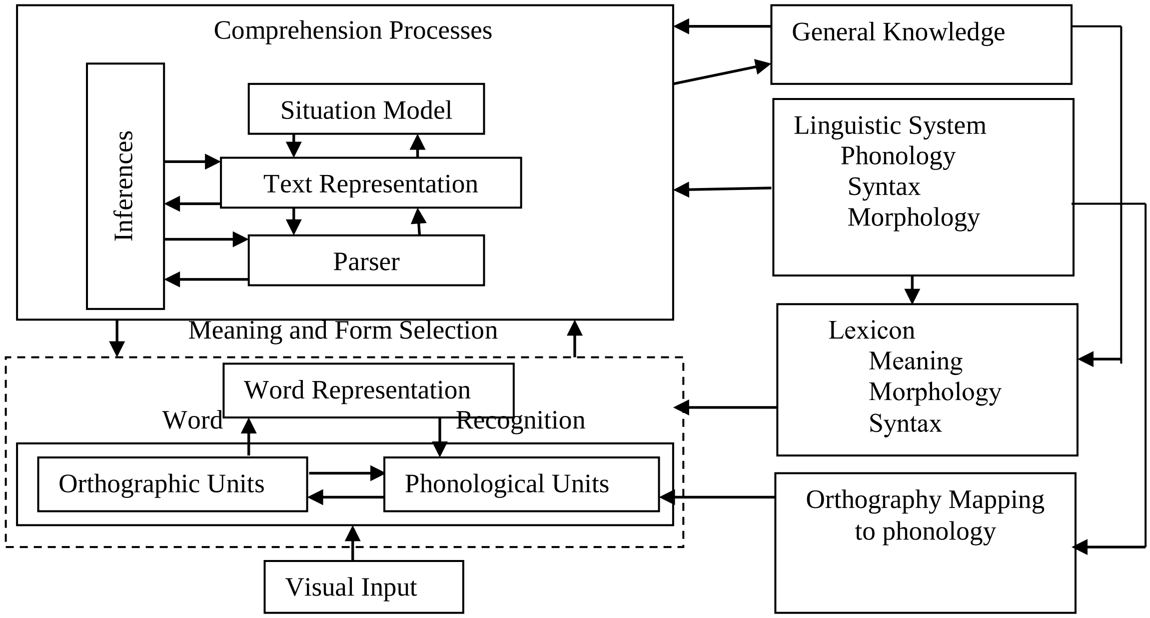
Relating it to the model of memory as described previously, Kintsch and Rawson's (2007) view follows the multidimensional model. In terms of semantic analysis this can be realized through the macro and micro structure of the text involved. However, the formation of text meaning towards reading comprehension depends on several other entities such as anaphor resolution (any linguistic means of referring to a previously mentioned text), topic identification of the text which depends on reader's previous knowledge stored in the LTM.

Making inferences with sufficient knowledge, comprehension monitoring and domain knowledge are also essential towards comprehension (Nation, 2007). Literal meaning of text is highly integrated with inference ability and is developed with knowledge. The ability to remember texts (literal memory) is related to the performance of global comprehension. The knowledge of the syntactical forms and the meaning of words are crucial to the development of comprehension. The development of the latter subject, also commonly referred to as vocabulary development or sometimes known as semantic knowledge, involves the active working memory, the ability to remember and retrieval of words within the sentence.

In addition, according to Kintsch and Rawson (2007) for deeper understanding to occur what is called as the situation model (i.e. the situation described in the text) must be involved and models such as the Latent Semantic Knowledge and, Vector Space Model may be useful to represent human knowledge in this context. As a conclusion text comprehension is a complex process involving various information and representation. To achieve comprehension, it is not merely adding up all of these *processes but instead it is the coordinated operation as a system.*

Alternatively, referring to the framework of Perfetti, Landi and Oakhill (2007) of comprehension, multiple levels of processes are involved: word level, sentence level and text level. Within these levels word recognition, parsing, referential mapping, sensitivity to story structure, comprehension monitoring and inference processes play important roles and Figure 2.1 represents this framework. Comprehension is not achievable without word recognition and meaning retrieval, while meaning retrieval is important in predicting comprehension.

The complete top level process of reading comprehension involves the cognitive skills related to situation model, text representation and parser with close relation to the ability of making inferences. The external factors involved are the general knowledge, linguistic system of the text i.e. the phonology, syntax and morphology. The top level process needs vital input from the bottom level, word representation through word recognition via the visual input. These relate to external factors such as lexicon and orthography mapping to phonology.



Source: Perfetti, Landi and Oakhill (2007)

Figure 2.1 Word recognition as part of reading comprehension process

2.2.1 Word recognition processes in reading comprehension.

Word Recognition is the process of retrieving a string of letters from text to what is stored in lexical memory or lexicon (Lupker, 2007). There are several model of word recognition studied since 1960s until today. Lupker (2007), stated four main models on word recognition. These are the word superiority effect, the word frequency effect, the semantic priming effect and the mask repetition priming effect. Theoretically, the word superiority effect and word frequency effect are two most stable theories on word recognition. This is partly due to the historical fact that these two theories started to be studied extensively since the 1960s. They are also two most well-known theories on word recognition and therefore, subsequent discussions are based on word superiority effect and word frequency effect as the focus of this study.

(a) The word superiority effect (WSE)

The experimentation work by Reicher (1969) and Wheeler (1970) can be seen as the first most influential work on WSE. This effect states that letters are easier to recognize in known words compared to non-words.

There are three models that explain word superiority effect (Larson, 2003):

- Word Shape model
- Serial Letter Recognition model
- Parallel Letter Recognition model

The Word Shape Model found that words are recognized as a whole through a few experiment. First, the experiment showed that the letter D was more accurately recognize in the word WORD compared to word ORWD (Reicher, 1969). Second, the experiment found lower case letter are read faster compared to uppercase (Fisher, 1975). Third, the experiment found that it is more difficult to read in alternating case (such as AlTeRnAtInG) compared to same case (McClelland, 1976).

In the Serial Letter Recognition model, it was proposed that word is read serially, letter by letter. It is based on the experiment by Sterling (1963) that found short words are observed to be faster recognized compared to long words. However, this particular theory fails to explain the word superiority effect that letters are easier to be recognized in known words rather than in isolation. In Serial Letter Recognition model, letters in in the third position of a word takes 3 times longer to recognize.

Lastly, the Parallel Letter Recognition model proposes that words are recognized simultaneously and the letter information is used to recognize the word. This was based on the experiment done by McClelland & Johnson (1977) demonstrated letters in pseudowords (nonwords that have phonetic regularity such as Mave and rint) are recognize faster than letters in nonwords (such as amve). This the strongest evidence that word recognition is cause by regular letter combination (the Parallel Letter Recognition model) and not word shape.

Later, McClelland & Rumelhart (1981) and Perfetti, Landi, & Oakhill (2007) modelled word recognition as involving both bottom up and top down processes. However, in Figure 2.4 a simplified model is illustrated from a bottom up perspective. The visual processing of word recognition occurs in three levels: the feature, the letter and the word and, this is known as the interaction activation model (McClelland & Rumelhart, 1981). The arrow indicates positive evidence which occurs when each adjacent node suggests each other's existence; and the circles indicate negative evidence when those nodes do not suggest each other's existence. Reed (2000) and Lupker (2007) pointed out that this model was intended to also explain the *word superiority effect.*

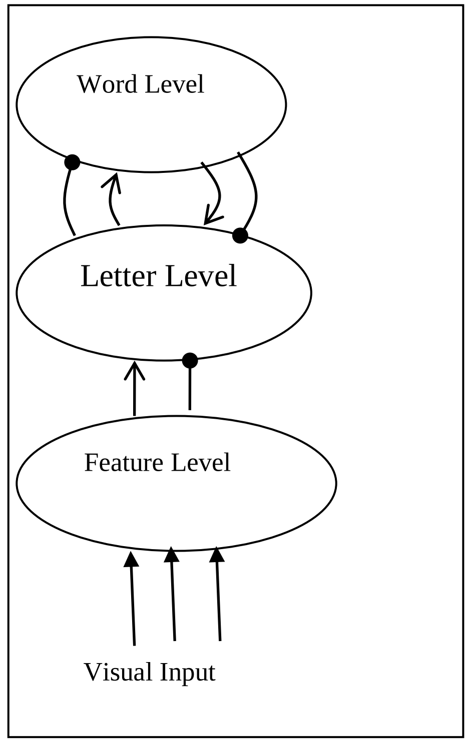


Figure 2.2: Interaction activation model of word recognition

When a literate sees an alphabet written on a paper, he will recognize it to be an alphabet and he can also recognize a word without going through it letter by letter, the word is recognized as a whole. Al-Haytam (2000) described that when a literate person sees a letter, he recognizes the letter form and knows for example '*a'* comes before '*d'*. The same process happens for a familiar word such as '*Allah'*, recognized as a whole without inspecting its letter one by one. However, if the word is unfamiliar, a literate person inspects it letter by letter until the meaning is perceived.

Research in WSE is conducted in languages that uses the Latin alphabets i.e. English {such as done by McClelland and Johnson (1977), Borowsky, Owen, & Masson (2002), Estes & Brunn, (1987)}, French and Italian and in nonalphabetic language such as Chinese. However, there is also evidence that WSE occurs in the Arabic language (Jordon, Peterson, & Almabruk, 2010) although there are differences between the Latin and the Arabic script.

(b) The word frequency effect (WFE)

Forster and Chambers (1973) conducted one of the first influential experimentation on WFE. This effect states that words that are seen more frequently are responded to more rapidly. There are numerous researches adopting this model such as done by Taft & Russell (1992), Segui, Mehler, Frauenfelder, & Morton (1982), Morrison & Ellis (1995), Oweini & Hazoury (2010) and Grainger (1990).

Based on these studies many researchers had come up with the development of frequently used words that are taught in early stage of reading especially to children (Oweini and Hazoury, 2010). This includes the following:

- Dolch list: 220 basic recognizable words that consist of conjunctions, prepositions, pronouns, adverbs, adjectives and verbs. Another 95 words for common nouns
- Fry list: Instant Word List where the first 100 words make up 50% of written material and the first 10 words of the 100 words make up 24 % of printed material. These ten words are: *the, of, and, a, to, in, is, you, that* and *it*.

These lists are for the English words and they are created since the 1930s. Recently, Oweini & Hazoury (2010) had also come up with the Arabic frequently used words for usage in the Arab world . However, Dewhurst, Brandt, & Sharp (2004) pointed that these frequently seen words are actually easier to recall than recognized. It is the rare words that are easier to be recognized. Experiments conducted regarding this issue, were based on pure and mixed lists. A pure list is one that contains all frequent or rare words and a mixed list contain a mixture of frequent and rare words. The pure list experiment leads to a recall advantage for frequent words list compared to the rare words list.

2.2.2 Reading comprehension problem and solution

Comprehension difficulties can be associated to lexical deficiency or decoding skills, which is mapping orthographical elements (of the text) to phonological elements (Nation, 2007). It is also known that sensitivity of readers to morphological information of a word relates to reading comprehension and word recognition. In addition, these readers are found to have faster reading time and thus demonstrate reading fluency (McCutchen, Logan, & Biangardi-Orpe, 2009). Reading with appropriate expression and intonation (prosody) can indicate fluency and this can only be achieved in the existence of reading comprehension. Fluency is achieved in automatic reading which is reading with speed in an effortless manner and lack of consciousness (Kuhn, Schwanenflugel, & Meisinger, 2010).

On the other hand, visual word recognition helps in the lexical processes so that our working memory resources can be devoted to comprehension. In addition, as reading and spoken language comprehension are related (Perfetti, Landi, & Oakhill, 2007), the ability to speak a language and developing this skill can also help to develop reading comprehension skill. Yaxley and Zwaan (2007) found that readers mentally simulate the visibility of an object during language comprehension thus linguistic simulation of the object properties is one of the ways that could help the reader to comprehend. Similar work by others (Zwaan, Madden, Yaxley, & Aveyard, 2004; Gosselin and Schyns, 2004; Richardson, Spivey, Barsalou, & McRae, 2003) also support this evidence. A related work such as by Haber and Myers (1982) found that there is greater accuracy in remembering pictograms compared to words. Another, interesting method for graphic representation of language can be seen in the ancient Egyptian hieroglyphic scripts. This script represents words directly by an analog schematic picture (Frost, 2007). However, this type of script called the ideogram is not efficient enough to be used as means of communication. Consequently, large number of hieroglyphic signs were alphabetic in nature.

Vocabulary deficiency can also impede reading comprehension (Harmon and Wood, 2008; Nagy and Scots, 2000; Lehr, Osborn and Hiebert, 2004). Vocabulary is knowledge of words and word meaning is sometimes also known as a lexicon. Vocabulary is important in achieving reading comprehension and it predicts academic achievement of children at schools (Silverman & Crandell, 2010). It is estimated that in order to comprehend a text, 95% of the words should be known (Fukkink, Hulstijn, & Simis, 2005). Word level processing therefore is an essential part of comprehension; it is the vital basic requirement before any higher level processes of comprehension can take place. Also, it is estimated that approximately

2000 to 3500 words should be known in a year for vocabulary development (Silverman & Crandell, 2010).

In setting up instructional procedures for building word knowledge, five aspects must be taken into consideration: word learning is incremental; word knowledge is multidimensional; words have often multiple meaning (polysemy); a word knowledge is interrelated to other words; word knowledge depends on the type of word involved (Nagy and Scotts, 2000). However, the best way in learning words is through its natural process of immersion in oral and written language.

There are several practices that can help vocabulary development in a classroom setting. They are as follows: acting out and illustrating words; analyzing words semantically; applying words in new contexts; defining words explicitly in rich context and word study (Silverman and Crandell, 2010). Other methods include semantic mappings, morphological illustration (Harmon and Wood, 2008; Nagy and Scotts, 2000), semantic organizers (graphic representation of text) (NICHD, 2006), cognate words exposure and computer technology as tool for teaching (NICHD, 2006; Korat, 2010; Lehr, Osborn and Hiebert, 2004).

For example in a classroom setting, word study can be done in a text-talk session, i.e. the teacher gives a descriptive sentence and the children give an appropriate word response. In using a computer technology as a tool, e-book, animation, hypertext can be included in the system. Morphological illustration can make students aware that words can be based on meaningful word parts and this includes knowledge of root words, prefixes and suffixes. Cognate words however, are words that originate from another language and the learner of a second language can easily use this known word in oral or written context.

Choosing words for instruction is also important (Lehr, Osborn and Hiebert, 2004). The priority should be based on the importance and these kind of words are usually cue words. Another priority in choosing word is the usefulness and frequency of the words. Beck, McKeown, & Linda (2002) categorized these words into three tiers:

- Tier one: Words that students are most likely to know
- Tier two: Words that appear widely in numerous written and oral text
- Tier three: Words that rarely appear in text

Words have relations to each other just as family members. There are words that mean the same (synonym), opposite (antonym), hear alike (homonym). Understanding the relations of words develops linkage and network of knowledge which is essential for enhancing one's reading skill. This is done through one's real world experience of the word. The reader's pre-existing knowledge is essential to figure out the meaning of unknown words encountered in a sentence (Nagy & Hiebert, 2011). This can be done by looking at the words around it. It is one of the ways reading comprehension is achieved once word recognition processed is established.

2.3 Qur'anic lessons

Learning the Qur'an can be applied to all Muslims since it is considered as an obligation to do so in the Islamic teaching (Yusoff, 2011). However, there are two major paths in acquiring this knowledge. One is taking a specific educational path in which an individual can pursue to be specialized in the Islamic field. For example it can be observed that scholars of Islam went through Islamic education in Islamic school or *madrasah* and then pursued Islamic studies in the tertiary level (Abdullah, 2003; Ahmad, 2003; Yusoff, 2010). This path is just the same as any other field of studies. In this particular path, one would learn to recite the Qur'an and understand the message and other related fields towards the understanding of the message of the Qur'an. Many would also memorize the whole Qur'an and then be called an '*alim* or an *ustadz* (knowledgeable) and *hafidz* (the one who memorized).

The second educational path which can be applied to the majority of the Muslims is in learning to recite the Qur'an and other obligations to perform everyday rituals. For example, one must learn how to recite the first chapter of the Qur'an since it is needed to do so to perform the five daily prayers (Gade, 2004).

In Qur'anic recitation classes emphasise is on correct pronunciation of the written text, thus the science of recitation is from the fact that the *Qur'an* should be preserved as the original form (Yusoff & Saidi, 2008). It is predominantly known as *Tajwid*, the art of recitation. In lessons of *Tajwid*, one learns to decode the *Qur'anic* orthography to the *Qur'anic* phonology. The *Qur'anic* orthography is very similar to those of the modern standard

Arabic orthography. Due to the very detail and precise rules of recitation in *Tajwid* lessons, there exist many pedagogical approaches on teaching *Qur'anic* recitation. The ones that are applied in Malaysia are such as the *Baghdadiyyah*, *al-Myassar*, *Qiraati*, *al-Iqra'*, *al-Barqy*, *Kalam* and *al-Hatta'iyyah* (Muhammad, 2003; Mohd, Adel, & Ahmed, 2003 and Humam, 2005).

A typical approach that is adopted in *Tajwid* lesson would be to recognizing the Arabic letters, the voweling marks (*Tashkil*), recognizing the *Qur'anic* letter phonemes and the pronunciation of those phonemes (Mohd, Adel, & Ahmed, 2003). It is almost a unified way with only slight variations. Houtsonen (1994) claimed that even Morocco adapt this method of teaching *Qur'an*.

Tajwid lessons are considered as very important. We even see attempts to preserve correct recitation by printing the *Qur'anic* verses with colour coded *Tajwid* rules or numbering systems (Dahlan & Sahil, 2007) for *Qur'an* reciters. Without knowledge of *tajwid*, Qur'anic verses are read inaccurately. Inaccurate recitation gives inaccurate meaning for the recitation. As it was discussed previously, to achieve reading comprehension, the word recognition step has to be achieved by knowing how to transform the orthographic form to the phonological sound. And if this step is missing, the higher level step of comprehension is impossible to achieve. Another reason why *tajwid* is emphasize is that the Arabic script is difficult and takes longer time to learn (Anonymous, 2006) because of context of character (joining or non-joining, different shape when written as beginning, middle or end of word), the dots and the diacritics writing system. The nature of the script is cursive and for a beginner, the continuous writing (without spacing in the Qur'anic verses) may deter the motivation to learn the language. Therefore the basis of learning *Tajwid* extensively is with solid argument and not purely spiritual.

Lessons of comprehending the words of the *Qur'an* which is also important had always been a separate lesson from learning *Tajwid*. Qur'anic *tafsir* lessons are not suitable for young children. Instead, the emphasis is on the memorization of the Qur'an to perform obligatory acts. It is also to later on ease the process of Qur'anic recitation and subsequently comprehension. In addition, obligatory lesson such as learning the five pillar of Islam and the six pillars of faith are also emphasize. This is also true for adults; Qur'anic lessons to increase Islamic knowledge are essentially attained throughout an individual life span. Through these lessons one acquire the knowledge of topic and theme which are important for the process of reading comprehension such as making inference, word knowledge and word relations.

Several solutions for comprehension of Qur'anic word literal meaning had been proposed. The work done by Abdelbaki (1983) proposed pictorial sketches for certain *ayat* (verses) of the Qur'an to aid in the comprehension of Qur'anic verses. This is suggested for those in the intermediate level of non-Arabic speakers. Abdulazeez (2004), takes the approach of teaching the meaning of Qur'anic words based on word frequency. The focus was also to benefit comprehension of recitation during prayer. Another approach is the attempt to produce word by word translations books of the Qur'an verses to help reciters learn the meaning of words instead of just giving meaning verse (sentence) by verse. These Qur'an translation are only popular in Malaysia for example since the year 2009. Lastly, there are lessons conducted specifically to learn Qur'anic such as the *Harfiah* Qur'anic lessons by

Hasim (2012). The focus of this lesson is to teach how to segment Qur'anic words to be able to identify word meaning and hidden pronouns in verbs to achieve Qur'anic verses comprehension.

Therefore subsequent §s highlight the issues relating to the Qur'anic terminologies, languages involve i.e. Arabic and Malay language and software for the purpose of reading comprehension.

2.4 The Qur'an structure and terminologies

The Qur'an as explained in § 1.1 is believed as the word of Allah revealed and then later on complied in a form of book. It is not like a normal book; it contain scientific facts yet it is not a scientific book; there are stories of the past and the future yet it is not a history book; it is most elegantly written yet not a book of prose or book of poetry (Sa'ari & Borhan, 2003). It is believed as a book of guidance for the humankind to lead a meaningful life in preparation of the life (Al-Qur'an, *Al-Isra', sura* 17: *ayah* 9).

There are many literatures on the *Qur'an* (Ali, 1997; Denffer, 1983) which contain information about the *Qur'an*. One can also read the book itself, either the Arabic version only such as printed by the government of Saudi Arabia in the King Fahd *Qur'an* Complex[1] of Medina or the Arabic plus the translation versions (Maulana, 2002) or only the translation version (Yusuf , 2001).

[1] http://www.Qur'ancomplex.com/default.asp?l=eng

The *Qur'an* contains 114 chapters or known as *sura*, with a total of 6236 *ayat* or verses and 77430 words (Dukes, Atwell, & Sharaf, 2010; Dukes & Habash, 2010). It is also divided into seven *Manzil*s or 30 parts called *juz* or into 60 *hizbs*. Each *hizb* is further divided into four. *Juz, manzil* and *hizb* are division or parts of the document (in approximately the same length) to be read within seven, 30 or 60 days for those who are capable of doing so. The 114 chapters are of types *Meccan* or *Medinan* (chapters revealed in *Mekah* or *Madinah*, located in the Saudi Arabia). Figure 2.3 shows the number of verses in each *suras*. Each *sura* contains minimum of three ayah and maximum of two hundred and eighty six *ayah*. One *ayah* can contain up to 161 words (referring to the longest verse, *sura* two, verse 282). The first *sura* is called Al-Fatihah (The Opening) and the last *sura* is called Al-Nas (The Mankind). The longest sura is called *Al-Baqarah* (The Cow) which is the second *sura* in the Qur'an.

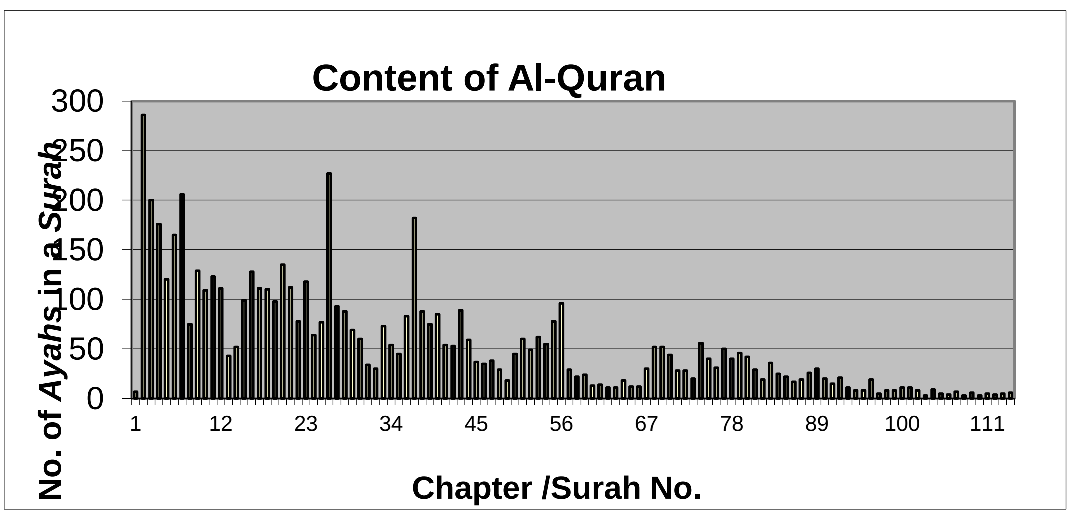


Figure 2.3: The Content of Al-Qur'an: 114 chapters and its corresponding quantity of verses/*ayahs*

2.5 The Qur’an and the Arabic language

Arabic is a Semitic language and the script originates from Nabataean Aramaic Script. It was used since the fourth century (Ager, 2010; Versteegh 2003; Dickins & Watson 2002). It is because of the desire to preserve the Qur’an in the whole discourse that new Arabic letters were created by adding dots during the seventh century as the original letters were causing problem of ambiguity (Omran, 1998; Denffer, 1983; Bauer, 1996).

2.5.1 Qur’anic word structure

The Qur’an is a classical book and the language is in the traditional Arabic known as *i’rab* (Dukes, Atwell, & Sharaf, 2010) . There are 28 main *Qur’anic* alphabets (shown in Table 2.1), identical both in the Modern Standard (MSA) and the traditional language[2]. There are also other alphabets which include ء (*‘*) pronounced as *hamza*, which can be used alone or with a carrier on أ(*A*), ؤ (*w)* or ئ (*Y*). Two others alphabets are the ta marbutah ة (*p*) and the alif maksura ى (*Y*) Qur’anic words are marked with diacritics (shown in Table 2.2) to give specific sound and meaning, avoiding ambiguity. The *fatha*, *damma* and *kasra* are short vowels while the *tanween fatha*, *damma* and *kasra* represent nunation (for doubled case endings). The *sukun* and *shadda* are syllabification marks. There can also be the combination of syllabification-short vowels and syllabification-double case ending, for example *Shadda-fatha* (بَّ) and *shadda-tanween damma* (بٌّ).

[2] Note that in this thesis the transcription of the Arabic letters used is from Buckwalter (2010)

Table 2.1. Phonetic transcription and transliteration for letters.

	Letter	Arabic	Transcription			Letter	Arabic	Transcription	
			Dukes (2010)	Buck-walter (2010)				Dukes (2010)	Buck-walter (2010)
1	*alif*	ا	*ā*	*A*	15	*ḍād*	ض	*ḍ*	*D*
2	*bā*	ب	*b*	*b*	16	*ṭā*	ط	*ṭ*	*T*
3	*tā*	ت	*t*	*t*	17	*ẓā*	ظ	*ẓ*	*Z*
4	*thā*	ث	*th*	*v*	18	*'ayn*	ع	'	*E*
5	*jīm*	ج	*j*	*j*	19	*ghayn*	غ	*gh*	*g*
6	*ḥā*	ح	*ḥ*	*H*	20	*fā*	ف	*f*	*f*
7	*khā*	خ	*kh*	*x*	21	*qāf*	ق	*q*	*q*
8	*dāl*	د	*d*	*d*	22	*kāf*	ك	*k*	*k*
9	*dhāl*	ذ	*dh*	*	23	*lām*	ل	*l*	*L*
10	*rā*	ر	*r*	*r*	24	*mīm*	م	*m*	*m*
11	*zāy*	ز	*z*	*z*	25	*nūn*	ن	*n*	*n*
12	*sīn*	س	*s*	*s*	26	*hā*	ه	*h*	*h*
13	*shīn*	ش	*sh*	*$*	27	*wāw*	و	*w*	*w*
14	*ṣād*	ص	*ṣ*	*S*	28	*yā*	ي	*y*	*y*

Source **(Dukes(a), 2009)** and (Buckwalter, 2010)

Table 2.2: Phonetic transcription for diacritics.

	Diacritic	Arabic	Tran-scription		Diacritic	Arabic	Tran-scription
1	*Tanween fatha*	ـً	*an*	5	*damma*	ـُ	*u*
2	*Tanween damma*	ـٌ	*un*	6	*kasra*	ـِ	*i*
3	*Tanween kasrat*	ـٍ	*in*	7	*shadda*	ـّ	(double)
4	*fatha*	ـَ	*a*	8	*sukūn*	ـْ	'

Source (Dukes(a), 2009) and (Rashwan, Al-Badrashiny, & Attia, 2010)

A verse is a sentence of the Qur'an, also known as *ayah*. As in all languages, an *ayah* can contain many words. In the English language and the Malay for example, a token can be found equivalent to a word in a sentence but never present as a phrase (which consists of several words). However, generally in the Arabic language a token can form a single phrase. This is referred to as a rich morphological language (Buckwalter, 2010). For example Figure 2.2 shows a token of Qur'anic Arabic word which forms a phrase.

you + we + created + and =

= وَ + خلق + نا + كم

Meaning: And We created you.

Figure 2.4: A token in Qur'anic Arabic which forms a complete *ayah*.

Source (Duke , 2010).

The structure of Qur'anic words can be categorized into sets of nominals called *ism'*, verbs called *fi'il,* and particles called *harf.* Nouns, pronouns, adjectives, and adverbs are all in the nominal category. The syntax of nominals is related to gender, adjectives, possessives, appositions, specification and numbers. Prepositions, conjunctions, interrogatives, and others are in the category of particle. The verbs are derived from a root word with 3, 4, or 5 consonants. The root word meaning is changed by inserting 10 affixed alphabets (10AL) which are ا, ه, ي, ن, و, م, ت, ل, ء, س (*A, h, y, n, w, m, t, l, ', s)*. These can be added in front (as prefix), in the middle (as infix), or the end (as suffix) of the root word forming another word with related meaning. For example, consider a root word كتب from 3 consonants ب (*b),* ت (*t),* ك(*k*)) which means "write". Adding one or a few of the 10AL to the root forms a stem that has different but related meanings: “writing", "book", "author", "library", or "office”. Nouns, proper nouns, and adjectives have roots while word particles have none. The meaning of particles usually remains the same and no derivatives are formed from them.

Consider the three consonant word فعل(with 3 consonant letters ف, ع, ل). Let ف (*f*) be the first consonant word, ع (*E*) the second consonant word, and ل (*l*) the third consonant word. There are also particles and pronouns that are concatenated to noun and verb tokens. These are added as suffixes or prefixes. Table 2.3 shows the related prefixes and suffixes common in the Qur’an.

Table 2.4 shows the different word forms that can be produced by adding one or a few of the 10AL. These words are taken from two main sources which are Ismail (2003) and Forbes (1868). The latter source is a very old book on Arabic grammar so that the grammar discussed would be close to the verses of the Qur’an, and the former is a more recent source yet at the same time it adopted a traditional approach to Arabic grammar.

Table 2.3 Particle and Pronouns Prefixes and Suffixes

Particles as prefixes		Particles/pronouns as suffixes	
Word token	Prefix	Word token	Suffix
الكتب	+لأ *Al*+	آبتك	+ك +*k*
للكتب	+لل *ll*+	أَبْنَآءَكُم	+كم +*km*
بكتب	+بِ *b*+	أَنفُسَهُمْ	+هم +*hm*
لكتب	+لِ *l*+	تَجِدُوهُ	+هُ +*h*
فَام	+فَ *f*+	أَقْرَرتمْ	+تـ م+ *tm*
وَعَىلـ	+وَ *w*+	وَعَدْناـ	+ناـ +*nA*
أَلَمَا	+أَ *A*+	وَعَمِلُوا	+او +*wA*
لَنُدْخِلَنَّهُمْ	+لَن *ln*+	جَعَلْنَهَا	+ها +*hA*
Others forms of the above		ٱتَّخَذْتُم	+تـ م+ *tm*
ٱلنَّاس	+ ٱلأ *Al* +		

Let X represents the 10AL added to the Arabic root word لعف *fEl* as prefix (Pre), suffix (Suf), or infix (In). The three main word pattern / word forms are as follows[3]:

- Root words with one to three added prefixes:
    - o لعف+$X_3X_2X_1$ *Pre1Pre2Pre3+fEl*
    - o لعف+ X_2X_1 *Pre_1Pre_2 +fEl*
    - o لعف+ X_1 *Pre_1+fEl*
- Root words with one to three added suffixes. However, the most common have up to a maximum of two added suffixes:
    - o X_3 X_2 X_1+فعل *fEl+Suf_1Suf_2 Suf_3*
    - o X_2 X_1+فعل *fEl+Suf_1Suf_2*

- Root words with one to two added infixes between the consonant alphabets. However, the most common type adds only one infix between each consonant:
    - o ل+X_2 ع+X_1+ ف *f+ In_1+E+In_2+ l*
    - o عل+ X_1+ ف *f+In_1+ El*

Lastly, the combinations of the three main word forms above are also possible. Table 2.4 shows 29 Arabic word patterns identified through literature.

[3] The notation used is representation of the Arabic alphabets in the English language, NOT in other languages.

Table 2.4: Arabic word patterns

No.	Word Pattern	X=prefix/infix /suffix	No.	Word Pattern	X=prefix/infix/ suffix
1	فعل + X_1	X_1 =ن ,ت, م, ا, أ, ي	12	عل +X_1+ ف	$X_{1=}$ ا , ي
	كتب+X_1	X_1=بِ , وَ,ل	13	عل +X_2X_1+ ف	$X_{1=}$وا
	مَا + X_1	X_1= فَ	14	فع +X_1+ل	$X_{2=}$ ا,و, ي
	فَلَا + X_1	X_1= أ	15	ف +X_1+ع +X_2+ل	$X_1X_{2=}$ وا
2	فعل + X_2X_1	$X_1X_{2=}$ ن ا $X_1X_{2=}$ ت م $X_1X_{2=}$ ن م $X_1X_{2=}$ ن ي $X_1X_{2=}$ ت ي $X_1X_{2=}$ ت ت $X_1X_{2=}$ ن ت		عل +X_2+ ف +X_1	$X_1X_{2=}$ ت ت $X_1X_{2=}$ ا ت $X_1X_{2=}$ ت م $X_1X_{2=}$ ا م $X_1X_{2=}$ ت ي $X_1X_{2=}$ ا ي $X_1X_{2=}$ ت ا
	كتب+X_2X_1	$X_1X_{2=}$ ل ا $X_1X_{2=}$ ل ل	16	عل +X_3+ ف +X_2X_1	$X_1X_2X_{1=}$ ا ت م $X_1X_2X_{1=}$ ا ت ي
3	فعل + $X_3X_2X_1$	X_1X_2 $X_{3=}$ ت س م X_1X_2 $X_{3=}$ ت س ي X_1X_2 $X_{3=}$ ت س ا	17	ل+X_2+فع +X_1	$X_1X_{2=}$ و إ $X_1X_{2=}$ ا إ $X_1X_{2=}$ ي ت $X_1X_{2=}$ وم $X_1X_{2=}$ ام $X_1X_{2=}$ اي $X_1X_{2=}$ وي
4	X_1+فعل	X_1=ت ,ة, ن, ي, ا,و	18	ل+X_3+فع +X_2X_1	X_1 $X_2X_{3=}$ ا ن ا
5.	X_2X_1+ فعل	X_1X_2= ا ت X_1X_2= ا و X_1X_2= م ت X_1X_2=ن ا	19	ل+X_3+فع +XX_2X_1	X_1 $X_2X_{3=}$ ا ت س ا
	X_2X_1 + word	X_1X_2= م ك X_1X_2= م ه	20	X_3+ل +X_2+فع +X_1	X_1 X_2 $X_{3=}$ ن ا ت
6.	$X_3X_2X_1$ + فعل	$X_1X_2X_3$=ا م ت	21	X_2+عل +X_1+ ف	X_2 $X_{1=}$ ة ا

Table 2.4 Continued

7.	X2+فعل +X1	X1 X2=ا ا X1 X2=ن ا X1 X2=ل ا X1 X2=ي ا X1 X2=ت أ X1 X2=ن ي X1 X2=ا ي X1 X2=ي ت X1 X2=ن ت X1 X2=ت ت X1 X2=ت م X1 X2=و ي	22	X3X2+عل +X1+ ف	X1 X2 X3= ن و ا X1 X2 X3= ت ا ا
8.	X3X2+ فعل+X1	X1 X2 X3=ن و ت X1 X2 X3=ن ي ت X1 X2 X3=ن ا ي X1 X2 X3=ن و ي X1 X2 X3=ن ا ت X1 X2 X3=ل ا ا X1 X2 X3=ا و ي	23	X4X3X2+عل +X1+ ف	X1X2X3X4= ن ا ت ا
9.	X4X3X2+ فعل+X1	X1X2X3X4= ن ا ن ت X1X2X3X4= ن ا ن ي	24	X3+عل +X2+ ف +X1	X1X2X3= ة ا م
10.	X_1+ word	$X_{1=}$ ةُ, كَ	25	X_2 + ل+X_1+فع	$X_1X_{2=}$ ة و $X_1X_{2=}$ ة ي $X_1X_{2=}$ ة ا
11.	X_2X_1+ word	$X_1X_{2=}$ ن ا $X_1X_{2=}$ ن و $X_1X_{2=}$ ن ي	26	X_3+ ل+X_2+فع +X_1	$X_1X_2X_{3=}$ ل ا ا $X_1X_2X_{3=}$ ن ا ي $X_1X_2X_{3=}$ ة و م
			27	X_4X_3 +ل+X_2+فع +X_1	$X_1X_2X_3$ $X_{4=}$ ن و و م $X_1X_2X_3$ $X_{4=}$ ن ا و م $X_1X_2X_3$ $X_{4=}$ ت ا و م
			28	$X_5X_4X_3$+ل+X_2+فع +X_1	$X_1X_2X_3X_4$ $X_{5=}$ ن ا ت و م
			29	X_3 + ل+X_2+ ع+X_1+ ف	$X_1X_2X_{3=}$ ن ا ا

2.6 The Malay language and the Qur'an

Malay is an Austronesian language and the earliest written script of the Malay language was written in the Indian script (Ager, 2010b). It is spoken in Malaysia, Indonesia, Singapore, Brunei and Thailand. There are one hundred and eighty eight million Malay speakers and one hundred and seventy million are from Indonesia.. However, when Islam was brought to the Malays through trade during the fourteenth century it had a strong influence on the local culture. Consequently, the Arabic script was used to write the Malay language following the language of the Qur'an. Many Arabic words are also originated from the Arabic language. However, during the seventeenth century, influenced by the English, the Arabic scripts were replaced by the Latin/Roman alphabets.

The Arabic version of the Malay script is still being used especially in Islamic religious schools and Islamic religious subjects. The script is called *Jawi* and contains all the twenty eight consonants of the Arabic letters plus six more giving the total of thirty four letters. Even though the original *Jawi* scripts did not have consonants letters following the Arabic language, now the consonants are part of the spelt words. They are included in the thirty four letters of the *Jawi* script (Nasrudin, Omar, Zakaria, & Liong, 2008).

There are many Malay words that are taken from the Arabic language. This is also the same case as in other languages such as Turkish, Urdu, Tajik and Afghani (Yusoff & Al-Qaisi, 2011). According to Karim (1996) there are about 1100 words from the Arabic language in the Malay dictionary (*Kamus Dewan* first edition, 1970s). Some Arabic words are taken purely in its original form – retaining meaning and pronunciation but some other words may have the influence of the Malay language on its pronunciation but retaining its original meaning (Yusoff & Adnan, 2009). There are also words with the same pronunciation but have a different meaning. For example as discussed by Rukaini (1996) the word *hakim*, *yakni* and *jahil* (from the Arabic language) retain the same meaning and pronunciation in the Malay language. The words *alim, kuliah* and *maktab* (from the Arabic language) retain the pronunciation but have different meaning in the Malay language. Also, the words which retain the same meaning but different pronunciation such as *asl, fikr, `ilm* (from the Arabic language) as compared to *asal*, *fikir* and *ilmu* (from the Malay language). However, the evolution of the *Jawi* writing had changed the way these words are written in the *Jawi* scripts. The original Arabic words which were written exactly like the ones in the *Jawi* script are no longer applicable. This imposes more problems in reading comprehension of the *Qur'anic* verses (Yusoff & Saidi, 2008).

2.7 Related software

Abdulazeez (2004), Abdelbaki (1983), Lazem (1993) and Yusuf (1992) had attempted on the teaching to understand *Qur'an* based on the most frequently occurring words. Abdulazeez (2004) and his team developed a software for this purpose although he started with designing lesson materials that were posted through email for interested individuals. The software can be downloaded from the UnderstandQuran.com (UQ) website[4]. This game software provides 15 *Qur'anic* word lists to be learned and memorized by the user with around 120 frequently used words in the Qur'an. It gives a word by word meaning to the Arabic words listed in the software. This includes breaking the words into segments when a token contains more then a word. Users need to answer questions based on the words provided. The software keeps track of the scores attained to monitor performance. Another such lesson can be obtained from Quraanic Lesson website (QL)[5]. The lessons provided in this website are similar to UnderstandQuran software with additional audio recitation feature provided. Both of these softwares are for learning Qur'anic words based

[4] http://www/understandquran.com
[5] http://quraaniclessons.com/index.asp

on Qur'anic words frequencies. One major limitation of these softwares is that the lessons designed are limited to questions set by the developer of the software. Once finished, users can only repeat and new questions are not included in the new version of the software. There is also no clear evidence for the effectiveness of both softwares since no reported user study can be found.

Another website related to understanding Qur'an is the Qur'an Corpus (QC)[6], it is a more comprehensive website about the Qur'an and also covers information on Qur'anic grammar. It also include free downloadable Qur'an ontology scripts, useful Java codes and links related to downloadable Qur'anic Arabic and translation databases in many different languages (Dukes , 2009a). Although word frequency is provided in this website, it is not designed for learning purposes; it is more towards providing information to users. A limited visualization view of the Qur'an ontology is also provided though its usefulness in context of understanding the Qur'an is not clear. Another type of Qur'anic software is the Zekr Qur'an (ZQ) software (Saboorian, 2010) and the Tanzil.info (TI) website that provides a computer interface for reading or reciting the Qur'an using a computer. This type of software is very typical and many others develop similar ones providing the facility to search words in the Qur'an, navigate through the book *sura* by *sura*, *ayah* by *ayah*, provides audio recitation from various famous reciters and translations in various languages. However, there are not many Qur'an softwares that provide the search function based on the root word in the Arabic language such as the the Tanzil.info. However, these softwares do not provide the support for learning the Arabic language other then providing the translation. Visualization is usually not provided to aid in the learning process.

[6] http://corpus.quran.com

There are also other English medium sofwares developed mainly for the purpose of improving one's vocabulary. One such software is the Ultimate Vocabulary (UL) software (eReflectSoftware, 2009). It provides usage examples, word visualization, translation in other languages, test, and games and it prepares interested individuals for exams such as GRE and TOEFL. The Free Reading (FR)[7] website is a website developed specifically for teaching children to read in English (Toppo, 2007). Although it is not a software system, the website is developed in a systematic manner providing activities, instructions, guidelines and resources on how to enhance children's reading skills. Word Recognition is part of its activity apart from offering Letter Sound, Letter Writing, Phonological Awareness and Sounding Out activities. Readily made available are the resources that can be downloaded and used with specific instruction on how to conduct a game / teaching activity. Table 2.5 shows the summary of features in all the softwares mentioned above.

Table 2.5: Comparison of features in Qur'anic software

	Feature \ Systems	UQ	QL	QC	ZQ	TI	UL	FR
1	Word Segmentation	√	√	√				
2	Visualization			√			√	
3	Score Tracker	√	√				√	
4	Word by Word Meaning	√	√	√			√	√
5	Educational purposes	√	√	√			√	√
6	Games	√					√	√
7	Frequently Used Words	√	√	√			√	
8	Audio		√		√	√	√	√
9	Translation in other language	√	√	√	√	√	√	
10	Root word search					√	√	
11	Qur'an based	√	√	√	√	√		
	TOTAL FEATURES	8	8	7	3	4	9	4

[7] http://www.freereading.net/index.php?title=FreeReading_Media

	SUPPORTED							

2.8 Summary

Many Muslims can read fluently without comprehension. This peculiarity is related to the complexity of the Arabic word morphological structure. In addition to this, in these recent years the emphasis on the Arabic language is not strong enough and the *Jawi* script based on the Arabic script that had been used for many centuries is replaced by the Latin script in the formal education system. While reading comprehension relates to many coordinated processes involving our memory, the SM, STM and LTM, teaching to read with comprehension is not an easy task. Learning word meaning does not necessarily achieve comprehension and that frequently used words based on the WFE and WSE can be beneficial in determining word-list to use in language learning.

3.0 INFORMATION VISUALIZATION FOR COMPREHENSION IN QUR'ANIC VERSES

In § 2.6.2, it was pointed out that semantic maps and semantic organizer can be useful in teaching vocabulary. This is where information visualization can play its role. Semantic maps are ways to show the connection of word meaning to another word and semantic organizers are using images as part of text in helping towards reading comprehension.

The goal of visualization can be seen as to transform data into a perceptually efficient visual format (Ware, 2000). There are mainly two aspects in information visualization. One aspect deals with the concept of quantity which involves numerical data such as size, time, magnification, height, length, dimensionality, scale and others. Familiar techniques involved would be plotting graphs such as scatter plot and bar charts. Another aspect is on position and connectivity. Visualizing geographical information, hierarchical data and network diagram or scattered objects in the spatial space falls into this pattern. Nodes and trees such as hyperbolic trees also fall into this pattern. The two aspects are sometimes combined together to provide more effective view of the information involved. The rational for visualizing information is to provide clues to the users on what is available to create potentially useful insight to the information. This is related to the information processing stages discussed in §2.5.

Information visualization concepts are not just about visualizing the data or information. The fact that dynamic interaction and exploration of the visual display are part of the system, contributes significantly to effective operation of interpreting data. Operation such as zooming, panning, selection and rearrangement would contribute to effectiveness of how one integrates the pre-existing knowledge and the new one subsequently resulting in the understanding of certain concepts (Spence, 2001). It can be seen as a mental exercise that facilitates the cognitive process.

Therefore in this chapter, discussion is focused on word visualization concepts and information visualization based on aspects of quantity and connectivity. The latter discussion includes related visualization software on Qur'an and the comparison between perpendicular and parallel plots.

3.1 Word visualization

Word visualization is more popular amongst educators of the English language for children in lower schools (Simmons, 2002; Claggett&Brown, 1992; Carrus, 2008). The words are drawn illustrating the meaning or written to show the structure. For example, Carrus (2008) investigated methods of teaching vocabulary to second grade students and found that by creating visualization and definition of the vocabulary help children to learn the words better as shown in Figure 3.1.

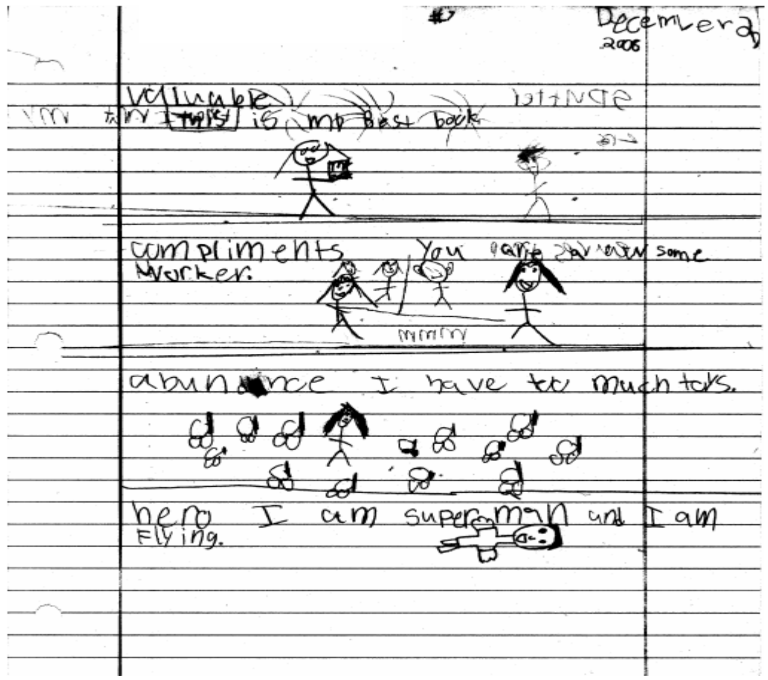


Figure 3.1: Drawing the meaning of the word (Source: Carrus, 2008)

Figure 3.2 (Simmon, 2002) shows a vocabulary card of the word inconceivable; the World Trade Centre was drawn as the background to illustrate the meaning and at the same time the word 'inconceivable' is written to show the structure which includes the prefix, root and suffix.

There is also work on visualizing word relations in electronic dictionary (Yoshie, Naoaki & Yoichi, 2009). This study found that word memorization using WorNet software (Miller, 1995) that visualizes word relations results in higher scores. Scurletis (2009) also supported this finding, that visualization of words helps in vocabulary retention and this can be achieved using supporting software such as the Webo word (http://www.weboword.com). There is also work on visualizing words by breaking them into syllables and then associating colors to them (Das, 2004). However, the purpose of this work is not clear except that it was mentioned to offer preliminary design concepts to envision words.

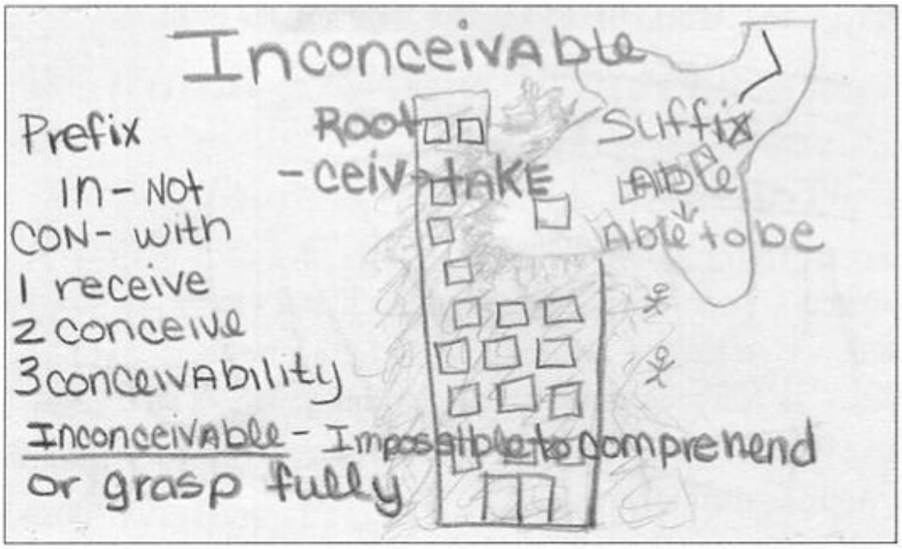


Figure 3.2: Visualizing word structure (prefix, root and suffix) (Source:Simmon, 2002)

3.2 Information visualization based on quantity and connectivity

In this §, information visualization based on quantity, connectivity and both are discussed. Word information for example, word count and word position in context of the structure of a document are useful to be visualized to assist users in finding those words in the document. Finding words helps in knowing certain issues related to the word. In the context of this study the word information helps to support WFE toward achieving word recognition in Qur'anic verses. In this case, quantity directly relates to word count and connectivity directly relates to word position. The act of finding, filtering and analysing information related to word information is seen to facilities the learning process towards word recognition.

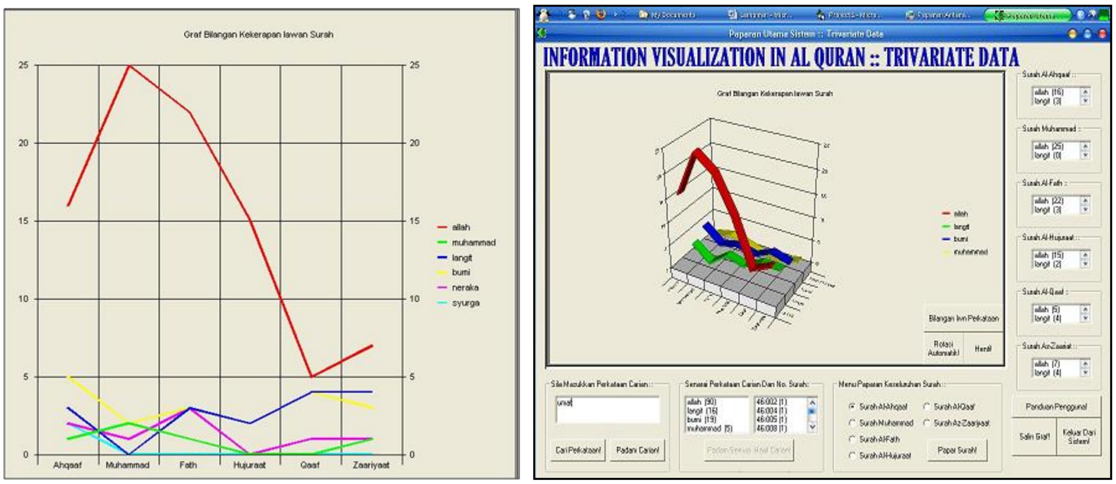
3.2.1 Quantity aspects in visualization

(a) Scatter or line plots

Scatter plot is one of the most popular techniques used to display quantity. Many literatures in information visualization extend the idea of scatter plot as a technique to display data for effective viewing such as discussed by Spence (2001). One classical example is John's Snow London cholera outbreak diagram as shown by Spence (2001). Also included were the Spotfire software by Ahlberg (1996) and the Influence Explorer by Tweedie, Spence, Dawkes, & and Su (1995). Other examples are such by Ahlberg & Wistrand (1995) who developed the IVEE software that uses starfield, an interactive scatter plot that allows users to zoom and pan, exploring the data to gain insight. As for Weippl (2001), he worked on content based relation of hypertext database, displaying scattered text in an information landscape. Other applications are such as the visualization of geographical data (such as NASA earth observation database), molecular biology project (Keim & Kriegel, 1996), financial data (such as stock prices and exchange rates), public policy (such as crime rates) (Heer, Kong, & Agrawala, 2009). Many of the targeted users are in the category of experts.

The related tools which are specific for plotting any data in the scatter plots are such as VisDB (Keim & Kriegel, 1996), VizRank (Leban, Zupan, Vidmar, & Bratko, 2006), Spotfire (Ahlberg, 1996; TIBCO Spotfire, 2010) and XGobi (Symanzik, Majure, & and Cook, 1996).

Qur'anic software based on the scatter plot had also been developed (Harun, Ainul, & Hamdzani, 2005). The search function for the system is in Malay particularly for *juz* 26. It displays the frequency of occurrence for the searched words in two and three dimensional forms.



(a) (b)

Figure 3.3: (a)The scatter plot in two dimensional display (left) (b)The plot in three dimensional display(right)

Figure 3.3 (a) is the scatter plot in two dimensional display, while Figure 3.3(b) is in three dimension. In actual fact, there is no third variable involved. The variables are the frequency of the search word related to another variable which is the *suras* in *juz* 26. The three dimensional plot can be rotated manually using the mouse or automatically rotated by pressing the rotation button.

(b) Bar charts

There are also many studies in information visualization that extended bar chart as the effective quantity interpreter such as discussed by Spence (2001). Examples are such as

Rao and Card (1995) on the Table Lens, the study on geographical location for supply distribution by Roth, Chuah, Kerpedjiev, Kolojejchick, & Lucas (1997), the design of electrical components by Spence *et*. *al*. (1995); Friendly (1994) used the mosaic to display Titanic data sets and Carey, Kriwaczek, & Rüger (2000) used radial visualization, a hybrid of bar chart for document searching and browsing.

The visual directory prototype is an interface developed to represent all *suras* of the Qur'an according to the number of *ayah*. It resembles an upside down bar chart, allowing the users to navigate to a particular *sura* by clicking on one of the bars. Figure 3.4 shows the bar chart for all the 114 sura and 30 *juz* of the *Qur'an*. The leftmost position is the position of the first *juz* and the rightmost is the 30th *juz*. All the corresponding *sura* are placed under each *juz* and visualized according to the amount of verses it contains; the longer the bar, the higher the amount of verses of the *sura*. It must be highlighted that if one *sura* starts from a particular *juz* it can continue into the next. This depends on how long that particular *sura* is. Taking an example from Figure 3.4, in the first *juz*, the small green bar shows the first *sura*, (called *al-Fatihah*) which contains only seven verses. The second *sura* which is called *al-Baqarah* contains 286 verses and it continues until the third *juz* (indicated by the same red bar). We can see that four colours are used to visualize the *Meccan* and *Medinan sura* with two colours for each type so that the start and the end of each *sura* can be clearly identified since two consecutive *sura* can be of the same type. Each bar can be clicked to navigate to the particular *sura*.

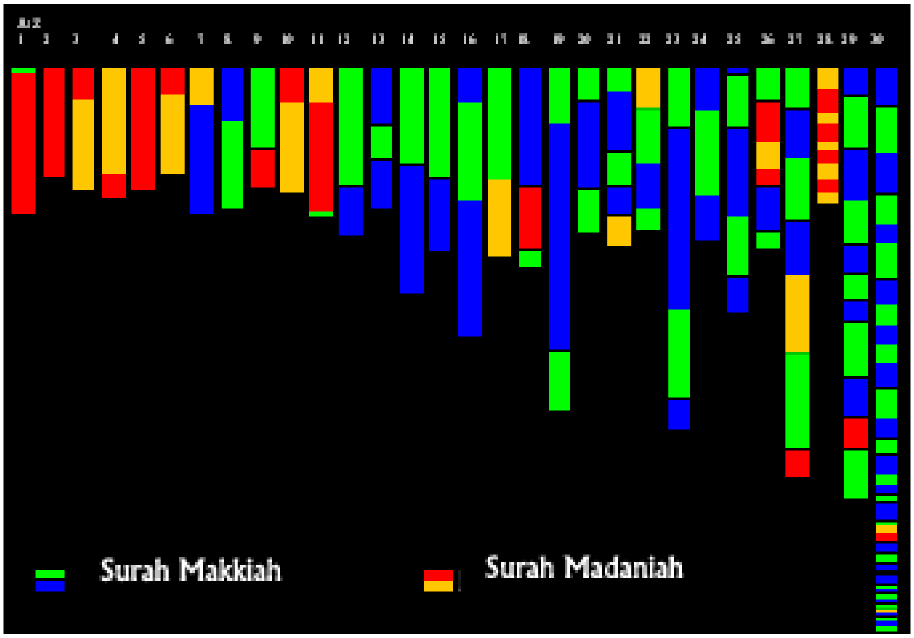


Figure 3.4: The visual directory approach

3.2.2 Connectivity aspect of visualization

(a) Hyperbolic

The hyperbolic visualization applies the connectivity aspect, representing data in the form of network and hierarchical trees. Work by Song, Curran, & Sterritt (2004) described the use of large hierarchical flex tree to visualize the directory structure of a personal computer. One of the most popular trees is the hyperbolic tree by Lamping, Rao, & Pirolli (1995) applied as the hyperbolic browser of the World Wide Web, and this idea was followed by the hyperbolic approach in visualization of *Qur'an* by Mashud (2004). The prototype system displays all *sura* of the Qur'an in hyperbolic tree representation, letting the users zooming-in a particular *manzil*, *juz*, *sura* and *ayah*. A particular *sura* can be searched by entering the *sura* name and *ayah* number. It also classifies the *sura* into *Medinan* or *Meccan*. Figure 3.5 shows the root view of the document structure using the hyperbolic approach with the root node at the center of the screen, which consists of seven *manzil* nodes (trunks of the tree). All *juz* nodes (branches of the tree), *sura* nodes (sub-branches) and *ayah* nodes (leaves) are distorted. Nodes with yellow colour represent *Manzils*, nodes with orange colour represent *Juz*, nodes with blue colour represent *sura* belonging to *Medinan* category and nodes with green colour represent *sura* belonging to *Meccan* category. Users can click on either one of these *Manzils* to show its detail.

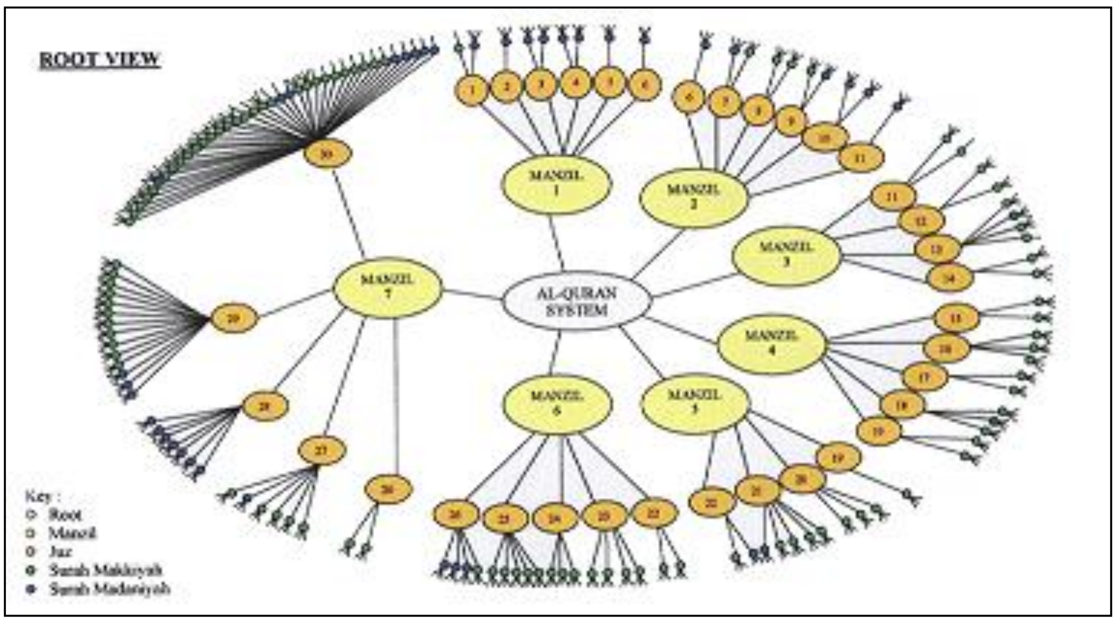


Figure 3.5: The root view of the document structure using hyperbolic approach

(Source: (Mashud, 2004))

(b) Others

A distorted view approach allows focus to a set of objects and distorting others not in the main view and at the same time provides the sense of positioning (the "where am I?" question). Examples are, the bifocal display technique applied to the Perspective Wall (Mackinlay, Robertson, & Card, 1991; Hartley, Churcher, & Albertson, 2000) and the fisheye view that involves focus plus context (Furnas, 1986). In this way, users are aware of all information that exist when dealing with large data or information and would not be lost in the experience.

Figure 3.6 which illustrates a network approach in relation to the Qur'an (Harun, Ainul, & Hamdzani, 2005) shows the search results for *Langit* and *Bumi* in Malay language. Users are able to compare the results of two searched words. The system keeps track of the word count of the search words in relation to each *sura* in *Juz 26.* Then, nodes are drawn according to the word count for each *sura*. Different node colour is use to differentiate the two search words.

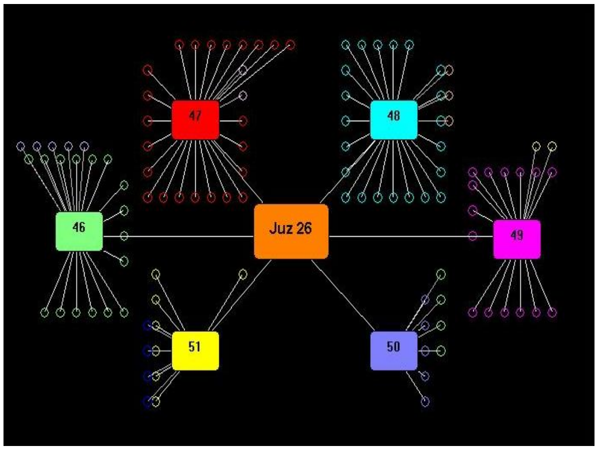


Figure 3.6: The network approach

3.2.3 Using both quantity and connectivity aspect of visualization

(a) Galaxy

Wise *et. al.* (1995) did research work to visualize document using the metaphor of Galaxies. The visualization display resembles starfields displaying clusters of document in the form of scatter plots points. Points within a corpus cluster tend to be related and the gist term for the cluster is displayed creating an overall view such as the stars in the sky. Further interaction with the interface results in change to the metaphor of galaxies to that of landscape. Here, users can choose a document of interest by looking at topics that are covered together represented by several elevations in a terrain. Another related work by Rennison (1994) was on Galaxy of News, visualizing large quantity of news stories displayed on three dimensional news of information space. The relevant searched topic is displayed with the biggest font and automatically linked with related articles.

A software developed in relation to the Qur'an based on both quantity and connectivity aspect is the galaxy prototype (Hashim, 2005). This closely followed the work by Rennison (1994). It provides a search function and displays the search results of *juz* 23 resembling the stars in the sky (the galaxy). When a particular word (in Arabic) is searched, all *suras* that contain the searched word will be displayed. Figure 3.7 shows the interface. For example the searched word (سبيل)in red is bigger in size whereas *sura* is in blue colour and smaller in size. On the right side of interface is where the translated name of the *sura* in Malay language. Once the *sura* is clicked it will be bigger in size and the search word appears smaller and blended in the background.

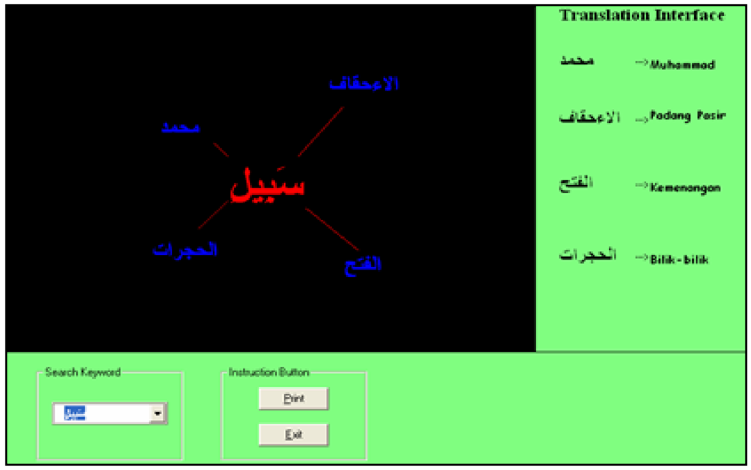


Figure 3.7: The display for the searched word سبيل

Figure 3.8: The display when *sura* Muhammad (محمد) is clicked.

This tells users which image to focus on. Figure 3.8 shows that once the *sura Muhammad* is clicked, the focus is now on the particular selected *sura* which displays the particular *ayah* that contains the searched word سبيل. The translation of the *ayah* is also provided in the right hand side of the interface. If the verse is selected, the whole *sura* will appear for users to recite. Users are also provided with the ‘Print’ function.

(b) Parallel plot

Parallel plots are relatively unknown to many users. It is a plot that can be used interchangeably with the scatter plot. It can be used in aspects of quantity and connectivity in an overall view. Several literatures discussed on work done with this plot (Inselberg, 2009; Keim and Kriegel, 1996; Chou, Lin, & Yeh, 1999).

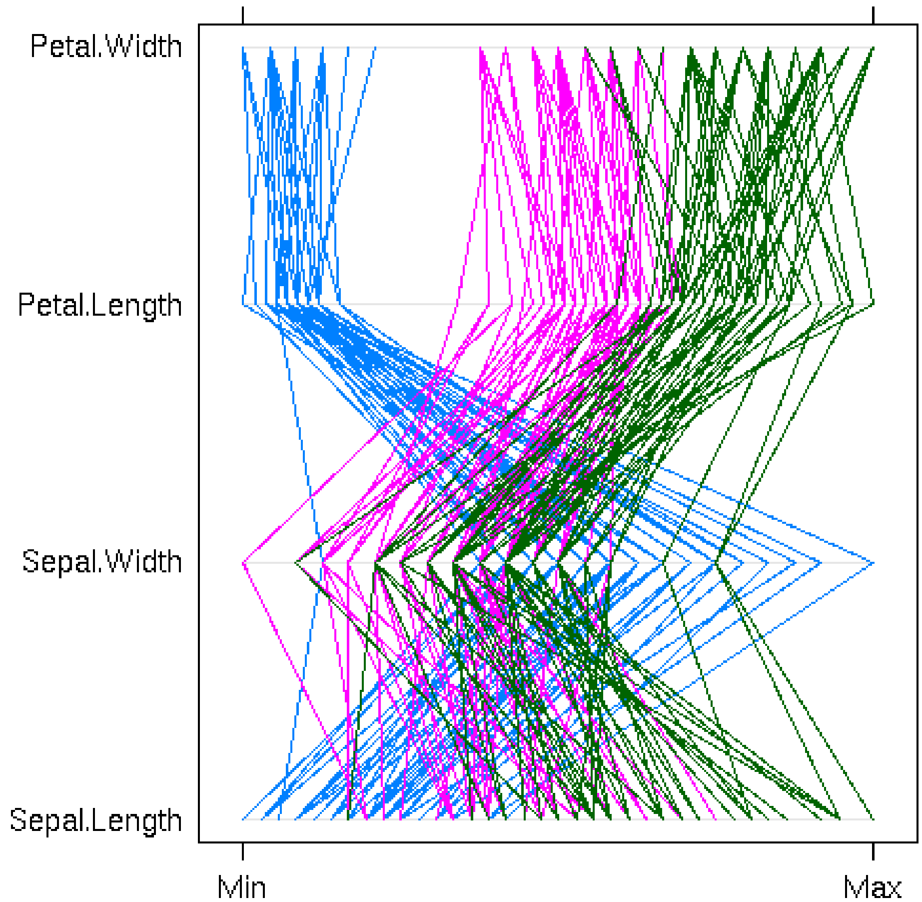


Figure 3.9: Parallel plot.

Examples of real application of parallel plots are in network attacks data (Choi, Lee, & Kim, 2009), land satellite data (Ge, Li, Lakhan, & Lucieer, 2009), climate analysis (Steed, Fitzpatrick, Jankun-Kelly, & Swan II, 2009), data envelopment analysis (Weber & Desai, 1996) and fuel injection simulation (Matkovic, Jelovic, Juric, Konyha, & Gracanin, 2005). The related tools for parallel plots usually include other types of visualization such as GGobi (Swayne, Lang, Buja, & Cook, 2003), ComVis (Matkovic, Jelovic, Juric, Konyha, & Gracanin, 2005) and Animator (Barlow & Stuart, 2004). Figure 3.9 shows an example of a parallel plot taken from Wilkinson & Wills (2005).

3.3 Visualization for multidimensional variables in parallel and perpendicular plots

In this § the comparison between two types of plots which are the parallel and the perpendicular plots is discussed when projecting multidimensional variables into 2 dimensional displays. A parallel plot can be categorized under the geometry projection type (Keim & Kriegel, 1996; Leban, Zupan, Vidmar, & Bratko, 2006; Ankerst, Keim, & Kriegel, 1996). The hyperbox (Alpern & Carter, 1991) and the hyperslice (van Wijk & van Liere, 1993) visualization are also under the geometry type. There are also many kinds of perpendicular plots (which use the Cartesian system) that had been developed for data analysis projecting multidimensional variables into 2 dimensional displays such as that of a pixel based type. The scatter plot and line charts are examples of pixel based visualization plots. The comparison between parallel coordinate and scatter plot is particularly chosen because of the principle of duality; the two techniques are somewhat interchangeable between each other and many literatures discuss and compare these two techniques (Xu, Hong, Li, & Song, 2007; Grinstein, Hoffman, & Pickett, 2002; Chou, Lin, & Yeh, 1999). This § will then continue with the discussion of the fundamentals of perpendicular and parallel coordinate plot. It is then followed by detail discussion on perpendicular plot and then parallel plot. Finally, the discussion is on the comparison of parallel coordinate and scatter plot matrix by Grinstein, Hoffman, & Pickett (2002).

3.3.1 Fundamentals of perpendicular and parallel plots

A point can be represented in the perpendicular plot in a Cartesian system as $p(x_1, y_1)$, with line *l* and distance *d* represented as the following equations:

$$\boldsymbol{l}: \boldsymbol{y} = \boldsymbol{mx} + \boldsymbol{b} \quad \textbf{– (3.1)}$$

$$\boldsymbol{d} = |\boldsymbol{m}\boldsymbol{x_1} - \boldsymbol{y_1} + \mathbf{b}| / \sqrt{\mathbf{1} + \boldsymbol{m}^{\mathbf{2}}} \,. \qquad \textbf{- (3.2)}$$

A simple representation of this concept can be illustrated, whereby the Cartesian system plane is represented as $\mathbb{R}^N$, where *N*=2 since it is 2-dimensional and the parallel plane $\mathbb{P}^2$ represents the $\mathbb{R}^2$ Cartesian system plane (Inselberg, 2009). A point in the $\mathbb{R}^2$ plane, p_r can be represented in the $\mathbb{P}^2$ plane as a line, ℓ_p. This is referred to as the principle of duality. Figure 3.10 (a) and (b) illustrate an example, point with (3,0) in $\mathbb{R}^2$ plane represented as a line connecting the corresponding points on the x and y axis drawn in parallel to each other. The position of the axis can be interchanged as shown in Figure 3.10(c).

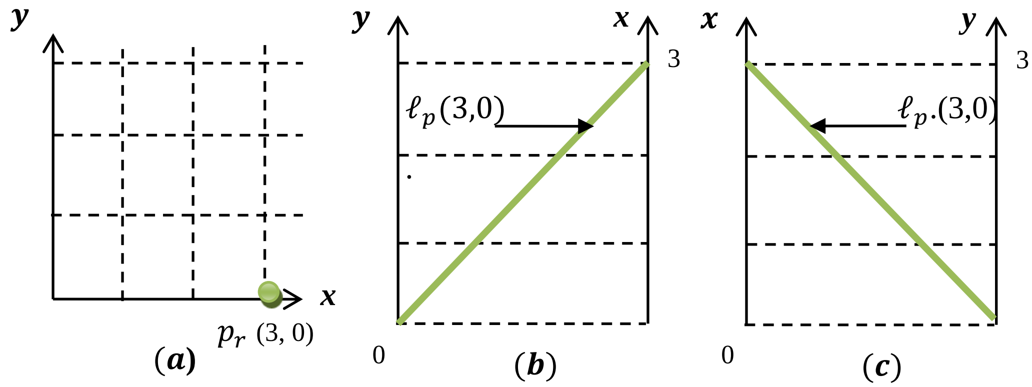


Figure 3.10: (a) a point (3,0) in the Cartesian $\mathbb{R}^2$ plane, (b) point (3,0) on the line $\boldsymbol{\ell_r}$ in the parallel ($\mathbb{P}^2$) plane, (c) interchangable axis position

Therefore, a line in the $\mathbb{R}^2$ plane, ℓ_r can be represented in the $\mathbb{P}^2$ plane as many lines, $\ell_{p(i-1),i}$ where *i* corresponds to a point in the line,

$$\boldsymbol{\ell_r}: \begin{cases} \boldsymbol{\ell_{p\,1,2}} : \boldsymbol{x_2 = m_2 x_1 + b_2} \\ \boldsymbol{\ell_{p\,2,3}} : \boldsymbol{x_3 = m_3 x_2 + b_3} \\ \qquad \ldots \\ \boldsymbol{\ell_{p\,N-1,N}} : \boldsymbol{x_N = m_N x_{N-1} + b_N} \end{cases} \quad \textbf{3.3)}$$

where *m* and *b* are values related to the slope and the point on the parallel lines. Figure 3.11 illustrates an example of the concept for a negative slope. Taking this further as a general point, the point of inter§ can be represented by

$$\boldsymbol{\bar{l}} : \left(\frac{\bar{d}}{1-m}, \frac{b}{1-m}\right), \boldsymbol{m \neq 1} \quad \textbf{- 3.4)}$$

where $\bar{d}$ is the distance between the two adjacent parallel axis and *m* is the slope of the line in the $\mathbb{R}^2$ plane (Inselberg, 2009). Figure 3.12 illustrates example of a positive slope.

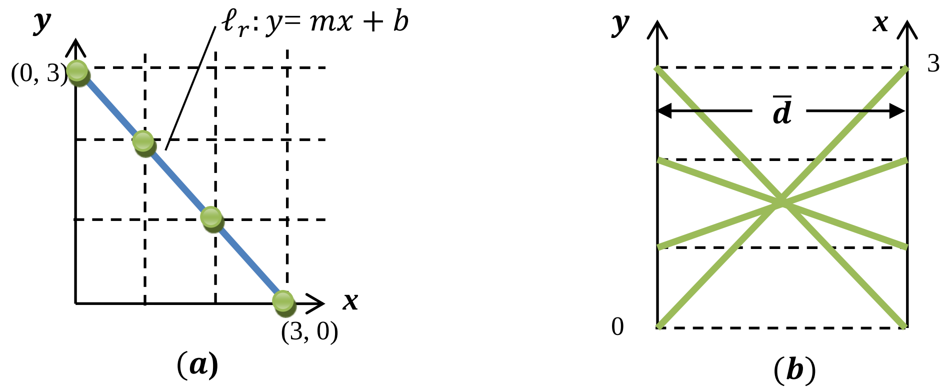


Figure 3.11: (a) negative slope line $\boldsymbol{\ell_r}$ in the Cartesian $\mathbb{R}^2$ plane, (b) line $\boldsymbol{\ell_r}$ in the parallel ($\mathbb{P}^2$) plane

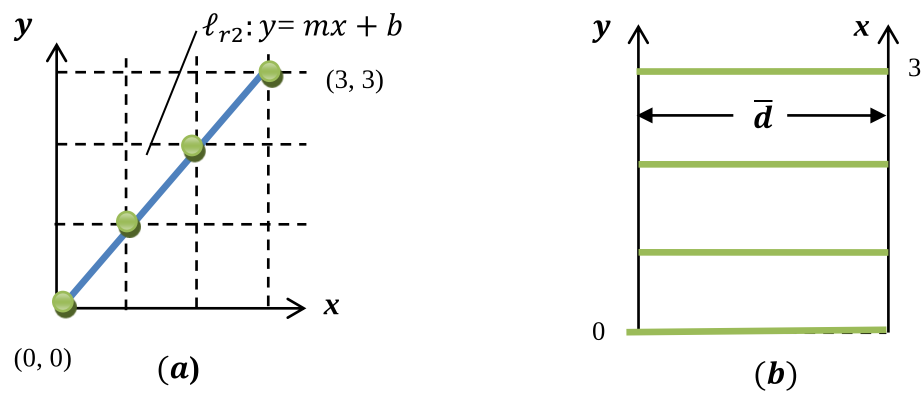


Figure 3.12: (a) positive slope line $\boldsymbol{\ell_{r2}}$ in the Cartesian $\mathbb{R}^2$ plane , (b) line $\boldsymbol{\ell_{r2}}$ in the parallel ($\mathbb{P}^2$) plane

When a third dimension is added in the $\mathbb{R}^2$ plane, a third line is added in the $\mathbb{P}^2$ plane as illustrated in Figure 3.13 (a) and (b) showing three dimensional Cartesian $\mathbb{R}^3$ plane and parallel $\mathbb{P}^3$ plane.

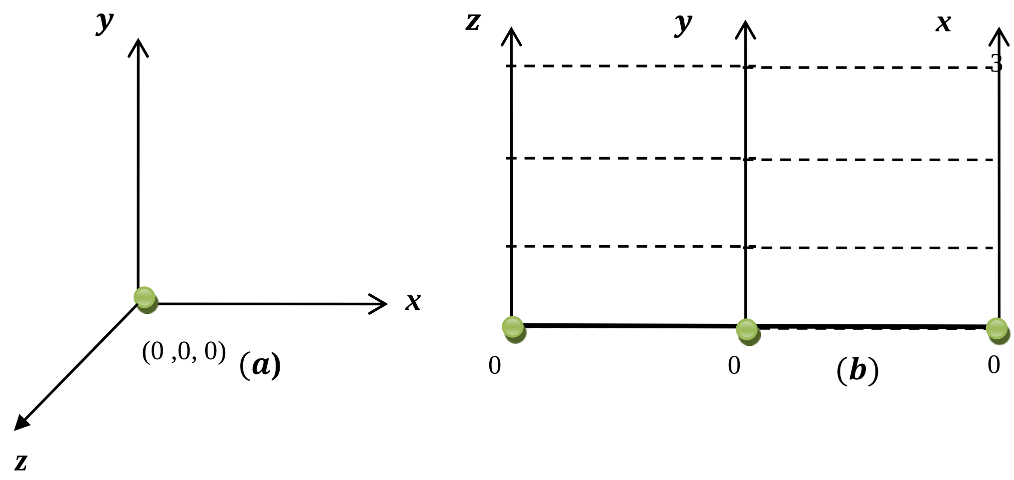


Figure 3.13: (a) 3 dimensional lines in the Cartesian $\mathbb{R}^3$ plane (b) 3 parallel lines represented in the parallel ($\mathbb{P}^3$) plane

Emphasis of the discussion in this § is therefore on the comparison between parallel coordinate (one type of geometry projection visualization technique) and pixel based plots under the family of scatter plot and line charts.

3.3.2 Perpendicular plots

Perpendicular plots are plots that use the x and y coordinates in a Cartesian system. As mentioned before, scatter plots and line charts are examples of perpendicular plots. These plots are usually used to represent two dimensional variables although multidimensional variable are also possible to be displayed by using techniques such as overlapping graphs (Leban, Zupan, Vidmar, & Bratko, 2006), increasing data display density for display space efficiency (Heer, Kong, & Agrawala, 2009) and small multiples (Heer, Kong, & Agrawala, 2009) techniques. These techniques give rise to new type of charts such as stacked graphs, horizon graphs (Heer, Kong, & Agrawala, 2009) and vizrank (Leban, Zupan, Vidmar, & Bratko, 2006).

Interpreting a scatter plot would involve identifying correlations, functional dependencies, clusters and other interesting relations (Keim & Kriegel, 1996; Yuan, Guo, Xiao, Zhou, & Qu, 2009). Sometimes these could be detected by looking at direction, forms, strength and outliers of the plot. For example, if the scatter plot forms a line with negative or positive slope, then a negative or positive relation could be implied between the attributes involved. These can be used for prediction. However, if the dimensionality increases, the number of different plots to be inspected increases by $m\ (m\text{-}1)/2$, where m is the number of attributes in the set of data (Leban, Zupan, Vidmar, & Bratko, 2006; Ahlberg, 1996). The data can be differentiated via colour, symbol or shape. Values of attributes are represented as points and only the position is effected within the perpendicular x, y axis and not its size, shape and colour (Leban, Zupan, Vidmar, & Bratko, 2006).

The disadvantage of the scatter plot is that only limited amount of data items can be visualized on the screen (Ankerst, Keim, & Kriegel, 1996). Analysis of the scatter plots for example can be difficult and time consuming since the x and y projection increases exponentially with the number of attributes and tend to have $O(N^2)$ complexity (Inselberg, 2009; Leban, Zupan, Vidmar, & Bratko, 2006). In a larger dataset the points can overlap and not perceived (Heer, Kong, & Agrawala, 2009; Leban, Zupan, Vidmar, & Bratko, 2006; Ankerst, Keim, & Kriegel, 1996). Bertini & Santucci (2005) estimated that a human cannot distinguish between two points closer than an area of 14 X 14 virtual pixels. This particular problem is called the occlusion problem, where occlusion is the number of identifiable points in relation to other visible points (Bertini and Santucci, 2006). In addition, sometimes data need to be compared across multiple charts (Heer, Kong, & Agrawala, 2009; Chou, Lin, & Yeh, 1999) which is difficult if the data set is large.

The advantage is that these plots already exist for such a long time and are popular (Yuan, Guo, Xiao, Zhou, & Qu, 2009) and many know how to interpret them easily. The perceptibility of the data characteristic is therefore considered reliable. Heer, Kong, & Agrawala (2009) referred this as the graphical perception.

There are several operations used with scatter plots especially to reduce cluttering and to enhance the occlusion percentage. Jittering and distortion can be used to promote occlusion (Bertini & Santucci, 2005). While jittering displaces the overlapped items around their original position so that they are visible, distortion presents more detail in less dense area and less detail in more dense area. Other operations involved are zooming, panning, filtering and selection detail-on-demand, brushing (highlighting) and 3D manipulation (Ahlberg, 1996).

3.3.3 Parallel plots

The parallel plot however, refers to variables that are multidimensional and these are arranged onto parallel axes equidistant from each other (Inselberg, 2009; Keim and Kriegel, 1996; Chou, Lin, & Yeh, 1999). Instead of representing attributes as points, it is represented by m-1 connected line segments (Leban, Zupan, Vidmar, &

Bratko, 2006) as also shown in § 3.3.1. There are several types of parallel coordinate plot such as the extended parallel coordinate (Hauser, Ledermann, & Doleisch, 2002), the extruded parallel coordinate (Rubik, 1997), scattering points in parallel coordinate (Yuan, Guo, Xiao, Zhou, & Qu, 2009) and 3-D parallel coordinate (Johansson, Forsell, Lind, & Cooper, 2008). All these different types of parallel coordinate are created as an attempt to solve some of the display problems related to the plot such as occlusion, identifying patterns relating to correlation and line cluttering display.

Interpreting parallel plots is also possible by looking at certain kinds of patterns. A positive correlation between two attributes shows horizontal parallel lines with minimum or no inter§ of lines between two adjacent vertical axes of the attributes. In contrast a negative one shows inter§ of lines on a point in between two vertical axes of the attribute concerned (Inselberg, 2009; Ledermann, 2002). Cluster can also be identified by their edges within any pair of adjacent axes being closed to a point which corresponds to a two-dimensional projected line in the Cartesian system (Chou, Lin, & Yeh, 1999; Artero, Oliveira, & Levkowitz, 2004). This can be seen since a point in the Cartesian system is represented by a line joined between two adjacent axes in the parallel plot. Figure 3.10 b) illustrates this.

The main disadvantage of parallel plots is that the display is considered cluttered and hard to interpret if there are too many dimensions and set of attributes to be shown but this would depend on the size of the monitor displaying the data (Leban, Zupan, Vidmar, & Bratko, 2006; Yuan, Guo, Xiao, Zhou, & Qu, 2009; Forsell & Johansson, 2007; Siirtola & Räihä, 2006). This problem is related to the occlusion problem which had been mentioned earlier in the discussion on perpendicular plot. Yuan, Guo, Xiao, Zhou, & Qu (2009) and Henley, Hagen, & Bergeron (2007) pointed out that interpretation of parallel coordinates is harder and requires expert knowledge although there is also evidence showing that parallel plots are not difficult to manipulate and can result in a more accurate interpretation of data even by novice users (Siirtola & Räihä, 2006; Siirtola, Laivo, Heimonen, & Raiha, 2009; Henley, Hagen, & Bergeron, 2007). Also, parallel coordinate is not suitable for visualizing large data sets which are more than the range of hundreds to thousands and the relationships between attributes can only be easily identified for two adjacent axes (Johansson, Forsell, Lind, & Cooper, 2008; Novotny & Hauser, 2006). Rearranging of the axes can overcome the mentioned problem.

The great advantage of these plots is that it is considered as more effectively consuming the display space and therefore increasing the amount of data a human can work with at the same time (Heer, Kong, & Agrawala, 2009). However, Johansson, Forsell, Lind, & Cooper (2008) found that 11 axes is the maximum threshold for a human to efficiently analyze the data. Also, N-dimensional properties can be visible from the screen (Inselberg, 2009).

The operations related to parallel plots are very similar to the scatter plots i.e. zooming, panning, filtering and selection detail-on-demand, focus+context, linking+brushing (Yuan, Guo, Xiao, Zhou, & Qu, 2009; Matkovic, Jelovic, Juric, Konyha, & Gracanin, 2005; Hauser, Ledermann, & Doleisch, 2002) and 3D manipulation. There is also operation such as rearranging the axes in order to permit analysis between two adjacent axes (Forsell & Johansson, 2007; Yuan, Guo, Xiao, Zhou, & Qu, 2009).

3.3.4 Comparative discussion

From the previous discussion on parallel and perpendicular plots, we can summarize the features of these two types of plots. Table 3.1 shows the summary of comparison. Both plots have similar features in various aspects apart from a few exceptions such as:

- the optimization of space (feature 4): an advantage for parallel plot,
- the perceptibility (feature 5): an advantage for perpendicular plot

Nevertheless, the advantages of each plot are more or less equal to each other which can lead to a design implication. In choosing either technique, one would end up having similar outcomes in terms of effectiveness and efficiency of the system. The usage context of the plots might just determine which of the plots to use. For example even though it was found by Siirtola, Laivo, Heimonen, & Raiha (2009) and Henley, Hagen, & Bergeron (2007) that the parallel plot resulted in a more accurate interpretation of data, Li, Martens, & Wijk (2008) found that for determining correlation, the scatter plot was superior.

Table 3.1: Summary of comparison of features in parallel and perpendicular plots

No	Features	Parallel plot	Perpendicular plot
1	Visualize multidimensional attributes	yes	yes
2	Attribute values representation	Lines joining parallel axes	Points within perpendicular axis
3	Differentiating attribute set	Colour	Colour, symbols and shapes
4	No of charts generated	1	m(m-1)/2, where m= number of attributes
5	Graphical Perception / perceptibility	Not good	Good
6	Targeted users	Novice to experts	Novice to experts
7	Pattern interpretation	Correlation, cluster	Correlation, cluster
8	Application	Wide	Wide
9	Type of visualization	Geometrical based	Pixel based
10	Data Cluttering	Avoidable via shading and brushing	Avoidable via multiple charts
11	Evidence of usability performance	yes	yes

Grinstein, Hoffman, & Pickett (2002) had also discussed in their lengthy article on setting up a standard benchmark for the evaluation of visualization for data mining. In doing so, five visualization techniques were compared using ten data sets (Simple Seven, Balloons, Contact Lenses, Shuttle O-rings, Monks problem, Iris Flower, Congress Voting, Liver Disorders, Cars and Wines) which can be accessed online. The five techniques compared are parallel coordinate, scatter plot matrix, survey plot, circle segments and radviz.

They proposed that a few criteria should be addressed for testing the visualization techniques: scalability related to the processing time of the data, ease of integration of domain knowledge, ease of classification, dealing with incorrect data, high dimensionality, query functionality and summarization of results. However, their final comparison was based on the features illustrated by the data sets and how these can be efficiently visualized using the five different visualization techniques.

For the purpose of this research, the focus of discussion is on parallel coordinate and scatter plot matrix. Table 3.2 illustrates the conclusion by Grinstein, Hoffman, & Pickett (2002) on scatter plot and parallel plot visualization techniques. Yp stands for yes in parallel coordinate view and Ys stands for yes in scatter plot matrix. The unfilled space means that they are not applicable.

Table 3.2: Scatter plot and parallel coordinate comparison

No.	TASK **DATA SET**	See Outliers	See Clusters	Find Class Cluster	See All Important Features	See Some Important Features	See Possible Rule/ Model	See Exact Rule/ Model
1.	Simple Seven							
2.	Balloons							
3.	Contact Lenses							
4.	Shuttle O-rings	Ys Yp	Ys Yp	Ys Yp	Ys Yp	Ys Yp	Ys Yp	
5.	Monks problems		Yp		Yp	Yp	Yp	Yp
6.	Iris Flower	Ys Yp	Ys Yp	Ys Yp	Yp	Ys Yp	Ys Yp	
7.	Congress Voting							
8.	Liver Disorder	Ys Yp	Ys Yp					
9.	Cars	Ys Yp	Ys Yp	Ys Yp		Ys Yp	Ys Yp	
10.	Wine	Ys Yp	Ys Yp	Ys Yp		Ys Yp	Ys Yp	

Grinstein, Hoffman, & Pickett (2002).

In Table 3.2 parallel coordinate and scatter plot matrix are shown to be relatively at par to each other in context of showing the appropriate tasks. However, parallel coordinate can be used to see all important features in the Iris Flower data set while it is not possible in the scatter plot matrix. For ‘monks problems’ data, parallel coordinate is used and not scatter plot. While parallel coordinate covers the entire tasks listed above stating from seeing outliers to seeing exact rule/model (7 tasks overall), scatter plot matrix only covers 6 tasks within the tested data set.

3.4 Summary

In this chapter, the discussion on word visualization shows that words relate to illustration of words in the form of drawing and sketches to show word meaning and word structure. The next discussion was on information visualization based on quantity and connectivity. Under the aspects of quantity the scatter plots and bar charts visualization had been used to display data related to frequency values; while under the aspect of connectivity, the hyperbolic, distorted views and network display had been used to display positional information. The combination of quantity and connectivity aspects is also found in the galaxy and parallel plots. Information visualization for multidimensional data was also discussed based on parallel and perpendicular plots. Both scatter plot and parallel plot can be used to show most of the tasks such as showing outliers, see clusters, see important features and others with slightly more usage of parallel plot.

4.0 RESEARCH METHODOLOGY

The research methodology adopted for this study is not straight forward since many research were conducted in order to solve reading comprehension in Qur'anic verses. This chapter attempts to reveal the most suitable approach for Qur'anic verses comprehension focusing on word recognition.

There are three research phases involved in this study. First, the identification of techniques to aid Qur'anic verses comprehension focusing on word recognition. Second, the design and implementation phase for the purpose of evaluation of the identified techniques. Third, the phase for the evaluation of the techniques designed and implemented. Figure 1.2 shows the summary of the methodology.

4.1 The identification of techniques

The identification phase consists of literature investigation followed with investigating theories of reading comprehension and information theory through preliminary studies. From the literature review, reading comprehension and word recognition theories reveal that the processes needed to achieve reading comprehension are non-trivial, therefore word recognition which is an important part of word comprehension is the focus of this study. Two models of word recognition were adopted, the Word Superiority Effect and the Word Frequency Effect. Literature investigation also reveals that information visualization of word that may include vocabulary helps to achieve word recognition. There were several available visualization that had to be chosen and the criteria was to be able to display multidimensional variables.

Next is the investigation of theory through the preliminary studies. There were 3 studies conducted:

i) Study 1 is to find out the problems related to recitation of *Qur'an*

- The focus was on recitation because recitation is directly related to word recognition i.e. to be able to decode the orthographical form to the phonological form.
- As stated before, the Qur'an is written in the Arabic language and therefore the assumption is that an Arab speaker would have lesser problems of recitation compared to a non-Arab speaker. It is then worthwhile to compare the problems of recitation and the teaching and learning methods between the two above. The reason is for improvement in recitation quality for at least to the non-Arabic speakers.
- The cross-§al survey technique was chosen to conduct the above investigation so that the problems can be identified for teaching and learning methods for the population under the category of Arabic and non-Arabic speakers.
-

ii) Study 2 is to find out whether images used in Qur'anic verses have any significant effect on reading comprehension

- Abdelbaki, (1983) approach is slightly different from this study, although it is almost testing the hieroglyphic method of the ancient Egyptian discussed in Chapter 2.
- This study was to invesitigate the possible solution for Word Superiority Effect.

iii) Study 3 is on the contextual comparison of parallel and perpendicular plots

- These two plots are chosen in relation to Word Frequency Effect and the ability to represent multidimensional variables which involve quantity and connectivity, such as word count, word and word position. A brief analysis was also outlined to predict display representation issues.
- A relational concept was developed to ease the analysis of tasks using GOMS and later on use to describe the algorithm for implementing the PaCQ interface.
- Goals, Operators, Methods and Selection Rules (GOMS) analysis was chosen and therefore no participants are involved. GOMS is a task analysis technique for analysing tasks in terms of the time taken to finish a certain goal of task (John & Kieras, 1996). A high level goal task must be decompose into smaller subtasks until it reaches the smallest operator typed tasks such as clicking the mouse button, key-press, mental preparation and so on such as shown in Table 4.1. GOMS was chosen because it can accurately predict the time taken to finish certain computer interaction related tasks. The 'Goals' are the highest level task, 'Operators' are the lowest level task such as in table 4.1, 'Methods' refer to the way a goal/task is accomplish and there could be more one way for it to be accomplish. Such as the copy and paste task, that can be done either using the menu option in the tab or by pressing the 'CTRL' and 'V'/'X' keys in Microsoft Word. 'Selection Rules' refer to the 'Method' chosen depends on certain situation.

  The goals and scenario of tasks chosen are specific to the parallel and the perpendicular plot for finding information related to the variables involve for word count, word, word position and percentage. The time taken to accomplish any related goal to find related information in the parallel and the perpendicular plot can be used to indicate the effectiveness of the technique in accomplishing related tasks. The lesser the time taken, the more efficient the predicted task should be.

Table 4.1: GOMS estimated operation duration

Operation	Duration(sec)
Click-mouse-button	0.20
Shift-click-mousebutton	0.48
Cursor movement	1.10
Mental Preparation	1.35
Determine Position	1.20
Key-press	0.28

Source (John & Kieras, 1996)

Study 1 was conducted to investigate problems of recitation, while study 2 was to explore the possible solution to recitation without comprehension which was found in study 1. The result found in study 2 was not robust enough to be used as a solution hence the study 3 was conducted to explore possible solution for attraction of Sensory Memory.

4.2 Design and Implementation phase

The next phase was designing a hypothesized model solution. There were several steps taken to develop a hypothesized model:

- Decomposing the problem of reading comprehension focusing on word recognition based on literature review and preliminary studies

- Dealing with counting and stemming issues of Qur'anic words

Based on this model the PaCQ interface was developed with the embedded theoretical concepts. The hypothesis is that visualization with vocabulary tracking support word recognition in Qur'anic verses.

4.3 Evaluation of techniques phase

There are a few steps taken for the evaluation:

- The Experimental study was to test whether the PaCQ system has any significant effect in the word recognition performance.
    - First, the experimental study starts with a pre-test to obtain stratified participants which were then divided into the control and experimental groups. This is to make sure that there is almost equal number of participants with good and poor scores in the control and experimental group.
    - Second, the post-test was conducted to evaluate the word recognition performance via the Arabic Word Recognition Test (AWRT).
    - Third, the user satisfaction test was also conducted to find out the usability issues of PaCQ interface
    - Fourth, the test of parallel plot interpretability, conducted to confirm the third preliminary study on the effectiveness of the parallel plot in finding word related information. This information is crucial since the interactivity of users with this plot is hypothesized to help toward achieving reading comprehension.

- Finally interview sessions were conducted after modification of the PaCQ interface to confirm why errors occur while the participants interact with the PaCQ interface and to confirm whether PaCQ interface is a software tool useful for

4.4 Summary

The methodology adopted in this study consist of 3 phases; the identification of techniques, design and implementation and evaluation of techniques. In the first phase, the literature review combined with preliminary studies helped in finding the solution to the problem of word recognition. The second phase was on design and implementation. In this phase the proposed model of the hypothesized solution was designed, followed with the identification of the requirements of the PaCQ interface and then the implementation. The last phase was on the evaluation of the PaCQ interface to confirm whether the hypothesis of visualization with vocabulary tracking support word recognition of Qur'anic verse.

5.0 PRELIMINARY STUDIES

There are three preliminary studies done in relation to *Qur'an* recitation and visualization:

i. Study 1 is to find out the problems related to recitation of *Qur'an*

ii. Study 2 is to find out whether images used in Qur'anic verses have any significant effect on reading comprehension

iii. Study 3 is on the contextual comparison of parallel and perpendicular plots

5.1 Preliminary study 1: Problems related to recitation of Qur'an

The objectives

The objectives of the survey are listed as follows:

1. To identify the problems in reciting the *Qur'an* for non-Arabic speakers (Malays)
2. To identify the problems in reciting the *Qur'an* for Arabic speakers (Malays and Africans)
3. To identify methods of teaching/learning how to recite the *Qur'an*

The participants

A survey was conducted on 45 readers/reciters of the Qur'an to both Arabic and non-Arabic speakers consisting of 33 females and 12 males. The nationalities from 10 countries were involved – Malaysia, Indonesia, Singapore, Nigeria, Somalia, Kenya, Eritrea, Djibouti, Sudan and Tanzania. There were 25 Arabic speakers and 20 non-Arabic speakers. The non-Arabic speakers were mostly Malaysian and the rest were the Arabic speakers even though their actual native language may not be Arabic. Sudan is the only country with

Arabic as the native language. There were 4 respondents with age of lesser or equal to 20 years, 20 respondents in the age category of lesser or equal to 30 years, 14 respondents for age lesser and equal to 40 years and 6 respondents for age more than 40 years. One respondent did not state his age.

Table **5.1** shows the frequency of respondents according to age.

Table 5.1: Frequency of respondents according to age

Age	Frequency
<=20	4
<=30 years	20
<=40 years	14
>40 years	6
TOTAL	45

The survey questions

Appendix B1- B9 is the sample of the survey questions in three languages (Malay, English and Arabic). The respondents were asked questions related to where and how they acquire the knowledge of Qur'anic recitation and the problems encountered by them while recitation of the Qur'an.

Data analysis

The survey questions were comparatively analysed through these parameters:

- Age

- Arabic and non-Arabic speakers
- *Qur'an* recitation learning method (general and specific):
    - General method refers to passive acquirement of the knowledge of *Qur'anic* recitation
    - Specific method refers to active acquirement of the knowledge of *Qur'anic* recitation
- Problems encountered by respondents

The results

There are many problems faced by Qur'an reciters found in this survey, in relation to: fluency, searching verses (*ayah*) and chapters(*sura*), remembering Arabic letters, remembering rules of recitations (*Tajwid* rules), comprehending the words in the Qur'an, focusing while reciting the Qur'an and finding suitable time in reciting the Qur'an. The two main highest rated problem are understanding the *Qur'anic ayat* and remembering *Tajwid* rules. It is found that:

- there is significant difference between the Arabic (M=0.56 , SD=0.507) and the non-Arabic(M= 0.9, SD=0.308), $t(43)$=2.634 , $p<0.05$ (p=0.012) speakers in understanding the *Qur'anic ayat* .
- there is significant difference between the Arabic (M=0.36 , SD=0.90) and the non-Arabic(M= 0.75, SD=0.444), $t(43)$=2.764 , $p<0.05$ (p=0.008) speakers in understanding the *Tajwid rules*.
- However, there is no significant difference for other types of problems including the total problems faced by respondents below and above 30 years of age.

See details in Appendix C1- C14.

Discussion

This shows that there exists the problem of understanding *Qur'anic ayat* for non-Arabic speakers. This may be related to the fact that the respondents mainly learned *Qur'an* in a one teacher to a many students environment. Teachers may not have the time to go through all the lessons and time may have been devoted on the *Tajwid* rulings compared to word meaning. Even so, many only learned parts of the *Tajwid* rules.

It is also found that the respondents went through a certain learning process in Qur'an recitation lessons, which are: learn to recognize and pronounce Arabic letters, associate symbols and pronunciations of the letters in verses of *Qur'an*, recognize the Arabic letters in different locations of a word (e.g. How ب looks like if it is in the beginning, in the middle and in the end of a word), recite *sura Al-Fatihah* (the first chapter of the *Qur'an)* first before the thirtieth *juz'* (*Juz Amma*), memorize all recited *sura* (chapter), learn a portion of the rules of recitation (*Tajwid*), learn all of the rules of recitation(*Tajwid*), learn to recite *Qur'an* with melody and learn the meaning of the words in the *Qur'an*. The highest percentage is for learning to recognize and pronounce letters and the lowest percentage is learning the word meaning. Also, we can conclude that many respondents are more familiar with the *Juz Amma*.

5.2 Preliminary study 2: Test for text visualization in Qur'anic verses

Objective

The preliminary study 2 presents an experimentation to investigate the effectiveness of text visualization. Particularly, it is envisaged to solve the problem of users that can read in Arabic but cannot comprehend the meaning. The main objective of this experiment is to test whether there is any significant difference on the level of comprehension when images are used as part of the reading of *Qur'anic* verses.

Materials

The materials used in the experiment are the *Qur'anic* verses and their translation selected from the Chapter 114 or *Sura Al-Nas* of the *Qur'an*. The *Sura* consists of six *ayat* or verses. To achieve the objective, randomly selected readers were asked to read one of the following:

- *Sura Al-Nas* (114), *ayah* 1-6 tabulated with all the translation in text form (I_i), (See Table C.3 in Appendix C)
- *Sura Al-Nas* (114), *ayah* 1-6 tabulated with the translation in text form for the most frequently used words (I_{ii}), (See Table C.4 in Appendix C)
- *Sura Al-Nas* (114), *ayah* 1-6 tabulated with the translation in visual form for the most frequently used words (I_{iii}). Preposition/conjunction words are included and they are left in the form of text since it is quite impossible to represent visual image of the words, (See Table C.5 in Appendix C)
- *Sura Al-Nas* (114), *ayah* 1-6 only in the Arabic language, (I_c) (See Figure C.4 in Appendix C)

Participants

Participants were among those who can read the Qur'an but could not speak the Arabic language. The number of participants involved was 46 for instrument Ii, 36 for Iii, 27 for Iiii and 39 for Ic totalling to 148. There were 7 participants with highest education from the primary school, 20 with secondary school, 35 with Diploma, 67 with Bachelor, 13 with Master, 2 with PhD and 4 unknown. Most of the participants were Malaysian, only 1 from Yemen and Iran and 14 unknown nationalities. Among these participants 24 were in the age category 10-20 year, 94 in 21-30years, 8 in 31-40 years, 5 in 41-50 years and 51-60 year, 1 in 61-70 years and 11 unknown. There were 75 females and 72 males and 1 unknown.

The participants were asked to read the instruments I_i, I_{ii}, I_{iii} or I_c after which they had to answer a few questions. However, they were not given the instruments to refer to while answering the question. Appendix C13 shows the sample of the questions.

Procedure

At the start of the experiment, the participants were asked to answer the demographic questions one their age, gender, language spoken, experience, education level. They were then asked to read as fast as possible the instrument given to them. They were also advice to try and understand the words while reading. After all these had been completed they were asked to answer the questions.

The questions are open ended. If the participants could not answer the questions, they were asked to leave it blank. The questions are as the following:

(i) What are the activities described in the *Sura*? List them out.

(ii) Who are involved? List them out with some description (if possible).

The answers to the questions should be as follows:

(i) The activities involve are:

a) Say

b) Seek refuge

c) Whisper into the heart

d) Withdraws after whisper

(ii) Characters involve are:

a) Allah (the Lord and Cherisher, the King of ManKind)

b) Mankind

c) The Whisperer/ The Devil (Whispers evil into the heart)

d) *Jinns*

Result and discussion

Yaxley and Zwaan (2007) found that readers mentally simulate the visibility of an object during language comprehension and thus linguistic simulation of the object properties is one of the ways that could help the reader to comprehend. Similar work by others such as Zwaan, Madden, Yaxley, & Aveyard (2004), Gosselin & Schyns (2004) and Richardson, Spivey, Barsalou, & McRae (2003) also support this evidence. A related work such as by Harber and Myers (1982) found that there is greater accuracy in remembering pictograms compared to words. These theories suggest that images may be used as a basis to solve the problems encountered by readers of the Arabic document (the *Qur'an*, especially those who can read in Arabic but could not comprehend the meaning).

For the case of I_i, there is evidence that using Arabic text and translation leads to no significant difference from the expected results (See Appendix C8, Table C.6). For the cases of I_{ii} and I_{iii}, it was found that there is significant difference on the level of comprehension when images are used as part of the reading text resulting in higher comprehension level (See Appendix C8-10 for details) and this is consistent with what was found by Yaxley and Zwaan (2007), Zwaan, Madden, Yaxley, & Aveyard (2004), Gosselin & Schyns (2004) and Richardson, Spivey, Barsalou, & McRae (2003) although the experiment done by them was different in approach and context.

5.3 Preliminary Study 3: Contextual comparison of parallel and perpendicular plot

In this § the contextual comparison of parallel and scatter plot is discussed to investigate the effectiveness of these two plots on finding information related to *Juz Amma* especially in relation to time. GOMS (John & Kieras, 1996) analysis is used for this purpose and *Juz Amma* is chosen because it is one of the most frequently read parts of the Qur'an based on the preliminary study 1. It is also chosen because in this 30th part of the Qur'an there are 37 *sura*s and this can represent the whole Qur'an which contains 114 *sura*s. Before comparing the plots a relational concept is outlined for ease of discussion in later parts of this §.

5.3.1 The relational concept

Multidimensional variables can be displayed in both the parallel and perpendicular plots. In the perpendicular Cartesian plot when the variables increase, multiple charts can be used or different symbols of points can be shown in the same plot. In the parallel plot however, the parallel lines will increase to *N-1* where *N* is the maximum number of variables. Let p_r be a point that is representing a binary relationship between two variables v_{i-1} and v_i in the Cartesian plane. If there are N variables, the set of binary relationship R that can be plotted are denoted by *R*= $\{R_{v1v2}, R_{v2v3}, R_{v1v3}. \ . \ . \ . \ , R_{(1-(N(N-1)/2)(N(N-1)/2)}\}$ where there are *N* (*N*-1)/2 relationships. The corresponding plot in the parallel plane contains N axes line relation plotted in parallel denoted by *P*= $\{P_{v1}, P_{v2}, P_{v3} .... , P_N\}$ where the lines $l_{pv(i-1)v}$ are the points representations in the parallel plot. For each relation in *R*, there is a set of points with value v_{i-1} and v_i, $R_{vi(vi-1)}$= $\{(v_{i-1}, v_i), (v_i, v_{i+1}), (v_{i+1}, v_{i+2}), . . . (v_{k-1}, v_k)\}$ where k is the maximum number of data set and *v* and *i* are arbitrary numbers.

In this study, there are several variables sets:

- The Arabic word, *A*, *where* $A = \{A1 \cup A2\}$
  - $A1 \in all\ Arabic\ words\ in\ Sura1$
  - $A2 \in all\ Arabic\ words\ in\ Juz\ Amma$
- The Vocabulary words, V, where V⊆ A1 U A2
  - Note: the vocabulary for a user in the scope of the study is limited to only the Arabic words in *JuzAmma*.
- The word count, C where C = {c: c ∈ ℕ } and ℕ is the set of natural numbers.

- The corresponding *Sura*, S where S = {s: s $\in \mathbb{N}$, s=1 and78 $\leq$s $\leq$ 114}.
    - Note: *Suras* in *JuzAmma* are from *Sura* number 78 to 114 and the *Sura al-Fatihah* is *sura* 1.
- The corresponding Ayah, Y, where Y = {y: y $\in \mathbb{N}$, y $\leq$ 46}.
    - Note: the maximum number of *Ayah* / verse in *JuzAmma* is 46.
- The percentage of Arabic word, T, where T = { t: t $\in \mathbb{N}$, t $\leq$ 100}

The main relations between these variables can be written as:

- R_1 = {(c, a, s, y): There are c counts of Arabic words in *JuzAmma* and can be found in *Sura* s and *Ayah* y} where c $\in$ C, a$\in$ A, s $\in$ S and y $\in$ Y.
- R_2 = {(v, s, t): There is T percentage of the vocabulary words V in *Sura* S} where v $\in$ V, s$\in$ S, t$\in$ T.
- R_3 = { (c, s_1, s, t): There is T percentages of Arabic words in *Sura* S_1contain in other *Suras* S in *JuzAmma* and C count of Arabic words in the *Sura* S} where {s_1: P(s_1) [8]} where P(s_1) is exactly one *Sura* in *Juz Amma* and s$\in$ S, t$\in$ T.

5.3.2 Predictions and estimations

***(a)* The attribute:**

[8] P(s) is a property defined as stated above

Inspecting the relations R_1 , R_2 and R_3 above there are at most four attributes involved at this point. Therefore, in a parallel plot, the maximum line plot is P = 4 while in the perpendicular plot, the maximum relation to be plotted is R = 6, which means plotting 6 charts.

(b) The Arabic words (A):

There are about 3000 words in *Juz Amma* but the base words (without repetition words) are only around 800 words (See § 5.1.5). Therefore there are 800 lines to be plotted in the parallel plot while in the perpendicular plot there are 800 points to be plotted in maximum of 6 charts. The corresponding plots for both types of views are illustrated based on values in Table 5.2. Figure 5.1 (word-count, word-*sura*, word-*ayah* no.,count-*ayah no.*, *sura*-count, *ayah no.-sura* graphs), Figure 5.2 and Figure 5.3 are the related parallel and perpendicular plots.

Table 5.2: Example of Arabic word information in *Juz Amma*

Arabic word	count	Sura	Ayat	WordNo.
واد	2	79	16	4
واد	2	79	9	6
وتر	1	89	3	4
وجوه	5	80	40	2
وجوه	5	80	38	1
وجوه	5	88	8	1
وجوه	5	88	2	1
وجوه	5	83	24	3
وحوش	1	81	5	3

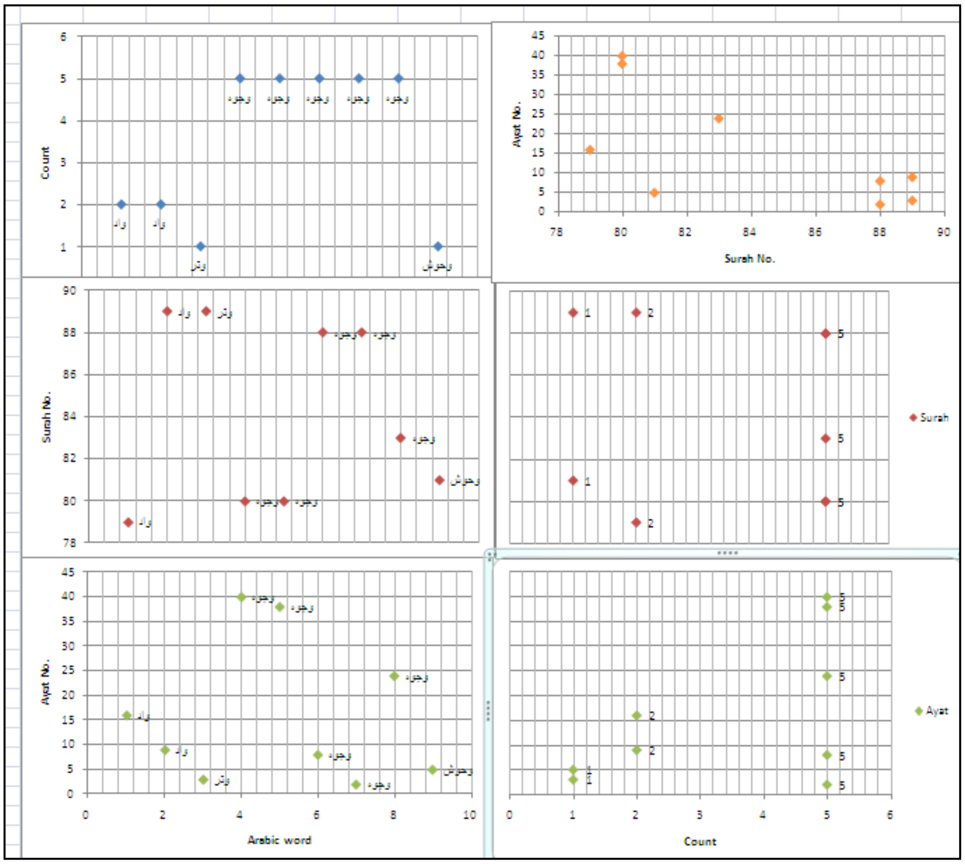


Figure 5.1: The corresponding perpendicular plot (scatter plot matrix) for Table 4.2.

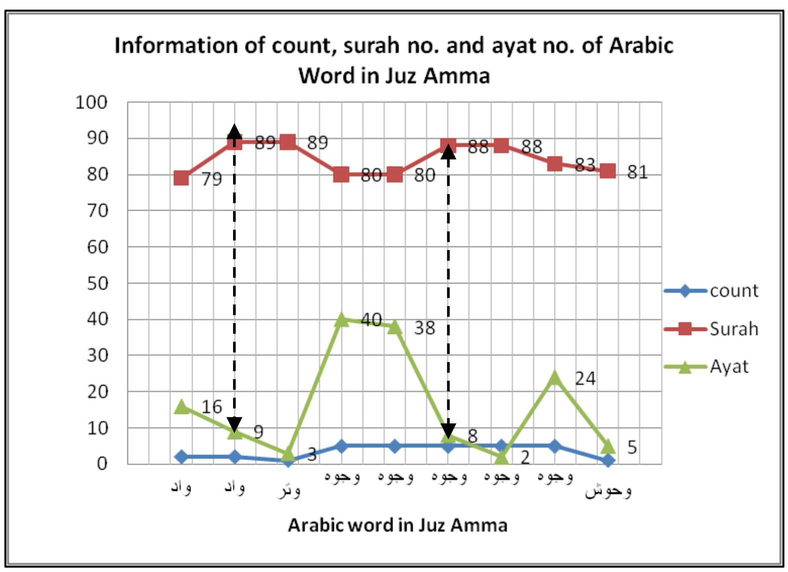


Figure 5.2: The simplified view of the perpendicular plot (scatter plot matrix)

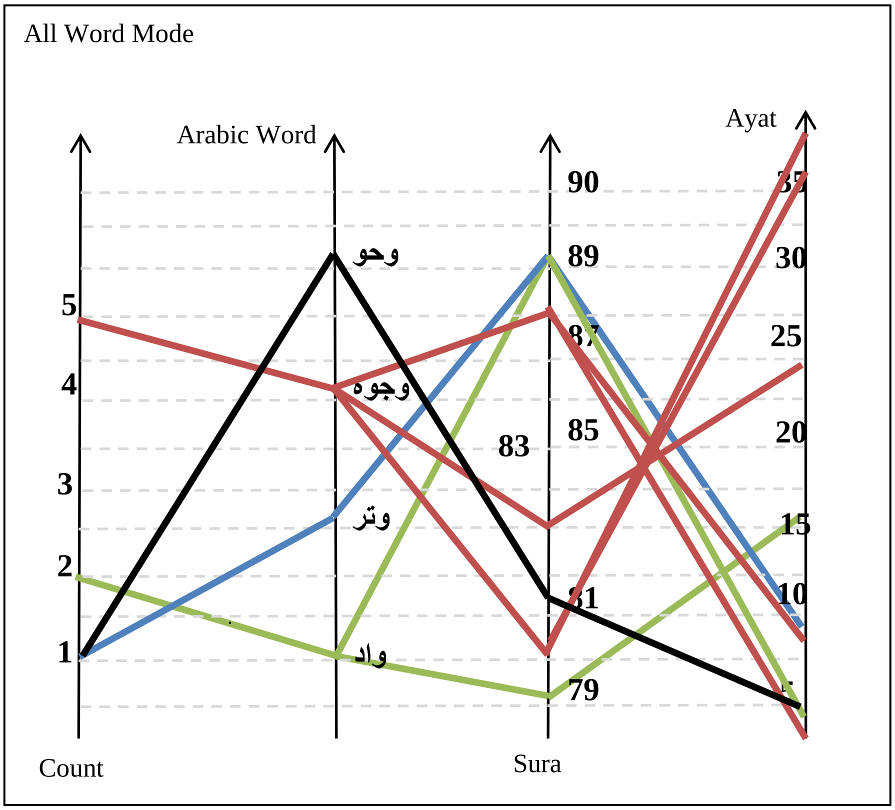


Figure 5.3: The parallel plot view of Table 4.2[9]

Plotting a scatter plot with 37 *sura*s- percentage relation $\boldsymbol{R_2}$ as in Figure 5.4a (with Table 5.3 values) would not be too problematic since 37 symbols could easily fit in any typical screen resolution. Figure 5.4 b) shows the corresponding parallel plot. However, only limited vocabulary words can be fitted in the scatter plot for quick inspection.

Plotting a scatter plot to compare the percentage of words in one *sura* with another to view relation R_3, a total of almost 37 $\times$37 = 1369 charts needed to be inspected. It is also possible to compare the values as in Figure 4.6 and the plot would involve

[9] Note that each colour represent each word in Table 4.2 and this applies to other plots in this chapter

the 1369 symbols aligned next to each other to fill up the entire horizontal or vertical screen. In this case distinction between two symbols would be impossible unless different colour is used. In addition, there are 2 other related attributes that should be compared to, the *sura* and the percentage which will add to the cognitive load of the user in interpreting the plot. This problem would also occur for plot as in Figure 5.2 to view relation R_1, since there will be around 800 words to be plotted horizontally.

Table 5.3: Vocabulary, *sura* and percentage values

Vocab. Words	**Sura**	**Percentage**
وحوش, وتر	**89**	**4**
وحوش, وتر	**80**	**2**

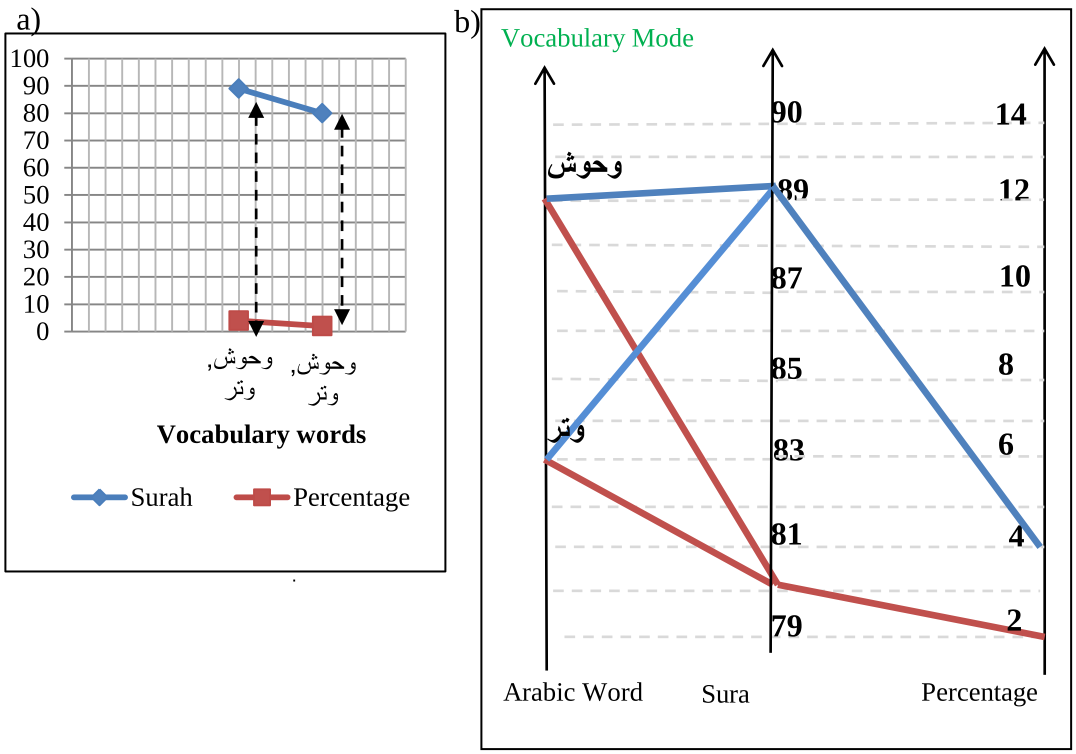


Figure 5.4: a) Vocabulary graph in scatter plot and b) Vocabulary graph in parallel coordinate

A size of a symbol such as in Figure 5.5 is estimated to be around 2 *mm* which is equivalent to 0.08 *inches* in a 17 *inch* monitor at the *resolution* of 1024 $\times$ 1028. Then the size of a symbol in *pixel* = *PPI (pixel per inch)* $\times$ 0.08 = $\frac{\sqrt{1024^2+1028^2}}{17}$ $\times 0.08 = 6.8$. Therefore, the number of symbols that can fit into a screen in the horizontal or vertical direction would range around 118 to 280 symbols for 800 to 1900 screen resolution (the typical screen resolution).

Taking the previous point further, Bertini & Santucci (2005) also stated that the human eyes can only distinguish two points above the angle of 1°, which is approximately 0.0057 inches (for a display distance of 20 inches) and this is about 14×14 virtual pixel. Therefore for 800 symbols, there are 800× 6.8 = 5440 overlapping pixels and for 1369 symbols, there are 9300 overlapping pixels.

Table 5.4: Count, Base *Sura*, *Sura* and Percentage

Count	**Base *Sura***	***Sura***	**Percent.**
25	**80**	**79**	**15**
		81	**7**
		83	**16**
		85	**30**
		87	**3**
		89	**4**
		90	**7**
12	**86**	**79**	**15**
		81	**12**
		83	**23**
		85	**10**
		87	**9**
		89	**17**
		90	**25**

Getting back to the original discussion, comparing the same attributes to be plotted in the parallel plot, Figure 5.6 shows an example with Table 5.4 as the tabulated values. To view relation $\boldsymbol{R_3}$, 36 lines emerge from one point in a base *sura* axis and these lines are connected to another 36 points in *sura* axis. This occurs for all 37 points in the base *suras*. In this case, there is a problem of $37 \times 37 = 1369$ lines intersecting each other. In viewing relation $\boldsymbol{R_1}$, the inter§ of 800 word relations is also problematic. However, in viewing the relation $\boldsymbol{R_2}$ such as Figure 5.4, the problem of intersecting line is bearable for low vocabulary percentage but as the vocabulary words grow, it is hard to detect what words are actually being added to the list. Therefore, it is also impossible to interpret the plot without involving operations such as zooming, filtering, brushing and selecting. However, scrolling or panning techniques would not be involved.

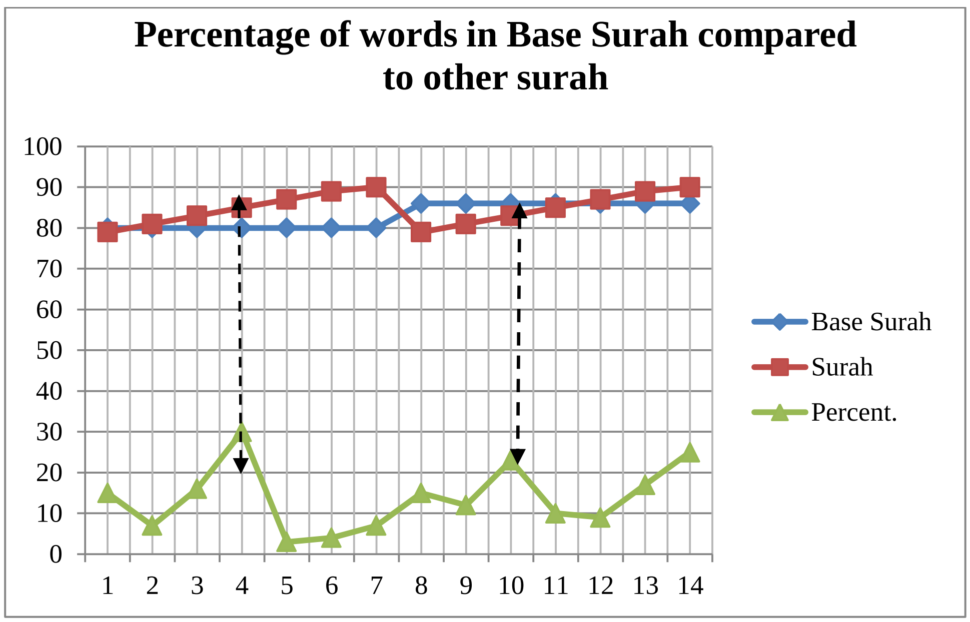


Figure 5.5 Scatter plot view for percentage word view on comparing base *sura* to other *suras*

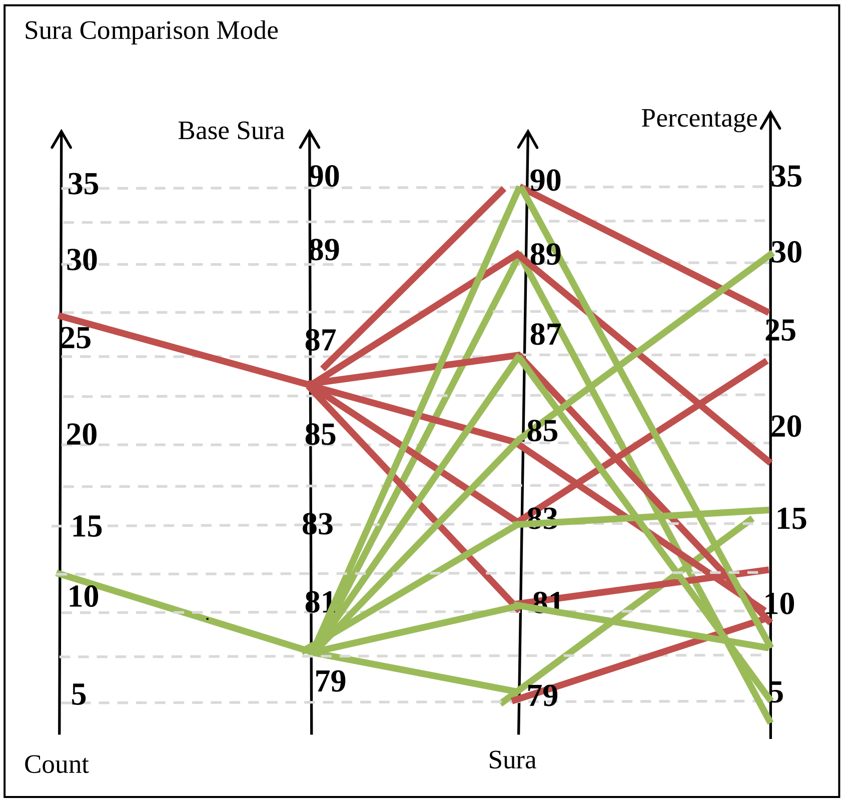


Figure 5.6: Parallel plot for *sura* comparison mode

Table 5.5: Comparison between parallel and perpendicular plot

No.	Relation to View	Data display in Parallel Plot (‖)	Data display in Perpendicular Plot (⊥)
1	R_1	800 intersecting lines of Arabic word	800 symbols align next to each other
2	R_2	37 intersecting lines to show percentage attributes in relation to the vocabulary word.	37 symbols covering the perpendicular area
3.	R_3	1369 intersecting lines of base *sura* and *sura* comparison percentage	1369 symbols align next to each other

Table 5.5 shows the summary of the comparison of parallel and perpendicular plots in context of this project. For viewing relation $R_1$and R_3 both types of plots are problematic but in the parallel plot panning and scrolling can be avoided. In viewing relation R_2 the parallel plot is problematic only if the vocabulary words grow to occupy all the vertical space but in the perpendicular plot even a small list of words can give rise to problems of identifying the words. At this point, the parallel plot seems a better technique to adopt compared to the perpendicular plot.

5.4 Using GOMS analysis to compare plots

Before starting the GOMS analysis of parallel and perpendicular plots, the scale used in both types of plots should cater for the maximum value of each attribute. The plot which contains all attributes with relation R_1, R_2 and R_3 is assumed to be projected in the two types of plots. In other words, an overview plot which contain 800 Arabic words with its corresponding count, *sura no.* and *ayah no.* for relation R_1, vocabulary words and the *sura* no. with its corresponding percentage for relation R_2, count, base *sura*, *sura* compared and the percentage for relation R_3. Figure 5.5 and Figure 5.6 are the simplified version of the full view plot. The y axis scale of the Figure 5.2 will have to be increased up to 300 since the maximum value is for the count attribute, whereas in the parallel plot, each count axis, Arabic word axis, *sura* no. axis and *ayah* no. axis can have different values according to the maximum attribute value (1 to 300 for count axis, 1 to 800 for Arabic words axis, 78 to 114 for *sura* no. axis and 1 to 46 for *ayah* no. axis).

GOMS analysis for relation $\boldsymbol{R_1}$Table 5.6 shows the tasks to be carried out in order to carry out the goal of relation $\boldsymbol{R_1}$. The lowest level tasks involved are given below.

Table 5.6: Tasks for relation $\boldsymbol{R_1}$

Relation	Scenario/ Goal	
R_1	What is the word with the highest count and where can this word be found in *Juz Amma*?	R1G: Find the word with the highest count and where can this word be found in *Juz Amma*
		S1:Find word
		S1.1:Locate/identify count axis/ symbol
		S1.2: Locate the highest count
		S1.3: Locate the corresponding word
		S2: Find where this word can be found in *Juz Amma*
		S2.1: Locate the corresponding *sura no.*
		S2.1.1: Locate/identify *sura* axis/symbol
		S2.1.2: Read the corresponding *sura* no.
		S2.2: Locate the corresponding *ayah no.*
		S2.2.1: Locate/identify *ayah* axis/symbol
		S2.2.2: Read the corresponding *ayah* no.

In a *scatter plot* these tasks may involve other subtasks:

- Identify count symbol:
    - This can easily be accomplished by looking at the legend of the graph -1
- Locate the highest count:
    - This may involve other subtasks. In this case, users may be confronted by display that is very cluttered; the highest point for the symbol attribute may appear the same as the second highest point. Therefore, users may need to zoom in to the particular point of interest which may be located anywhere along the horizontal axis (see Figure 5.2 for verification of this point of argument). Consequently, the subtasks involved may include the following:
        - Select/located the area with the highest point -2

- Vertical zoom into the area -3
- Read the highest count value -4

▪ Compare to next highest value (may involve scrolling)

- Scroll to the other potential highest point -5
- Read the potential highest point – 6
- Interpret the comparison -7

- Locate the corresponding word:
  - This involves the identification of the word with the highest count. As discussed previously, the words can overlap each other intensely that even if the vertical point had been zoomed in, the horizontal line would still be cluttered. Therefore, the subtasks involved may be the following:
    - Click/mouse over on the highest point -8
      - View the displayed word on the screen -9
- Identify *sura* / *ayah* no. :
    - This can easily be accomplished by reading the legend of the graph -10
- Read the corresponding *sura* / *ayah* no.:
    - There are many corresponding *sura*/ *ayat* no. on the same vertical line corresponding to the highest count point and the Arabic word involved in this task scenario. Users may find it difficult to identify all these *sura* / *ayat* no. Furthermore, it is not possible to identify exactly for each *sura* no., the corresponding *ayat* no. All that will be

traceable via perpendicular plot will only be the list of *sura* no. and the list of *ayat* no.. Therefore, the subtask involved is as follows:

- Ask the system to display out the specific *sura* / *ayat* no. (which need to be further decomposed in other subtasks) -11

In the *parallel plot* the same tasks as in the paragraph above are involved as listed in Table **5.6**:

- Instead of identifying symbol, users need to locate the count axis -1
- Locate the highest count (users may be confronted by cluttered display). However, it is not a problem to compare with the next highest point since the points are displayed vertically instead of horizontally in the perpendicular view. Therefore, the subtasks involved include the following:
  - Select/locate the area with the highest point -2
    - Zoom into the area – 3
    - Read the highest count value -4
- Locate the corresponding word;
  - at this point the words in the parallel plot may overlap each other. Therefore, the subtasks involved are as follows:
    - Click the highest count value -5
      - This action can be connected to a filtering function. Once this value is clicked the system can automatically display the corresponding attributes which are the word, the *sura* no. and the *ayah* no.. All

other related values can be faded out or temporarily filtered off the display and screen consequently simplifying all subtasks that comes after this.

- View the displayed word on the word axis
  - Locate the word axis -6
  - Read the word – 7
- Locate *sura* / *ayah* axis - 8

- Read the corresponding *sura*/ *ayah* no.. At this point in the parallel plot there will also be many possible *sura* and *ayah* related to the highest word count but it is possible to read all the *sura* and *ayah* no.. -9

Therefore, we find that to achieve a goal related to R_1, the number of subtasks involved in the perpendicular plot is more than 11 while in the parallel plot, there are 9 subtasks. At this point the parallel plot is observed as more efficient technique to adopt compared to the perpendicular plot.

5.4.1 GOMS analysis for relation $\boldsymbol{R_2}$

Table 5.7 shows the tasks to be carried out in order to carry out goal of relation $\boldsymbol{R_2}$ with the lowest level tasks involved as written in the bulleted points below. In a perpendicular plot, since there are only 37 symbols across the x axis (horizontal display) the task listed is not difficult to accomplish. Refer to Figure 5.4.

Table 5.7: Tasks for relation $\boldsymbol{R_2}$

Relation	**Scenario/ Goal**	
R_2	What is the sura with the highest percentage based on the vocabulary words in the list?	R2G: Find the *sura* with the percentage of the vocabulary words in the list.
		S1: Find highest percentage
		S1.1: Locate/identify percentage axis/symbol
		S1.2: Locate/identify the highest percentage value
		S2: Find the corresponding *sura*
		S2.1: Locate the *sura no.*

The following tasks do not involve any further subtasks:

- Identify percentage symbol – 1
- Locate the highest percentage value- 2
- Locate the sura no. -3

In the parallel plot, 3 tasks are involved; identify percentage axis, locate highest percentage value, locate the *sura* number. Therefore, regarding relation R_2, the subtasks to achieve the related goal are the same for the parallel and the perpendicular plots.

5.4.2 GOMS analysis for relation R_3

Table 5.8 :Tasks for relation $\boldsymbol{R_3}$

Relation	**Scenario/ Goal**	
R_3	What is the highest percentage of words in one *sura* compared to another and what are the *sura*s?	R3G: Find the highest percentage of words in one *sura* compare to another and the corresponding *suras*.
		S1: Find the highest percentage of words
		S1.1: Locate/identify percentage axis/symbol
		S1.2: Locate the highest percentage value
		S2: Find the corresponding *suras*
		S2.1: Identify the base *sura*
		S2.1.1: Locate/identify the base *sura* axis/symbol
		S2.1.2: Locate/identify the base *sura* no.
		S2.2: Identify the compared *sura*
		S2.2.1: Locate/identify the compared *sura* axis/symbol
		S2.2.2: Locate the compared *sura* no.

Table 5.8 shows the tasks to be carried out in order to carry out the goal of relation R_3 (R3G), the lowest level tasks involve are discussed below. In a *scatter plot* these tasks may involved other subtasks:

- Identify the percentage symbol:
  - as discussed previously; this can be done easily by looking at the legend. -1
- Locate the highest percentage value. The subtask involved is similar to identifying the highest count value in relation R_1, but this time the overall display view is more condensed. This is due to the 37×36 possibilities of comparison between *suras*. Therefore the subtasks involved are as follows:
  - Select/locate the area with the highest percentage -2
  - Zoom vertically into the area above -3
  - Read the highest percentage value -4

- Compare the highest value to the next highest value (may involve scrolling)
    - Scroll to the other potentially highest percentage -5
    - Read the potentially highest percentage – 6
    - Interpret the comparison -7
- Identify base *sura* symbol – 8
- Identify base *sura* no.. As there are 37×36 comparisons between *sura*, as mentioned earlier, the x axis may be very cluttered. Therefore, this particular task would be very similar to locating the highest percentage value which involves 5 subtasks. – 9-13
- Identify the compared *sura* symbol - 14
- Identify the compared sura no.: subtasks are very similar to identify base *sura* no., therefore this involves 5 subtasks. 15-19

In the parallel plot, the subtasks involved are as follows:

- Locate the percentage axis -1
- Locate the highest percentage value: in this case, since the scale of percentage is from 0-100, a detectable value, users can follow the sequence of task as in relation R_1 to locate the highest count, which involve 3 subtasks – 2-4
- At this point users can use a filter function such as clicking on the highest percentage value that displays the corresponding attributes (the base *and* compared *sura*). This action automatically reduces the subsequent subtasks - 5
- Locate the base *sura* axis -6
- Locate the base *sura* no. -7

- Locate the compared *sura* axis -8
- Locate the compared *sura* no. - 9

Therefore, we found that to achieve a goal related to relation R_3, the number of subtasks involved in the perpendicular plot is 19 while in the parallel plot, it is 9. Again, in this case, the parallel plot is a more efficient technique to adopt compared to the perpendicular plot.

5.4.3 Discussion

The comparison between the two plots through literature shows very similar outcomes in terms of features that the plots can support as shown in § 3.4. In terms of visualizing multidimensional attributes, attribute representations, pattern interpretation, application and data cluttering, both plots have similar ratings. However, in general, perpendicular plots are more popular and have a better perceptibility to users. Although many assumed (as in the § 3.3.2) that the parallel plot is better for novice users there are empirical evidence that the performance of the novice users using parallel plot is better. The comparison by Grinstein, Hoffman, & Pickett (2002) also shows that the two plots cover the same amount of related tasks. However, the parallel plot was used to analyze the Monk problems dataset while the scatter plot (perpendicular plot) was not. The Monk problem dataset is multivariate in characteristic with 432 instances and 7 attributes[10] which is similar to the Qur'anic dataset used in this study with slight variation i.e. with

[10] http://archive.ics.uci.edu/ml/datasets/MONK%27s+Problems

approximately 800 instances and 4 attributes . This leads to an extra task that could be covered by the parallel plot.

In the last part, the contextual comparison between two plots using GOMS analysis shows that to achieve goals related to relations R_1, R_2 and R_3, the estimated subtasks involved are fewer in the parallel plot compared to the scatter plot for relation R_1 and R_3. While for the relation R_2 the subtasks involved are the same. GOMS analysis involves time estimation as shown in Table 4.1. Although in this analysis, no timing calculation was done, it is observed that critical operations do not exceed the non-critical ones. The critical operations would be the ones involved cursor movement, mental preparation and determining position. Therefore, it is sufficient to only consider calculating the subtasks. As a conclusion it is observed that the parallel plot is a better plot to be adopted in the context of finding information in *Juz Amma*.

5.5 Summary of the preliminary studies

From preliminary study 1, it is found that the two highest rated problems in learning to read the Qur'an are understanding Qur'anic words and remembering Tajwid rules. This is mainly due to type of Arabic language (traditional) in Qur'an, those who can understand Arabic (especially of other types) may also not able to interpret the way it is written due to reasons such as indirect meanings and expressions, generality rather than specific facts for some parts besides the language type with different structure. Also there are special rules that need to be applied while reading the Qur'an aloud which is not applied while reading modern standard Arabic text.

In the second preliminary study there is significant improvement in reading comprehension scores when images are used as part of the text in the Qur'anic verses. This result is consistent with other findings such as by Yaxley and Zwaan (2007) that found readers mentally simulate the visibility of an object during language comprehension. Similar work by others such as Zwaan, Madden, Yaxley, & Aveyard (2004), Gosselin & Schyns (2004) and Richardson, Spivey, Barsalou, & McRae (2003) also support this evidence.

Finally from the third preliminary study it is found that the parallel plot involves fewer tasks in finding information in *Juz Amma*. An important aspect is mainly due to the filtering operation that only involves a mouse click in the parallel plot while in the scatter plot this is not possible without involving other operations.

6.0 INTEGRATED LEARNING FOR QUICK RECOGNITION

In this chapter, problems of reading comprehension for *Qur'anic* verses are analysed and a solution is proposed. As mentioned in the introduction chapter § 1.1 and 1.2, the problem of reading comprehension while reading the Qur'an is common in the non-Arabic speaking community. This was also found in the preliminary study 1 discussed in § 5.1.

Reading without comprehension of *Qur'anic* verses is mainly related to the problem in understanding the intended meanings of the written messages in the verses in depth. Understanding the words of the *Qur'an* in depth need special knowledge such as when and how the *ayah* (verses) were revealed, highly qualified and skilled in the Arabic language and possess other related knowledge. Therefore special educational path must be pursued in order to achieve it. However, one may still comprehend the clear and simple messages of the *Qur'an.* For the non-Arabic speakers these messages can be understood by reading the translation of the Qur'anic verses which are widely available in many languages. However, even though the translations are already available for many years, many still do not comprehend what had been recited in the Qur'an.

The problem of depth can be explained based from Bloom's taxonomy level of learning where the first two levels of learning are classified as *knowledge* followed by *comprehension* (Krathwohl & Bloom, 2002). Generally, *knowledge* is simpler to handle compared to *comprehension* in the second level of the taxonomy. Other higher levels of learning are even more complex to be dealt with. Revised anatomy of the cognitive domain has made minor but significant modifications of the higher levels to include *creating* at the highest level. Apart from *comprehension* at the second level, a higher, contextualised level of understanding needs evaluation of ideas. In analysing the cognitive processes of reading comprehension it is found that many coordinated processes are involved such as described in § 2.1. Although word recognition is one of the lower levels of the reading comprehension process, it is an important element which is required before achieving the higher hierarchy.

The focus of this research is on solving word recognition problem in Qur'anic verses. Therefore, this chapter covers discussion on word recognition in Qur'anic verses with respect to visualization for WFE and WSE, the approach taken for quick recognition, counting and stemming Qur'anic words and strategy of reading comprehension.

6.1 Word recognition in Qur'anic verses

It can be seen in the preliminary study 1 in § 5.1 that *tajwid* lesons are emphasized. The basis for this phenomenon is the motivation and obligation to preserve the Qur'anic verses in Its original form. In § 2.3 it was discussed that accurate recitation leads to accurate comprehension of the meaning. This can only occur when the reading process is automatic, i.e. the transformation of the text to sound does not impeded the process of reading, the ability to read a word without going through the letter by letter spelling process and this should be linked to recalling the word meaning. In the context of Qur'anic recitation, a fluent reciter may already have the ability to recognize the words but without recalling the meaning. Therefore, time and effort can be spent to learn the meaning of the words after each recitation. Two important factors to consider within the word recognition theory are the *word superiority* (WSE) and *frequency effect* (WFE) as mentioned in § 2.2.1: whereby a letter is easier to be recognized for known words compared to non-words; while the word frequency effect shows that frequent words are responded to more rapidly. In developing the software system in context to facilitate learning, the focus is how to design the interface so that the information presented is readily assimilated by users or learners (Ertmer & Newby, 1993). In the context of word recognition in Qur'anic verses, the interface should help to achieve WSE and WFE.

6.1.1 Visualization for WFE

(a) Qur'anic word frequency

The word frequency effect is in relation to the implication that the Qur'an contains many repeated words since many non-Arabic speakers have the ability to recite the Qur'an even though they do not comprehend the language. It is also based on the discussion in§ 2.1, for any information to go through to the LTM, the stimulus must attract the attention of one of the sensory system. Therefore, a type of stimulus could be introduced to Qur'an reciters to achieve word frequency effect. The reciters can be supported with learning activities in a classroom setting or software tools such as mentioned in § 2.2.2 as suggested by the National Reading Panel of the United States (NICHD, 2006). A list of most frequently used words in Qur'anic verses can be complied such as the Dolch and the Fry's list discussed in § 2.2.1. These words can be taught in classroom settings or incorporated in learning softwares for children or adults.

(b) Attracting the visual memory

Thus, it is suggested that visualization of word frequency can be applied as a stimulus in attracting the attention of the visual memory. For example, users might be more interested in learning words that occur more frequently. When users see the visualized information of word frequency in the Qur'anic verses, they might wonder what those words are and where those words can be found. At this point users start to search for the word and try to learn the meaning of the word. Subsequently, a learning process occurs and if this process is repeated, the process from SM to STM to LTM happens. Something that attracts attention of the visual sensory does not necessarily be nice to look at, it can also be with the opposite characteristic just as the phenomenon of focusing on one human face (may be an ugly one) in a crowded room explained **in § 2.1**. There are many techniques used in information visualization in various application domains that can be categorized into quantity and connectivity aspects as explained in § 3.1. Therefore, the visualization technique chosen to attract attention of the sensory memory must support both quantity and connectivity i.e. multidimensional data variables. Mentioned in §5.3, there are at least 4 variables identified for the relational concepts and that the parallel plot was found to be a more efficient visualization technique compared to the scatter plot.

6.1.2 Visualization for WSE

The word superiority effect implies that those that are fluent in their recitation of the Qur'an recognize most of the words. As discussed before in § 2.4.1, the process of recognizing Arabic words is not a straight forward path such as in English because Arabic words have complex morphological structure especially with regards to the following:

- Arabic tokens can contain more than a word, i.e. a phrase which usually contains preposition and pronouns
- Arabic words especially verbs can exist in many forms based on their root words by adding prefixes, suffixes and also infixes

Thus, in learning to recognize Qur'anic Arabic words, one should be exposed to various word forms and stems that originate from a root. It is then considered essential to visualize the Arabic word structures such as in § 3.1. In this way, users can learn to recognize words in many forms. These words can include nouns, pronouns, prepositions and conjunctions as well as verbs.

(a) Cognate words and translations

Usually one learns a language by learning the words and improving one's own set of vocabulary by attending a class. Sometimes, a determined person may also self-learn a language by reading a related book. One of the most popular ways of learning a language for a traveller to a certain country is to buy a simple word by word translation book.

The Arabic language is indeed taught in schools such as practised in Malaysia (Wikipedia, 2012). However, the vocabulary does not cover most words from the Qur'an partly because the Qur'an uses classical words that are not used in the Modern Standard Arabic language. However, the comprehension process is supported by other means such as Qur'anic lessons in schools and mosques such as mention in § 2.3. The knowledge of topic and theme exists to form the higher level of reading comprehension, the bottom level knowledge of transforming text to sound also exist. The only missing link is recalling process of the meaning of those words.

Part of this solution to a large extend was to adopt the Arabic based writing scripts by most nation of the Muslim world. For example, the Urdu script for writing Indian language and *Jawi* script for writing the Malay language was originated from the Arabic script. In addition, many of the Arabic words have been adopted as part of the local language. Therefore, this explains the reason of existence of cognate words such as discussed in § 2.6.

As mentioned in § 2.6, although there are many Arabic words in the Malay language sometimes these words do not necessarily carry the same meaning and pronunciation. This poses a problem while learning the words in the Qur'an. Although this problem is not easy to solve, for this research, the translated meanings of all Arabic words are provided. This includes all meanings of particles, pronouns, verbs and other types of words. As this research focus is targeted to non-Arabic

Malay speakers, the chosen languages of translation used are Malay and internationally recognised English language.

Even though, there exist many Arab-Malay translation verse by verse, it is not word by word translation. The readers may have some ideas on the verse but this knowledge is not transferable. They may have the retaining skill but not transferable skill. Therefore it is considered that verse by verse translation is not optimum for supporting word recognition. Verse by verse translation supports comprehension after recitation. Although some verses can be comprehended by memorization, it still can be argued that the process of memorization is aided by word recognition of one or two keywords of the verse. Instead, word by word translation is seen to be better for the transferable skill to be used as a mean of attaining reading/recitation with comprehension. Therefore, the approach taken to solve the issue is to segment the verse translation, i.e. map each Qur'anic Arabic word to the translated meaning. Figure 5.1 illustrates the example taken from the first *ayah* 1 of *sura* 114 in *Juz Amma*. It is observed that each segmentation is usually mapped onto more than a word of Malay or English. There is also the issue of words that have different meanings in different context. Both contextual meaning and general meaning can be used.

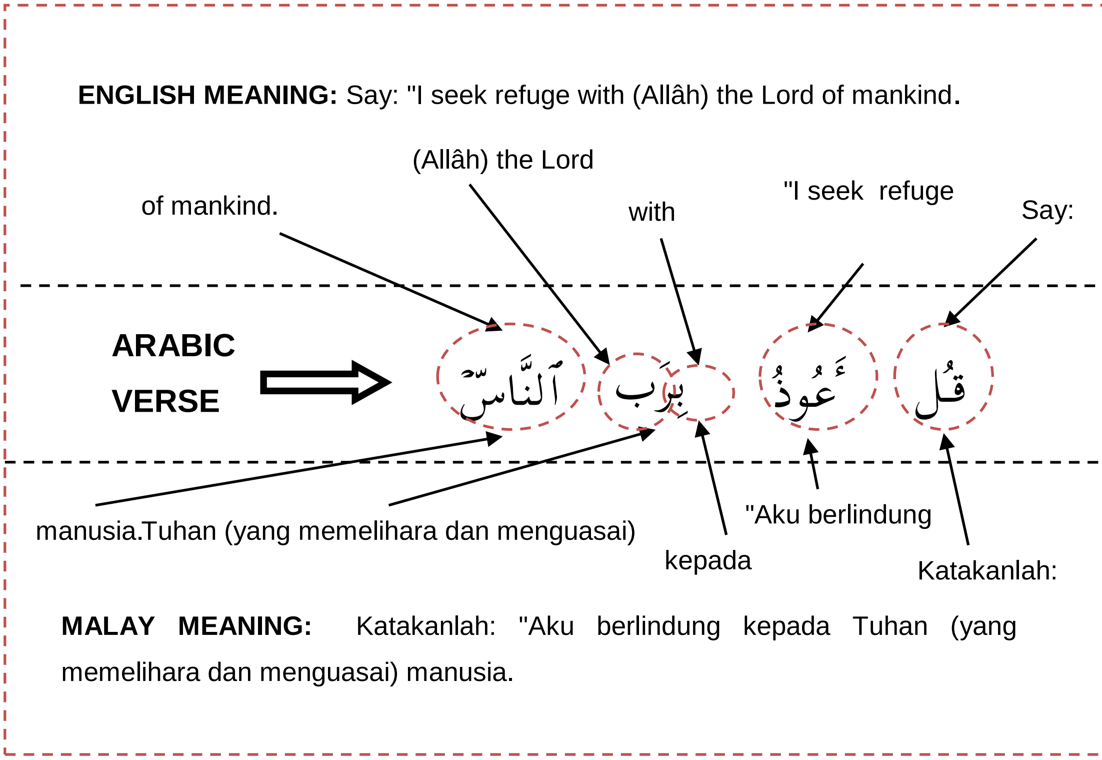


Figure 6.1: Segmentation of the translated meaning of the Qur'anic Arabic verse (*ayah*/verse 1, *sura* 114)

(b) Supporting the rehearsal action

It was mentioned in § 2.2.2 that in order to comprehend a text, 95% of the words should be known. Supporting word recognition visualization through a vocabulary tracking system with the cognate and meaning of words is therefore considered a suitable logical approach to provide reinforced learning. In § 2.2.2, it was mentioned that semantic mapping and organization can help in the vocabulary development, and in the preliminary study 2 in § 5.2 it was found that there is evidence that using images in Qur'anic text can improve comprehension compared to when only the Qur'anic text is used. In addition, it is also known that most people of 18 years and above are visual learners, i.e. they learn better if pictures, diagrams or symbols are shown to them (Felder & Silverman, 1988).

If users have the options to keep track, be reminded and shown their own vocabulary with respect to all the words contain in the Qur'an, then it should lead to a better learning strategy of Arabic words in the Qur'an. However, mere static image visualization is not as helpful as compared to when the information is transformed into a perceptually efficient visual format (Ware, 2000). For example, the interactivity operations that are tightly coupled with the visualization fields are for users to gain insight of a particular data to be analysed. The act of exploring, interpreting and making deduction of data related to vocabulary visualization therefore is hypothesized to help individuals in their semantic word detection ability (to recognize and remember the meaning of words), i.e. a means of rehearsal action for the information to go from visual memory to the LTM. This is also related to active learning as explained in § 2.1 It is also predicted that when there are elements of intra communication within the cognitive thinking of a user, eventually users not only end up enhancing their word recognition ability but also their reading comprehension ability. The effect of word relation, predicting words within the context in which it appears in a sentence with the addition of existing knowledge of theme or topic of the *sura* helps in the whole integrated process that should be considered in the implementation.

6.1.3 The approach taken for quick recognition

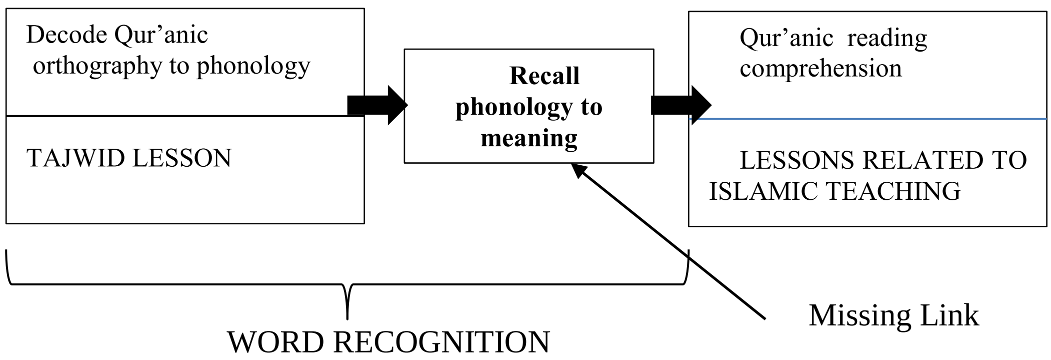


Figure 6.2: The missing link to achieve word recognition

As a summary, a proposed model such as illustrated in Figure 1.1 is hypothesized to fill in the missing link illustrated in Figure 6.2 to solve the word recognition problem in Quranic verses eventually leading to reading comprehension. This is solved by means of visualization to achieve WSE and WFE with the mechanism of cognitive information processing i.e. for the information to go through from SM to STM. WSE is supported by visualization linking to the vocabulary tracker, cognate words and word segmentation to achieve semantic word detection. While the WFE, is supported by the parallel plot in visualizing the word count and word position. The parallel plot visualization is hypothesized to attract the visual memory and actions related to semantic word detection such as explained in the previous § are hypothesized to help in the learning process of word recognition leading to reading comprehension through reinforced learning. Both the semantic word detection and the parallel plot are linked together as part of the integrated solution. This model is proposed to achieve quick recognition.

6.1.4 Counting and stemming Qur'anic words

In this § counting and stemming Quranic words are addressed. There are 77,430 words in the Qur'an and analysis of information gathered through Qur'an.corpus.com by Dukes (2009) shows that there are about 5155 non-repeated words (base-words) amounting to only 6.7 % of all the words in the Qur'an. However, counting words in the Qur'an is not an easy task since Qur'anic words have complex morphological structure. The above information was deduced from the table of word count for the Qur'an provided by Dukes (2009).

Previously in § 2.3.1 it was mentioned that a root word can exist in many forms. Therefore in order to count words in the Qur'anic verses, a methodology need to be specified since the counting algorithm can be based on the root or stem of the word. For example, the word فعل *f E l* can exist in other forms such as يفعل *yfEl*, فاعل *fAEl*, فعال *fEAl* (details are as described in Table 2.4). If root word counting is adopted then all forms of the word increases the word count, if not only exact match of the word pattern contributes to the word count. In this research we adopt root word counting.

6.1.5 Qur'anic word count of *Juz' Amma*

This § covers discussion on the frequency analysis of words in *Juz Amma*: the words occurring less than 10 times, more than 10 times, more than 40 times, as well as Malay-Arab words and comparison of base and all words in *Juz Amma*.

(a) Word frequency analysis of *Juz Amma*

Method: Computer algorithm written in Java language and SQL statements are written in this research to conduct the word frequency analysis based on the database consisting of all the words in *Juz Amma*. However, there are still errors on the exact count of the words allowing for an estimate of the frequency analysis. Therefore, numbers used in this § tend to be rounded numbers instead of the exact values. Apart from using computer algorithm, words with complex morphological structure are also analyzed manually.

Summary of algorithm: The algorithm used to do the frequency analysis is a simple matching string algorithm. Only exact words are detected. All suras in *Juz Amma* were read as an input file. Each word in the file was put into a Vector list, and then another list is produce by stemming word particle such as و (*w) and* pronouns such as ك م + +*km* before each item in this new list is compared to every other item. If exact match is found the counter is increased. However, for counting prefix and suffix the original list is used, for example, to count the word و (*w*), the input string is checked, if the first character contains و (*w)* the counter is increased by 1. Taghva, Elkhoury, & Coombs (2005) and many others uses this type of algorithm. This approach is used for both prefix and suffix. For words with more complex morphological structure, i.e. the words are counted manually based on the root such as يفعل *yfEl,* فاعل *fAEl,* فعال *fEAl,* are counted as 3 for the word فعل *f E l.* The manual check is subject to counting error if the root word is not detected correctly and tend to be counted as unique or with their exact string match. This is how the rounding is done.

Overall Analysis: There are 3456 words in *Juz Amma*, and from these words, there are only 815 distinct base words encompassing about 27% of all the words in *Juz Amma*. There are about 640 words occuring once or twice, 70 words occuring three to four times, 2 words occuring one hundred to one hundred and thirty times and one word, the word و, the most repeated word in *Juz Amma* occurs almost 280 times.

Table 6.1 shows the distribution of number of occurences for the base words. Comparing low and high number of occurences, generally there are less number of base words which occur more frequently (high occurence). Hence the range for high number of occurence is increased to group the base words. The average number of occurence (AV_o) is given by:

$$\mathbf{AV_o} = \frac{\sum_{i=1}^{815} f i}{815} = \frac{3456}{815} = 4.24$$

Where f_i = frequency of each word

Q_3, Upper Quartile = No. of occurrences 101-130 = 2

Q_2, Median = No. of occurrences 61-80 = 6

Q_1, Lower Quartile = No. of occurrences 5-6 = 41

Table 6.1: Word occurences in Juz Amma

No. of occurences	Range[11]	No. of base words	Average[12]
1-2	1	640	320
3-4	1	70	35
5-6	1	41	20.5
7-9	2	15	5
10-15	6	16	3.2
16-21	6	15	3.0
22-40	18	5	0.26
41-60	19	4	0.20
61-80	19	6	0.30
81-100	19	0	0
101-130	29	2	0.07
131-290	159	1	0.00625
TOTAL		815	

Note: See Appendix I1-I8 for the list of all word frequency

Malay-Arab words: There are 40 identified Malay words that are also *Qur'anic* Arabic words in *Juz Amma*. These words were identified manualy by going through the list of all words in *Juz Amma*. This list was obtained form the algorithm discussed earlier in this §. In the previous §, this type of words are called cognate words. Although it is optimal to use cognate words in *Juz Amma* with the same meaning and the same phonological sound for this research project, words with different phonological sound with the same meaning are also involved. Table 6.2 shows the list of the identified cognate words. Most of these words occur less than 10 times and only three of these words occur more than 10 times but less than 25 times.

[11] Range is calculated base on the no of occurenses
[12] Average= (No of base word)/Range + 1

Table 6.2: Cognate words in Qur'anic Arabic and Malay

Arabic word	Malay word	Arabic word	Malay word	Arabic word	Malay word	Arabic word	Malay word
قبر	Kubur	ساعة	Hari berbangkit	جهنم	Jahannam	الله	Allah
قدر	Kadar	شيطان	Syaitan	حسب	Hisab/per iksa	آخرة	Akhirat
قلم	Kalam	صبح	Subuh	حسد	Hasad	أحسن	Ihsan/ sebaik-baik
كفر	Kafir	صبر	Sabar	حق	Hak	خبار	Khabar/ berita
كرمة	Mulia/ karamah	صالح	Saleh	ذكر	Zikir	أهل	Ahli
ملائكة	Malaikat	عصر	Asar	رزق	Rezeki	امن	Iman
موت	Maut	عالم	Alam	رفع	Rafa'/me ninggikan	إبراهيم	Ibrahim
موسى	Musa (Nabi)	عذاب	Azab	زيتون	Buah zaitun	إله	Ilah
يتيم	Yatim	عمال	Amal	سأل	Soal	إنسان	Insan
يقين	Yakin	غيب	Ghaib	سلام	Kesejahte raan/ Salam	تين	Buah tinn

Comparison of Base Words and All Words in *Juz Amma*: Figure 6.3 shows the comparison of base words and all words in *Juz Amma*. If we observe the percentage of base words (dark bar), there are 1.6% words occurring more than 40 times, 4.4% occurring between 10 times and 40 times and very high percentage of 94% in the other category (< than 10 times), showing a decrease in percentage with increment of word frequency. However, if we compare the same category of words from all words in *Juz Amma* (lighter bar), the highest percentage of words (43.8%) is for the category of words occurring less than 10 times and although there are only a few base words in the category of word frequency occurence > 40, they make up, up to

36.4 % of all word because the words repeat many times as for example the most repeated word و occurs almost 290 times.

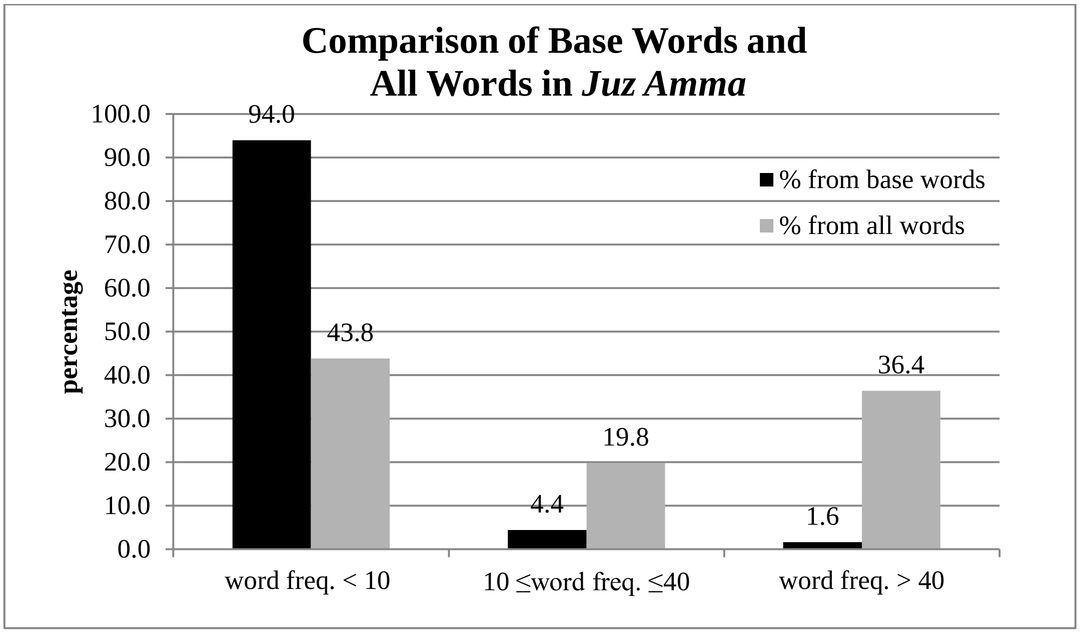


Figure 6.3: Comparison of Base Words and All Words in *Juz Amma*

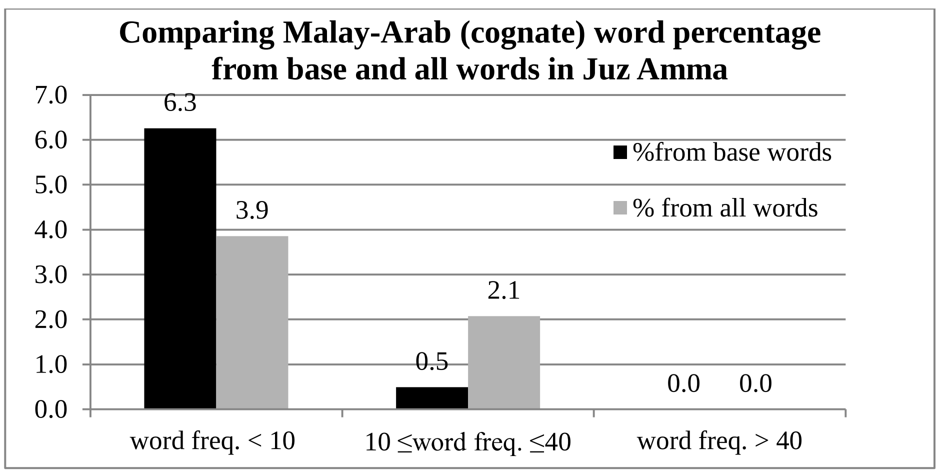


Figure 6.4: Percentage of Malay-Arab words in *Juz Amma*

There are only 6.8% (6.3+0.5) of Malay-Arab (cognate) words from base words of *Juz Amma*. The percentage of Malay-Arab (cognate) words are 6% (3.9+2.1) from all words in *Juz Amma* as shown in Figure 6.4.

6.1.6 Approach on the strategy of reading comprehension

The main components which constitute the elements of visualization for reading comprehension need to be identified and properly linked. The strategy of learning to comprehend *Juz Amma* can be developed based on reading comprehension theories and frequency analysis of *Juz Amma*. Word structure analysis and word meaning need to be taken into account. In the previous §, it was observed that although the two categories of words (the words occurring at least 10 times and, words occurring more than 40 times) are only 6% (4.4+1.6) of base words in *Juz Amma*, they consume 56.2% (19.8+36.4) of all words in Juz *Amma*. Also, observation showed that these words are easy to remember such as word particles and pronouns.

In addition the Malay-Arab words facilitate the learning due to the bilingual cognate words having the similar pronunciation and meaning as well as the effect of word relation between adjacent words and discourse of *Juz Amma* can help in the comprehension. This is one of the factors hypothesized to contribute to the semantic word detection as shown in the integrated model. Although, there is the problem of comprehension of the Arabic language, the knowledge of the theme or topic written in the Qur'anic *suras* is usually wider and can be one's prior knowledge. This is due to many medium of conveying the message of the Qur'an through Qur'anic lesson classes, in the main stream media such as television and radio as well as personal reading habits. Therefore, in cases when unknown words are found while reading the Qur'an, existing knowledge on the theme and topic helps in the process of comprehension by relating the meaning in context of a particular sentence. Therefore it is worthwhile to teach *Qur'an* reciters of *Juz Amma* those words.

6.1.7 Stemming algorithm in Quranic verses

In deciding on which methodology to adopt for stemming algorithm, the related issue that is crucial as a determinant is the fact that the main objective is to support word recognition. To learn words in the stem form means that there are more words to learn and more words to process, however this approach gives better meaning of words in the context of a sentence. For quicker learning and simpler implementation root word form is better. Here in this study, the approach is to adopt the algorithm based on root words. Next is to decide on how this should be implemented. There are two studies conducted in this research as in the following §s. The first is a case study on the 30th part of the Qur'an (*Juz Amma*). The second is an actual implemented algorithm based on the word patterns identified in Table 2.4 for *sura 29*. This algorithm is referred to as the Rule-Based Stemming Engine (RES) (Yusof R, Zainudin, & Yusoff, 2010). Both studies were done on a dual 2.4-GHz processor with 2.99 GB of memory.

(a) Case study on *Juz Amma*

The Case Study was conducted using *Juz Amma* consisting of short *suras* form chapters 78 to 114 of the *Qur'an*. There are approximately 3000 words or 2000 over word tokens in *Juz Amma* which are taken as the dataset to conduct the case study. These words are stored in a MySQL database and queries on the related tables were made based on the test cases below using Regression Expression to find the percentage of correctness in removing prefixes and suffixes.

The tested cases are divided into 2 categories, *i.e.* removing one of ا, ه, ي, ن, و, م, ت, ل, ء, س (*A, h, y, n, w, m, t, l, ', s)* (10AL - see § 2.3.1 for clarification) and pronoun/ particle. The statements are written based on following; + ي is equivalent to *y*+, where *y* is the transcription of the Arabic letter ي (see Table 2.1). The + sign indicates that ي *y* is a prefix since other letters can be added after the + symbol:

i. Prefixes of only one or two alphabets

   i. + ي *y*+, + ت *t*+, +م *m*+ (in the category of 10AL)

   ii. + ل ل *ll*+, + فَ *f*+ , +و *w*+ (in the category of particles)

ii. Suffixes of one or two alphabets

   i. ا ن + +*An*, و ن+ +*wn*, ي ن+ +*yn* (in the category of 10AL)

   ii. ت م + +*tm*, ك م + +*km* and ك + +*k* (in the category of pronouns)

(b) Implementing word pattern identifier

The word pattern identifier algorithm is implemented in this research using a Rule-based Stemming Engine (RSE). The stemming process is to determine the root word of the input token from Dukes (2008) Qur'an Corpus which contains the Arabic Qur'anic words, the word information (word, *ayah* and *sura* number) and also the part-of-speech tagset which is written in roman scripts. This tagset include the root word of all words in a verse, words such as particles does not contain a root word. Our algorithm caters for three letter root word since this is the most common type.

The main reference for the implementation of this word pattern identifier is the table 2.4 § 2.5.1. The word pattern identified in this table were further categorised into number of characters (4, 5, 6, 7 and 8). The RSE algorithm is created based on this categorization. After each the stemming process, each word left is compared to the root word information in the tagset written in roman script. Therefore each Arabic stemmed words must be transliterated using the table 2.1 shown in §2.5.1 to determine whether the stemming had been done correctly.

The algorithm is as follows:

i) Get word tokens from the corpus.

   a. Separate word token from word tag set into two vector lists

      - Qur'an word list (Arabic script)
      - Part-of-speech tag set list (Roman script)

ii) Remove particle prefixes (+لِ *l*+,+فَ *f*+ ,+بِ *b*+ , +وَ *w*+,+أ *A*+)

iii) Remove diacritization from all word tokens

iv) Remove prefixes (the prefixes in table 2.4 § 2.5.1 excluding the ones in ii)

v) Remove suffixes (the suffixes in table 2.4 § 2.5.1)

vi) Determine length of word, stem the words by removing one of 10AL as in table 2.4 in § 2.5.1 and put them into a new vector list

   a. If word length = 1, 2, 3 or >8[13], return the word without stemming

   b. If word length = 4, use algorithm to detect word pattern of length 4

      - Delete one of 10AL
      - return the three letter root word

   c. If word length = 5, use algorithm to detect word pattern of length 5

[13] Our algorithm assumes that there should be no words with length > 8

- Delete two of 10AL
- return the three letter root word

d. If word length = 6, use algorithm to detect word pattern of length 6

- Delete three of 10AL
- return the three letter root word

e. If word length = 7, use algorithm to detect word pattern of length 7

- Delete four of 10AL
- return the three letter root word

f. If word length = 8, use algorithm to detect word pattern of length 8

- Delete five of 10AL
- return the three letter root word

vii) In a stemmed word of length three, replace middle vowel letters to و *(w)*[14]

viii) Transliterate generated root words after stemming and compare to tag set

a. Take the new list containing the stemmed Arabic words

b. Transliterate the list above so that the comparison of the root words written in roman script from the other list (described in i-a) can be done

ix) Compute accuracy of RSE

a. Compare list i-a and list in viii-b

b. Increase count if elements in the 2 lists in a) above is the same

c. Compute percentage by dividing count to the size of list above and multiplying with 100.

[14] This is not right in all cases, at this moment this step result in a higher accuracy rate.

(c) Result

In this §, result of the test cases and RSE algorithm is shown. These result are importance in determining the design of PaCQ interface for counting words.

Result of case study

Figure 6.5 shows the result of the experiment in the case study when alphabets according to the test cases outlined in the § 2.5.1 are to be removed. The graph shows two information, the occurrences and the percentage of the incorrect stemming for words with the specified suffixes and prefixes.

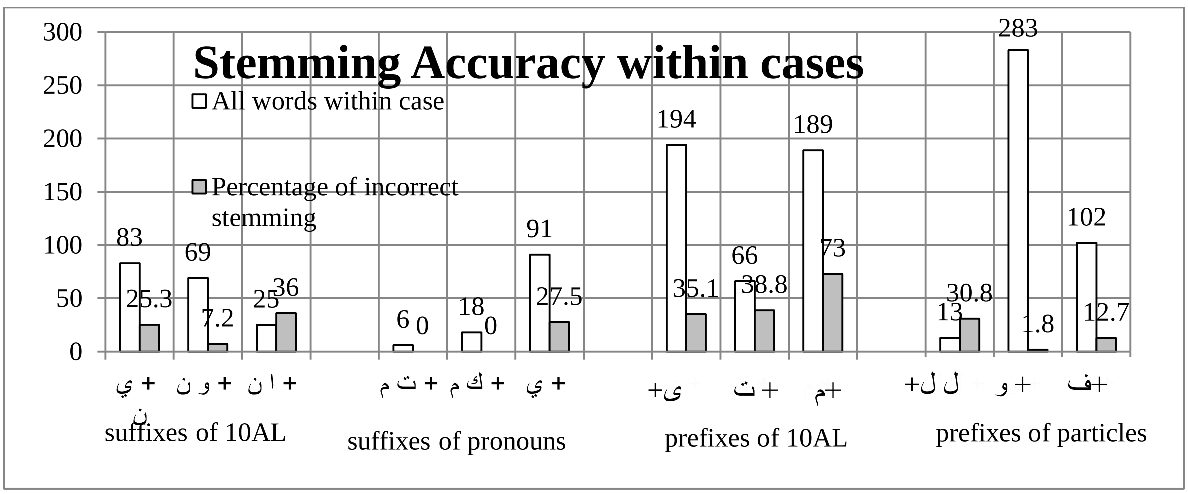


Figure 6.5 .Stemming accuracy within cases

The following prefixes, +ي *y*+, +ت *t*+ and +م *m*+(in the category of 10AL) occurrences are 194, 66 and 189 times respectively. Removing these prefixes will cause 35.1%, 28.8% and 73% of incorrect stemming. This means that some of the words which contain the respective alphabets are part of the whole word. Some of the examples of these words are such as يوم *ywm*, تبث *tbt* and من *mn*.

Prefixes in the category of particles, + ل ل *ll*+, +و *w*+, + فَ *f*+ occur 13, 283 and 102 times and the percentage errors of stemming are 30.8%, 1.8 % and 12.7 % respectively. Some word examples are لله *llh*, وجه *wjh* and فصل *fSl*. Interestingly, it was observed that removing +و *w*+ causes low stemming error.

Suffixes in the category of 10AL ا ن + +*An*, و ن+ +*wn*, ي ن+ +*yn* occur 83, 69 and 25 times respectively with percentage errors of 25.3%, 7.2% and 36%. Some words examples are كان *kAn*, زيتون **ytwn* and يقين *yqyn*.

Finally, in the category of pronouns suffixes ت م+ +*tm*, ك م + +*km* and ك + +*k* occur 6, 16 and 91 times repectively. The percentage errors are 0, 0 and 27.5 % respectively. Some words examples are ذلك **lk*, إياك *Ayk*.

It can be observed that prefix +م *m*+ has the highest percentage of stemming error of 73%. This is partly due to the fact that there are many words containing من *mn* and ما *mA* words. This error can be overcome if detection of these words is to be incorporated in the stemming algorithm. The case study also shows that there tends to be more error for removing the 10AL in both the prefixes and the suffixes and this is called over-stemming. It is shown otherwise for words in the category of pronoun and particles both for the prefixes and suffixes.

Although in the context of Modern Standard Arabic, related algorithm suggested by Taghva, Elkhoury, & Coombs (2005) stop words such as من *mn* and ما *mA* should be removed. However, Al-Shammari & Lin (2008) pointed out that the removal of these type of words results in loss of information that will be useful to identify noun and verbs. As in the test cases, a total of 314 words are found to be over-stemmed and this is about 10% of words in *Juz Amma*.

Result of implementing word pattern identifier

The second experimentation started with inputting word tokens including the tagset into RSE. The word tokens from *sura* 29 were first fed into RSE to find out the accuracy percentage.

Table 6.3: Performance measure of RSE Stemming

Number of Tokens including Stop words (935 tokens)			
Unanalyzed	Errors	Fault Rate (%)	Accuracy (%)
41	376	41.0	59.0
Number of Tokens excluding Stop words (585 tokens)			
350	187	34.0	66.0
Average = (59.8+66)/2= 62.5%			

Table 6.3 shows the efficiency of RSE with 41 unanalyzed words, whereby it achieves 59% accuracy for all tokens including stop words and 66% accuracy for tokens excluding stop word. In average RSE achieves 62.5% accuracy.

(d) Design implication

It is necessary for the stemming algorithm to be near 100% accurate for counting the root word. Incorrect stemming leads to incorrect word count information. This may lead to less optimum learning for Arabic word recognition that have complex morphological structure which subsequently impedes reading comprehension in Qur'anic verses. Therefore for the purpose of this project, a database containing information of root words and sub words will be used as part of the system as reference to this information can be obtained from translated Qur'an or from a website such as Qur'an corpus by (Duke, 2009b) or the Qur'an complex as mentioned previously.

6.2 Summary: On requirements for the visualization system

There are several main requirements identified to design and implement the visualization system for word recognition in Qur'anic verses. The system needs to:

i. Visualize the Arabic words for WSE (requirement i – REQ i
   a) The word segments
   b) The vocabulary which includes cognate words in Malay and Qur'anic Arabic
   c) This includes the vocabulary tracker

ii. Visualize the frequency of Arabic words for WFE – REQ ii
   a) This includes the word position

iii. Link the visualization fields and the vocabulary tracker to the verses of the Qur'an in *Juz Amma* – REQ iii
   a) This includes the meaning of the Arabic words

iv. Store related information in a database designed to access information for word recognition – REQ iv

v. Provide functions for users to interact with the visualization fields – REQ v

The next chapter discusses the system design and implementation of these requirements translated into more detailed classes.

7.0 DESIGN AND IMPLEMENTATION OF PaCQ

In this chapter, the design and implementation of the system called PaCQ (Parallel plot for Comprehension of Qur'an) developed in this research is discussed. The system is aimed to solve Qur'anic Arabic reading comprehension problem by focusing on Arabic word recognition and the retrieval of its meaning through vocabulary tracking.

This chapter starts with the architectural design of PaCQ, the functional modules and interface. Finally the implementation and algorithm is highlighted according to the designed components.

7.1 Architectural design of PaCQ interface

There are 19 classes related to objects identified for the PaCQ interface to support the 5 main requirements as listed in §, 6.2. These classes can be categorised into several components and packages as follows:

- PaCQ User Interface (PUI) component –for REQ v
- Arabic Word Visualizer (AWV) component – for REQ i
- Parallel Coordinate Graph (PCG) component – for REQ ii
- Intermediate Interface (II) component – for REQ iii
- The Database package – for REQ iv

The PUI component provides users interface for PaCQ interface system. It is through this component users interact with PaCQ. The AWV component provides functions relating to the Qur'anic Arabic words. Through this component the Qur'anic Arabic words are retrieved from the database and then displayed in Othmaniy font. The II component serves as an intermediate between the PUI component and the database that may involve other intermediate dialog windows. The component PCG is needed to process the parallel plot graph which is plotted based on the Qur'anic Arabic word information from the database. Finally, the DatabaseConnection class is the interface component that provides the facility for other components to access information from the database or to store information to the database.

Table 7.1 shows the classes related to the components and packages stated above. Figure 7.1 shows the component diagram of PaCQ interface designed in a 3-tier architecture.

Table 7.1: The container classes of the PaCQ interface components

	Components/Package	Container Classes
1.	PUI	i. PaCQInterfaceSystem ii. DisplayData
2.	AWV	i. WordInformation1 ii. Sura iii. OthmanCharDisplayer iv. WordInformationDialog v. SearchWordDialog
3.	II	i. Menu ii. SaveInformation iii. OpenVocabulary iv. VocabularyList
4.	PCG	i. Graph ii. Analyse iii. WordInformation2 iv. WordInformationList v. MalayArabWord vi. PercentageInformation vii. ListColor
5.	Database	i. DatabaseConnection

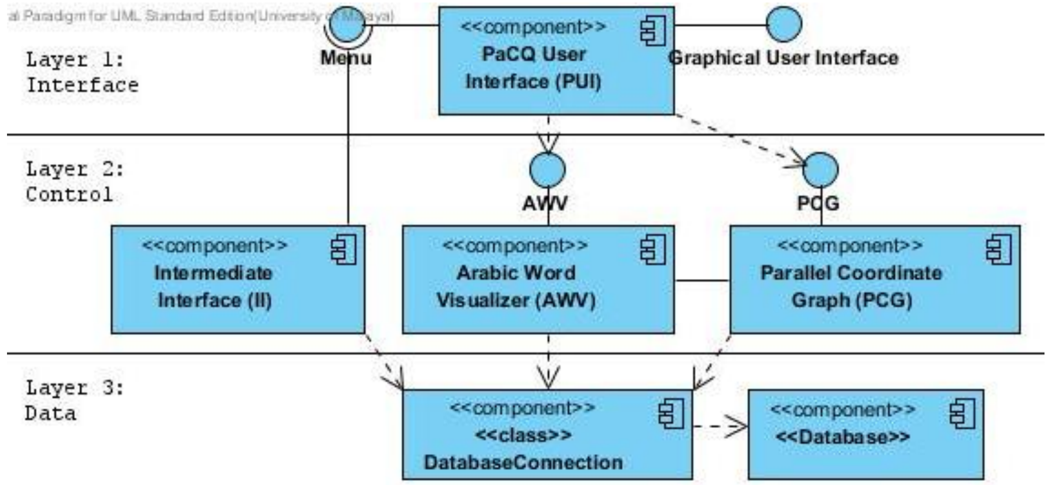


Figure 7.1:Component Diagram of PaCQ interface in a 3-tier architectural design

The overall architectural design of the PaCQ interface can be viewed as a three-tier architectural design. Only adjacent layers can access functions between each other. The first layer pertains to users interface, the second the control and the third the database.

The database used in this research is based on MySQL community server version 5.0.67. It contains 4 tables with schemas as the following and as listed in Table 7.2

- sub_root_words: RootWord(suraNumber, ayahNumber, wordNumber, subwordNumber, subwordOt, subwordAr, subwordAr2, root, position, English, Malay)
- Qur'an: Qur'an(suraNumber, ayatNumber, wordNumber, pageNumber, wordOthman, wordArabic, wordOthman2, wordArabic2, position)
- meaning: WordMeaning(suraNumber, ayatNumber, wordNumber, English, Malay)
- juz_amma_suras: Sura(suraName, suraNumber)

Table 7.2: The database information for PaCQ

No	Table Name	Schema Name	ColumName	DataType	Flags
1	sub_root_ word	RootWord	suraNumber -the Sura number	Integer	Unsigned
			ayatNumber -the ayah number	Integer	Unsigned
			wordNumber -the word token number	Integer	Unsigned
			subwordNumber -the subword number within a word token	Integer	Unsigned
			subwordOt -the subword in Othmaniy font	VarChar(120)	Binary Column Charset:ucs2 Column Collate:ucs2_general_ci
			subwordAr - the subword in the form of Unicode character set	VarChar(120)	Binary Column Charset:ucs2 Column Collate:ucs2_general_ci
			subwordAr2 - the subword in the block form downloaded from Qur'an complex website	VarChar(120)	Binary Column Charset:ucs2 Column Collate:ucs2_general_ci
			root - the root word	TinyInt(1)[15]	Unsigned
			Position - contain the sura, ayah and word number	VarChar(45)	Binary Column Charset:ucs2 Column Collate:ucs2_general_ci
			English - the English translation of the Arabic word	VarChar(255)	Binary Column Charset:ucs2 Column Collate:ucs2_general_ci
			Malay - the Malay translation of the Arabic word	VarChar(255)	Binary Column Charset:ucs2 Column Collate:ucs2_general_ci

[15] The column root indicate whether the word is a root word or other particles such as pronouns, prepositions or conjunctions. The integer number 1 indicate it is root, while 0 is non root.

Table 7.2 continued.

2	meaning	WordMeaning	suraNumber	Integer	Unsigned
			ayatNumber	Integer	Unsigned
			wordNumber	Integer	Unsigned
			English	VarChar(255)	Binary Column Charset:ucs2 Column Collate:ucs2_general_ci
			Malay	VarChar(255)	Binary Column Charset:ucs2 Column Collate:ucs2_general_ci
3	Qur'an	Qur'an	suraNumber	Integer	Unsigned
			ayatNumber	Integer	Unsigned
			wordNumber	Integer	Unsigned
			wordOthman	VarChar(120)	Binary Column Charset:ucs2 Column Collate:ucs2_general_ci
			wordArabic	VarChar(120)	Binary Column Charset:ucs2 Column Collate:ucs2_general_ci
			wordArabic2	VarChar(120)	Binary Column Charset:ucs2 Column Collate:ucs2_general_ci
			position	VarChar(45)	Binary Column Charset:ucs2 Column Collate:ucs2_general_ci
4	juz_amma_ suras	Sura	suraName	VarChar(255)	Binary Column Charset:ucs2 Column Collate:ucs2_general_ci
			suraNumber	Integer	Unsigned

* Note: Explanation of each ColumnName entry is the same as in the sub_root_word

In Table 7.2, we can see that the data type is either Integer or VarChar. Particularly, for the VarChar type the character set used is ucs2 which is the Unicode character set using 16 bits per character. This character supports the Basic Multilingual Plane, i.e. to encode the Arabic character that is being used in PaCQ interface. The ucs2_general_ci collation is used so programming codes for string comparison can be done on each Arabic character in the ucs2 character set. Pure Arabic Unicode character is not used for displaying *suras* in the Qur'an since this is a non-standard form of Qur'an writing, the *Othmaniy* font is the standard form. *Othmaniy* font can also help in remembering *Tajwid* rules (this should help in overcoming the problem mentioned in preliminary study 1). This font uses coded symbol that can help to remember *Tajwid* rules such as the diacritic for *tanween fatha* (ً), if the fatha (َ) is written perfectly as one on top of the other (such as written on top of the letter ض in the word قَرْضًا), it is read without nasalization. However, if the *fatha* (َ) is written slightly spaced out one from the other (such as written on top of letter ن in the word حَسَنًا) is read with nasalization. However, using the existing *Othmaniy* form that we have, causes some word and letter sequence problems (see Figure 7.2) Therefore we used the *Othmaniy* font for the Qur'an that was downloaded from the Qur'an complex website in Madinah to overcome sequence problems and remembering *tajwid* rule problem.

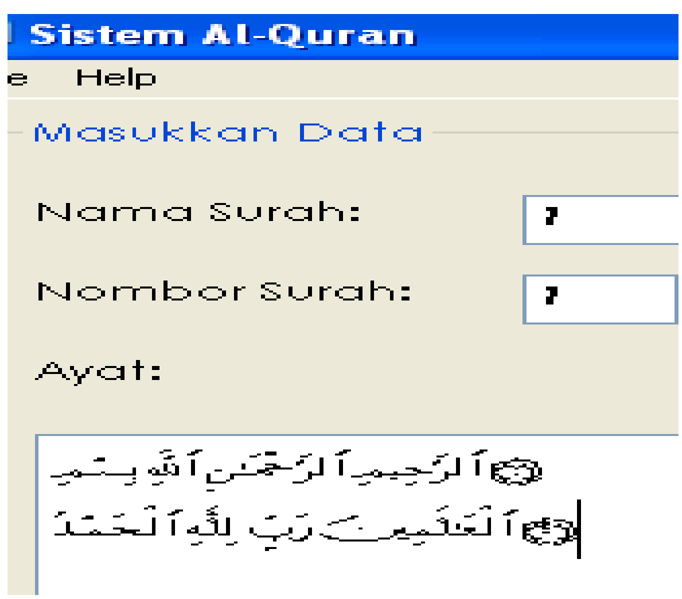


Figure 7.2: Word sequence problem using the existing Othmaniy font

7.2 The functional modules and interfaces

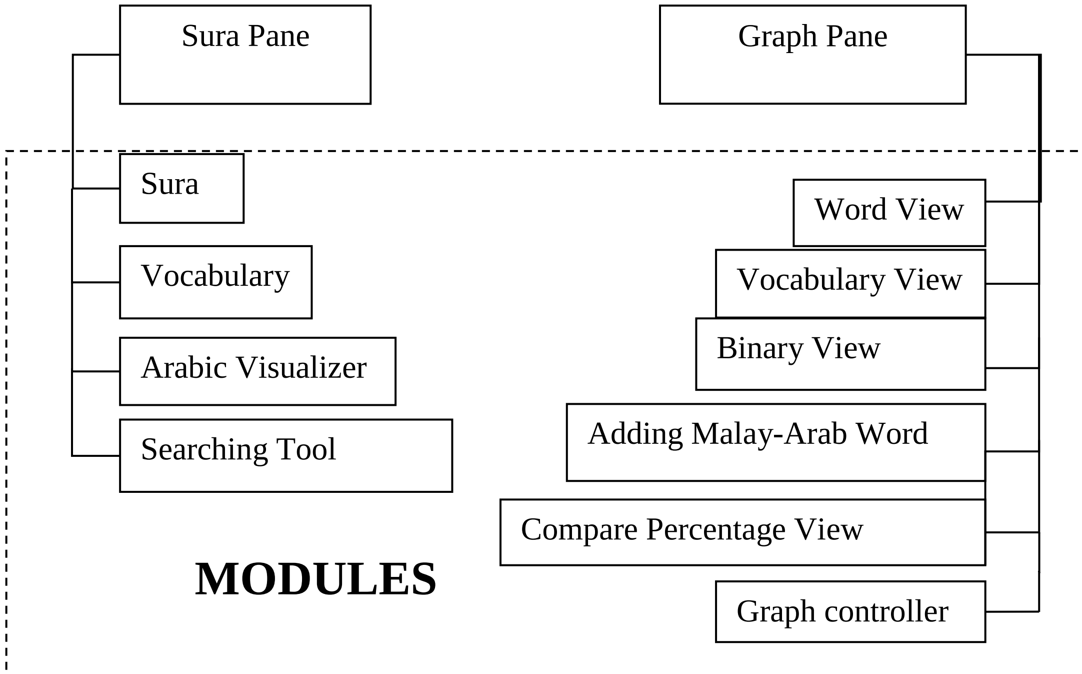


Figure 7.3 : Modules of PaCQ interface

The PaCQ interface contains two main parts, the *Sura* and the Graph Pane. Figure 7.3 shows the overall modules of the PaCQ interface and Figure 7.4 shows the window interface. The *Sura* Pane contains several modules:

i) *Sura* module :

- This is for displaying and choosing/searching *sura* in the 30th *Juz.* Figure 7.3 illustrates how this can be achieved by using the combo box and *sura* panel (refer to CBX and SP)

ii) Vocabulary module:

- This is for adding, saving and opening the vocabulary.
- Figure 7.5 shows how this can be achieved through the Add Vocabulary button and the menu item for saving and opening the vocabulary list.

- Another part of this module is the vocabulary highlighter, where words added to the vocabulary list are highlighted in each of the *sura* being displayed in SP.

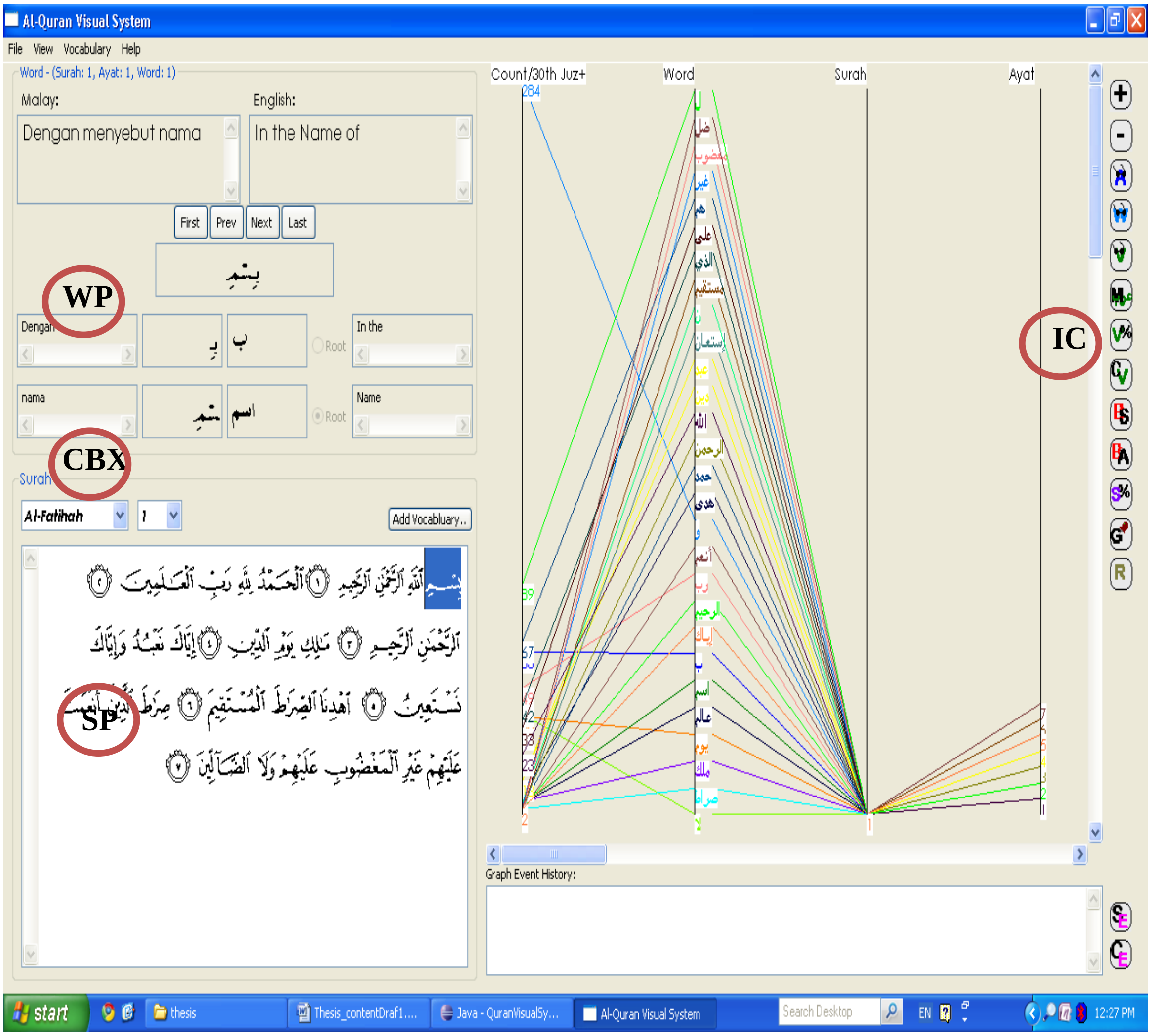
iii) Arabic word visualizer module:

- This is for displaying segmented Arabic words and its meaning. Figure 6.5 shows WP, the word panel segmenting the Arabic word

iv) Searching word tool:

- This is for searching words in Arabic, Malay language and English.

- Figure 7.6 illustrates the related dialog window. Searching is done by selecting the menu option shown in Figure 7.6a). When the searching option is clicked, a dialog window appears such as in Figure 7.6a) and c) where users can search the word via Arabic word, Arabic root word, Arabic sub word, Malay and English words. Once a searched word is entered the result is shown as in Figure 7.6 c) where the word *langit* (this Malay word means sky in English) is searched. When the Arabic word option is chosen, the text entered is searched from the Arabic word column in the database. Similarly, if the 'Root word' option is chosen, the text entered is searched from the root column, if 'Subword' then the subword column, if 'Translation (Malay)' then the Malay column and if 'Translation (English)' then the English column.



WP:Word Panel, **CBX**: Combo box, **SP**: Sura Panel, **IC**: Icons

Figure 7.4: Overview of the main PaCQ interface divided into the *Sura* and the *Graph* Pane

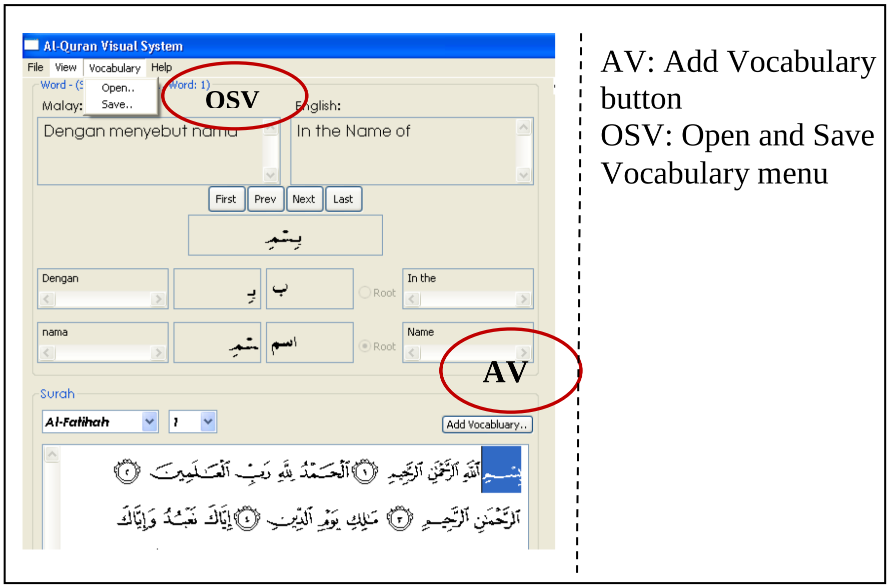


Figure 7.5: Adding, Saving and opening vocabulary

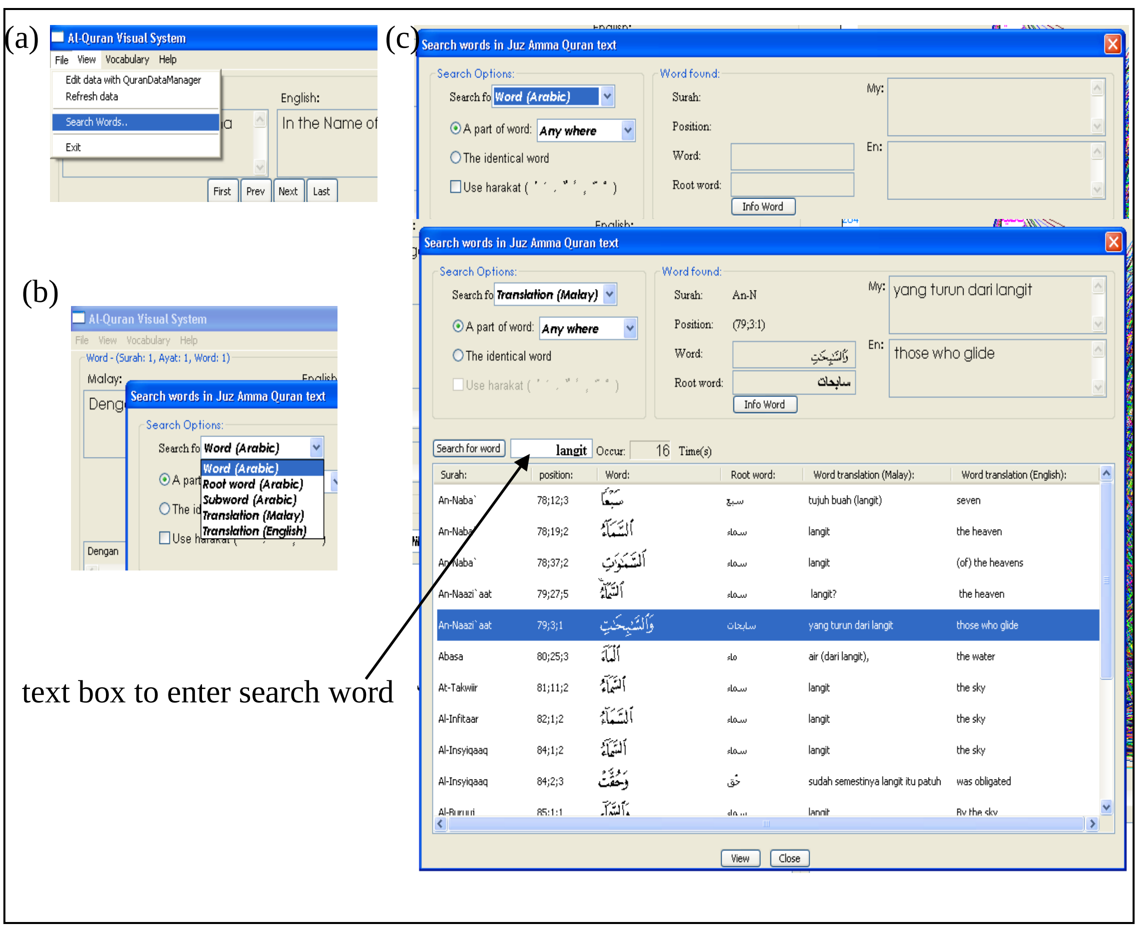


Figure 7.6: (a) The Search word option from menu (b) Several methods of searching (c) The result upon entering a search

The Graph Pane contains functions that can be divided into several module categories:

i) Word view module:

This is for viewing all words for overall global view and viewing a particular word. Figure 7.7 a) shows the view of all words and Figure 7.7 b) shows the view when the word بسم is selected from the *Sura* Panel.

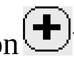
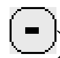
ii) Binary view module: for viewing word-*sura* axes and word –*ayah* axes

Figure 7.8 shows the word-*sura* view and word ayah binary views (Figure 7.8a) illustrate the binary view of words بِ, اسم, الله with *sura* axis and Figure 7.8b) illustrate the binary view of words بِ, اسم, الله with *ayah* axis).

iii) Vocabulary view module:

This is for adding, deleting the vocabulary list, and viewing the percentage of the vocabulary list within each *Sura* in the 30th *Juz* Figure 6.9 a) illustrate the interface when the word بسم is selected and added to the vocabulary list by clicking Add Vocabulary Button, and the Vocabulary list dialog window appears. Figure 6.9 b) shows the interface when the word الله is selected and added to the vocabulary list. Figure 7.9 c) shows the percentage view of the words بِ, ا سم, الله when added to the vocabulary list. It shows the percentage of these three words containing in each *sura*. When words are added to the vocabulary list, they are automatically highlighted in the surah panel for easy recognition. The list can be deleted by clicking the  icon.

iv) Adding Malay-Arab word module: for adding the Arabic words that have similar phonology and meaning in the Malay language to the vocabulary list. Figure 7.10a) shows the vocabulary view when the Malay-Arab word is added while Figure 7.10b) shows the corresponding vocabulary percentage view after adding the Malay-Arab word. Compare percentage view module: for viewing the percentage comparison of all words in one *sura* with all other *sura* in the 30th *Juz*. Figure 7.11a) shows the overall comparison percentage view between *sura*. Figure 7.11b) shows the view of percentage when a particular base *sura* is clicked. Figure 7.11c) shows the view when a particular percentage is chosen from percentage axis.

v) Graph controller module: for zooming in (can be achieved by clicking icon ) and out (can be achieved by clicking icon ) the graph; see Figure 7.12, refreshing the graph (can be achieved by clicking icon ), and pinning graph (can be achieved by clicking icon ).

Refer to Appendix D1- D6 for the detail interface manual.

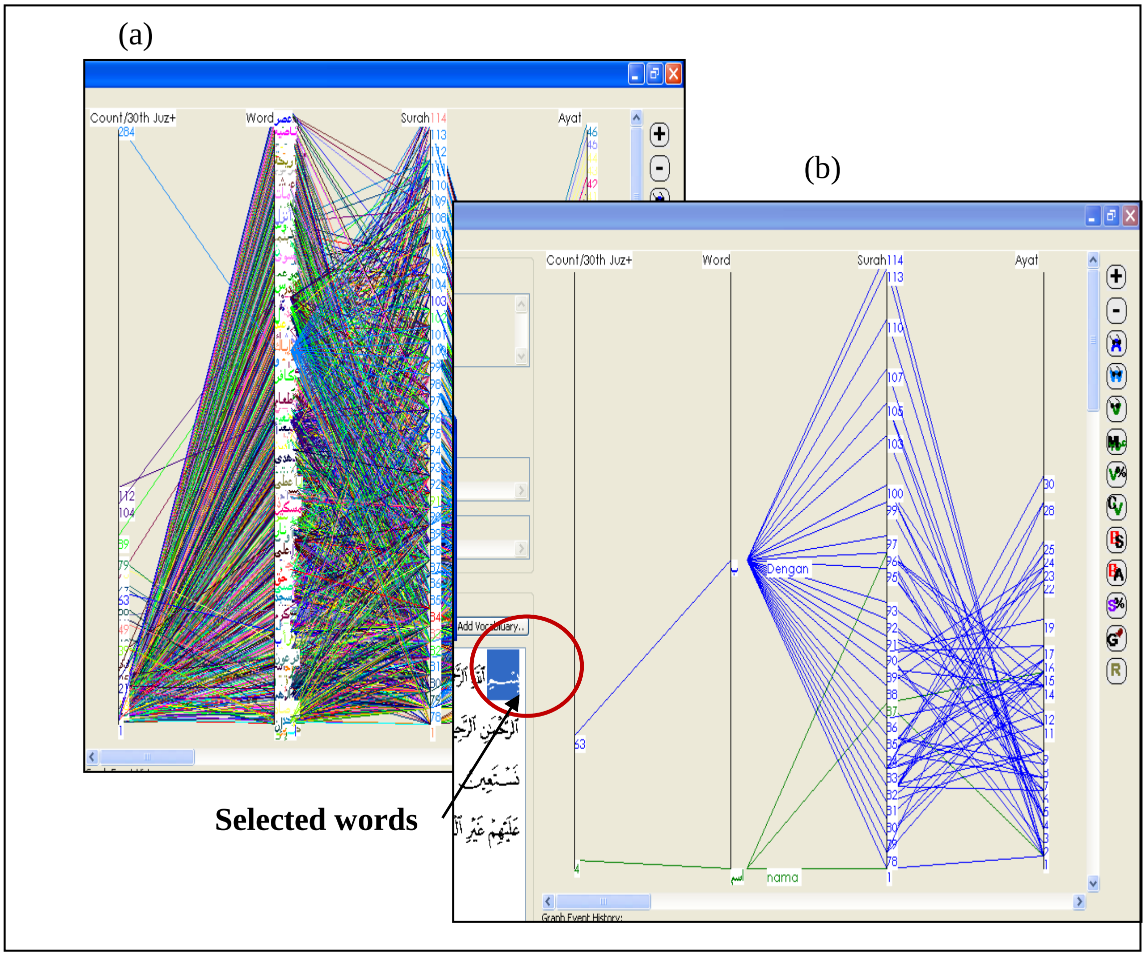


Figure 7.7: (a) Graph Area view for all words (b) Graph Area view when a particular word is selected

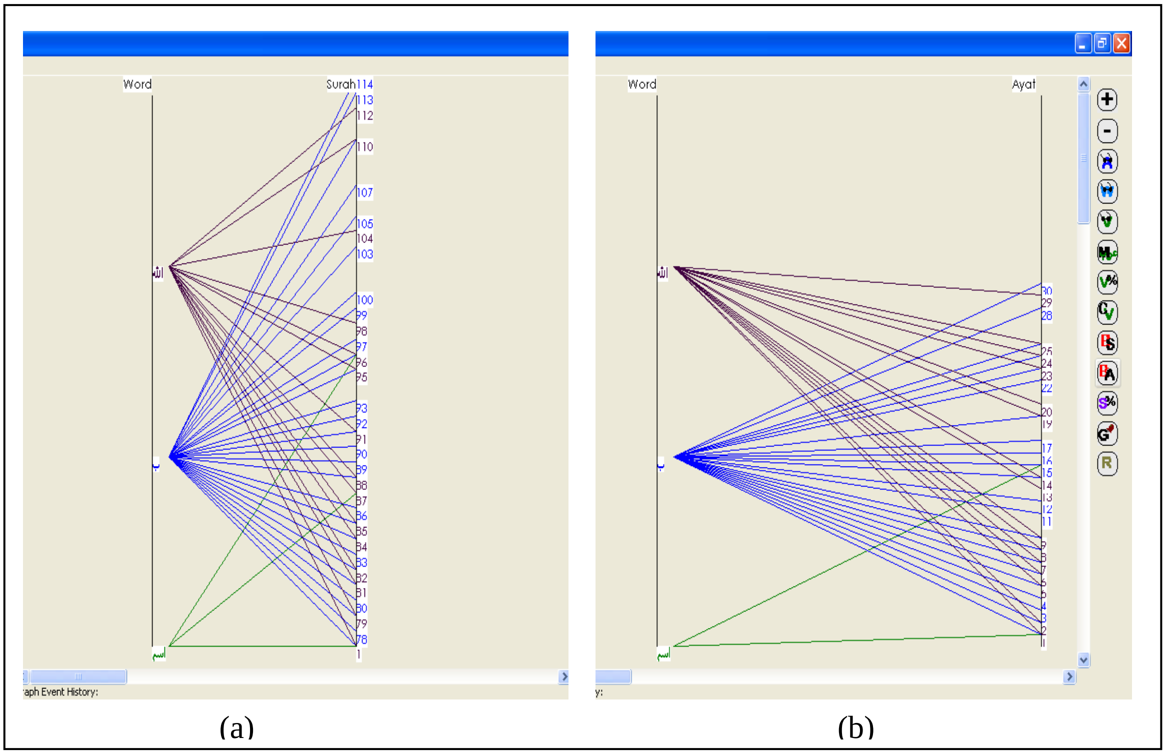


Figure 7.8: (a) Binary view of words with *sura* axis (b) Binary view of words with *ayah* axis

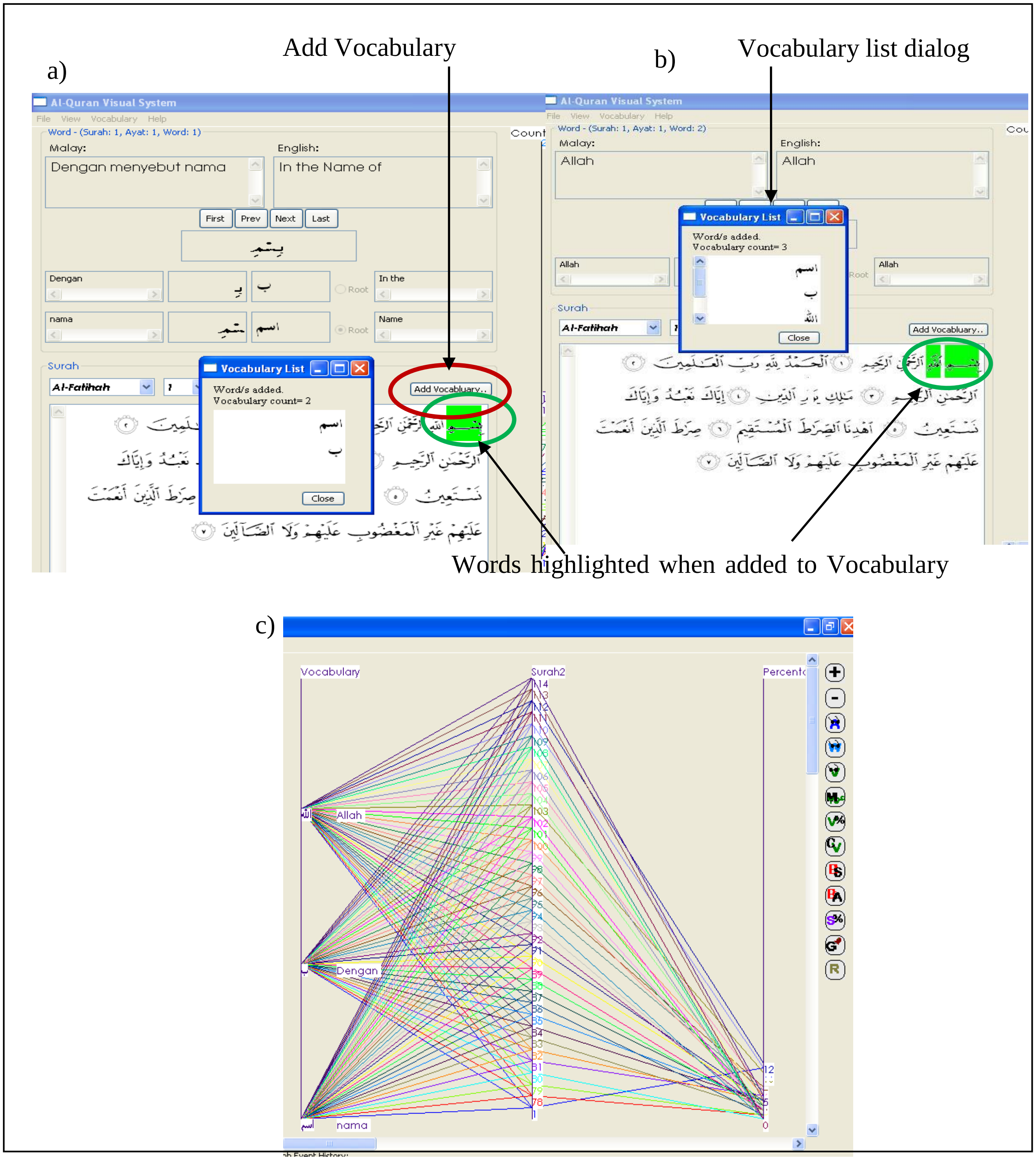


Figure 7.9:(a) Selecting and adding a word .to the vocabulary list (b) Selecting another word to the list (c) The percentage view of the words

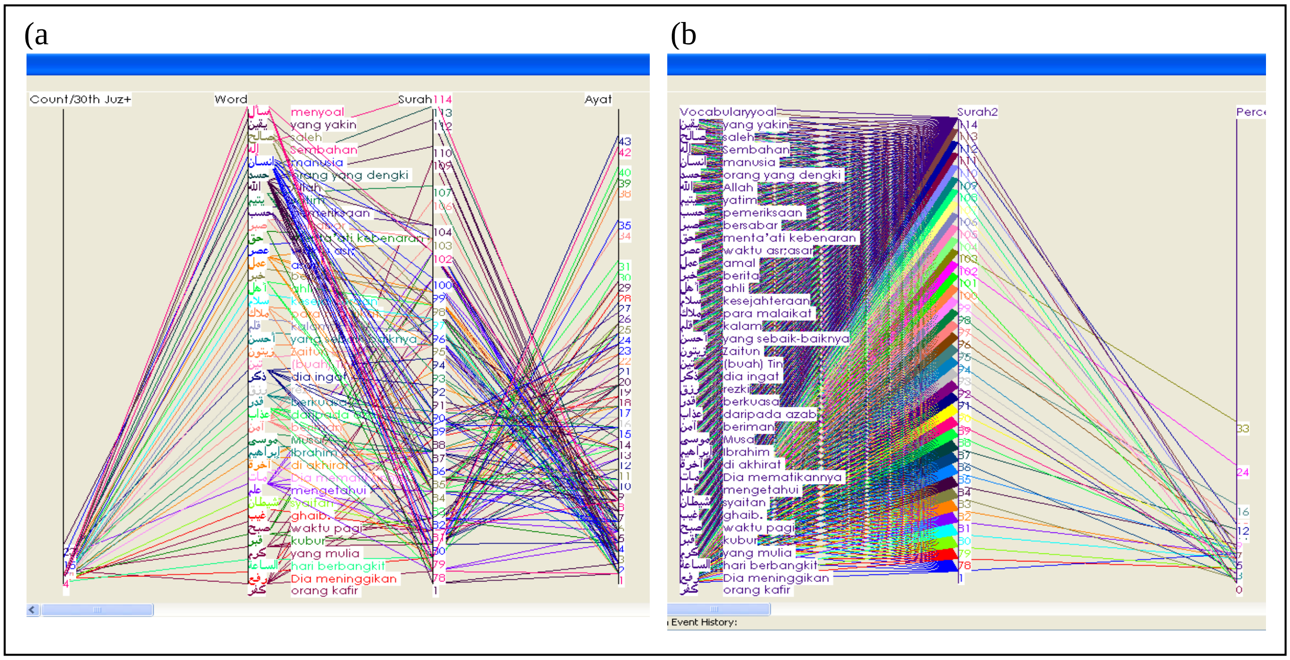


Figure 7.10: (a) The Malay-Arab word view (b) The percentage view of Malay-Arab word

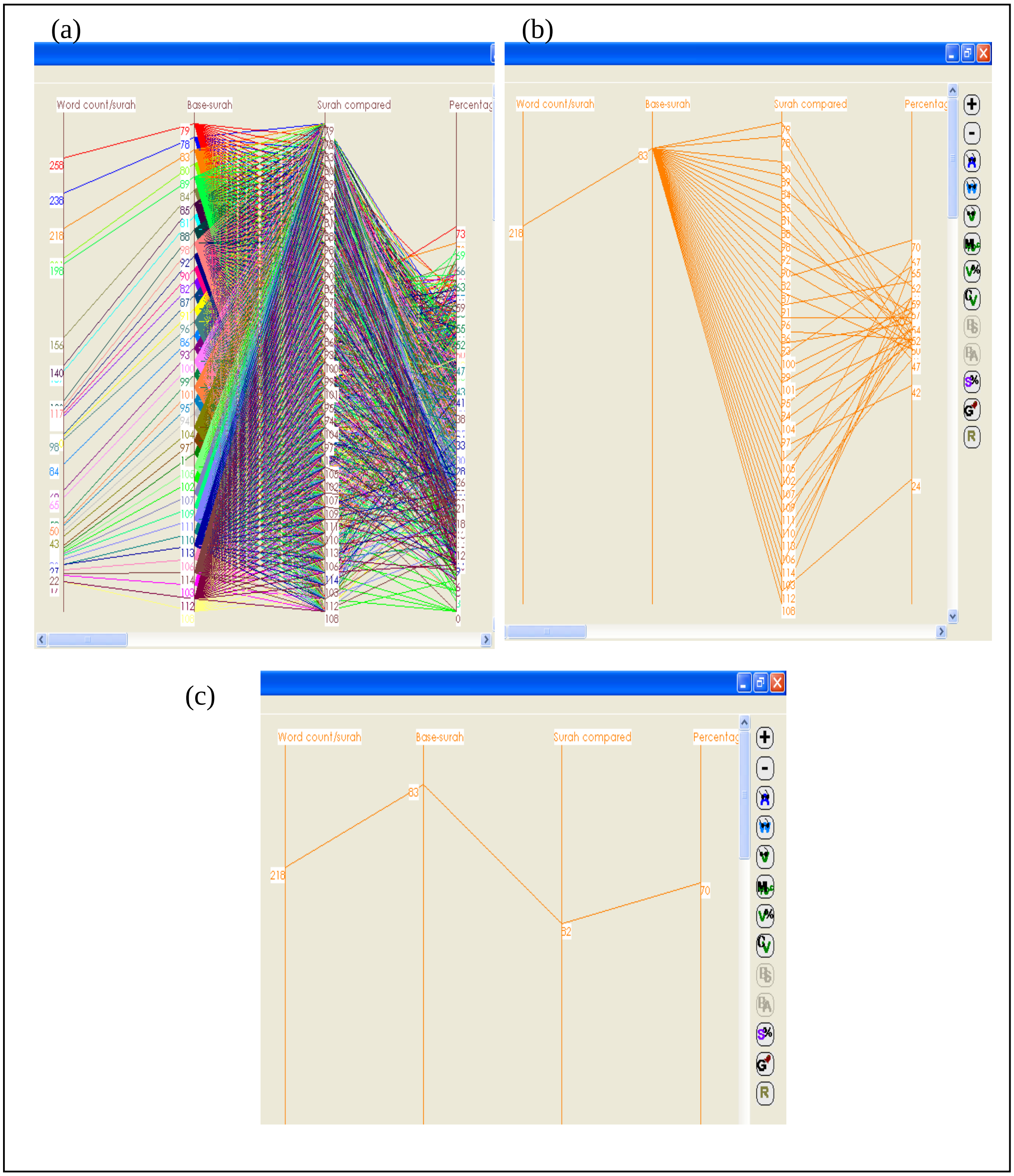


Figure 7.11:The comparison of percentage views a) overall b) a *sura* c) a percentage

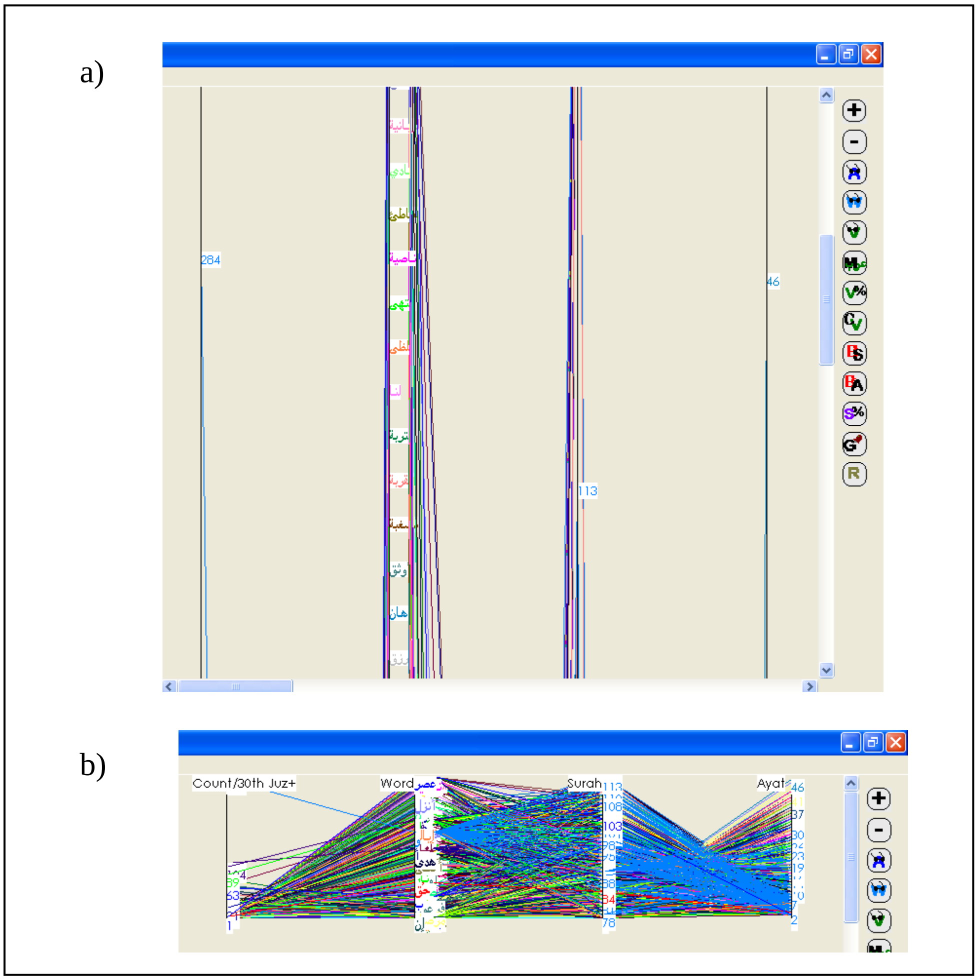


Figure 7.12 a) Graph in zoom-in view b) Graph in zoom-out view

The PaCQ interface has usability features incorporated to improve data exploration performance:

- **Automatic Brushing**: a set of attributes related to an Arabic word having the same colour (for example, the Arabic word and all the corresponding word count, the *sura* and the *ayah* are of the same colour). Therefore, users can easily detect the area of interest while performing data exploration. See Figure 7.7b).
- **Binary view**: to be able to detect interesting patterns such as positive or negative correlations; two axes should be placed adjacent to each other; the binary view function lets

users place the *sura* and the word axis and also the *ayah* and the word axis adjacent to each other (see Figure 7.8a) and b)).

- Filtering Operation: the Graph Area is set to be a clickable canvas so that users can select particular data of interest while performing data exploration. For example if a user double clicks on a particular percentage in the percentage axis of the compare percentage graph, all corresponding word count, the base *sura* and the *sura* compared are shown in the graph. Other data items are filtered out (see Figure 7.11 c))

- Highlighting: If a word is added to the vocabulary list, it is highlighted (see Figure 7.9).

- Zooming in and out: the Graph Area can be zoomed in and out to facilitate data exploration (see Figure 7.12).

7.3 The Implementation and algorithms

The PaCQ interface is implemented on a Dell, Vostro 1200 (T5670) notebook with Intel Core 2 Duo CPU and 1GB Memory. The software environment is based on Windows XP SP3 with Eclipse version 3.3.2 as the Integrated Development Environment. The Java language is used as the implementation language based on Java SE Runtime Environment version 1.6.0_02. MySQL is used as the database.

In the previous chapter 5, § 5.3.1, the declarations of the related variables are defined for A, A1, A2, V, C, S, Y and T.

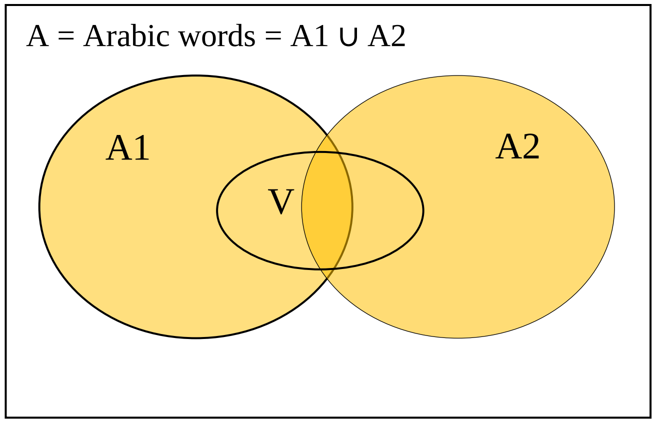


Figure 7.13: Venn diagram of A inrelation to A1, A2 and V for Arabic words

In addition:

- A can be as follows:
- $A \equiv M$ and $A \equiv E$, *where* $M = \{m \in$ the translated elements of A in Malay$\}$

and E= $\{ e \in$ the translated elements of A in English$\}$

- $A \equiv Ot$ *and* $A \equiv Md$ *where* $Ot = \{ot \in$ the font style of a in *Othmaniy*$\}$ *and*

$Md = \{md \in$ the font style of elements of A in Mushaf al-Madinah$\}$

- The *sura* name, $SN = \{sn \in$ *sura* name in *Juz Amma* $\cup$ *Sura1*$\}$
- The word number, W *where* $W = \{w: w \in \mathbb{N}^+, 1 \leq w \leq 30\}$
    - This is because the maximum number of words in one *ayah* is 30
- The sub word, B where $B \subseteq A$ (see also Figure 7.14):
    - A root word, $R \subseteq B$ where $R = \{r \in$ all root words of elements in $A\}$
    - A pre-fix word, $PRE \subseteq B$ where $PRE = \{pre \in A\}$
    - A suffix word, $SUF \subseteq B$ where $SUF = \{suf \in A\}$
- The sub word number, BN where $BN = \{ bn : bn \in \mathbb{N}^+, 1 \leq bn \leq 5\}$
    - Note that a word can contain up to 5 sub words.

- The word count in each *sura*, $WC = \{ w_c : w_c \in \mathbb{N}^+ , 1 \leq w_c \leq 283\}$
    - Note that the maximum of word count is 283 for the word و

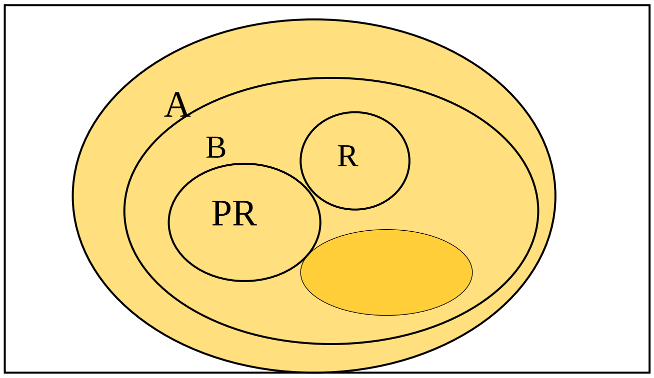


Figure 7.14 : Venn Diagram of A in relation to B, PRE, SUF and R

The main relations are as stated before in § 5.3.1

- R_1 = {(*c*, *a*, *s*, *y*): "There are *c* counts of Arabic word *a* in *Juz Amma* and can be found in *Sura s* and *Ayah y*"} where $c \in C$, $a \in A$, $s \in$ S and $y \in Y$.

- R_2 = {(*v*, *s*, *t*): "There is *T* percentage of the vocabulary words *V* in *Sura S*"} where $v \in V$, $s \in S$, t $\in T$.

- R_3 = { (c, s_1, s, t): There is *T* percentages of Arabic words in *Sura* S_1 compared to other *Suras S* in *Juz Amma* and *C* count of Arabic words in the *Sura S*} where $\{s_1$: P(s_1) } and P(s_1) is exactly one *Sura* in *Juz Amma* and $s \in S$, t $\in T$.

7.3.1 Algorithm based on PacQ User Interface (PUI) component

The PUI component is used to display the related information to users:

- Call the PaCQ Interface System class
- Load the *Qur'anic* Arabic data from *A1* and the related graphical user interface

- Get AWV component related classes

- Load the Qur'anic Arabic data in the form of parallel coordinate graph and the related graphical user interface
    - Get PCG component related classes

7.3.2 Algorithm based on Arabic Word Visualizer (AWV) component

The AWV component deals with the Arabic words information control functions. When the system is loaded, by default A1 is displayed in the *sura* panel. The default word/ *ayah*, Y is y =1. Users are able to view the meaning of each block of Arabic word a.

- Word Panel Operation:
    - Click next button in the Word Panel:
        - For $\forall A1$, get elements of M and E : Let this operation = Nx
            - For each element of A1 get $\forall$ *element of BN* in A, M and E
    - Click previous button will result in an inverse Nx operation, $\overline{Nx}$
    - Click last button, get last element of B
    - Click first button, get the first element of B
- Sura Panel Operation:
    - Choose combo box with one of element S or SN:
        - For chosen S, get $\forall$ corresponding element A, where Y =1 until Y=$Ymax$
- Search Words (SR) Operation:
    - Choose Arabic word option

        - Get $A = SR$, by comparing to ∀ elements of A
        - Then get the corresponding elements of SN, S, Y, BN, M, E and C ...(1)
        - Choose one of the listed elements of A, show in the Word Information panel
    - Choose Arabic root word option
        - Get $A = SR$, then compare to ∀ elements of A and if found get the corresponding R, then do as in (1)
    - Choose Arabic sub word option
        - Get $A = SR$, then compare to ∀ elements of A and if found get the corresponding B, then do as in (1)
    - Choose Malay translation option
        - Get $M = SR$, then compare to ∀ elements of M and if found get the corresponding M, then get the corresponding elements of SN, S, Y, BN, E and C
    - Choose English translation option
        - Get $E = SR$, then compare to ∀ elements of E and if found get the corresponding M, then get the corresponding elements of SN, S, Y, BN, M and C
- For adding vocabulary to the vocabulary list:
    - Get word clicked by user and put it in the vocabulary list
    - Highlight words in the list in the *Sura* Panel

7.3.3 Algorithm based on Parallel Coordinate Graph (PCG) component

When the PaCQ interface is started, apart from loading the elements of *A1*, the PaCQ interface also loads the parallel coordinate graph:

- First, ∀ elements of *B* will be accessed from the database and translated to a **list** of words that contains corresponding elements of *A*, *C*, *S*, *Y*, *W* - ...**(2)**
- Draw *C*, *A*, *S* and *Y* vertical axes (for relation R_1) equally spaced out with specific *x* coordinate for each: x_c, x_a, x_s, and x_y.
- Draw the full word list graph
- ▪ For $item_1$ until $item_{\max}$, from the word list in (2) above, a line is drawn in between x_c and x_a; x_a and x_s; x_s and x_y using the related values of elements in *A*, *C*, *S* and *Y*.
- The y coordinate changes for each value of elements in *A*, *C*, *S* and *Y* and the string label (of each value) is also drawn and positioned according to the y coordinate
- Elements of *A* values are arranged according to the index of word base list (non-repeated words) generated from word list (2).
    - The y_a coordinate =

      Graph Area Height (GH) - ($\frac{GH}{baselist\ size} \times index$)
    - Elements of *C* are according to their values
    - y_c coordinate = *GH* - ($\frac{GH}{C\ max} \times \text{the corresponding } C$)
    - Elements of *S* are according to their values
        - if *S* ==1: y_s coordinate = *GH*
        - if S ≥ 78: y_s coordinate= *GH* - ($\frac{GH}{Smax - S + 1} \times S - 77$)

- Elements of W are according to their values

- y_w coordinate= GH - ($\frac{GH}{Wmax+1} \times w$)

There are also other types of related graphs:

- The one word graph or the graph that shows a selected word information in the *sura* panel that can be associated to B, the sub word
  - Draw the graph for the selected word by using a filtered list that contain only the selected word
    - Compare selected B with ∀ words in list (2)
    - If found in list (2), get all corresponding A, C, S, Y, W and put in another list... **(3)**
    - Use list in (3) to draw graph

- The Vocabulary Graph: draw the graph using the vocabulary words list added by users (4)
- The Binary View Graph:
  - To view A and S axes, draw the graph by filtering C and Y axis and all related values
  - For viewing A and Y axis: draw the graph by filtering C and A axis and all related values
- The Vocabulary Percentage Graph (for relation R_2):
  - Get users vocabulary list in (4),
  - Compare ∀ element of V to each element of A in ∀ element of S and compute the corresponding T

- Put these in another vocabulary list
- Put also the corresponding elements of E and M in the list
- Draw the graph

- The All Percentage Graph (for relation R_3):
  - Compute the T value each elements of S in comparison to ∀ other elements of S
  - Put these in a list as defined in R_3 (5)
  - Draw the graph
    - Draw C, S_1, S and T vertical axis (for relation R_3) equally spaced out with specific x coordinate for each: x_c, x_{s1}, x_s, and x_t.
    - The y coordinate changes for each element of C, S_1, S and T values and the string of each values is also drawn and positioned according to the y coordinate. The values of elements S_1 and S are arranged according to A.
      - Values of elements in *S1*
        - The y_a coordinate =

          $GH - (\frac{GH}{s1max-78+4} \times surah\ index)$
      - Values of elements in S, y_s is the same as the above
      - Values of elements in C
        - y_{wc} coordinate =

          $GH - (\frac{GH}{Wcmax} \times w_c)$
        - t values
          - y_w coordinate= $GH - (\frac{GH}{100} \times t)$

The filtering of item based on mouse clicks algorithm are as the following:

- Get the clicked item coordinate
- Compare to the formula of the *x* and *y* coordinates
- Identify the item
- Filter the appropriate list according to the item above
- Redraw the graph

Other related algorithms:

- Colour information is added to the each list item
- Zooming in and out occur by altering the Graph Area Width and Graph Area Height
- The Malay-Arab word is identified and stored in a list of array. To draw:
- ▪ Put the array in the vocabulary list
- ▪ Call draw operation

7.3.4 Algorithm based on Intermediate Interface (II) component

The II component is responsible to provide the intermediate facility to execute operation related to other components through the menu option:

- Choose search item
    - ▪ Call the search dialog window
- Choose Save Vocabulary item
    - ▪ Call the Save Vocabulary window
        - Get user id and ∀ elements of *V*

- Store into database

- Choose Open Vocabulary item
  - Call the Open Vocabulary window
    - Get user id
      - put corresponding element of *V* in the vocabulary list
      - translate the list into parallel coordinate graph
- Choose to view Word Panel
  - Show Word Panel window above the *Sura* Panel
- Choose to hide Word Panel
  - Hide Word Panel window above *Sura* Panel
- Choose to exit system
  - Close and exit all PaCQ interface window

7.3.5 Algorithm based on DatabaseConnection class

This class is responsible for getting all related information from the database.

- First, establish connection with database
- To fill *sura* in the *Sura* Panel, search for element *S* requested
  - Get ∀ corresponding elements of *Y*, *W*, *Ot*, *MD* from table *Qur'an*
- To fill the elements of *SN* in the combo box list
  - Get ∀ all elements of *SN* and *S* from table *juz_amma_sura*
- To highlight selected elements of *A* in Graph
  - Get the corresponding elements of *S*, *Y* and *C*
- To get elements of *E* and *M*
  - Get the elements of *S*, *Y* and *C* from table meaning

- To get elements of *B* that are searched by a user, filter diacritics if needed
    - Get elements *of S, Y* and *C* from table sub_root_words
- To get elements of *A* that are searched by a user, filter diacritics if needed
    - Get ∀ corresponding elements of *S*, *Y*, *C*, *MD*, *E* and *M* that match the prefix, suffix and infix of elements of A in database from tables *sub_root_words*, *Qur'an*, meaning and *juz_amma_suras*.
- To get elements of *E* or *M* that are searched by a user
    - For elements of *E*: Get ∀ corresponding elements of *S*, *Y*, *C*, *OT*, *MD*, *E* and *M* that match the prefix, suffix and infix of elements of *E* in database from tables *sub_root_words*, *Qur'an*, meaning.
    - For elements of M: Get ∀ corresponding *S*, *Y*, *C*, *OT*, *MD*, *E* and *M* that matches the prefix, suffix and infix of elements of *M* in database from tables *sub_root_words*, *Qur'an*, meaning and *juz_amma_suras*.
- For plotting graph purposes,
    - Get ∀ elements of *BN* from table sub_root_words
    - Add elements of *BN* and the corresponding *S*, *Y*, *C*, *E* and *M* in a word List
    - Count all elements of *BN* in the word List and add the information in the same word List

7.4 Summary

This chapter discusses the design and implementation of the PaCQ interface. The design § identifies the classes involved and they are arranged into component diagrams which are organized into a 3-tier architectural style. There are 19 classes distributed into 6 components/ packages; the top level contains the PaCQ User Interface (PUI) component, the second level contains the Intermediata Interface (II), Arabic Word Visualizer (AWV) and Parallel Coordinate Graph (PCG) component and the third layer contains the DatabaseConnection package with a direct connection to the database.

Next the design of the database is discussed and subsequently followed by the functional properties of the PaCQ interface relating them to the screen shots of the system. The database contains four tables; the sub_root_word, the meaning, the Qur'an and the *juz_amma_suras* tables. The functional modules are divided into the Sura Pane module and the Graph Pane module. The Sura Pane module contains the Sura, Vocabulary, Arabic Visualizer and the Searching Tool. The Graph Pane, has the Word View, the Vocabulary View, the BinaryView, the Adding Malay-Arab Word View, the Compare Percentage View and the graph controller View.

Finally the algorithms involved in implementing the PaCQ interface are outlined based on the components described above.

8.0 EXPERIMENTATION AND EVALUATION

The PaCQ interface is specifically built to incorporate visualization with vocabulary tracking mechanism based on the hypothesis that it supports Arabic word recognition in the reading process. An experimental study was conducted on PaCQ to test this hypothesis. Apart from word recognition test, the user evaluation of PaCQ and the user evaluation on interpreting the parallel coordinate graphs were conducted. Specific discussions are regarding the following:

- the description of the experimental study carried out on the PaCQ interface
    - Arabic Word Recognition Test
    - User Satisfaction Survey
- the discussion on the results and findings of the above
- the description and results on the study of Interpreting Parallel Plot
- the interview after modification

8.1 Experimental study

8.1.1 The objectives of experimental study

The main aim of this experimental study is to test the hypothesis that visualization with vocabulary tracking mechanism supports Arabic word recognition in the reading process. The objectives of this experimental study are as follows:

i) To investigate whether the PaCQ interface has any significant performance impact to the non-Arabic speakers in recognizing the Arabic words in the following categories:

a. UIG only (within-participants) that will relate the performance of the AWRT:

- for each sets of questions in set 1, 2 and 3 before and after using the PACQ interface (see next part for the detail description)

b. Comparison between Control group (CG) and/or the Interface User Group (UIG) i.e. the experimental group (between-participant). This relates to the performance of the AWRT:

- for each set of questions in set 4, 5 and 6 (see next part for the detail description), and
- for the average score of questions in set 4, 5, and 6.

ii) To investigate the average score for the age group less than 35 years old (LEG) and the age group of at least 35 years old (MOG)

iii) To investigate how does time to complete the AWRT influence the scores

However, each set relates to components of the integrated model discussed in § 6.1.3. Each set of the AWRT in the post test is linked to the model as follows:

- Set 1A – WFE without emphasize of the Parallel Plot (PP)- (SM to STM)
- Set 2A – WSE with PP (SM to STM)
- Set 3A –Cognate Word for WSE with PP (SM to STM)
- Set 4- – WFE without emphasize of the Parallel Plot (PP) (SM to STM)
- Set 5- – Cognate Word for WSE with PP (SM to STM)

- Set 6 – Reinforce learning for Reading Comprehension in the LTM

8.1.2 Participants

In selecting the participants for the experimental study, 90 volunteers of 19-60 years old, participated in the study. 35 of the participants were 19-25 years old, 9 of them were 26-30 years old, 6 of them were 31-35 years old, 13 of them were 36-40 years old, 10 of them were 41-45 years old and 17 of them were more than 45 years old. Only those who can read but could not comprehend the meaning of the Qur'an were selected. The education background of participants are as follows: 4 with primary school as highest education, 18 with SPM (fifth form- secondary certificate), 8 with STPM (sixth form - secondary certificate), 8 with diploma, 34 with Bachelor, 12 with Master and 6 with PhD. Figure 8.1 shows the formal education and the perception of Arabic language of participants. Most participants were in the medium level of Arabic proficiency. This is the case for all education background except for the primary and STPM groups. There were 48 females and 42 males participants. There are two groups of participants the control group and the participants that had to go through the PaCQ interface (experimental group).

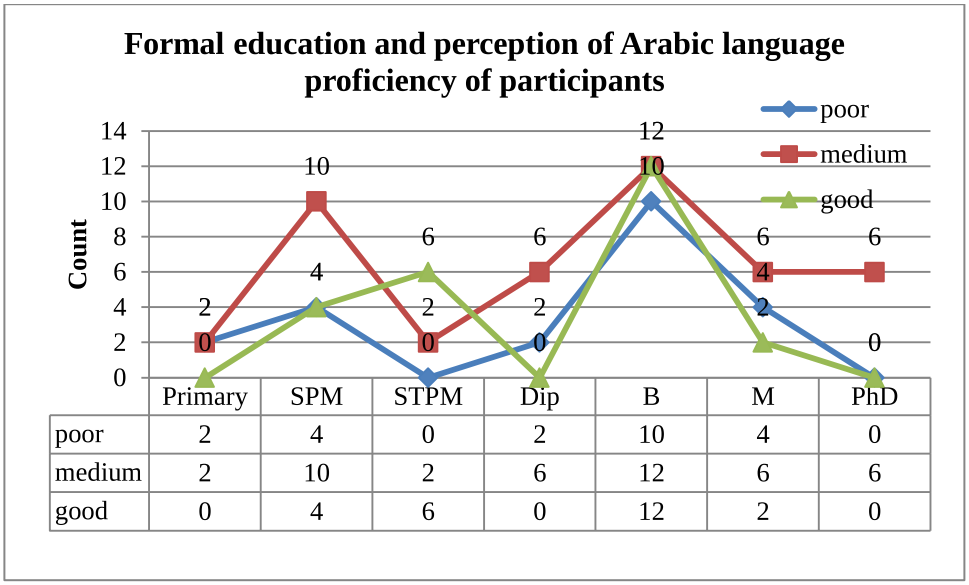


Figure 8.1: Formal education and perception of Arabic language proficiency of participants

The participants mentioned above went through the pre-test. They were carefully divided into several groups according to the pre-test order. This is discussed in § on pre-test ordering.

In the post-test, the same participants were involved. However, they were divided according to their scores in answering the pre-test into the control and experimental group. This is discussed in § on selection of post-test, post-test for the control group and experimental group

8.1.3 Experimental procedure

Both in the pre-test and post-test participants were given an Arabic Word Recognition Test (AWRT). The tests are related to three main word categories:

i) The most frequently occurring words

ii) All words containing in one *sura*

iii) The Malay-Arab words (cognate words)

Participants can only refer to each AWRT set of the above category while doing the test. This was an important rule applied in the entire test conducted. This include participants that used the PaCQ interface, they can only refer to the current interface involve in the AWRT not the whole PaCQ interface.

In order to test whether or not the PaCQ system has any significant impact on the Arabic word recognition, a number of tasks involving the functions of the PaCQ interface are created for the participants. Assessment activity is conducted via an Arabic Word Recognition Test (AWRT) sessions involving questions to be answered after the learning process. The AWRT was designed using selected *suras* from *Juz Amma* and also *sura al-Fatihah*. An experimental design approach is used in this experiment involving a control and experimental group. The control group did the AWRT without going through the PaCQ interface while the experimental group did the AWRT after using the PaCQ interface.

The experiment is divided into two parts:

i) The pre-test, to test the knowledge level of participants. This information is then used to divide participants into the experimental and control group

ii) The post-test, to test effect of usage of the PaCQ interface on the Arabic word recognition test.

The subsequent § discusses the experimental procedure according to the following:

- The pre-test

- The pre-test ordering
- Selection of the post-test (match-pair)
- The post-test
- Post-test for the control group
- Post-test for the experimental group
- The user satisfaction form

The Pre-test

The pre-test is to capture the participant's ability to recognize Arabic words and retrieve the meaning based on individual's memory. The pre-test are named Set 1,2 and 3, containing Arabic Word Recognition questions. Participants were told not to refer to other sets while answering questions for each set. For example in answering set 1, only set 1 can be referred to. Before the actual pre-pest, a pilot test was conducted on 3 people. Mistakes related to the printed text and procedure inefficiencies were detected.

Each participant was assigned to a set of forms which contain (see Appendix E):

a. The consent form for agreeing on carrying out this experiment – Appendix E 5-8

b. The information on the experiment carried out– Appendix E 3-4

c. The ID of the participant (to be filled by the experimenter) – Appendix E 9

d. Questions on the demography of the participants (to be filled by the participants)- Appendix E 9

e. Questions related to Arabic Word Recognition Test (to be filled by the participants) (S0-S3) – Appendix E 11-30

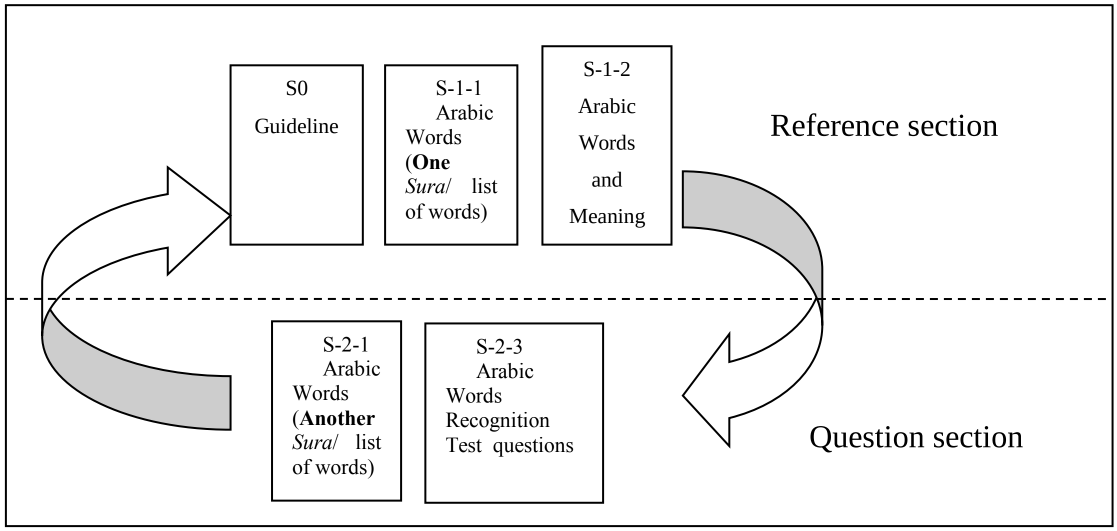


Figure 8.2: Pre-test procedure

Figure 8.2 illustrates the pre-test procedure. Each set contains the reference and question §s. The reference § is used to answer the question §. There are three sets of questions for AWRT, set 1, 2 and 3. The first set relates to the most frequently occurring words, the second relates to learning words in one *sura* and the third relates to cognate words. Each set contains five §s: the Arabic words or a selected *sura*, the meanings of the Arabic words or the selected *sura*, a second selected *sura* in Arabic, the meaning of this second selected *sura* (serves as a marking scheme) and lastly the question sheet to write answers to the meanings of each word in the second selected *sura*. The second selected *sura* must contain words from the first §s. Table 8.1 contains the detail description of the documents involved.

Table 8.1: Documents for the pre-test

No.	THE PRE-TEST DOCUMENTS
1	The general guideline of Arabic word recognition – Appendix E 11
2	**SET 1** : The questions related to the most frequently occurring words (Set 1)- Appendix E 12-18
	Set 1-1-1: the list of 14 words (2 verb words, 2 noun words, 10 words either pronouns, preposition or conjunctions)
	Set 1-1-2: the meaning of the 14 words listed above
	Set 1-2-1 : the *Sura* 107 containing the Arabic words
	Set 1-2-2 : the meaning of all words in *Sura* 107- the answer schema for the purpose of marking
	Set 1-2-3: the question sheet that contains all words in *Sura* 107 and blank spaces for the participant to fill in the meaning of the Arabic words in Malay language and their general understanding of the meaning of each *ayah.*
3	**SET 2:** The questions related to learning all words in one *Sura* (Set 2) – Appendix E 19-24
	Set 2-1-1: the *Sura* 1 containing the Arabic words
	Set 2-1-2: the meaning of all words in *Sura* 1
	Set 2-2-1 : the *Sura* 109 containing the Arabic words
	Set 2-2-2 : the *Sura* 109 containing the Arabic words with the meaning – the answer schema for the purpose of marking
	Set 2-2-3: the question sheet that contains all words in *Sura* 109 and blank spaces for the participant to fill in the meaning of the Arabic words in Malay language and their general understanding of the meaning of each *ayah.*
4	**SET 3:** The questions related to Malay-Arab words (Set 3) - Appendix E 25-30
	Set 3-1-1: the *Sura* 103 containing the Arabic words
	Set 3-1-2: the meaning of all words in *Sura* 103
	Set 3-2-1 : the *Sura* 95 containing the Arabic words
	Set 3-2-2 : the *Sura* 95 containing the Arabic words with the meaning – the answer schema for the purpose of marking
	Set 3-2-3: the question sheet that contains all words in *Sura* 95 and blank spaces for the participants to fill in the meaning of the Arabic words in Malay language and their general understanding of the meaning of each *ayah.*

Before starting the experiment, participants were briefed on the general guideline (appendix E 11), that a block of phrase in Arabic can contain several words and there are 10 letters of the Arabic alphabets that can be added to a verb word (in front, middle or end) to indicate slightly different meanings.

ii) Participants were asked to do set 1, 2 or 3 (S-1, S-2 or S-3) in a specified order

iii) For each set S, participants were asked to do the following steps (in order) as shown in Figure 8.2:

a. Read the material S-1-1 (containing only Arabic words) and learn the meaning of each word presented in the *Sura* / paper given from material S-1-2.

b. Read material S-2-1.

c. Answer questions related to AWRT of the *Sura* / paper given from material S-2-3. Participants were given the opportunity to refer back to material S-1-1, S-1-2 and S0 (the guideline).

The pre-test ordering

Table 8.2: The pre-test ordering

Gender/Age of Participant/ Targeted no	**Order 1**	**Order 2**	**Order 3**	**Order 4**	**Order 5**	**Order 6**	**SUB-TOTAL**
Male/ (<35 years)/ 25	1-2-3	1-3-2	2-1-3	2-3-1	3-1-2	3-2-1	20
Female (<35 years)/ 25	1-2-3	1-3-2	2-1-3	2-3-1	3-1-2	3-2-1	24
Male (>=35 years)/ 25	1-2-3	1-3-2	2-1-3	2-3-1	3-1-2	3-2-1	22
Female (>=35 years)/ 25	1-2-3	1-3-2	2-1-3	2-3-1	3-1-2	3-2-1	24
TOTAL							90

Table 8.2 shows the order of the test given to the participants. There are three sets of questions as explained earlier. The order of the AWRT given to the participant is as in the sequence 1-2-3, 1-3-2, 2-1-3, 2-3-1, 3-1-2 or 3-2-1 (See Appendix E 12-30). This is to nullify any confounding effects[16] that could be associated to the answers given by the participants if the ordering was not taken into consideration. Each order is targeted to four types of groups, Male and Female below 35 years and

[16] Confounding effect in this context is the performance of participants that is caused by the sequence in which the AWRT were given to them.

at least 35 years of age. Each group is arranged to consist 20-24 participants so to have a total of 90. This pre-test took about one month after which the answer scripts were marked to obtain the scores of the participants and results were analysed. Table 8.2 shows 20 males for the category below 35, 22 males for above 35 of age and 24 females for both age groups.

Selection of the post-test (match-pair)

The post-test is to find out whether there is any significant difference in the scores of AWRT between those who used the interface (UIG) and those who did not use the interface (CG). The AWRT pre-test answers were marked and the scores of the participants were recorded. Based on these marks, participants were divided into either the control or the experimental group by matching pairs of scores. Equal number of participants in the categories of above average (>=45%) and below average (< 45%) scores were targeted for both the experimental and the control groups to avoid any bias results. Table 8.3 shows this process of selecting participants in various categories according to experimental or control group, age, gender and also AWRT scores.

45 % of the pre-test score were chosen to be the cutting point between the above average and the below average since it was observed that the participant distribution would not be balanced if 50% was taken, differing only by 5% in normalizing the sample.

Table 8.3: The selection process of participants for the post-test

Experimental Group (45 participants)	Male/ (<35 years)	Above average AWRT score (>=45 %)
		Below average AWRT score (<45 %)
	Female (<35 years)/	Above average AWRT score (>=45 %)
		Below average AWRT score (<45 %)
	Male (>=35 years)/	Above average AWRT score (>=45 %)
		Below average AWRT score (<45 %)
	Female (>=35 years)	Above average AWRT score (>=45 %)
		Below average AWRT score (<45 %)
Control Group (45 participants)	Male/ (<35 years)	Above average AWRT score (>=45 %)
		Below average AWRT score (<45 %)
	Female (<35 years)/	Above average AWRT score (>=45 %)
		Below average AWRT score (<45 %)
	Male (>=35 years)/	Above average AWRT score (>=45 %)
		Below average AWRT score (<45 %)
	Female (>=35 years)	Above average AWRT score (>=45 %)
		Below average AWRT score (<45 %)

The post-test

The post-test involves one set of general guidelines of Arabic word structure (the same as the one given in the pre-test) and three sets of the AWRT questions. The control and the experimental group consist of participants that had gone through the pre-test. The control group had to only answer paper-based questions (the AWRT, similar to the one in the pre-test) while the experimental group had to go through a training session using the PaCQ interface and also answer the same paper-based questions given to the control group.

Post-test for the control group

The post-test procedure for the control group is similar to the pre-test shown in Figure 8.1. There are three sets of questions, however, with slightly lesser documents involved (the three sets of questions for AWRT post-test, set 4, 5 and 6). The first set relates to the most frequently occurring words, the second relates to cognate words and the last relates to all words that had been introduced to the participants from set 1-5 (which include words in the pre-test). Each set contains three §s: a list of Arabic words and the meanings, a selected *sura* in Arabic with the meanings (serves as a marking scheme) and lastly the question sheet to write answers to the meanings of each word in the selected *sura*. The selected *sura* must contain words from the first §s.

Table 8.4: Post-test order for the control group

Gender/Age of Participant	Order 1	Order 2	Sub-total
Male/ (<35 years)	4-5-6	5-4-6	10
Female (<35 years)	4-5-6	5-4-6	9
Male (>=35 years)	4-5-6	5-4-6	12
Female (>=35 years)	4-5-6	5-4-6	14
TOTAL			45

Table 8.4 shows the post-test order of the control group. There are 10 males and 9 females below 35 years. While for the category 35 and above, there are 12 males and 14 females. Table 8.5 contains the detail description of the documents involved.

Table 8.5: Documents for the post-test in the control group

No	THE POST-TEST DOCUMENTS IN CONTROL GROUP
1	The general guidelines of Arabic word recognition – Appendix F 4
2	SET 4: The questions related to the most frequently occurring words (Set 4)- Appendix F 5-8
	Set 4-1: the list of 4 Arab words with meanings in Malay language (nouns, pronoun, preposition or conjunctions)
	Set 4-2: *Sura* 108 with the meaning of all words - the answer schema for the purpose of marking
	Set 4-3: the question sheet that contains all words in *Sura* 108 and blank spaces for the participant to fill in the meaning of the Arabic words in Malay language. It is also a test of participants' general understanding of the meaning of each *ayah* in *Sura 108.*
3	SET 5: The questions related to Malay-Arab words (Set 5) - Appendix F 9-12
	Set 5-1: the list of Malay-Arab words with meanings in Malay Language
	Set 5-2: *Sura* 102 with the meaning of all words - the answer schema for the purpose of marking
	Set 5-3: the question sheet that contains all words in *Sura* 102 and blank spaces for the participant to fill in the meaning of the Arabic words in Malay language. It is also a test of participants' general understanding of the meaning of each *ayah* in *Sura 102.*
4	SET 6: The questions related to all words that had been introduced to the participants during the AWRT (from set 1-5). These questions (set 6) also assess the ability of participants in recognizing the Arabic words and retrieving the meanings from memory. - Appendix F 13-15
	Set 6-2: *Sura* 110 with the meaning of all words and the answer schema for the purpose of marking
	Set 6-3: the answer sheet that contains all words in *Sura* 110 and blank spaces for the participant to fill in the meaning of the Arabic words in Malay language. It is also a test of participants' general understanding of the meaning of each *ayah* in *Sura* 110.

The participants were also asked to record the start and end time of each AWRT (Set 4-3, 5-3 and 6-3).

Post-test for the experimental group

The experimental group had to answer 2 sets of the AWRT questions and a PaCQ evaluation form after using the PaCQ interface. The PaCQ interface was installed in Vostro 1310 laptops. The Windows Systems was set at a 1280 ×800 pixel resolution and highest (32 bit) color quality. The experimental group interacted with the system using the mouse. The session was conducted in a maximum of 5 persons per session in a lab environment.

Approximately 20 sessions of the post-test were conducted until all the targeted participants had gone through the training. Each post-test session took approximately 90-120- minutes and ended up within one month. The experimental group post-test was divided into three parts and were given the following:

i) approximately 60 minutes of training to use the PaCQ interface before the AWRT was conducted. The user manual of PaCQ interface was given to the group. Details of this training session are given in the next §. – Appendix D

ii) two sets of test given to the participants after the training session as the following:

- the same AWRT set of questions as the pre-test from Set 1-3 (given straight after the training session took place), see § 8.1.4 - Appendix G 2-11
- the new set of questions similar the control group, set 4-6 (given after performing tasks associated to the PaCQ interface) –Appendix G 14-23

iii) a user satisfaction form of the PaCQ interface- Appendix G 24

- covering aspects such as user ratings on word recognition, vocabulary building, task satisfaction, efficiency and effectiveness of visualization interface, various functions of the visualization interface and other opinion of the users.

After the training session, the post-test was continued with answering the same sets of questions as in the pre-test after using the PaCQ interface. In this set, the participants were asked to learn and remember the meaning of the Arabic words, then add them to the vocabulary list which is part of the PaCQ interface storage function. The Arabic words to be added are exactly the same as the ones in the pre-test. The question set was given a different name set 1, 2 and 3 A&B (see Appendix G 2-11). This is to avoid confusion of the data in the analysis stage. Table 8.6 shows the details description of the documents involved.

Straight after the AWRT set 1A-3B (as above) had been completed, participants were given the choice to stop or continue the testing process. Most of the participants continued, a few stopped and came back another time to complete the test. In this part 3 of the test, the participants were asked to learn and remember the meaning of the Arabic words using the PaCQ interface by referring to Set 4, 5 and 6 (see details in Table 8.6. This is done through the *Sura* Area, whenever an Arabic block word/phrase is selected in the *Sura* Textbox, the meaning of each segmented word in the block will be shown in the Word Panel above the *Sura* Textbox. Then the words can be added to the vocabulary list. The ‘add Malay-Arab Word icon’ can also be used to add those words in the vocabulary list. The AWRT questions are the same as in the post-test in the control group.

Table 8.6: Documents for the post-test in the experimental group

NO	THE POST-TEST DOCUMENTS IN EXPERIMENTAL GROUP
	PART 1
1	User Manual of PaCQ interface Appendix D
	PART 2 AWRT DOCUMENT
2	The general guidelines of Arabic word recognition – Appendix G1
3	Set 1A: the list of 14 Arabic words to be added to the vocabulary list (Appendix G 2)
4	Set 1B: the question sheet that contains all Arabic words in *Sura* 107 and blank spaces for the participant to fill with participants' general understanding of the meaning of each *ayah/verse.* (Appendix 4-5)
5	Set 2A: the *Sura* 1 Arabic words with vocabulary and word percentage (Appendix G 6)
6	Set 2B: *Sura* 109 question sheet (Appendix G 7-8)
7	Set 3A: the *Sura* 103 Arabic words (Appendix G 9)
8	Set 3B: *Sura* 95 question sheet (Appendix G 10-11)
	PART 3 AWRT DOCUMENTS
9	Set 4-1: the list of 4 Arabic words to be added to the vocabulary list (Appendix G 16)
10	Set 4-3: the question sheet that contains all words in *Sura* 108 and blank spaces for the participant to fill in the meaning of the Arabic words in Malay language with participants' general understanding of the meaning of each *ayah/verse.* (Appendix G 17-18)
11	Set 5-1: the list of Malay-Arab words added to the vocabulary list by clicking the Malay-Arab icon (G Appendix 19)
12	Set 5-3: *Sura* 102 question sheet (Appendix G 20-21)
13	Set 6-3: *Sura* 110 question sheet (Appendix G 22-23)
	PART 4 USERS OPINION
14	User satisfaction form (Appendix G 24-25)

The participants were also asked to record the start and end time of each AWRT (Set 4-3, 5-3 and 6-3).

Table 8.7: The post-test order

Gender/Age of Participant	Order 1	Order 2	Sub-total
Male/ (<35 years)/ 25	4-5-6	5-4-6	10
Female (<35 years)/ 25	4-5-6	5-4-6	15
Male (>=35 years)/ 25	4-5-6	5-4-6	10
Female (>=35 years)/ 25	4-5-6	5-4-6	10
TOTAL			45

Table 8.7 shows the order of the test given to the experimental group participants in the post-test. There are three sets of questions as explained in the question part. The order of the AWRT given to the participant is 4-5-6 or 5-4-6. Only set 4 and Set 5 were interchanged since set 6 is meant to see whether the participants could recognize and retrieve the words and the meaning that had been introduced to them in the AWRT before this point of time. Each order is targeted to four types of groups, Male and Female below 35 years and 35 years and above. There are 10 males and 15 females below 35 years. In the category of 35 and above there are 10 males and females.

The user satisfaction form

At the end of the AWRT, the participants were asked to fill in the user satisfaction form for the word recognition and vocabulary development in Qur'anic Arabic language (Appendix G 24). The questions asked were on two main issues:

i) whether the tasks they performed while using the PaCQ interface are meaningful to them, have characteristic of being achieving satisfaction

ii) whether the various functions provided in the PaCQ interface are good enough to perform the tasks. Such as the 'View All', 'View Vocabulary', 'View Binary', 'View vocabulary percentage', and others listed in the user manual.

8.2 Results

8.2.1 The user performance results

Before extracting the results using SPSS software version 17, the data sets to be analyzed need to be tested for normality. Table 7.8 shows the Kolmogorov-Smimov and the Shapirpo-Wilk tests of normality. All data sets are significantly different from the normal distribution except for the Average AWRT Score for Set 4,5 6 (post-test). This research uses the non-parametric test for the data sets which contain non-normal distribution. The original type of data is in scale form since the AWRT test scores are just the marks of the participants answering the test. Before proceeding to the non-parametric test, the data are changed to ordinal type of data.

Table 8.9 describes the transformation. The scale data for AWRT scores on set 1, 2, 3, 4, 5 and 6 are transformed into a ranked form between 1 to 5. Score from 0-20% is ranked with number 1 for very poor, 21-40% is ranked 2 for poor and so on. Data related to the time taken to complete set 4, 5 and 6 are transformed into a ranked form between 1 to 5. Time taken from 1-4 minutes is ranked very fast, 5-9 minutes is ranked fast and 10-14 minutes is ranked average and so on.

Table 8.8: The Kolmogorov-Smimov and the Shapiro-Wilk tests of normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
AWRT Set 1 Score Percentage	.184	44	.001	.913	44	.003
AWRT Set 2 Score Percentage	.189	44	.000	.862	44	.000
AWRT Set 3 Score Percentage	.175	44	.002	.853	44	.000
AWRT Set 1 Score Percentage for PaCQ Group	.130	44	.060	.922	44	.005
AWRT Set 2 Score Percentage for PaCQ Group	.301	44	.000	.747	44	.000
AWRT Set 3 Score Percentage for PaCQ Group	.140	44	.031	.920	44	.005
AWRT Set 4 Score Percentage(Post-test)	.177	44	.001	.916	44	.004
AWRT Set 5 Score Percentage(Post-test)	.330	44	.000	.804	44	.000
AWRT Set 6 Score Percentage(Post-test)	.135	44	.043	.913	44	.003
Postest_set4_T	.226	44	.000	.850	44	.000
Postest_set5_T	.169	44	.003	.826	44	.000
Postest_set6_T	.183	44	.001	.862	44	.000
Average time to complete test	.143	44	.025	.885	44	.000
Average AWRT Score for Set 4,5 6 (Post-test)	.070	44	0.200*	.989	44	.941

Table 8.9: The transformation of data from scale to ordinal

Data type	Transformed data	Description
1. Scale data for AWRT scores on set 1,2,3, 4,5 and 6	Ordinal data with rank between 1 to 5	1- very poor (scores 0-20%) 2- poor (scores 21-40%) 3- average (scores 41-60%) 4-good (scores 61-80%) 5- very good (scores 81-100%)
2. Scale data for the time to complete set 4, 5 and 6 of AWRT 3. Scale data for the average time to complete set 4, 5 and 6 of AWRT	Ordinal data with rank between 1 to 5	1- very fast (time 1-4 minutes) 2- fast (time 5-9 minutes) 3- average (10-14 minutes) 4- slow (15-19 minutes) 5- very slow (20- 24 minutes)

The parametric independent t-test is used to analyse the Average AWRT Score for Set 4, 5 6 (post-test). Significance level of 0 .05 are used for all statistical tests. In the results § the control group (CG) refers to those participants who did the Arabic Word Recognition Test (AWRT) referring only to paper printed information while the experimental group (UIG) refers to participants who used the PaCQ interface for the AWRT.

There are four test used to analyse the data:

- Wilcoxon test
    - this test is for testing paired samples of non-parametric data
- Mann-Whitney U test
    - This test is for testing independent samples of non-parametric data
- the independence t-test

- This test is for testing independent samples of parametric data
- the Pearson correlation test
  - this test is for testing correlation between variables

The results are as follows:

1. A Wilcoxon test was conducted to evaluate whether the AWRT scores for set 1, 2 and 3 after using PaCQ interface showed any improvement in the scores. The results indicate significant difference for all sets 1, 2 and 3 with details as follows:

   - The result for set 1 of the AWRT scores after using PaCQ is significantly higher with $z = -2.27$, $p < 0.05$ ($p = 0.023$). The mean rank of Set 1 AWRT scores after using the PaCQ interface is 18.14, while the mean rank of the Set 1 AWRT scores before using the PaCQ interface is 13.36.
   - The result for set 2 after using PaCQ is significantly higher with $z = -3.89$, $p < 0.05$ ($p = 0.000$). The mean rank of Set 2 AWRT scores after using the PaCQ interface is 11.25, while its mean rank before using the PaCQ interface is 6.00.
   - The result for set 3 of the AWRT scores after using PaCQ is significantly higher with $z = -2.24$, $p < 0.05$ ($p=0.025$). The mean rank of Set 3 after using the PaCQ interface is 12.56, while its mean rank before using the PaCQ interface was 12.33.

2. The Mann-Whitney U test was conducted to test the significance for set 4, 5 and 6 between CG and the UIG. It is found that there is significant difference on the

average rank in the AWRT ranked scores for set 4 only and no significant difference for set 5 and set 6.

- Set 4 AWRT scores for the UIG is higher (*Mdn*=5) than that of the CG (*Mdn*=4) with U=723, p = 0.012, Z=-2.505.
- Set 5 AWRT scores for the group that uses PaCQ (*Mdn*=3) do not differ from the control group (*Mdn*=3), with U=949, p = 0.543, Z=-0.543
- Set 6 AWRT scores for the group that uses PaCQ (*Mdn*=3) do not differ from the control group that (*Mdn*=2), with U=831 p = 0.133, Z=-1.501.

3. The Mann-Whitney U test was conducted to test the significance for 'Age' of AWRT Scores for set 1, 2 and 3. It was found that there is no significant difference (for all sets) on the average rank of the scores of AWRT ranked scores between LEG (age < 35 years) and MOG (age >= 35 years).

- Set 1 AWRT scores for age LEG (*Mdn*=4) do not differ from MOG (*Mdn*=4), with U= 250.0, p = 1, z=0.0
- Set 2 AWRT scores for age group LEG (*Mdn*=5) do not differ from MOG (*Mdn*=5), with U= 198.5, p = 0.167, z=-1.383.
- Set 3 AWRT scores for age LEG (*Mdn*=4) do not differ from MOG (*Mdn*=4), with U= 234.0, p = 0.697, z=-0.389

4. The Mann-Whitney U test was conducted to test the significance for Age of AWRT Scores for set 4, 5 and 6 in the Control group (CG). It is found that there is

no significant difference (for all set) on the average rank of the scores of AWRT ranked scores between LEG and MOG.

- Set 4 AWRT scores for LEG (*Mdn*=5) do not differ from MOG (*Mdn*=3.5), with U= 194.5, p = 0.208, z=-1.26.
- Set 5 AWRT scores for LEG (*Mdn*=3) do not differ from MOG (*Mdn*=3), with U= 221.5, p = 0.523, z=-0.639.
- Set 6 AWRT scores for LEG < 35 years (*Mdn*=3) do not differ from MOG (*Mdn*=2), with U= 234.0, p = 0.759, z=-0.306.

5. The Mann-Whitney U test was conducted to test the significance for ‘Age’ of AWRT Scores for set 4, 5 and 6 in the PaCQ group (UIG). There is no significant difference (for all sets) on the average rank of the scores of AWRT ranked scores between LEG and MOG.

- Set 4 AWRT scores for LEG (*Mdn*=5) do not differ from MOG (*Mdn*=4.5), with U= 222.0, p = 0.504, z=-0.668.
- Set 5 AWRT scores for LEG (*Mdn*=3) do not differ from MOG (*Mdn*=3), with U= 185.0, p = 0.122, z=-1.548.
- Set 6 AWRT scores for LEG (*Mdn*=3) do not differ from MOG (*Mdn*=2.5), with U= 185.0, p = 0.123, Z=-1.544.

6. The Mann-Whitney U test was conducted to test the significance for the average time to complete set 4, 5 and 6 between CG and UIG. The average time to complete set 4, 5 and 6 for UIG (*Mdn*=2) are faster compared to the CG (*Mdn*=2), with U=542.0, p = 0.006, z=-2.769.

7. The Independence t-test was conducted to test the significance difference of the Average AWRT percentage scores for set 4, 5 and 6 between CG and UIG. The Average AWRT scores for set 4, 5 and 6 for UIG (M =64.44, SD =21.58) are higher compared to CG (M =54.85, SD =23.55), t (88) =-2.014, $p<0.05$ (p=0.047).

8. The Mann-Whitney U test was conducted to test the significance for the average time to complete set 4, 5 and 6 for age <35 and age 35 and above. The average time to complete set 4, 5 and 6 for the age < 35 (*Mdn*=2) is faster compared to the age group 35 and above (*Mdn*=2), U=542.0, p = 0.005, z=-2.785.

9. The Mann-Whitney U test was conducted to test the significance for the average time to complete set 4, 5 and 6 in CG for age <35 and age 35 and above. The average time to complete set 4, 5 and 6 for the age < 35 (*Mdn*=2) is faster compared to the age group 35 and above (*Mdn*=2), U=104.0, p = 0.002, z=-3.144.

10. The Mann-Whitney U test was conducted to test the significance for the average time to complete set 4, 5 and 6 in the UIG for age <35 and age 35 and above. The average time to complete set 4, 5 and 6 for the age < 35 (*Mdn*=2) does not differ significantly from the age group 35 and above (*Mdn*=3), U=175.5, p = 0.922, z=-0.098.

11. For the test of correlation between the perceived personal knowledge (participant's opinion of their knowledge level of Arabic language) and the observed level of knowledge (the scores obtained for the AWRT) is found that these two knowledge were positively correlated, Pearson's r (86) =0.443, $p < 0.001$.

8.2.2 The user satisfaction results

The users of the PaCQ interface were asked to rate the statement of tasks and the functions provided (shown below). The rating 1: for Strongly Disagree, 2: for Disagree, 3: for Neutral, 4: for Agree and 5: for Strongly Agree. The percentage of the rating and the average are shown in Table 8.10. For the positive type questions (1, 2 and 3), the average ratings are at least 3.5, showing agreement with the maximum average rating is 4.2 for the meaningfulness of the task. The negative type questions (4, 5) have average rating less than 3, showing disagreement on non-usefulness and difficultly to perform task effectively and efficiently. The average rating was 1.7 for non-usefulness of task and 2.7 for difficulties in performing the tasks. For functions related questions, the average rating are at least 3.5 with 10 out of 14 questions have the average rating of at least 4. Functions related to *Sura* no. and name Drop Down List, the Arabic-Word meaning boxes and the *Sura* text box, have maximum average rating of 4.2 while the minimum average rating is 3.5 for the Interface and Binary View function.

Table 8.10: Result of the user satisfaction test

Tsks Related questions	**1- Strongly Disagree %**	**2- Disagree %**	**3- Neutral %**	**4- Agree %**	**5- Strongly Disagree %**	**Total %**	**Average Rating (1-5)**
1. Satisfaction of Task	2.2	8.9	22.2	48.9	17.8	100.0	3.7
2. Meaningfulness of Task	0	2.2	15.6	40.0	42.2	100.0	4.2
3. Interface Good Enough for Task	2.2	15.6	31.1	35.6	15.6	100.0	3.5
4. Non-Usefulness of Task	55.6	31.1	6.7	4.4	2.2	100.0	1.7
5. Difficult to Perform Task Effectively and Efficiently	11.1	35.6	33.3	13.3	6.7	100.0	2.7
Function Related questions							
1. Zooming Function	0	6.7	28.9	48.9	15.6	100.0	3.7
2.View All Word Function	0	4.4	33.3	51.1	11.1	100.0	3.7
3. View Word Function	0	6.7	17.8	57.8	17.8	100.0	3.9
4. View Vocabulary Function	0	4.4	6.7	64.4	24.4	100.0	4.1
5. Add Malay-Arab word Function	0	0	15.6	57.8	26.7	100.0	4.1
6. Percentage of Vocabulary and Compare Sura	0	4.4	20.0	48.9	26.7	100.0	4.0
7. Add and Clear Vocabulary Function	0	2.2	20.0	57.8	20.0	100.0	4.0
8. Save and open Vocabulary Function	0	2.2	20.0	51.1	26.7	100.0	4.0
9. Binary Views Function	2.2	8.9	31.1	48.9	8.9	100.0	3.5
10. First, Previous, Next and Last Buttons	0	2.2	22.2	51.1	24.4	100.0	4.0
11. Sura No. And Name Drop Down List	0	2.2	11.1	46.7	40.0	100.0	4.2
12. The Arabic-Word meaning boxes	0	4.4	4.4	53.3	37.8	100.0	4.2
13. The Sura text box	0	4.4	8.9	48.9	37.8	100.0	4.2
14. Filtering Information by clicking	0	6.7	20.0	44.4	28.9	100.0	4.0

Apart from the user's rating result, the comparison of these ratings were analyzed between the participants below the age of 35 years and participants with the age of 35 and above. The Mann-Whitney U test was conducted and it is found that for all ratings there is no significant difference except for task no. 5 related question (on difficulty of performing the task efficiently). The participants with age below 35 (*Mdn* =2) have less difficulties in performing the tasks given to them compared to those of age 35 and above (*Mdn* = 3), where U= 146.0, $p<0.05$ (p=0.013), r=-0.37.

8.3 Discussions on the performance of AWRT

The PaCQ interface was developed in this research to support Qur'anic Arabic word recognition. The main objective of conducting the experimental study is to find out whether the PaCQ interface had any significant performance effect on the users. The performance of AWRT is based on the meaning retrieval of the Arabic words. The results show several findings as follows:

i. The AWRT score percentage rank for participants after using PaCQ interface is significantly higher compared to that before using PaCQ. This implies that the integrate components of the hypothesized model have positive effect on the word recognition performance.

ii. There is significant difference in the AWRT score percentage rank for Set 4 AWRT scores in which its value for UIG is higher (*Mdn*=4) than that of CG. In Set 4 the focus of word recognition was the frequently occurring Arabic words such as pronouns, conjunctions and prepositions. These types of words are written as prefix or suffix to another word. If users can identify these words correctly, then users are more likely to segment word structure and identify other word which is attached to these prefixes and suffixes. For example in the word لِرَبِّكَ the letter word لِ (means for) and كَ (means your) are prefix and suffix for the word رب which means God. If users could not identify word لِ and كَ the next time the users come across the word رب or رَبِّكَ they might not be able to recognize these words which means 'God' or 'your God' respectively. This result also implies that without the emphasize on the parallel plot, the WFE can be achieved.

iii. There is no significant difference for the Set 5 between the UIG and CG, this implies that WSE has not been achieved through the visualization

iv. There is no significant difference for the Set 6 between the UIG and CG, this implies that the process of retention in the LTM through reinforce learning has not been achieved

v. the average AWRT percentage scores for set 4, 5 and 6 for the UIG, show significant difference in the rank score compared to that of CG, this supports find number i)

vi. The average rank time to complete set 4, 5 and 6 for the UIG are significantly faster compared to that of CG. This result indicates that the PaCQ interface does have a positive effect on the time for the users to recognize the words.

vii. Another important issue to discuss is on the perceived personal previous knowledge and the observed level of knowledge. The perceived personal knowledge refers to what the participants' perception of their own knowledge of the Arabic language while the observed level of knowledge refers to the scores of the AWRT that are attained by the participants above the expected outcome. This means that the participants recognize the Arabic words even though no information was provided on those words. It was found that these two variables correlate positively with each other.

viii. The performance in terms of AWRT scores of the LEG and the MOG are found not to differ significantly both in the CG and the UIG. However, the average time for these two age groups differ significantly, whereby in the CG participants age 35 and below have a faster completion time then those age group 35 and above. For the UIG the average time are found not to differ in both the age groups. Therefore there is evidence showing that the gap between the two age groups in the UIG in completing the word recognition tasks had been narrowed to a point that there is no significant difference in the time between the two age groups and performance wise, the two groups did not differ significantly.

8.4 Discussion on the user satisfaction survey

Overall rating for the user satisfaction survey is well above average. However, the lowest rating for PaCQ is on the interface and the binary view functions. At the same time there is obviously some problems related to efficiency in performing tasks for the age group 35 and above. These problems may be related to the font size and colour of certain words appearing in the Graph and on difficulty on finding a particular word from the vocabulary graph when many words had started to be added in the list. However, despite the above problems, there is still evidence showing positive effect on scores performance and time to recognize the Arabic words using PaCQ.

8.5 Study on interpreting parallel plot graph

A different study was also conducted to evaluate the performance of users interpreting the parallel plot in the PaCQ interface. This study was carried out separately from the post-test described earlier firstly, because of timing factors. The post-test already took around an hour in the condition that participants are already cognitively overloaded which is not best for generating accurate result of interpreting the parallel plot. Secondly, although suitability of parallel plot has been discussed in chapter 4 using GOMS analysis but no participants were involved. This study should determine the results of user performance in practice. The following §s describe the conducted study and the results found.

8.5.1 The participants

15 participants were chosen from staff and students of the Faculty of Computer Science and Information Technology of University of Malaya. 7 of them were female and 8 of them were male. The age of the participants were between 20 – 40 years with 6 of them aged below 25 years old and 9 aged 25 and above. Their computing skills were in majority good; only 4 of them considered themselves

average. They are divided into 3 main categories according to their education level as below with 5 participants in each category:

- High school academic qualification faculty staffs (SPM or equivalent)
- Bachelor degree students
- Master degree students

The participants chosen were among adults who have at least average computing skills so that they are able to do the test without too much difficulty. Since PaCQ is meant for those who can read/recite the Qur'an but could not speak the Arabic language, the test were conducted among adult participants since many in age category below then 20 may still be learning how to read the Qur'an.

Participants were given a questionnaire with 7 questions based on the parallel coordinate graph in the PaCQ interface. The questions are related to the 3 relations described in Chapter 5 and 7, R_1(word count and word position information), R_2(vocabulary percentage information) and R_3(comparing percentage of content of words between *sura*). The questions also incorporate instructions on how to interact with the parallel coordinate graph. See Appendix H 1-3 for the sample of questionnaire.

PaCQ was set at a 1280 ×800 pixel resolution and highest (32 bit) color quality. Participants interacted with the system using the mouse via a laptop Dell Vostro 1310 or 1210. The study was conducted in various conditions, sometimes simultaneous sessions with up to a maximum of 5 participants per session and sometimes individually in a session. The questionnaires were given to the participant/s and they were told to answer the questions as they go through the instruction (in the question) one by one.

8.5.2 Objectives and results of the study

The objectives of this study are as the following:

1. To evaluate the graphical perceptibility of parallel plot before and after using PaCQ interface

2. To investigate the correlation between the score and the time taken to complete the task.

3. To investigate the performance of interpreting the parallel plot graph based on the questions given:

a. The performance related to the percentage score of the questions

b. The performance related to the time taken to complete the task

The following results are obtained:

The graphical perceptibility: The participants were asked to rate what they thought of the parallel plot graph showed to them in the PaCQ interface. The rating is from 1 to 5 with 1 for "very hard to understand" and 5 for very easy to understand. Table 8.11 shows the description of the scale. While

Table 8.12 shows the mean and standard deviation of the participant's rating in interpreting the graph. It can be observed that the mean rating at the start time is 1.8 and at the end time is 3.2 with standard deviation of 0.68 and 0.86 respectively. The start time refers to the time at which the participant started to read the instructions of the question paper. While the end time refers to the time at which the participants finished answering all questions.

Table 8.11: Description of rating scale of parallel plot graph

Rating scale	Description
1	Very hard to understand
2	Hard to understand
3	Average/Neutral
4	Easy to understand
5	Very Easy to understand

Table 8.12: Graphical perceptibility of parallel coordinate

	Perception of Parallel Coordinate at start	Perception of Parallel Coordinate at end
Mean	1.80	3.20
Std. Deviation	0.676	0.862

The Correlation: The Pearson correlation between percentage of the total score and the time taken to complete task are found to be non-significant, with $r(13)= 0.457$; $p>0.05$. Table 8.13 shows this result. The scatter plot from Figure 7.3 also shows no linear relation between these two variables.

Table 8.13: Pearson Correlations

		Percentage of the total score	Time Taken to complete task
Percentage of the total score	Pearson Correlation	1	0.457
	Sig. (2-tailed)		0.087
Time Taken to complete task	Pearson Correlation	0.457	1
	Sig. (2-tailed)	0.087	

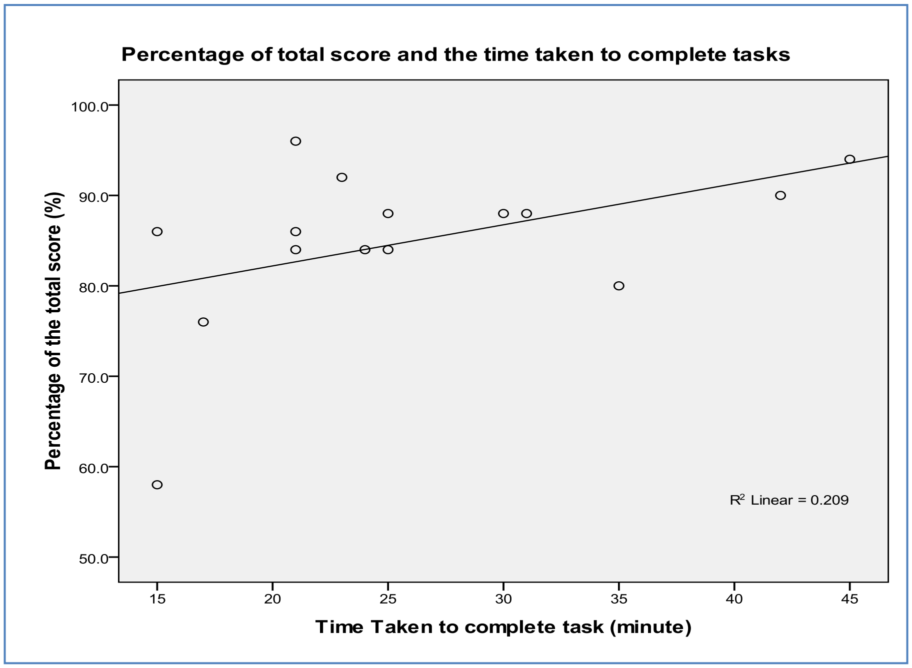


Figure 8.3: Percentage of total score and the time taken to complete task

The performance related to the percentage score: Table 8.14 shows the mean percentages of SPM, Bachelor and Master degree as 1.8, 2.2 and 2.2. From the one way ANOVA test, there is no significant difference in the percentage level scores between the three educational level of participants (SPM/equivalent[17],Bachelor degree, Master degree), $F(2,14) = 0.727$, $p > 0.05$ and Table 8.15 shows the detail results.

[17] SPM means Sijil Pelajaran Malaysia which is the secondary school level exams taken in the 5th form, while equivalent means those with slightly higher or lower level than the SPM level such as Diploma degree or Certificates obtain from specific courses taken.

Table 8.14: The means and standard deviation of 3 categories of participants for percentage level score

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
SPM/ Equivalent	5	1.80	0.837	0.374	0.76	2.84
Bachelor Degree	5	2.20	0.447	0.200	1.64	2.76
Master Degree	5	2.20	0.447	0.200	1.64	2.76
Total	15	2.07	0.594	0.153	1.74	2.40

Table 8.15: One way ANOVA percentage level score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.533	2	0.267	0.727	0.503
Within Groups	4.400	12	0.367		
Total	4.933	14			

i. *The performance related to the time taken to complete the task*: Table 8.16 shows the means and standard deviation of the time taken to complete task. SPM/Equivalent, Bachelor Degree and Master degree are 18.2, 35.6 and 24.2 respectively means to complete task. It is found from the one way ANOVA test (see Table 8.17), that there is significant difference in the time taken to complete the tasks between the three educational level of participants, $F(2,14) = 12.64$, $p < 0.05$. As in Table 8.18, the Scheffe's test shows that the bachelor degree participants take more time to complete tasks compared to the other two groups. There is no significant difference in the time taken to complete between the Master degree and the SPM/Equivalent group.

Table 8.16: The mean and standard deviation of 3 categories of participants for time taken to complete task

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
SPM/ Equivalent	5	18.20	3.633	1.625	13.69	22.71
Bachelor Degree	5	35.60	8.112	3.628	25.53	45.67
Master Degree	5	24.20	3.701	1.655	19.60	28.80
Total	15	26.00	9.071	2.342	20.98	31.02

Table 8.17: One way ANOVA for time taken to complete parallel plot related tasks

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	781.200	2	390.600	12.641	.001
Within Groups	370.800	12	30.900		
Total	1152.000	14			

Table 8.18: Comparison of time taken to complete task between participants in the 3 levels of education

	(I) Education Level	(J) Education Level	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Scheffe	SPM/ Equivalent	Bachelor Degree	-17.400*	3.516	.001	-27.20	-7.60
		Master Degree	-6.000	3.516	.271	-15.80	3.80
	Bachelor Degree	SPM/ Equivalent	17.400*	3.516	.001	7.60	27.20
		Master Degree	11.400*	3.516	.023	1.60	21.20
	Master Degree	SPM/ Equivalent	6.000	3.516	.271	-3.80	15.80
		Bachelor Degree	-11.400*	3.516	.023	-21.20	-1.60

Other results: It is found that 11 out of the 15 participants interpreted the number 8 as 3 or 5 based on the question 2 and question 5. 2 of them from the SPM/ Equivalent level, 5 of them in the Bachelor level and 4 of them from the Master level.

8.5.3 Discussion on the results of interpreting parallel coordinate graph

The mean rating for the graphical perceptibility at the starting point when the participants were using the system is low (1.8); in other words the participants found that the parallel coordinate graph was hard to understand. This is also consistent with the findings of other researches on parallel coordinate (Leban, Zupan, Vidmar, & Bratko, 2006; Yuan, Guo, Xiao, Zhou, & Qu, 2009; Forsell & Johansson, 2007; Siirtola & Räihä, 2006). However, at the end of the session the average rating increases to 3.2, meaning that once the participants know how to interact with the graph, they found that it is much easier than what they have expected. The data shown to users can be interpreted accurately even by novice users, and this was also found by Siirtola & Räihä (2006), Siirtola, Laivo, Heimonen, & Raiha (2009) and Henley, Hagen, & Bergeron (2007) as discussed in § 8.5.2.

Investigation was also done on whether the time taken to complete the task has any correlation with the percentage score attained by the participants. Results show that there is no signifcant relation and the scatter plot between these two variables shows no linear relationship between them as despicted in Figure 8.3. However, the performance measures related to the mean time taken to complete the task between the Bachelor degree group and the other two groups (the SPM/equivalent and Master level) are significantly different as shown in Table 8.18. There is no significant difference in the mean time taken to complete the task between the SPM/equivalent and the Master level groups. This indicates that the higher the education level does not necessarily determine a faster speed of completion.

The performance related to the percentage of scores between the three levels of education groups is found to be not significant. Again, this indicates that higher the education level does not necessarily determine a better performance in percentage score. The three group participants are all considered as non-expert (novice) users of parallel plot. Here, there is evidence showing that non-expert users can also interpret the parallel coordinate graph accurately in the context of this project contrasting the arguments of Yuan, Guo, Xiao, Zhou, & Qu (2009) and Henley, Hagen, & Bergeron (2007) stating that expert users are needed to interpret the graph.

The problem encountered by users mistakenly reading the number 8 as number 3 may be due to the design adopted in the PaCQ interface. As in Figure 8.4, the number 8 is so closely positioned to the vertical axis, contributing to reading error by the participants.

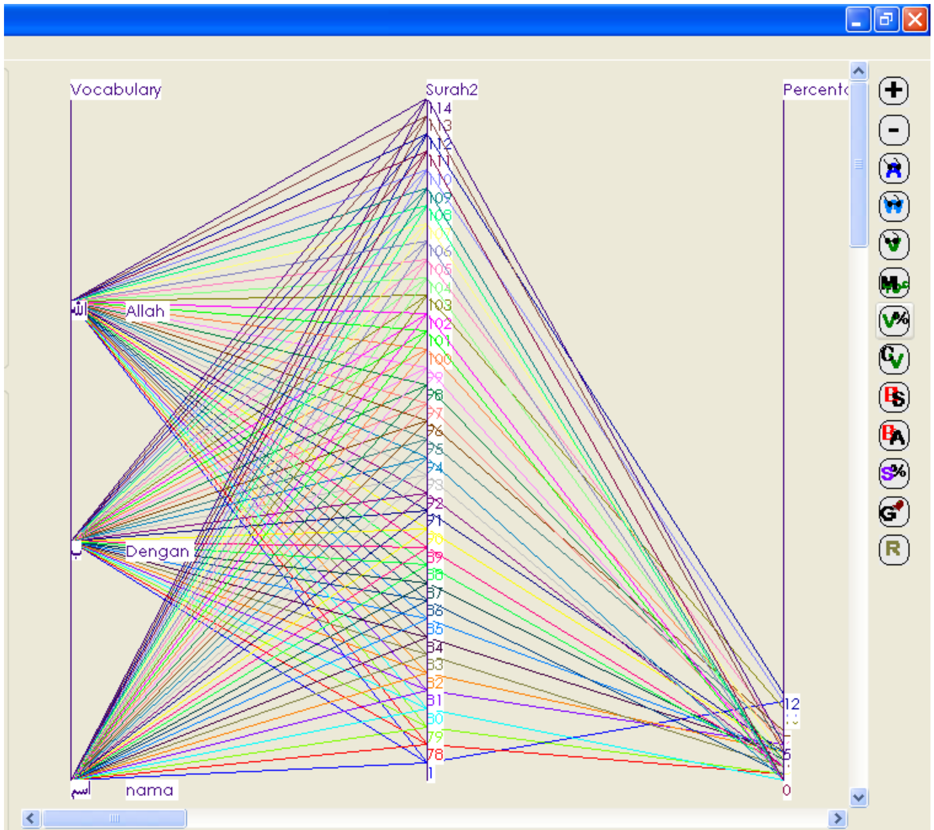


Figure 8.4: The *Sura* axis showing the *sura* numbers in the vocabulary percentage graph

8.6 Modification and retesting via user's opinion

Minor modification of the PaCQ was done on 3 things: the font size (previously size 10 increased to size 12), increase of text font size after zooming in (previously, text size remain static when zoomed in) and tooltip of text when cursor is mouseover on the cluttered information. The findings in this chapter were further tested via interview sessions. The objective of this interview was to reconfirm issues related to PaCQ interface design, the perception on parallel plot and to get their opinion on the usefulness of the PaCQ as means to learn words for recognition and subsequently comprehension.

Participants: Four people were involved. 3 of them were aged between 20-28 years and one of them was 40 years old. There were 2 females and 2 males. 3 of them considered themselves as intermediate level Arabic speakers and one of them considered himself as advanced. Two of them possess intermediate computing skill level and two of them were advanced. Their level of education were Diploma, Bachelor and Master (2 of them).

Procedure: The participants were briefly introduced to the PaCQ interface and particularly mentioned that the purpose of the software is to help non-Arabic speaker to learn Arabic words in the Qur'an for word recognition and subsequently reading comprehension. They were asked a few questions with three main issue:

i. Learning: Do you think that the software is useful in helping to learn words for word recognition and reading comprehension in Qur'anic verses?

ii. Parallel Plot: What do you think of the parallel plot before and after explanation of its usage? Do you have any suggestions regarding this issue?

iii. Function: Can you read the *sura* number on the bottom of the *graf*? Do you think that the zooming function and the text tooltip are helpful in reading the cluttered information?

Result: Table 8.19 shows the summary of participant's response to the questions asked during the interview. Response to question 1 shows that all four participants agree that PaCQ is very useful for the purpose of word recognition and if use many times can help in comprehension. PaCQ also helps us to focus on unknown words. They responded that the software is also useful for memorization and research on the Qur'an. The word count information is useful to be displayed to strategies learning of words

Table 8.19:Participant's respond to question

	Participant 1
Q1	- PaCQ is very useful for the purpose of word recognition and if use many times can help in comprehension. -It is also helpful for researchers of Qur'an
Q2	- parallel plot give cluttered view, however if know its usage it is not hard to use - suggest to give viewing information in list form - visualization is useful in this software but suggest other forms
Q3	- *sura* number from 80-89 can be read correctly - font size use if sufficient for viewing - binary view, zooming and tooltip are useful to make clear the cluttered information view
	Participant 2
Q1	- PaCQ is useful for the purpose of word recognition and if use many times can help in comprehension. - it helps us to focus on unknown words since we can identify them easily
Q2	Parallel plot give cluttered view however if know its usage it is not hard to use - visualization is useful in this software but suggest other forms
Q3	sura number from 80-89 can be read correctly - font size use if sufficient for viewing - binary view, zooming and tooltip are useful to make clear the cluttered information
	Participant 3
Q1	- PaCQ is useful for the purpose of word recognition and if use many times can help in comprehension. - useful for people who tries to memorize the Qur'an - word count is useful information to help strategies learning
Q2	- parallel plot give cluttered view, however if know its usage it is not hard to use - show word count and word graph only, position in another graph
Q3	sura number from 80-89 can be read correctly - font size use if sufficient for viewing - binary view, zooming and tooltip are useful to make clear the cluttered information
	Participant 4
Q1	- PaCQ is useful for the purpose of word recognition and if use many times can help in comprehension. - useful for people who tries to memorize the Qur'an - word count is useful information to help strategies learning
Q2	- parallel plot give cluttered view, however if know its usage it is not hard to use - show word count and word graph only, position in another graph
Q3	sura number from 80-89 can be read correctly - font size use if sufficient for viewing - binary view, zooming and tooltip are useful to make clear the cluttered information

Response to question 2 shows that the participants perceive the parallel plot as very cluttered, however if its usage is known it is not hard to use. Even though they found that the visualization is useful, they suggested other viewing techniques such as not displaying all four axis(word count, word, aura and ayah) at the same time.

Response to question 3, shows that all participants read the number 8 correctly. The zooming, binary and tooltip function was found to be useful to make clear the cluttered information.

8.7 Summary and Conclusion

An experimental study was conducted to test the hypothesis that visualization with vocabulary tracking mechanism supports Arabic word recognition in the reading process.

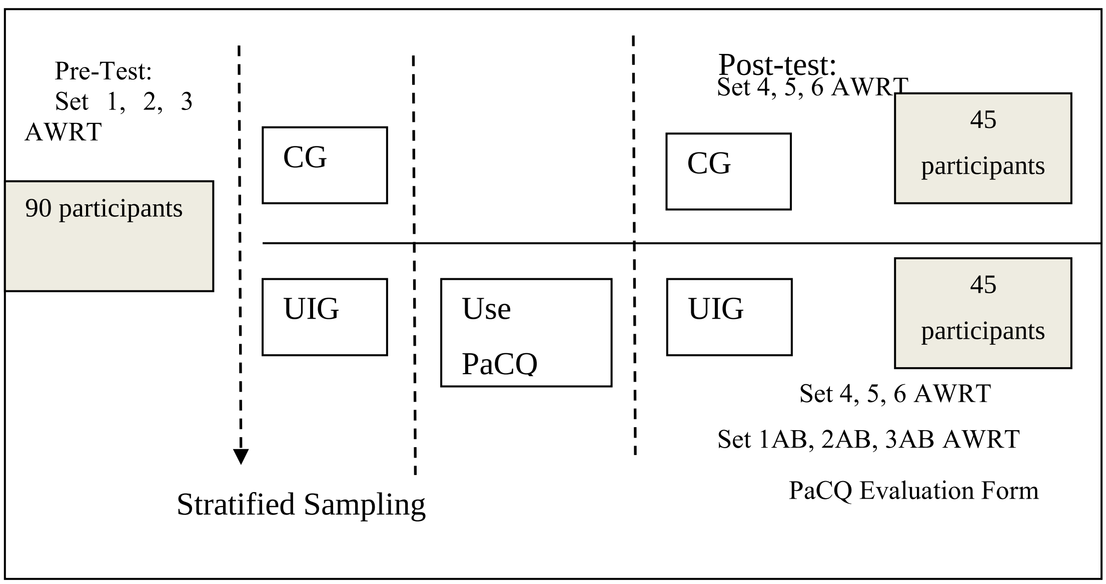


Figure 8.5: The experimental set-up

Figure 8.5 shows the experimental set-up. The score of the pre-test (Set 1, 2, 3 AWRT) is used to divide the participants into the CG (Control Group) and the UIG (User Interface Group, which is the group that uses the PaCQ interface) by stratified sampling. This is done so that, there are equal number of participants with below and above average scores in the CG and the UIG. In the post-test both groups were given another set of questions (Set 4, 5, 6 AWRT). The UIG was also given the same questions as in the pre test after using PaCQ. In addition, they had to also fill in the PaCQ evaluation form.

The results show several significant findings:

- the AWRT score percentage rank for participants after using PaCQ interface is significantly higher compared to that before using PaCQ.
- set 4 AWRT scores for the UIG is higher than that of the CG.
- the average AWRT percentage scores set 4, 5 and 6 for the UIG, show significant difference in the rank score compared to the CG.
- the average rank time to complete set 4, 5 and 6 for the UIG is significantly faster compared to the CG
- participants' perception of their own knowledge of the Arabic language correlates positively with the expected scores
- there is significant difference in the timing for age group below 35 (faster) and above 35 (slower) in the CG. However, there is no significant difference in the UIG for both age groups, leading to evidence showing improvement in the time taken to complete the AWRT in the UIG for the age 35 and above with no difference in the score both in the CG and UIG.

These finding also highlighted a few issues regarding the effectiveness of the component of the hypothesized model. All of the AWRT procedure emphasized on the parallel plot except for test set 1A and set 4. Both of these two sets are meant to test the WFE and it was found that the score of the CG and UIG were significantly different, indicating that the WFE was achieved without the emphasize on the parallel plot. From this finding it can be deduced that the Semantic word visualization is the stronger component in the visualization model compared to the parallel plot. This may be due to the difficulties faced by the participants while the interacting with the parallel plot since the graphical perceptibility of the plot is relatively low.

The interview session found that all participants agree that PaCQ is useful for learning Arabic words for word recognition and if used many times can subsequently help in comprehension of Qur'anic verses during recitation. In terms of the PaCQ interface, users rated fairly well for the whole system although a few items should be considered for improvement. Although they found that the visualization is useful to help the learning process, result of the interviews suggest that other visualization type was favoured. This result correlates with the result found in the experimental study. The graphical perceptibility of parallel plot was found to be low. However, the participants were found to be able to interpret parallel plot relatively well. There is no significant difference in the percentage scores of the participants with different education levels.

In conclusion, there is evidence showing improvement in the performance of users using the PaCQ interface in aspects of the performance score and the timing taken to complete the AWRT. The hypothesis that visualization with vocabulary tracking mechanism supports Arabic word recognition in the reading process can therefore be accepted, although the stronger visualization component is the Semantic Word Visualization.

9.0 CONCLUSION AND FURTHER RESEARCH

In this chapter the conclusion of this thesis is presented. It begins with revisiting the aim and objectives of this research, then generally outlining all the tasks that have been carried out to test the hypothesis based on several research questions in the research findings §. The following §s are then on the research contribution, outcome and subsequently the research challenges. Based on these discussions, future research recommendation is finally outlined.

9.1 Research findings

The aim of this research is to solve reading comprehension problem for non-Arabic speakers in Qur'an reading. To achieve reading comprehension many processes and factors are involved. Therefore, the approach of this research is to investigate and support word recognition process as an important phase of the solution of reading comprehension. The hypothesis is that visualization with vocabulary tracking mechanism supports Arabic word recognition in the reading process. This hypothesis was tested by an experimental study using the PaCQ interface specifically developed to support Word Superiority Effect (WSE) and Word Frequency Effect (WFE) with two main components which are, the visualization and the vocabulary tracking mechanism. The findings pertain to the research questions and objectives are stated below

Several research questions are answered:

i) What are the underlying problems of comprehension amongst *Qur'an* reciters?

This question is answered through literature investigation in chapter 2 in which it can be concluded that the emphasis is on teaching how to read the Qur'an in many parts of the Muslim world rather than reading with comprehension. However, understanding the content of the Qur'an relates to classes and lectures on the interpretation of the Qur'an which are usually conducted separately from the reading classes. Learning how to comprehend the Qur'an, however, would relate to learning the Arabic language which is a rich morphological language. In addition the problems are also identified in § 2.3 and Chapter 5. Firstly, the preliminary study 1, shows that understanding/comprehending the Qur'an while recitation and remembering *Tajwid* rules are the two main problems. Secondly, the problem of comprehension was found to be related to the Arabic writing script and the complex morphological structure of the Arabic language. In § 5.1.3 it was concluded that the missing link to achieve Qur'anic word recognition is the retrieval of meaning.

ii) How does visualization support word recognition based on word superiority effect and word frequency effect?

In chapter 2, literatures related to the learning process, reading comprehension and word recognition are outlined based on the information processing theory. It is concluded that, WSE and WFE are two main theories in achieving word recognition as part of important basic component in achieving reading comprehension.

WSE states that letters are easier to recognize in known words compared to non-words and this theory extends to the fact that words are recognized as a whole not letter by letter. Since cognate words are words that exist in two languages in concern (in this case Arabic and Malay language), these word should help in the process of learning/knowing the Arabic words to achieve word recognition (Lehr, Osborn, & Hiebert, 2004). However, the difficulty lies in the fact that the writing scripts are different, comparing the Malay Latin script and the Arabic script but the Malay *Jawi* script which is based on the Arabic script shed some light to this issue. Otherwise, readers of the Qur'anic verses must be helped to realize that there exist Malay-Arab words in Qur'anic verses although sometimes the phonological sound can differ very slightly.

Learning to recognize Arabic root words is considered difficult because of the complex morphological structure. It is a step toward word recognition. There are 10 alphabets in Arabic language that can be added to root words leading to different meanings in context of a sentence. These alphabets can be added as prefixes, infixes and suffixes. In addition, there are also particles and pronouns words that can be added either as prefixes or suffixes to an Arabic word. It is hypothesized that visualization of the word structure with the meaning can help to support WSE. It was explained that the learning process involves the information processing stages. Information need to go through to the LTM via the Sensory memory and the STM. For quick and effective learning the sensory memory must be attracted so that the information can be passed on to the STM. Later on the information is passed to the LTM via rehearsal process. Therefore active learning with attraction and rehearsal process must be supported.

WFE states that words that are seen more frequently are responded to more rapidly. It is therefore logical to show the word frequency of the words in Qur'anic verses in the reading strategy. It is also found that the Qur'an contains many repeated words and this can be used as a source of attracting attention as addressed in research question iv). In chapter 5, it is mentioned that many people are visual learners, and therefore it is sensible to attract the attention via the visual memory through information visualization. Visualization of frequently occurring words to support WFE is therefore chosen to be used to attract the visual memory.

Vocabulary tracking can be a means of supporting the rehearsal process. For example one can explore, retrace, recall and compare the vocabulary learning achievement using a vocabulary tracker. Cognate words of the Malay-Arab language can also be used for quicker learning process. Visualization of the vocabulary can also help to remind the learners of their own set of known words. Interaction with the visualization elements can therefore activate active learning, thus supporting the rehearsal process.

Both WSE and WFE are hypothesized to help in the word recognition process. It is also later on foreseen that once many words are part of the individual vocabulary system, i.e. are recognized by the individual, word relation process comes into role to further contribute to reading comprehension. In addition, existing knowledge of the theme or topic discussed in the *sura*s help in the transfer knowledge discourse in for example *Juz Amma*. However, this process occurs only when frequent rehearsal takes place either through the PaCQ interface or other means of acquiring knowledge such as done in classes, mosques, listening to lectures in the television and radio. This is when the word information has gone into the LTM and combined with the pre-existing knowledge.

iii) What are the suitable techniques to visualize word information to support word superiority effect and word frequency effect?

The second preliminary study discussed in Chapter 5 shows that there is evidence of significant difference on the level of comprehension when images are used as part of the reading text. In the literature review from § 2.3 there were also attempt to visualize the Qur'anic verses in pictorial form. The hieroglyph ideogram of the ancient Egypt discussed in § 2.2.2 also showed the usage of this approach. However, mere static image visualization is not as helpful as compared to when the information is transformed into a perceptually efficient visual format. For example, the interactivity operations that are tightly coupled with the visualization fields are for users to gain insight of a particular data to be analysed. The act of exploring, interpreting and making deduction of data related to vocabulary visualization therefore is hypothesized to help individuals to recognize and recall the meaning of words, thus supporting the rehearsal process as explained in research question vi).

WSE and WFE are important parts in word recognition. The WSE can be supported by visualizing the segmented word structure. In addition, words in the vocabulary list can be highlighted to users as a reminder that those words should already be known and recognized by the users. The WFE can be supported by visualization of the word information i.e. word count and word position in relation to *suras* in the Qur'an. At least 4 variables have been identified: word count, word, *sura* position, *ayah* position. Scatter plot or parallel plot are the two most suitable techniques since they can be used to visualize multidimensional variables. In the preliminary study 3 of chapter 4 the comparison between parallel and perpendicular plots is discussed for this purpose. It was shown through literature by Grinstein, Hoffman, & Pickett (2002) and GOMS analysis that the parallel plot is a better approach to be used in the context of this research. The visualization of word frequency with respect to their position in the Qur'anic verses does not directly relate to Qur'anic word comprehension. The visualization helps to attract the attention of the users to learn more on those words with the higher frequency and therefore strategizing on a quicker word recognition process.

iv) How should Arabic words be counted since the word can exist in many forms?

As explained in chapter 6, the Arabic words can be counted in its stem forms or the root form. It is chosen in this research to count words based on the root form. To learn words in the stem form means that there are more words to learn and more words to process, however this approach gives better meaning of words in the context of a sentence. For quicker learning and simpler implementation root word form is better. Here in this study, the approach is to adopt the root word form. Therefore, root words need to be identified. The case study and the Rule-Based Stemming Engine implemented show that the accuracy rate is not satisfactory to be used and it needs further improvement. Therefore, a database containing the root word information is used based on the resources provided by Dukes (2008).

v) What are the platforms or architecture to implement the solution found based on the above questions?

Chapter 6 addresses this question in detail. A three tier architecture is implemented using JAVA SE Runtime Environment version 1.6.0_02 and Eclipse version 3.3.2 as the Integrated Development Environment. MySQL 5.0.67 community version is used as the database. The developed system is called the PaCQ interface and contains functions that support the hypothesis that visualization with vocabulary tracking mechanism supports meaningful word recognition.

vi) What are the performance measures used for system evaluation ?

Arabic Word Recognition Tests are designed to evaluate the PaCQ interface. The tests are based on verses from *sura Al-Fatihah* and *suras* in *Juz Amma*. An experimental study was set up starting with a pre-test of Arabic Word Recognition Test (AWRT) to all the targeted participants. Then stratified sampling was used to divide the participants into the control and the experimental groups for the AWRT post-test. PaCQ was tested for any improvement in the word recognition percentage scores and the time taken to complete word recognition test. PaCQ interface was also evaluated based on user satisfaction test.

The test on parallel plot interpretability was to find out the ability of the users to interpret the plot and their perception of the plot. The interview session done after the modification of PaCQ was to recapture issues related to the function of the parallel plot and the usefulness of the visualization of word frequency.

vii) Is there any improvement on the word recognition performance supported by an integrated system based on visualization with vocabulary tracking mechanism?

Chapter 7 shows several performance improvements for those who used the system. The AWRT score is higher after using the PaCQ interface compared to before using the system, the average AWRT scores is higher for the UIG, the average rank time to complete the AWRT is also faster for the UIG and there is evidence showing improvement in the time taken to complete AWRT for the UIG with the age 35 and above. It can also be concluded that the semantic word visualization component was a stronger part of the model embedded in the PaCQ compared to the parallel plot visualization. This finding suggests that the parallel plot should be replaced with other visualization technique.

9.2 Concluding Remarks

In the PaCQ interface users are able to learn words in the Qur'anic verse to achieve word recognition. This is supported through visualization for WSE and WFE. Semantic word visualization through vocabulary, cognate words and word segmentation are to support WSE. While the parallel plot visualization of the word count, word position and percentage information are to support WFE. Users for example can start to explore the more frequent words which can be found easily via the parallel plot visualization. The information of the word frequency can easily attract the visual memory compared to list of statement written in text form. They are able to strategies their learning by starting to learn the frequently occurring words shown through the parallel visualization. New learned words can be added to their personal vocabulary list. Alternatively, users can also learn the words as they read a *sura* or *ayah* through the semantic word visualizer and subsequently add those words involved in the personal vocabulary list. The growing vocabulary list can be used to further strategies the learning by analysing, interpreting and deducing the vocabulary percentage in relation to all *suras*. The next step is to focus on unknown words with the aim of recognition users can actively decide which *sura* to read and learn so that many words in the Qur'anic verse can be recognized. Other functions such as comparing percentage, searching, and filtering, also helps in the rehearsal process of learning and therefore provide active learning. All these operations are done through interaction of the visualization fields in the PaCQ interface which include the act of exploring, interpreting, deducing, comparing, searching and filtering. These interaction if repeated frequently not only make users able to achieve word recognition in Qur'anic verses but subsequently reading comprehension.

In answering the research questions, all the objectives of this research are achieved in relation to reading comprehension problem, word recognition, visualization, information processing, system development, experiments and performance measures. The PaCQ interface was developed incorporating visualization with a vocabulary tracking mechanism for supporting Arabic word recognition. This is particularly shown in Chapter 8, the design and implementation chapter. Then in chapter 9 it is shown that the experimental set up was designed to investigate the performance as regards to word recognition. The evaluation of the performance is based on meaning retrieval of the Arabic words. The results show that the performance of users using the PaCQ interface has improved in aspects of the performance score and the timing taken to complete the AWRT with the Sematic visualization as the stronger component.

9.3 Significant contributions

(a) Research contribution

This research has contributed on three field of studies cognitive science, information processing and visualization are as follows.

Contribution to cognitive science field:

i) word recognition using visualization:

Research work here is on word recognition and vocabulary for Qur'an reading comprehension. System development such as the PaCQ interface for learning to recognize Qur'anic Arabic words with incorporation of information visualization theory is not known to exist elsewhere. Cognitive research in reading comprehension always uses non-words as a comparison to known words. In this research the participants were assessed using somewhat similar concept expect that it is applied in real world situation i.e reading without comprehension, instead of using non-words reading situation which rarely occurs.

There are already softwares developed to address the issue of understanding Qur'an using frequently used words (Abdulraheem, 2010). The PaCQ interface also adopts this strategy as part of its component through linkage with reading comprehension theory. Therefore, the results of the experimental study conducted on the PaCQ interface in this research can serve as scientific evidence that such effort is worth taken into consideration where the issue of comprehending/understanding *Qur'anic* verses is addressed.

ii) application of vocabulary tracking system to support *Qur'anic* Arabic word recognition

In this research the development of a vocabulary tracking system is feasible and relevant since Qur'an contains fixed number of words and this is one of the contributions to research. Although there exists vocabulary software that tracks vocabulary achievement, it does not relate to the various reading material of the user.

Contribution to information processing field:

This research contributes towards a rule-base stemming engine algorithm in Arabic word pattern analysis. Studying the literature, the Khoja stemmer and BAMA (Buckwalter, 2007) can be considered as the two most frequently cited Arabic stemmer engines. The Khoja stemmer extracts roots from a word based on a prepared list of function words by a straightforward pattern matching procedure. BAMA also includes a prepared list of several types: the lexicon of prefix, the suffix, and the stem on which word entries are checked upon. It deals with word stems rather than roots. The work of Al-Shammari and Lin (2008), Taghva, Elkhoury, and Coombs (2005), Yagi and Yaghi (2004), Larkey, Ballesteros, and Connell (2002) addresses issues related to Arabic stemming. Mainly, discussions are related to algorithms that can give the most accurate result when searching or retrieving Arabic words. The literature reveals that the most common algorithm on stemming includes

i) removing diacritics in Arabic words;

ii) removing the suffixes and prefixes which may be particles or added alphabets;

iii) normalization of the vowels (concerning the spelling variation representing a few Arabic letters to another to ease the deduction of roots from word stems or tokens);

iv) and comparing word entries to pre-existing lists or lexicons of stems, roots, prefixes, or suffixes.

Triliteral Root Extraction(TRE) stemmer engine developed by Al- Shalabi, Kanaan, and Al-Serhan (2003) developed an engine called (TRE) assign different weights to vowel and consonant letters. Consonants were assigned zero weight while the letters in the word سألتمونيها (where all Arabic affixes are formed from the combination of these letters) were assigned different weights. Letters with the lowest weight are assigned as root letters. Sawalha and Atwell (2008) developed the Voting algorithm, the most common roots generated from the three stemmer engines Khoja, BAMA, and TRE for a given word were chosen as the root word.

Our stemming word pattern identifier algorithm called the RSE discussed in § 5.1.7 uses the algorithm that include the four common ones discussed above and also the use of the letters in the word سألتمونيها such as done in TRE (Al-Shalabi, Kanaan, & Al-Serhan, 2003) can be considered unique and therefore a contribution to the field of Arabic Stemming. Although it is not 100% accurate, it can be made available for others to reuse for improvement, towards better algorithm for stemming words in the Qur'anic verses.

Contribution to visualization field:

Application of parallel plot visualization technique in the PaCQ interface was based on GOMS analysis. Parallel plot is known to be used only by expert users to interpret data such as discussed in Chapter 3. However, application of parallel plot to novice users had not been tested upon since many data interpretation using parallel plot involves experts. The interpretability of the parallel plot in PaCQ interface was tested and it is found that there is no significant difference of the scores in interpreting the plot between novice and experts with higher education background.

(b) Outcome

The outcome of this research is this thesis documentation on an integrated application of theories in cognitive science, reading comprehension, information visualization and information processing to solve a reading comprehension problem for non-Arabic speakers while reading *Qur'anic* verses. In addition to this a software system called the PaCQ interface was developed to test the above concepts and the theoretical model and currently no such software is available in the market. In addition, vocabulary tracker serves as meaningful word recognition technique.

9.4 Research challenges and lessons learnt

Resources and technology

One of the major barriers is the Arabic language which has different structure from more popular languages such as English with more easily found research references. In depth knowledge in Arabic is particularly useful for constructing the database dictionary used as part of the PaCQ interface. Several reference books were used apart from hiring and consulting an Arabic language expert to construct the word form patterns and the database dictionary. There are also various types of Arabic language which are the Modern Standard

Arabic, the Classical form and the dialectal form. Apart from these factors, prefix, suffix and infix with the writing convention from right to left adds to the difficulties and complexities that have to be dealt with.

The technology related in supporting the Arabic language especially related to the font of the *Qur'anic* verses is another challenge. As the writing system in Arabic language is from right to left, different from the conventional left to right, displaying the Qur'anic Arabic font on the interface was a problem. This was overcome by using the font downloaded from the Qur'an complex website in Madinah (www.Qur'ancomplex.com).

The assessment

The survey in preliminary study 1, could have been designed in other ways. Firstly, investigation on the problems faced in Tajwid lesson that contributed to deficiency of the word recognition process should have been explored. Secondly, the potential techniques that can be used to display quantification involved in the PaCQ interface was not found through survey in the preliminary study 3. The step taken was only through literature review and analysis through GOMS which has the disadvantage of not accurately tailored for the purpose of Qur'anic word recognition.

Relevant theories were investigated, selected and adapted to link the PaCQ system development. Conducting the experimental study was a challenge having to find the right 90 participants based on age and gender and then stratifying them based on the scores they achieved in the pre-test.

Realization of the integrated model found in this study was done through the PaCQ interface. However, the test conducted did not specifically test for each component of the model. For example, the tightly linkage of the vocabulary component to the visualization made it hard to test those component separately.

The assessment of word recognition performance is based on the ability to retrieve the meaning of the Arabic words in Malay or English language. A word can have several meanings depending on the context of the text. It is therefore, practical to associate a word to one possible meaning even though within the actual context the meaning may not be suitable. Visualization in PaCQ and scrutiny of word relations between adjacent words and transfer skill of word recognition to another *sura* with prior discourse knowledge on the *sura* help to achieve contextual meaning of words.

The assessment of the AWRT did not rule out the possibility of the meaning being memorized by a good memorizer. However, we found that very few participants score very well in the AWRT indicating that the possibility above can be ruled out.

There were problems associated with the font size and some bugs in PaCQ. However, the PaCQ interface was successfully constructed to test the concepts involved in this research.

9.5 Recommendation for further research

There are several research recommendations for future work the author suggests. One is to replicate the experimental study of the AWRT using the scatter plot approach to a different set of groups. An ANNOVA test can be used to compare the results between the PaCQ users, the scatter plot users and the control group.

Two, other visualization approaches that cater for multidimensional variables can be explored. Although this research adopted the parallel plot, it may not be the best approach to attract the sensory memory. Another view point in this context may suggest that instead of visualizing multiple variables all at the same time, one or two of the variables can be used to feed a back-end component system that could act as an agent in telling the users certain information through visualization of a simpler and acceptable technique such as the bar chart. For example the agent could tell the user the frequency of a word in the particular *sura* and the frequency of the word in comparison to the whole Qur'an.

Three, another replication study can be conducted based on an improved interface and a more accurate database. The improved interface should provide the following:

- include all possible meanings of an Arabic word in context of the verses concerned.
- layout of the Arabic words and meanings designed to support older people in using the PaCQ interface.

Context-aware AWRT can be conducted. The assessment on context-aware word recognition can be conducted based on perception, interpretation, time taken to complete task, score and most importantly measuring comprehension of Qur'anic verses after several rehearsal action such as done in the AWRT of the experimental group. Interview with observation of a few participants should also be included in assessing the effectiveness of PaCQ in supporting word recognition and subsequently reading comprehension.

Lastly, instead of counting words in the Qur'an based on the root, the stem of the roots can be used. This is because an Arabic word with its stem carries more accurate meaning. However, this may slow down the process of word recognition.

9.6 Summary

Finally, as a summary this chapter has gone through the achievements and findings in the research, research outcome, challenges and recommendation for further research. This research has achieved in producing an output (the PaCQ interface) that can be used to support reading comprehension for the Qur'an amongst non-Arabic speakers through word recognition with visualization and vocabulary tracking mechanism.

References:

Abdelbaki, I. (1983). Teaching Arabic through the Holy Quran to other than Arabic speakers, the medium level. *Master Thesis for the International Institute of Khartoum for Arabic Language for non-Arabic speakers of The Arab League Educational, Cultural and Scientific Organization*.

Abdulazeez, A. (2004). *Understand Qur'an the easy way.* Hyderabad: Understand Qur'an Academy.

Abdullah, M. (2003). Aliran Tafsir Ibn 'Arabi: Satu Tinjauan Awal [Tafsir Ibn 'Arabi: An early investigation]. *Al-Bayan: Journal of Al-Qur'an & Al-Hadith*, 69-80.

Abdulraheem, A. (2010). *Understand Quran - The Easy Way - Free Online Courses*. Retrieved 2012, from Understand Quran Academy: http://understandquran.com/

Ager, S. (2010a). *Arabic alphabets, pronounciation and language.* Retrieved July 19, 2010, from Omniglot writing systems and languages of the worlds: http://www.omniglot.com/writing/arabic.htm

Ager, S. (2010b). *Malay language, alphabets and pronounciation.* Retrieved July 19, 2010, from Omniglot writing systems and languages of the worlds: http://www.omniglot.com/writing/malay.htm

Ahlberg, C. (1996). Spotfire: An information exploration environment. *ACM SIGMOD Record*, (pp. 25-29).

Ahlberg, C., & Wistrand, E. (1995). IVEE: An Information Visualization & Exploration Environment. In N. a. Gershon (Ed.), *Information Visualization 1995 Proceedings* (pp. 66-73). Atlanta, GA, USA: IEEE.

Ahmad, A. R. (2003). Al-Baghawi[433H -516H]: Seorang Tokoh Tafsir. *Al-Bayan: Journal of Al-Qur'an & Al-Hadith*, 57-68.

al-Haytham, I. (2000). Historical perspective,. In d. J. A. Kennedy, *Perception and perceptual process* (A. Sabra, Trans., p. ii). North Holland.

Ali, R. (1997). *Pengantar Ulum Al-Qur'an [Ulum Al-Qur'an]* (1 ed.). Kuala Lumpur: Pustaka Salam Sdn Bhd.

Alpern, B., & Carter, L. (1991). The hyperbox. In G. M. Nielson, & L. Rosenblum (Ed.), *Proceedings of the 2nd Conference on Visualization '91* (pp. 133-139). San Diego, California: IEEE Visualization. IEEE Computer Society Press.

Al-Shalabi, R., Kanaan, G., & Al-Serhan, H. (2003). New approach for extracting Arabic roots. *Proceeding of International Arab Conference on Information Technology (ACIT'2003).* Alexandria, Egypt.

Al-Shammari, E. T., & Lin, J. (2008). Towards an error-free Arabic stemming. *Proceeding of the 2nd ACM Workshop on Improving Non English Web Searching*, (pp. 9-16). Napa Valley, California, USA.

Ankerst, M., Keim, D. A., & Kriegel, H. P. (1996). Circle Segments: A Technique for Visually Exploring Large Multidimensional Data Sets. *Proc. Visualization '96.* San Francisco, CA.

Anonymous. (2003, June 11). *Word Relationships.* Retrieved August 26, 2011, from Academic Support: http://academic.cuesta.edu/acasupp/as/507.HTM

Anonymous. (2006). *Problems of Arabic Typefaces*. Retrieved August 2012, from Arabic Type Today: http://www.arabictype.com/blog/articles/problems-of-arabic-typefaces/

Artero, A. O., Oliveira, M. C., & Levkowitz, H. (2004, October 10-12). Uncovering Clusters in Crowded Parallel Coordinates Visualizations. *IEEE Symposium on Information Visualization*, 81-88.

Bagozzi, R. P., & Silk, A. J. (1983). Recall, Recognition and the measurement of memory for print advertisements. *Marketing Science*, *2*(2), 95-134.

Barlow, N., & Stuart, L. J. (2004). Animator: A Tool for the Animation of Parallel Coordinates. *Proceedings of the Eighth International Conference on Information Visualisation (IV'04)* (pp. 1-6). Plymouth, UK: IEEE Computer Society.

Bauer, T. (1996). Arabic Writing. In D. P. W., *The World's Writing Systems* (p. 559). New York: Oxford University Press.

Beck, I. L., McKeown, M. G., & Linda, K. (2002). *Bringing words to life: Robust vocabulary instruction.* New York: Guilford.

Belaid, A., & Choisy, C. (2008). Human Reading Based Strategies for off-line Arabic Word Recognition. In D. Doermann, & S. Jaeger (Ed.), *Proceedings of the 2006 Conference on Arabic and Chinese Handwriting Recognition* (pp. 36-56). Berlin, Heidelberg: Springer-Verlag.

Bertini, E. and Santucci, G. (2006, Jun). Give chance a chance: modeling density to enhance scatter plot quality through random data sampling. *Information Visualization*, *5*(2), 95-110.

Bertini, E., & Santucci, G. (2005). Improving 2D scatterplots effectiveness through sampling, displacement, and user perception. *Proceedings of Ninth International Conference on Information Visualization*, (pp. 826-834).

Bertini, E., & Santucci, G. (2005). Is it darker? improving density representation in 2d scatter plots through a user study . *Proc. of SPIE Conference On Visualization and Data Analysis.* San Jose Marriott and San Jose Convention Center, San Jose, California, USA .

Bertini,E.; Dell'Aquila, L.; Santucci, G. (2006). Reducing InfoVis cluttering through non uniform sampling, displacement, and user perception. In R. F. Erbacher, J. C. Roberts, M. T. Gröhn, & K. Börner (Ed.), *Proceedings Vol. 6060, Visualization and Data Analysis 2006,, 6060.* San Jose, California, USA.

Borowsky, B., Owen, W. J., & Masson, M. E. (2002). Diagnostics of phonological lexical processing: Psedohomophone naming advantages, disadvantages, and base-word frequency effects. *Memory and Cognition, 30*(6), 969-987.

Buckwalter, T. (2007). Issues in Morphological Analysis. In A. Soudi, A. V. den Bosch, & G. Neumann (Eds.), *Arabic Computational Morphology* (pp. 23-41). Springer.

Buckwalter, T. (2010). *Arabic Transliteration*. Retrieved September 13, 2010, from Buckwalter Arabic Transliteration: http://qamus.org/transliteraiton.htm

Caravolas, M. (2007). The Nature and causes of Dyslexia in Different Languages. In M. Snowling, & C. Hulme (Eds.), *The Science of Reading: A Handbook.* (p. 336). Oxford , UK: Blackwell Publishing.

Carey, M., Kriwaczek, F., & Rüger, S. M. (2000). A Visualization Interface for Document Searching and Browsing. *Proceedings of CIKM 2000 Workshop on New Paradigms in Information Visualization and Manipulation.* Washington, DC.

Carrus, J. (2008). *Visualizing Vocabulary.* Research Report.

Chad, S. A., Fitzpatrick, P. J., Jankun-Kelly, T., Yancey, A. N., & Swan II, E. J. (2009, July 7). An interactive parallel coordinates technique applied to a tropical cyclone climate analysis. *Computers & Geosciences, 35*(7), 1529-1539.

Choi, H., Lee, H., & Kim, H. (2009, July). Fast detection and visualization of network attacks on parallel coordinates. *Computers & Security, 28*(5).

Chou, S.-Y., Lin, S.-W., & Yeh, C.-S. (1999, June). Cluster identification with parallel coordinates. *Pattern Recognition Letters, 20*(6), 565-572.

Claggett, F., & Brown, J. (1992). *Drawing Your Own Conclusion: Graphic Strategies for Reading, Writing, and Thinking.* Portsmouth: Heineman.

ColHeart, M. (2007). Modeling Reading: The Dual Route Approach. In M. Snowling, & C. Hulme, *The Science of Reading: A Handbook* (pp. 6-33). Oxford UK: Blackwell Publishing.

Dahlan, M., & Sahil, A. (2007). *Al-Qur'an Rasm Uthmani, Bertajwid and Disertai Makna [Uthmani Al-Qur'an, Tajwid and Meaning Included]*. Batu Caves, Selangor, Malaysia: Pustaka Haji Abdul Majid.

Das, S. (2004). *Visualizing words*. Retrieved 2011, from http://www.social.cs.uiuc.edu/class

Denffer, A. V. (1983). *'Ulum Al-Qur'an, an introduction to the Sciences of the Qur'an.* Leicester, UK: The Islamic Foundation.

Dewhurst, S. A., Brandt, K. R., & Sharp, M. S. (2004). Intention to learn influences the word frequency effect in recall but not recognition memory. *Memory and Cognition*, *32*(8), 1316-1326.

Dewhurst, S., Brandt, K., & Sharp, M. (2004). Intention to learn influences the word frequency effect in recall but not in recognition memory. *Memory & Cognition, 32*, 1316-1325.

Dorot, D., & Mathey, S. (2010). Visual word recognition in young and older adults: A study of cohort effects for lexical variables. *Revue europeenne de psychologie appliquee, 60*, 163-172.

Duke, K. (2008). *JAVA API - Uthmani script*. Retrieved 2010, from The Qur'anic Arabic Corpus: http://corpus.quran.com/java/uthmaniscript.jsp

Duke, K. (2009a). *The Quranic Arabic Corpus - Word by Word Grammar, Syntax and Morphology of the Holy Quran*. Retrieved 2012, from The Quranic Arabic Corpus: http://corpus.quran.com/

Dukes(a), K. (2009). *Phonetic Transcription*. Retrieved June 2010, from Qur'an Corpus: http://corpus.quran.com/documentation/phonetic.jsp

Dukes, K. (2009). *Computational Analysis of the Quran through Traditional Arabic Linguistics (PhD presentation).* Retrieved August 3, 2010, from http://corpus.quran.com

Dukes, K. (2009b). *Phonetic Transcription*. Retrieved 2012, from http://corpus.quran.com/documentation/phonetic.jsp: http://corpus.quran.com/

Dukes, K., & Buckwalter, T. (2010). A Dependency Treebank of the Quran using Traditional Arabic Grammar. *Proceeding of 7th international conference on Informatics and Systems.* Cairo, Egypt.

Dukes, K., & Habash, N. (2010). Morphological Annotation of Quranic Arabic. *Proceeding of Seventh International Conference on Language Resources and Evaluation.* Valletta, Malta.

Dukes, K., Atwell, E., & Sharaf, A. M. (2010). Syntactic Annotation Guidelines for the Quranic Arabic Treebank. *Proceeding of the seventh International Conference on Language Resources and Evaluation.* Valletta, Malta.

eReflectSoftware. (2009). *UltimateVocabulary*. Retrieved August 3, 2010, from http://www.ultimatevocabulary.com/

Ertmer, P. A., & Newby, T. J. (1993). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly*, *6*(4), 50-70.

Estes, W. K., & Brunn, J. L. (1987). Discriminability and bias in the word-superiority effect. *Perception & Psychophysics, 42*(5), 411-422.

Eviatar, Z., & Ibrahim, R. (2007, February). Morphological Structure and Hemispheric Functioning: The Contribution of the Right Hemisphere to Reading in Different Languages. *Neuropsychology, 21*(4), 470–484.

Fang, S., Lwin, M., & Ebright, P. (2006). Visualization of unstructured text sequences of nursing narratives. *Symposium on Applied Computing-Proceedings of the 2006 ACM symposium on Applied computing* (pp. 240 - 244). New York, NY, USA: ACM.

Farid, M., & Grainger, J. (1996). How Initial Fixation Position Influences Visual Word Recognition: A Comparison of French and Arabic. *Brain and Language, 53*, 351-368.

Felder, R. M., & Silverman, L. (1988). Learning and Teaching Styles in Engineering Education. *Engr. Education*, *78*(7), 674-681.

Fisher, D. F. (1975). Reading and visual search. *Memory and Cognition, 3*, 188-196.

Forbes, D. (1868). *Grammar of the Arabic Language.* London: WM H Allen & Co.

Forsell, C., & Johansson, J. (2007). Task-based evaluation of multi-relational 3D and standard 2D parallel coordinates. *Proceeding of IS&T/SPIE's International Symposium on Electronic Imaging, Conference on Visualization and Data Analysis* (pp. 1-12). San Jose, USA, 2007), ; 64950C-1–12.: SPIE: Bellingham and IS&T.

Forster, K., & Chambers, S. (1973). Lexical Access and Naming Time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627-635.

Friendly, M. (1994). Mosaic Displays for Multi-Way Contingency Tables. *Journal of the American Statistical Association (JASA), 89*(425), 190-200.

Frost, R. (2007). Orthographic systems and Skilled Word Recognition Processes in Reading. In M. Snowling, & C. Hulme (Eds.), *The Science of Reading: A Handbook* (p. 275). Oxford , UK: Blackwell Publishing.

Frost, R. (2007). Orthographic Systems and Skilled Word Recognition Processes in Reading. In M. Snowling, & C. Hulme (Eds.), *The Science of Reading: A Handbook.* (pp. 272-295). Oxford, UK: Blackwell Publishing.

Fukkink, R. G., Hulstijn, J., & Simis, A. (2005). Does Training in Secong-Language Word Recognition Skills Affect Reading Comprehension? An Experimental Study. *The Modern Language Journal, 89*(1), 54-75.

Furnas, G. W. (1986). Generalized Fisheye Views. In M. Mantei, & P. Orbeton (Ed.), *Proceedings of the ACM CHI 86 Human Factors in Computing Systems Conference* (pp. 16-23). Boston Massachusetts: ACM.

Gade, A. M. (2004). *Perfection makes practice: learning, emotion, and the recited Qur'ān in Indonesia.* York, USA.: University of Hawaii Press.

Garcia, E. (2006). *Miislita*. Retrieved July 5, 2010, from Vector Models based on Normalized Frequencies, Improving Term Weights with Normalized Frequencies: http://www.miislita.com/term-vector/term-vector-4.html

Ge, Y., Li, S., Lakhan, V. C., & Lucieer, A. (2009). Exploring uncertainty in remotely sensed data with parallel coordinate plots. *International Journal of Applied Earth Observation and Geoinformation, 11*(6), 413-422.

Ghali, N. (2007). *Arabic Grammar Unravelled.* Canada: Fun with Arabic; Bilingual edition.

Gosselin, F., & Schyns, P. (2004). A picture is worth thousands of trials: rendering the use of visual information from spiking neurons to recognition. *Cogni. Sci., 28*, 141-146.

Grainger, J. (1990). Word Frequency and Neighborhood Frequency Effects in Lexical Decision and Naming. *Journal of Memory and Language, 29*, 228-244.

Grinstein, G. G., Hoffman, P. E., & Pickett, R. M. (2002). Benchmark development for the evaluation of visualization for data mining. In U. Fayyad, G. G. Grinstein, & A. Wierse (Eds.), *Information Visualization in Data Mining and Knowledge Discovery* (pp. 129-176). San Francisco, CA: Morgan Kaufmann Publishers.

Haber, R., & Myers, B. (1982). Memory for pictograms, pictures, and words separately and all mixed up. *Perception, 11*(1), 57-64.

Hanley, R. (2007). Learning to Read in Chinese. In M. Snowling, & C. Hulme (Eds.), *The Science of Reading: A Handbook* (p. 316). Oxford, UK: Blackwell Publishing.

Harmon, J. M., & Wood, K. D. (2008). *Research Summary: Vocabulary teaching and learning across disciplines.* Retrieved March 15, 2011, from http://www.nmsa.org/Research/ResearchSummaries/VocabularyTeaching/tabid/1728/Default.aspx

Harun, M. R., Ainul, M. S., & Hamdzani, K. S. (2005). Information visualization in Al-Qur'an:Network and connectivity, scatter plot, trivariate data. *Bachelor of Computer Science Dissertation, University of Malaya, Malaysia*.

Hashim, H. (2005). Galaxy of Qur'an visualization. *Bachelor of Computer Science Dissertation, University of Malaya, Malaysia*.

Hasim, N. (2012). Kelas Pengajian Bahasa Arab Harfiah Al-Quran [Class Lesson on Harafiah Al-Quran Arabic Language]. Selangor, Malaysia.

Hauser, H., Ledermann, F., & Doleisch, H. (2002). Angular brushing of extended parallel coordinates. *Proceedings of the IEEE Symposium on Information Visualization* (pp. 127-131). IEEE Computer Society.

Heer, J., Kong, N., & Agrawala, M. (2009). Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations. *Proceedings of the 27th international Conference on Human Factors in Computing Systems* (pp. 1303-1312). Boston, MA, USA: ACM, New York, NY.

Henley, M., Hagen, M., & Bergeron, R. (2007). Evaluating Two Visualization Techniques for Genome Comparison. *Proceedings of the 11th International Conference of Information Visualization*, (pp. 551-558).

Houtsonen, J. (1994). Traditional Quranic Education in a Southern Moroccan Village. *International Journal of Middle East Studies, 26*(3), 489-500.

Huitt, W. (2003). *The information processing approach to cognition.* Retrieved July 19, 2010, from Educational Psychology Interactive: http://www.edpsycinteractive.org/interact.html

Humam, A. (2005). *Iqra', cara cepat belajar membaca al-Qur'an [Iqra', the fast way to learn reading the Qur'an].* Shah Alam, Selangor, Malaysia: Darul Kitab Sdn. Bhd.

Hunt, R., Ellis, H., & Ellis, H. H. (2004). *Fundamentals of cognitive psychology* (7 ed.). New York, USA: McGraw Hill.

Husain, A. (1974, December 9). *The Message of Qur'an.* Retrieved January 20, 2011, from http://muslim-canada.org/atharhusain.pdf

Ibrahim, R. (2006, November). Do Languages with Cognate Relationships have Advantages in Second Language Acquisition. (D. P. Robertson, & D. J. Jung, Eds.) *The Linguistics Journal*, *1*(3), 66-96.

Inselberg, A. (2009). *Parallel Coordinates: Visual Multidimensional Geometry and its Applications* (XXVI ed.). Springer.

Ismail, M. R. (2003). *Bahasa Arab Mikro, Binaan Kata dan Fungsi Makna. [Micro Arabic Language. Word Form and Semantic Function]*. Kota Bahru: Percetakan Darul Naim.

Ismail, M. R. (2008). *Kosa Kata Bahasa Arab Mikro, Teknik dan Aspek [Micro Arabic Vocabulary, Techniques and Aspects]*. Kota Bahru, Kelantan, Malaysia: Arabic Language and Consultancy.

Johansson, J., Forsell, C., Lind, M., & Cooper, M. (2008). Perceiving patterns in parallel coordinates: determining thresholds for identification of relationships. *Information Visualization, 7*(2), 152-162.

John, B. E., & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: comparison and contrast. *ACM Trans. Comput.-Hum. Interact.*, *3*(4).

Jordon, T. R., Peterson, K. B., & Almabruk, A. A. (2010). Revealing the superiority perceptibility of words in Arabic. *Perception*, *39*, 426-428.

Karim, N. S. (1996). Pendahulauan [Introduction]. In N. S. Karim (Ed.), *Unsur Bahasa Asing Dalam Bahasa Melayu [Foreign language elements in the Malay Language]* (1 ed., p. xi). Kuala Lumpur, Malaysia: Dewan Bahasa dan Pustaka.

Keim, D. A., & Kriegel, H. (1996). Visualization Techniques for Mining Large Databases: A Comparison. *IEEE Trans. on Knowl. and Data Eng.*, *8*(6), 923-938.

Keng-Pong, F. (1997). Islamic fundamentalism and female empowerment among the Muslims of Hainan Island, People's Republic of China. In J. Brink, & M. J., *Mixed Blessings: Gender and Religious Fundamentalism Cross Culturally* (p. 41).

Kettani, H. (2010). Proceedings of the 8th Hawaii International Conference on Arts and Humanities. Honolulu, Hawaii.

Kintsch, W., & Rawson, K. A. (2007). Comprehension. In M. Snowling, & C. Hulme (Eds.), *he Science of Reading: A Handbook* (pp. 209-226). Oxford, UK: Blackwell Publishing.

Korat, O. (2010). Reading electronic books as a support for vocabulary. story comprehension and word reading in kindergarten and first grade. *Computer and Education*, *55*, 24-31.

Krathwohl, D. R., & Bloom, B. S. (2002). *Taxonomy of educational objectives: the classification of educational goals, Volume 2.* London: Longman.

Kuhn, M. R., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning Theory and Assessment of Reading Fluency: Automaticity, Prosody, and Definations of Fluency. *Reading Research Quarterly*, *45*(2), 230-251.

Lamping, J., Rao, R., & Pirolli, P. (1995). A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 401-408). New York, NY, USA: ACM Press/Addison-Wesley Publishing Co.

Larkey, L. S., Ballesteros, L., & Connell, M. E. (2002). Improving stemming for Arabic information retrieval: light stemming and co-occurrence analysis. *Proceedings of the 25th Annual international ACM SIGIR Conference on Research and Development in information Retrieval* (pp. 275-282). New York, NY: ACM Press.

Larson, K. (2003, July). The Science of Word Recognition or how I learned to stop worrying and love bouma. *Proceeding of ATypI conference.* Vancouver.

Larson, K. (2004, July). *The Science of Word Recognition or how I learned to stop worrying and love bouma*. Retrieved March 15, 2011, from http://www.microsoft.com/typography/ctfonts/wordrecognition.aspx

Lazem, I. a. (1993). The terms of the thirtieth part of the holy Quran, A lexical and semantic study. *Master thesis for The International Institute of Khartoum for Arabic Language for non-Arabic speakers of The Arab League Educational, Cultural and Scientific Organization*.

Leban, G., Zupan, B., Vidmar, G., & Bratko, I. (2006, September). VizRank: Data Visualization Guided by Machine Learning. *Data Min. Knowl. Discov. , 13*(2), 119-136.

Ledermann, F. (2002). *Parvis*. Retrieved July 19, 2010, from http://www.mediavirus.org/parvis/

Lee, D. L., Chuang, H., & Seamons, K. (1997). Document ranking and the vector-space model. *IEEE Software, 14*(2).

Lehr, F., Osborn, J., & Hiebert, E. H. (2004). *A Focus on Vocabulary*. Retrieved March 19, 2011, from Pasific Resources for Education and Learning: http://www.prel.org/products/re_/ES0419.htm

Li, J. X. (2004). Visualization of high-dimensional data with relational perspective map. *Information Visualization*, 49-59.

Li, J., Martens, J.-B., & Wijk, J. J. (2008). Judging correlation from scatterplots and parallel coordinate plots. *Information Visualization, 9*, 13-30.

Lupker, S. J. (2007). Visual Word Recognition: Theories and Findings. In M. Snowling, & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 39-60). Oxford , UK: Blackwell Publishing.

Macaluso, E. (2012). Spatial Constraints in Multisensory Attention. In M. M. Murray, & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes*. Boca Raton (FL): CRC Press.

Mackinlay, J. D., Robertson, G. G., & Card, S. K. (1991). The Perspective Wall: Detail and Context Smoothly Integrated. In S. P. Robertson, G. M. Olson, & J. S. Olson (Ed.), *Proceedings of the ACM CHI 91 Human Factors in Computing Systems Conference* (pp. 173-179). New York, NY, USA: ACM Press.

Mao, W. W., & Chu, W. (2007). The phrase-based vector space model for automatic retrieval of free-text medical documents. (W. M. Aalst, B. Benatallah, F. Casat, F. Curbera, & E. Verbeek, Eds.) *Data & Knowledge Engineering, 61*(1), 76-92.

Mashud, I. (2004). Modeling Software Requirements on Visulaizaiton of Al-Qur'an System. *Master Dissertation, University of Malaya, Malaysia*.

Matkovic, K., Jelovic, M., Juric, J., Konyha, Z., & Gracanin, D. (2005). Interactive visual analysis and exploration of injection systems simulations. *Visualization, 2005. VIS 05* (pp. 391- 398). IEEE.

Maulana, M. (2002). *he Holy Qur'an with English Translation and Commentary*. Belgium: Ahmadiyya Anjuman Ishaat.

McClelland, J. L. (1976). Parallel Letter Identifiaction. *Journal of Experimental Psychology, 2*(1), 80-91.

McClelland, J. L., & Johnson, J. C. (1977). The role of familiar units in perception of words and non-words. *Perception and Psychophysics, 22*, 249-261.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of Basic Findings. *Psychological Review*, *88*, 375-407.

McCutchen, D., Logan, B., & Biangardi-Orpe, U. (2009). Making Meaning: Children's Sensitivity to Morphological Information During Word Reading. *Reading Research Quarterly, 44*(4), 360-376.

Mohd, A. Y., Adel, M. A., & Ahmed, K. M. (2003). Keberkesanan Iqra' sebagai kaedah pembelajaran membaca al-Qur'an [The Effectiveness of Iqra' as a method of learning to read the Qur'an]. In A. A. Hameed, M. A. Adel, A. N. Mohamed, & M. M. Muhammad (Eds.), *Isu Semasa Pengajian Qur'an dan Sunnah [Current Issues in the study of Qur'an and Sunnah]* (1 ed., pp. 101-114). Kuala Lumpur, Malaysia: Fakulti Pengajian Qur'an dan Sunnah.

Morrison, C. M., & Ellis, A. W. (1995). Roles of Word Frequency and Age of Acquisition in Word Naming and Lexical Decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 116-133.

Muhammad, M. M. (2003). Kaedah Mudah Belajar Membaca Al-Qur'an Bertajwid (Kalam) [The Easy method of learning to read al-Qur'an with Tajwid (Kalam)]. In A. A. Hameed, M. A. Adel, A. N. Mohamed, & M. M. Muhammad (Eds.), *Isu Semasa Pengajian Qur'an dan Sunnah [Current Issues in the study of Qur'an and Sunnah]* (1 ed., pp. 115-130). Kuala Lumpur, Malaysia: Fakulti Pengajian Qur'an dan Sunnah.

Nagy, W. E., & Hiebert, E. H. (2011). Toward a Theory of word selection. In M. L. Kamil, D. P. Pearson, E. B. Moje, & P. P. Aefflerbach, *Handbook of Reading Research* (Vol. 4, pp. 388-404). Oxon, UK: Routledge.

Nagy, W. E., & Scott, J. A. (2000). Vocabulary Processes. In M. L. Kamil, P. D. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of Reading Research* (Vol. 3, pp. 574-593). Mahwah New Jersey : Lawrence Erbaum Association.

Nasrudin, M., Omar, K., Zakaria, M., & Liong, C. Y. (2008). Handwritten Cursive Jawi Character Recognition: A Survey. *Computer Graphics, Imaging and Visualisation, 2008. CGIV '08. Fifth International Conference* , (pp. 247 - 256). Penang.

Nation, K. (2007). Children's Reading Comprehension Difficulties. In M. Snowling, & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 248-266). Oxford, UK: Blackwell Publishing.

NICHD. (2006, August 31). *Findings and Determinations of the National Reading Panel by Topic Areas*. Retrieved July 20, 2010, from http://www.nichd.nih.gov/publications/nrp/findings.cfm

Novotny, M., & Hauser, H. (2006). Outlier-Preserving Focus+Context Visualization in Parallel Coordinates. *IEEE Transactions on Visualization and Computer Graphics, 12*(5), 893-900.

Oweini, A., & Hazoury, K. (2010). Towards a sight word list in Arabic. *Int Rev Educ, 56*, 457-478.

Perfetti, C. A., Landi, N., & Oakhill, J. (2007). Modeling Reading: The Acquisition of Reading Comprehension Skill. In M. Snowling, & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 227-247). Oxford , UK: Blackwell Publishing.

Pikulski, J. J. (2010). *Teaching Word-Identification Skills and Strategies: A Balanced Approach.* Retrieved August 3, 2010, from http://www.eduplace.com/rdg/res/teach/index.html

Rahman, H. A. (1996). Pengaruh Sintaksis Bahasa Arab Dalam Bahasa Melayu [Influence of Arabic language syntax in the Malay language]. In N. Safiah (Ed.), *Unsur Bahasa Asing Dalam Bahasa Melayu [Foreign language elements in the Malay Language]* (pp. 49-64). Kuala Lumpur:, Malaysia: Dewan Bahasa dan Pustaka.

Rao, R., & Card, S. K. (1995). Exploring large tables with the table lens. *Proceeding of Conference on Human Factors in Computing System* (pp. 403 – 404). New York, NY, USA: ACM .

Rashwan, M., Al-Badrashiny, M., & Attia, M. A. (2010). Stochastic Arabic Diacritizer Based on a Hybrid of Factorized and Un-factorized Textual Features. *IEEE Transactions on Audio, Speech, and Language Processing (TASLP), 19*, 166-175.

Reed, S. K. (2000). *Cognition, Theory and Application* (5 ed.). Stamford, USA: Wadsworth, Thompson Learning.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology: General, 81*(2), 275-280 .

Rennison, E. (1994). Galaxy of news: an approach to visualizing and understanding expansive news landscapes. *ACM symposium on User interface software and technology*, 3 - 12.

Richardson, D., Spivey, M., Barsalou, L., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cogni. Sci., 27*, 767-780.

Rosly, M. A. (2004). Qur'an visual directory. *Bachelor of Computer Science Dissertation, University of Malaya, Malaysia*.

Roth, S., Chuah, M., Kerpedjiev, S., Kolojejchick, J., & Lucas, P. (1997). Towards an information visualization workspace: Combining multiple means of expression. *Human Comp. Interac., 12*, 131-185.

Rubik, G. (1997). *Extruded Parallel Coordinates.* Retrieved July 19, 2010, from http://www.cg.tuwien.ac.at/~rubik/extruded.htm

Rukaini, A. R. (1996). Morfem Bahasa Arab Dalam Bahasa Melayu [Arabic Language Morpheme in the Malay Language]. In N. Safiah (Ed.), *Unsur Bahasa Asing Dalam Bahasa Melayu [Foreign language elements in the Malay Language]* (pp. 37-38). Kuala Lumpur: Dewan Bahasa dan Pustaka.

Sa'ari, Z. C., & Borhan, J. T. (2003). Al-Qur'an: The Miracle of Miracles. *Al-Bayan: Journal of Al-Qur'an & Al-Hadith*, 43-56.

Saboorian, M. (2010). *Zekr - Quran Study Software for Windows*. Retrieved 2012, from The Zekr Project: http://zekr.org/quran/en/quran-for-windows

Sackney, L., & Mergel, B. (2007). Contemporary Learning Theories, Instructional Design and Leadership. In M. B. John, F. W. Charles, & K. Patricia (Eds.), *Intelligent Leadership, Constructs for Thinking Education Leaders* (pp. 67-98). Netherlands: Springer.

Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. *Commun. ACM, 18*(11).

Şaul, M. (1984). The Quranic School Farm and Child Labour in Upper Volta. *Journal of the International African Institute, 54*(2), 71-87.

Sawalha, M., & Atwell, E. (2008). Comparative Evaluation of Arabic Language Morphological Analysers and Stemmers. *Proc. of COLING 2008 22nd International Conference on Comptational Linguistics.*

Scurletis, G. (2009). *Visualization and Vocabulary Retention*. Retrieved 2011, from Think map: http://www.visualthesaurus.com/cm/wordshop/1840

Segui, J., Mehler, J., Frauenfelder, U., & Morton, J. (1982). The word frequency effect and lexical access. *Neuropsychologia, 20*(6), 615-627.

Seymour, P. H. (2007). Early Reading Development in European Orthographies. In S. M. C. (Ed.), *The Science of Reading: A Handbook* (p. 298). Oxford , UK: Blackwell Publishing.

Siirtola, H., & Räihä, K.-J. (2006). Interacting with parallel coordinates. *Interacting with Computers,, 18*(6), 1278–1309.

Siirtola, H., Laivo, T., Heimonen, T., & Raiha, K.-J. (2009). Visual Perception of Parallel Coordinate Visualizations. *Proceeding of 13th International Conference on Information Visualisation*, (pp. 3-9).

Silverman, R., & Crandell, J. D. (2010). Vocabulary Practices in Prekindergarten and Kindergarten Classrooms. *Reading Research Quarterly, 45*(3), 318-340.

Simmons, E. (2002). Visualizing Vocabulary. *The Quarterly-Summer* , 13-17.

Snowling, M., & Hulme, C. (Eds.). (2007). *The Science of Reading: A Handbook* (1 ed.). Oxford , UK: Blackwell Publishing.

Sobel, C. P. (2001). *The cognitive sciences: an interdisciplinary approach.* Michigan.: Mayfield Pub. Co.

Solso, R. L., MacLin, M. K., & MacLin, O. H. (2008). *Cognitive psychology* (8 ed.). USA: Pearson/A and B.

Song, H., Curran, E. P., & Sterritt, R. (2004). Multiple foci visualisation of large hierarchies with FlexTree. *Information Visualisation(3) 1,* 19 - 35.

Soucy, P., & Mineau, G. (2005). Beyond tfidf weighting for text categorization in the vector space model. In L. Kaelbling, & A. Saffiotti (Ed.), *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI-05)* (pp. 1130–1135). Edinburgh, Scotland: Professional Book Center.

Spence, R. (2001). *Information Visualization.* New York, US: Addison Wesley and ACM Press.

Spence, R., Apperley, M., Rankin, P., Colgan, L., Tweedie, L., Su, H., et al. (1995). Interactive qualitative design. *IEE Colloquium on Design Systems with Users in Mind: The Role of Cognitive Artefacts.*

Sperling, G. (1963). A model for visual memory tasks. *Human Factors, 5*, 19-31.

Steed, C. A., Fitzpatrick, P. J., Jankun-Kelly, T., & Swan II, J. E. (2009). North Atlantic Hurricane Trend Analysis using Parallel Coordinates and Statistical Techniques. *Computers & Geosciences, 35*(7), 1-4.

Swayne, D. F., Lang, D. T., Buja, A., & Cook, D. (2003). GGobi: evolving from XGobi into an extensible framework for interactive data visualization. *Comput. Stat. Data Anal., 43*(4).

Symanzik, J., Majure, J. J., & and Cook, D. (1996). The linked ArcView 2.1 and XGobi environment—GIS, dynamic statistical graphics, and spatial data. In S. Shekhar, & P. Bergougnoux (Ed.), *Proceedings of the 4th ACM International Workshop on*

Advances in Geographic information Systems (pp. 147-154). New York, NY: ACM,.

Taft, M., & Russell, B. (1992). Pseudohomophone Naming and the Word Frequency Effect. *The Quarterly Journal of Experimental Psychology, 45A*(1), 51-71.

Taghva, K., Elkhoury, R., & Coombs, J. (2005). Arabic Stemming Without A Root Dictionary. *Proceedings of the international Conference on information Technology: Coding and Computing (Itcc'05). 1*, pp. 152-157. Washington, DC,: TCC. IEEE Computer Society.

TIBCO. (2010). *Analystics Products*. Retrieved July 23, 2010, from TIBCO Spotfire: http://spotfire.tibco.com/products/overview/analytics-products.aspx

Toppo, G. (2007). *Free online materials could save schools billions.* Retrieved 2012, from USA Today: http://www.usatoday.com/news/education/2007-11-06-freereading_N.htm

Tweedie, L., Spence, B., Dawkes, H., & and Su, H. (1995). The influence explorer. In I. Katz, R. Mack, & L. Marks (Ed.), *Proceeding of Conference Companion on Human Factors in Computing Systems* (pp. 129-130). New York: ACM .

van Wijk, J. J., & van Liere, R. (1993). HyperSlice: visualization of scalar functions of many variables. In D. Bergeron, & N. G. (Ed.), *Proceedings of the 4th Conference on Visualization '93* (pp. 119-125). Washington DC: IEEE Visualization, IEEE Computer Society.

Versteegh, K. (2003). *The Arabic Language.* Edinburgh: Edinburgh University Press.

Ware, C. (2000). *Information Visualization: Perception for Design.* San Francisco, CA.: Morgan Kaufmann.

Watson, J. C. (2002). *The Phonology and Morphology of Arabic.* Oxford: Oxford University Press.

Weber, C. A., & Desai, A. (1996, April 5). Determination of paths to vendor market efficiency using parallel coordinates representation: A negotiation tool for buyers. *European Journal of Operational Research, 90*(1), 142-155.

Weippl, E. (2001). Visualizing content based relations in texts. *Proceedings of the 2nd Australasian conference on User interface* (pp. 34 – 41). Washington, DC, USA.: IEEE Computer Society.

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology, 1*(1), 59-85.

Whitaker, B. (2009, June 18). *The Arabic language*. Retrieved July 19, 2010, from www.al-bab.com: http://www.al-bab.com/arab/language/lang.htm

Wightwick, J., & Mahmoud, G. (2005). *Easy Arabic Grammar*. Hampshire: McGraw-Hill.

Wikipedia. (2012). *Education in Malaysia*. Retrieved 2012, from Wikipedia, the free encyclopedia: http://en.wikipedia.org/wiki/Education_in_Malaysia

Wilkinson, L., & Wills, G. (2005). *The grammar of graphics.* New York, USA: Springer.

Wise, J., Thomas, J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., et al. (1995). Visualizing the non-visual: Spatial analysis and interaction with information from text documents. *Proceedings of Information Visualization '95 Symposium* (pp. 51-58). IEEE Computer Society.

Wortham, S. (2003). Article 563. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science* (Vol. 1, pp. 1079-1082). New York: Nature Publishing Group.

Xu, Y., Hong, W., Li, X., & Song, J. (2007). Parallel Dual Visualization of Multidimensional Multivariate Data. *Proceeding of IEEE International Conference on Integration Technology, 2007. ICIT '07.*

Yagi, S. M., & Yaghi, J. (2004). Systematic Verb Stem Generation for Arabic. In A. Farghaly, & K. Megerdoomian (Ed.), *Proceedings of the Workshop on Computational Approaches to Arabic Script-Based Languages, COLING 2004*, (pp. 23-30). Geneva, Switzerland.

Yang, C. C., Chen, H., & Hong, K. (2002). Exploring the World Wide Web with Self-Organizing Map. *"Proceedings of the International World Wide Web Conference"*, 7-11.

Yaxley, R., & Zwaan, R. (2007). Simulating visibility during language comprehension. *Cogni. Sci.*, *105*, 229-236.

Yeap, W. K., Reedy, P., Min, K., & Ho, H. (2005). Visualizing the Meaning of Texts. *IV-Proceedings of the Ninth International Conference on Information Visualisation* (pp. 883 - 888). Washington, DC, USA: IEEE Computer Society.

Yoshie, K., Naoaki, N., & Yoichi, T. (2009). Effect of visualization of relation among words in electronic English-Japanese dictionary. *Proceeding of the IASDR 2009.* Souel, Korea.

Yuan, X., Guo, P., Xiao, H., Zhou, H., & Qu, H. (2009, November/December). Scattering Points in Parallel Coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 1001-1008.

Yusof R, J. R., Zainudin, R. B., & Yusoff, Z. M. (2010). Quranic words stemming. *Arabian Journal of Science and Technology, 35*(2C).

Yusoff, M. A., & Adnan, M. A. (2009). Perkataan Arab dalam Bahasa Melayu: Satu Tinjauan Dari Aspek Morfologi [Arabic Words in Malay Language: An Investigation from aspect of morphology]. *Journal of Usuluddin*, 305-324.

Yusoff, Z. M. (2011). Tadabbur Al-Qur'an: Satu Tanggungjawab yang terabai [Pondering Al-Qur'an: A neglected responsibility]. *Prosiding Seminar Keagungan Wahyu [Proceeding Seminar of the Greatness of Revelation]*, (pp. 7-11). Kuantan, Pahang.

Yusoff, Z. M., & Abdullah, A. H. (2010). Syed Abdillah Ahmad Al-Jufri (2003) Dan Tafsir Pelita Al-Qur'an: Satu Pengenalan [Syed Abdillah Ahmad Al-Jufri (2003) And Tafsir Pelita Al-Qur'an: An Introduction]. *Al-Bayan: Journal of Al-Qur'an & Al-Hadith*, 1-20.

Yusoff, Z. M., & Al-Qaisi, M. A. (2011). *The Influence of Arabic Language on Other Languages:*. Kuala Lumpur: Academy of Islamic Studies, University Malaya.

Yusoff, Z. M., & Saidi, M. (2008). Keupayaan bacaan Al-Qur’an dalam kalangan pelajar tingkatan empat: Kajian di beberapa buah sekolah menengah terpilih di negeri Terengganu [Ability of Al-Quran recitation among form four students: Investigation in a few selected secondary schools in Tereng. *Al-Bayan: Journal of Al-Qur’an & Al-Hadith*, 53-86.

Yusuf, A. (2001). *An English interpretation of the Holy Quran.* Lahore, Pakistan: Fine Offset Printing.

Yusuf, L. M. (1992). The design of a method of teaching Arabic of related purposes (the understanding of Quran and Hadith), for primary level in Malaysia. *Master Thesis for The International Institute of Khartoum for Arabic Language for non-Arabic speakers of The Arab League Educational, Cultural and Scientific Organization*.

Zwaan, R., Madden, C., Yaxley, R., & Aveyard, M. (2004). Moving words: dynamic representations in language comprehension. *Cogni. Sci.*, *28*(1), 611-619.