MULTIMODAL INTERACTIVE E-LEARNING: AN EMPIRICAL STUDY

Marwan N. K. Alseid

PhD

Department of Computing
School of Computing, Informatics and Media
University of Bradford

2009

MULTIMODAL INTERACTIVE E-LEARNING: AN EMPIRICAL STUDY

An experimental study that investigates the effect of multimodal metaphors on the usability of e-learning interfaces and the production of empirically derived guidelines for the use of these metaphors in the software engineering process

Marwan N. K. Alseid

Thesis submitted for the degree of Doctor of Philosophy in Software Engineering

Supervised by: Professor Dimitrios I. Rigas

Department of Computing
School of Computing, Informatics and Media
University of Bradford

2009

بسم الله الرحمن الرحيم

In the Name of Allah, the Most Compassionate, the Most Merciful

DEDICATION

This thesis is dedicated to my wife the twin of my soul, and our dear sons Hamza, Omar and Zaid for their constant love, encouragement and patience.

Marwan Alseid

ABSTRACT

This thesis investigated the use of novel combinations of multimodal metaphors in the presentation of learning information to evaluate the effect of these combinations on the usability of e-learning interfaces and on the users' learning performance. The empirical research described in this thesis comprised three experimental phases. In the first phase, an initial experiment was carried out with 30 users to explore and compare the usability and learning performance of facially animated expressive avatars with earcons and speech, and text with graphics metaphors. The second experimental phase involved an experiment conducted with 48 users to investigate their perception of avatar's facial expressions and body gestures when presented in both the absence and presence of interactive e-learning context. In addition, the experiment aimed at evaluating the role that an avatar could play as virtual lecturer in e-learning interfaces by comparing the usability and learning performance of three different modes of interaction: speaking facially expressive virtual lecturer, speaking facially expressive full-body animated virtual lecturer, and two speaking facially expressive virtual lecturers. In the third phase, a total of 24 users experimentally examined a novel approach for the use of earcons and auditory icons in e-learning interfaces to support an animated facially expressive avatar with body gestures during the presentation of the learning material. The obtained results demonstrated the usefulness of the tested metaphors to enhance e-learning usability and to enable users to attain better learning performance. These results provided a set of empirically derived innovative guidelines for the design and use of these metaphors to generate more usable e-learning interfaces. For example, when designing avatars as animated virtual lecturers in e-learning interfaces, specific facial expression and body gestures should be incorporated due to its positive influence in enhancing learners' attitude towards the learning process.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

3D	3 Dimensional
ANOVA	Analysis Of Variance
AVI	Audio Video Interleave
AVLBG	Auditory-enhanced Virtual Lecturer with Body Gestures
CBT	Computer Based Training
CD-Rom	Compact Disk-Read only memory
CMS	Content Management Systems
CV	Critical Value
CV	Controlled Variable
df	degree of freedom
DT	Difficult Task
DV	Dependent Variable
ED	End of Definition
E-learning	Electronic-Learning
ES	End of Statement
ET	Easy Task
FSLSM	Felder-Silverman Learning Style Model
H#	Hypothesis
HIK	High Important Keyword
ICT	Information and Communication Technology
IV#	Independent Variable
LCMS	Learning Content Management Systems
LIK	Low Important Keyword
LMS	Learning Management Systems
MIK	Medium Important Keyword
MMELP	MultiModal E-Learning Platform
MT	Moderate Task
NOR	Number of Requirements
Q#	Question
RL	Recall
RN	Recognition
S#	Statement
SD	Start of Definition
SS	Start of Statement
SUS	System Usability Scale
TVLFE	Two Virtual Lecturers with Facial Expressions
VB	Visual Basic
VLBG	Virtual Lecturer with Body Gestures
VLFE	Virtual Lecturer with Facial Expressions
VOELP	Visual Only E-Learning platform
WAV	Waveform Data

ACKNOWLEDGEMENT

First and foremost, I would like to thank ALLAH, the most merciful and the most gracious for giving me the strength and determination during the course of this research and throughout my life.

I would like to express my thanks and gratitude to the Applied Science University in Jordan for the scholarship they granted me to accomplish my PhD study.

Deep and sincere thanks to my supervisor Professor Dimitrios Rigas for his continual support throughout. I am highly appreciating his invaluable guidance and advice. He always encouraged and challenged me to reach a high quality work. I thank him not only for the professional supervision he offered me but also for his friendship.

I would like also to acknowledge the sacrifice, patience and love of my wife and children. Heartfelt and special thanks to my mother, my father-in-low and mother-in-low, my brothers and sisters, and my brothers and sisters in low who always included me in their prayers and supplications. In particular, I would like to express my deepest thanks and gratitude to Atef and Ma'en. Without their support, I would not have been able to start this work.

Last but not least, deep thanks go to my relatives, friends and colleagues for their support, help and motivation throughout the PhD period.

Chapter 1

Thesis Introduction

1.1 Introduction

In the last few years, Information and Communication Technology have shown accelerated developments in computer networks and hardware bringing easier and faster access to a large volume of information over the Internet. These developments opened the door widely to offer more opportunities to obtain knowledge in different disciplines through electronic learning (e-learning). Most of user interfaces heavily use visual stimuli to communicate information and this could result in overloading users' visual channel [1, 2] and missing important information being communicated [3]. The reviewed literature demonstrated the significance of incorporating both visual and auditory metaphors to enhance Human-Computer Interaction process. The inclusion of both visual and auditory metaphors in computer interfaces could contribute to reducing the amount of information delivered by specific sensory channel [4] and increasing the volume of communicated information [5] as well as allowing different information to be conveyed using different interaction metaphors [6]. In e-learning interfaces, multimodality has shown to be useful in enhancing the usability and users learning performance [7-13]. However, the need for additional research to integrate multimodal metaphors in e-learning applications is still highlighted. Multimodal interaction may help to alleviate some of the difficulties that e-learning users often encounter such as the lack of personal interaction [14-16].

This thesis investigates the use of multimodal interaction metaphors to provide audio-visual presentation of the learning information in e-learning interfaces. The experimental work undertaken within this investigation is aimed at exploring the influence of speech sounds (recorded), non-speech sounds (earcons and auditory icons) alongside avatars as virtual lecturers with facial expressions and body gestures on the usability and learning performance in e-learning systems. The main question is whether the inclusion of these interaction metaphors can improve the usability and learning in e-learning interfaces. The second question is related to the contribution of each of these metaphors to the expected improvement. In addition, how would the users evaluate the use of these metaphors when incorporated in e-learning interfaces? Finally, does it make a difference between one avatar and two in these interfaces? The following sections explain the aims and objectives of this thesis, the overall hypothesis and the method used to fulfil the aims. The chapter, finally, presents the thesis contribution and outlines its structure.

1.2 Aims

On overall, this research aims to investigate the effect of multimodal interaction metaphors on the usability and learning performance in e-learning interfaces, and to produce a set of empirically derived guidelines for the design and implementation of multimodal e-learning interfaces. The multimodal metaphors tested in this research involved auditory and audio-visual metaphors. The auditory metaphors consisted of recorded speech, earcons and auditory icons whereas the audio-visual metaphors incorporated the use of speaking avatars with human-like animated facial expressions and body gestures. These metaphors were investigated in different combinations to communicate learning information to users in order to evaluate its implications on the usability (in terms of efficiency, effectiveness, memorability, user satisfaction) and learning performance of e-learning interfaces.

1.3 Objectives

In order to achieve the aforementioned research aims, three experimental studies were conducted. In the first study, usability and learning performance of two different interface versions of experimental e-learning platform were explored and compared. The first interface was based on text and graphic metaphors in the presentation of the learning content and was referred to as Visual Only E-Learning Platform (VOELP). In comparison, the second interface incorporated a combination of graphics, earcons, recorded speech and speaking facially expressive avatar and was referred to as MultiModal E-Learning Platform (MMELP). The first study also investigated both interfaces with respect to tasks of different complexity and types. In other words, the required tasks were of increasing complexity (easy, moderate and difficult) and of two types; recall and recognition. The second study was dedicated to evaluate usability and learning performance of three different modes of human-like avatars when involved as virtual lecturers in the presentation of learning material in e-learning interfaces. The first presentation mode was named speaking Virtual Lecturer with Facial Expressions (VLFE) whereas the second presentation mode was named speaking Virtual Lecturer with facial expressions and Body Gestures (VLBG). The third mode however was based on Two Virtual Lecturers with Facial Expressions (TVLFE). The third study investigated whether the addition of non-speech sounds such as earcons and auditory icons could complement virtual lecturer presentation by communicating supportive information related to the delivered learning content. In addition to efficiency, effectiveness and user satisfaction, memorability of the added sounds were evaluated.

1.4 Overall Hypothesis

The overall hypothesis tested in this research was formulated as follows:

Compared to the use of text and graphics, the use of multimodal interaction metaphors will enhance the usability (in terms of efficiency, effectiveness, user satisfaction and memorability) and improve user's learning performance in e-learning interfaces.

1.5 Research Method

The method used to carry out this research included a literature survey, an initial experimental study and two further experiments. The data collection process was based on experimental observations and questionnaires. Experimental observations helped in gathering the data related to efficiency, memorability, effectiveness (or learning performance). However, data related to users' satisfaction and ratings was obtained by their responses to questionnaires. In all experiments conducted in this research, participating users were of different ages, backgrounds and gender. Also, they were undergraduates and postgraduates at the University of Bradford. Upon completion of each experiment, the obtained results were analysed and discussed. It is important to mention here that these results are content dependent and they are based on the learning material used. Finally, main conclusions were drawn and empirical guidelines for the design and implementation of multimodal e-learning interfaces were derived. The activities involved in this research method are illustrated in Figure 1 and described in the following subsections.

Literature Survey: The first step in this research was to review several relevant topics in the literature such as e-learning, multimodal interaction and multimodal e-learning systems. This review provided insights into the underlying theoretical background of

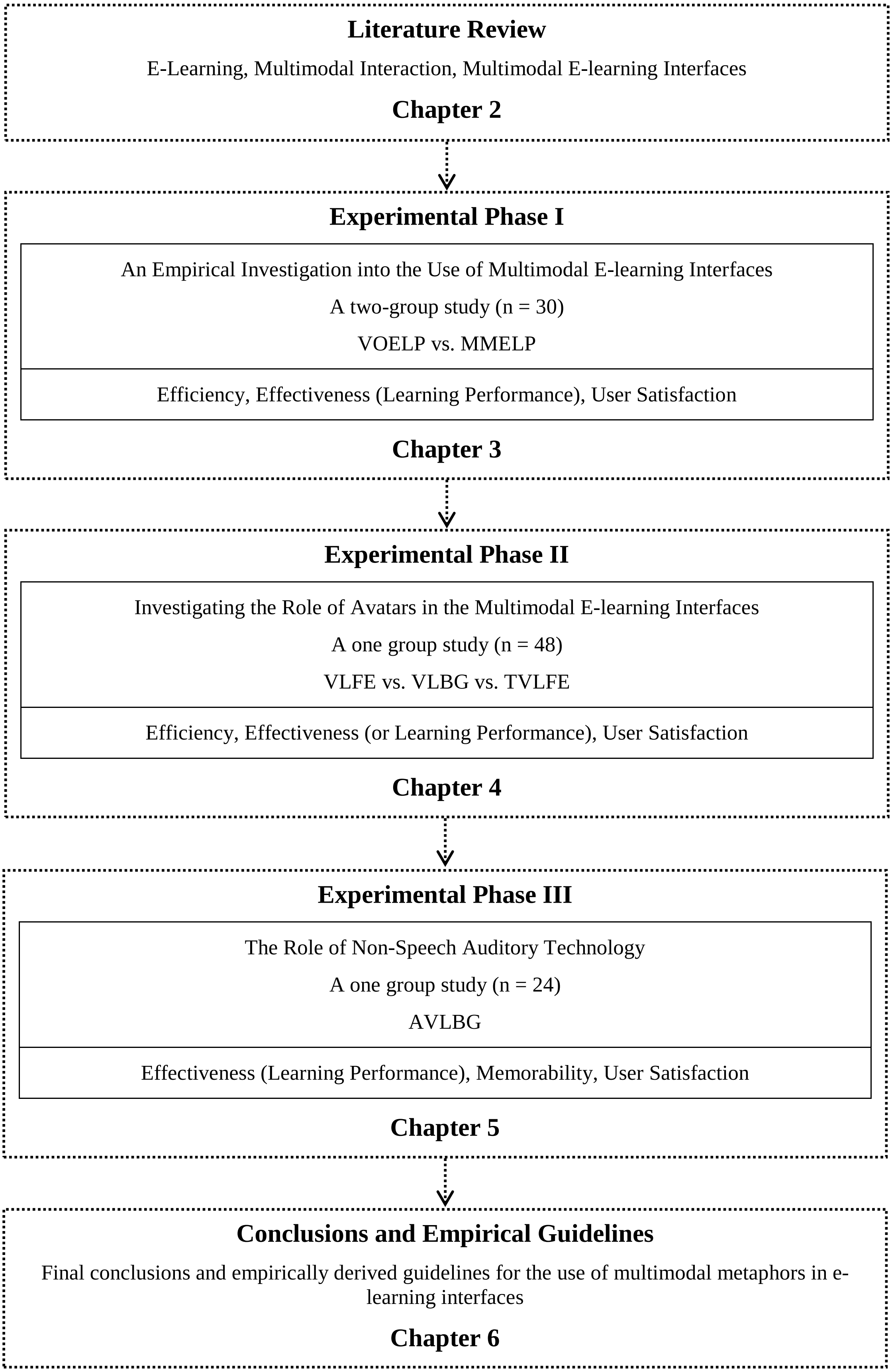


Figure 1: Structure of the thesis and experimental phases undertaken in this research

e-learning including e-learning definitions, pedagogical principles, trends and growth of e-learning as well as the hardware and software technologies used in e-learning. The review also covered multimodal interaction and related experimental findings. Finally, issues related to multimodal e-learning interfaces and examples of multimodal e-learning systems were reviewed.

First Experiment: This experiment represented an initial investigation of multimodal e-learning interfaces by performing an empirical study that was aimed to evaluate and compare efficiency, effectiveness as well as learning performance and user satisfaction of two different e-learning interfaces. Two independent groups of 15 users (n=30) were involved to perform common tasks and to answer questions related to the presented learning content. These questions were of two types; recall and recognition, and of increasing complexity. The first group of users; control, was provided with a typical e-learning interface with text and graphic metaphors. The second group; experimental, was provided with an interface that combined multimodal metaphors such as earcons, recorded speech and speaking facially expressive avatars. Both e-learning interfaces communicated the same information which was about three class diagram examples. The results of this experiment formed the basis to design and conduct the second experimental study in this research. On overall, the first experiment was designated to confirm findings of the literature survey and to carry out an initial evaluation to obtain an overall impression and understanding about the procedure and test criteria.

Second Experiment: This was carried out to investigate the use of avatars as virtual lecturers in e-learning interfaces. The aim of this experiment was to evaluate the usability aspects (efficiency, effectiveness and user satisfaction) and learning performance of three different e-learning interfaces recruited in the presentation of three common lessons about class diagram notation. The first interface incorporated the talking head of facially expressive avatar whilst the second interface used the speaking avatar with full body animation. The third interface, however, was based on talking heads of two facially expressive avatars. The experiment was aimed also to obtain users' views and evaluation of individual facial expressions and body gestures used in the tested interfaces in both the absence and presence of interactive context. The three experimental e-learning interfaces were tested dependently by one group of 48 users assigned to accomplish the required tasks and to answer recall and recognition questions in relation to the communicated learning content.

Third Experiment: The results obtained from the second experiment highlighted the need to design and perform a further experiment to explore the role that non-speech sounds could play in providing auditory messages to support and complement the speech of full-body animated virtual lecturer during the presentation of learning information. As an extension of the VLBG interface used in the second experiment, the experimental e-learning interface tested in the third experiment incorporated earcons and auditory icons to communicate auditory signals related to the most important parts of the presented learning material and referred to Auditory-enhanced Virtual Lecturer with Body Gestures (AVLBG). A total of 24 users were assigned to test the effectiveness, memorability, user satisfaction and learning performance of AVLBG by performing the experimental tasks and responding to recall and recognition questions about the delivered information.

Conclusions and Guidelines: In the final step of this research, the obtained results from the three experimental studies were discussed as a whole to draw final conclusions and to derive a set of guidelines to design and implement multimodal interfaces for e-learning systems. These guidelines are suggested to enhance usability and learning performance in e-learning interfaces.

1.6 Thesis Contribution

The research reported in this thesis contributes to the literature in both the e-learning and the multimodal Human-Computer Interaction. These contributions can be summarised in the following points:

- The thesis presents a novel approach for the investigation and methodology used in terms of combining multimodal metaphors to communicate learning information in e-learning interfaces. A set of three experimental studies were conducted to evaluate different combinations of multimodal metaphors when incorporated in e-learning interfaces. The obtained results demonstrated that the use of these metaphors could benefit in enhancing the usability in addition to enabling users to attain better learning performance. In other words, the thesis contributes to multimodal e-learning by providing different combinations of multimodal interaction metaphors that could be used in e-learning interfaces to enhance usability and learner performance. These combinations include: facially expressive avatar with recorded speech and earcons, and facially expressive avatar with recorded speech and full body gestures.

- The thesis also investigates users' evaluation of facial expressions and body gesture when used by avatars in both the absence and presence of interactive e-learning context, and suggests the adoption of specific expressions and gestures due to its positive influences on the usability of e-learning interfaces particularly in terms of enhancing users' satisfaction.

- Additionally, a novel application of earcons and auditory icons in e-learning interfaces is proposed to support the role of full-body animated facially expressive avatar in the communication of the learning material used in this research.

- Finally, the thesis suggests a set of empirically derived innovative guidelines for the design of more usable multimodal e-learning interfaces that could offers better learning for the users with respect to the communicated learning topic.

1.7 Thesis Outline

This thesis is structured in six chapters and a number of appendices. The following subsections describe these chapters and appendices.

Chapter 1: Introduction – This chapter provides an overall introduction to the thesis. The chapter briefly presents the research work carried out in terms of aims, objectives and the method followed in this thesis. It also outlines the thesis structure and its contribution to the research area in multimodal e-learning interfaces.

Chapter 2: Literature Review – This chapter reviews previous work in relation to multimodal e-learning systems and divided into three main sections; e-learning, multimodal interaction, and multimodal e-learning interfaces. The first section provides background information about e-learning such as definitions, advantages and limitations, underlying pedagogical principles, e-learning environments and technologies, and learning styles. Within the multimodal interaction section, the chapter provides the basic concepts of multimodal metaphors, usability evaluation and reviews several usability studies that highlighted the importance of multimodal metaphors in several computer applications. In the last section, design issues and relevant research into the use of multimodality in e-learning interfaces are discussed.

Chapter 3: Experimental Phase I: An Empirical Investigation into the Use of Multimodal E-learning Interfaces – This chapter reports an initial experiment performed to investigate the usability and learning performance of multimodal e-learning interfaces. This investigation was carried out empirically by assigning two independent groups of users to test two different versions of the experimental e-learning platform: text with graphics only (VOELP) and multimodal (MMELP). The obtained results were analysed and discussed in the light of the formulated hypothesis.

Chapter 4: Experimental Phase II: Investigating the Role of Avatars in the Multimodal E-Learning Interfaces – This chapter documents the second experiment that has been performed to investigate and compare the usability and learning performance of three different styles for incorporating avatars as virtual lecturers in e-learning interfaces. The chapter also explores users' views in regard to a set of 6 facial expressions and 10 body gestures used by virtual lecturers in both the absence and presence of interactive context.

Chapter 5: Experimental Phase III: The Role of Non-Speech Auditory Technologies – The 5th chapter evaluates the influence of non-speech sounds such as earcons and auditory icons when included in multimodal e-learning interfaces to provide auditory messages to indicate the key features of the learning content presented by virtual lecturer with full body animation.

Chapter 6: Final Conclusions and Empirically Derived Guidelines – The final chapter provides a summary of the experimental studies undertaken in this research, briefs the main conclusions and limitations drawn from the obtained results, and proposes a set of guidelines that could be utilised in the design of multimodal e-learning interfaces to enhance its usability.

Appendix A: First Experiment – Includes the questionnaire used during the first experiment in Chapter 3 (Appendix A1) and the three class diagram examples presented to users by the tested interfaces (Appendix A2). It also provides a frequency table with corresponding percentages for users' characteristics (Appendix A3), question answering time (Appendix A4), correctness of users' answers (Appendix A5) and their responses to the satisfaction questionnaire (Appendix A6).

Appendix B: Second Experiment – Presents the questionnaire of the experiment reported in Chapter 4 (Appendix B1), the textual content of the presented lessons (Appendix B2), and a frequency table with relative percentages (Appendix B3) for users' responses to the pre-experimental part of the questionnaire. It also provides users' evaluation for the facial expressions when demonstrated in the absence (Appendix B4 and Appendix B5) and presence (Appendix B6) of interactive context. The same data for body gestures are also presented (Appendix B7, Appendix B8 and Appendix B9). In addition, it presents time observations (Appendix B10) and correctness of users' answers to the required questions (Appendix B11) as well as their responses to the satisfaction questionnaire (Appendix B12). In the last part, users' answers to the post-experimental part of the questionnaire are shown (Appendix B13).

Appendix C: Third Experiment – Presents the questionnaire given to users of the third experiment described in chapter 5 (Appendix C1) and a frequency table with corresponding percentages for users' profiling (Appendix C2). It presents also the raw data for users' views in regard to the tested non-speech sounds in both the absence and presence of interactive context (Appendix C3). The raw data for the correctness of users' answers to the required learning evaluation question (Appendix C4) and memorability tasks (Appendix C5 and Appendix C6) are provided. The last part (Appendix C7) shows users' satisfaction data.

Appendix D: Technical Description – Includes the technical and design issues in relation to the development of the experimental e-learning interfaces used in the three experiments conducted in this research. In addition, this appendix presents samples of the key parts of the coding. This appendix is composed of three main subsections: D1, D2 and D3 related to the experimental e-learning platforms used in Chapter 3, Chapter 4 and Chapter 5 respectively.

Chapter 2

Literature Review: E-learning and Multimodality

2.1 Introduction

This chapter reviews and analyses the theoretical and practical research work in relation to the research carried out within this thesis. More specifically, it comprises three main sections; e-learning, multimodal interaction and multimodal e-learning. The first section presents introductory information about e-learning definitions, advantages, limitations and underlying pedagogical principles. It also provides an insight into the main components of e-learning environments and the technologies involved in e-learning process as well as the types of learning and human memory importance in the learning process. The second section provides the basic concepts of multimodal metaphors utilised in this research namely, visual metaphors, speech and non-speech sounds in addition to avatars, covering previous research studies in order to shed light on the significance of these metaphors in enhancing user to computer interaction in a variety of problem domains. The last section however focuses on the use of multimodal interaction metaphors in e-learning interfaces and on the research studies that highlights the usability and learning enhancement due to the utilisation of multimodality in e-learning.

2.2 E-learning

The continuous developments in information and communication technology (ICT) resulted in an accelerated development in educational technology and increased the demand for e-learning. More specifically, the Internet technology makes it easier and faster to access a large amount of educational material. The growth of e-learning market has been shown to be rapidly increasing. In 2000, it was estimated by Fortune Magazine that $22 billion would be invested by the end of 2003 in e-learning universal market [17]. Two years later, in 2002, this market reached $90 billion [18]. In Europe, it was predicted that the size of e-learning market would reach $4 billion in 2004 [19]. Also, it is expected that there will be about five million online learners within the next ten years [20]. However, the growth in web-based training in USA has been estimated to reach $11.4 billion in 2003 from $550 million in 1998 [21]. Based on the expectations of the International Data Corporation (IDC), the e-learning market will be between $21 billion and $28 billion by 2008 ([241] cited in [242]). Moreover, it has been reported by the global industry analysts that more than $52.6 billion will be spent on e-learning by 2010 and that the worldwide annual growth in e-learning market will be between 15% and 30% by that year according to the global strategic business report [243].

2.2.1 E-learning Definition

E-learning can be defined as a term that describes the learning process in which ICT could be utilised [22]. In e-learning, the learning material can be offered, circulated, and then accessed at anytime and anywhere by the employment of electronic technology such as computers, networks and communications [23-27]. These definitions were supported by Govindasamy [28] who stated that “e-learning includes instructions delivered via all electronic media including the Internet, intranets, extranets, satellite broadcasts, audio/video tape, interactive TV, and CD-Rom”. Other definitions were introduced and concentrated on the use of Internet technology in managing educational material, storing information about learners, and facilitating communication and cooperation among learners as well as monitoring their progress [29, 30]. In brief, e-learning is a learning method that is based on the use of information and communication technology [18] to enhance the learning experience in terms of learner's knowledge and performance [31].

2.2.2 Benefits and Limitations

There are several benefits of e-learning that have been reported by many authors. In comparison with traditional learning, e-learning offers better adaptation to individual needs, provides more flexible learning in terms of time and location, and facilitates monitoring learners' knowledge and skills [29]. Also, e-learning content could be easily and quickly accessed, updated, and then redistributed again so that all learners receive the same educational material in the same manner [32]. Furthermore, e-learning enables users to learn collaboratively [23, 33], and could enhance their motivation and interest in regard to the presented material [12]. E-learning could also be used to accommodate different teaching approaches [34] and learning styles [31, 35]. In addition to the aforementioned, learning becomes less expensive in terms of money and time [21] and less daunting to learners who can examine new things without being afraid of mistakes [36]. Also, the learning can be gained at learner's own pace [23, 35].

However, e-learning environments also have challenges and limitations. For example, users of e-learning are supervised only by parents or other adults and not by a teacher. Also, teaching methods and computer technology must be combined appropriately [37]. However, this technology including hardware and other ICT resources such as high-speed Internet connection is not always available and accessible [38] and as a result the delivery of e-learning content will be delayed [18]. Furthermore, learners are not always satisfied with computer-based learning [39] and experienced a lack of physical interaction with tutors [14] as well as among themselves [40, 41]. The lack of physical interaction means the absence of facial expressions and body gestures that could be used to convey important verbal messages [42]. Another limitation of e-learning is that e-learners need to have adequate computer skills to deal with e-learning applications [18, 43] and this could frustrate novice ICT learners [31, 44]. Hence, users' attitude towards e-learning must be enhanced and their accessibility to the needed technology should be facilitated [14].

2.2.3 Pedagogical Principles and E-learning

From a pedagogical perspective, it is not always the case that every e-learning solution presents high quality learning. E-learning is not just a means of knowledge delivery at any time and any where but also it is a means to "train the right people to gain the right skills or knowledge at the right time [28]. Therefore, there are fundamental pedagogical principles that must be considered to insure the successful implementation of e-learning. According to Govindasamy [28], pedagogical principles are "theories that govern the good practice of teaching" and its applicability can be extended from traditional learning to e-learning. These principles include different aspects of the learning process. For example, e-learning solutions should allow constructing the new knowledge based on previous experience and avoid swamping the learners with information. Also, e-learning is self-directed and therefore identifying learning competences, prerequisite knowledge and targeted goals as well is important to facilitate independent learning [31].

Another pedagogical aspect is related to evoking learner's intrinsic motivation. E-learning is an independent activity and learners tend to become frustrated. Therefore, they can be attracted and motivated by incorporating appropriate support and feedback in the design of e-learning [31]. In addition, the effectiveness of e-learning can be enhanced by accommodating differences in individual characteristics among learners [31]. So, a variety of pedagogical learning activities can be provided in e-learning to suit these differences [45]. Human brain absorbs knowledge in a gradual manner [31] therefore e-learning can be designed to enable the progressive assimilation of the presented learning content. According to Mayes and Freitas [46], this content should be decomposed into a hierarchy of small learning units and presented in a well-structured manner, supported by well-designed examples and multimedia elements such as text, graphics, sound and animation. Additionally, e-learning can be designed to simulate the social nature of traditional classroom in which students can learn from each other by facilitating social activities such as chatting, discussion forums and e-mail [31, 47]. Neglecting these pedagogical principles will result in weakening learning content, learners' performance and their attitude towards e-learning [28].

2.2.4 E-learning Environment

Different types of e-learning solutions have been developed varying from simple web page to enterprise e-learning system. To ensure the success of e-learning solution, consideration must be given to the needs of learners, instructors and other stakeholders in the learning process [43]. According to Henry [48], e-learning solution is composed of three main elements which are: content, technology and services. He stated that the content is the most significant one but how all these elements are integrated, introduced, implemented and managed is also important. The content incorporated all the material delivered including the material usually presented in classroom-based learning, and that customised for e-learning as well as any other knowledge the developer might offer. Design and development of this content must be consistent with the desired learning objectives therefore, a systematic five-phase approach to develop and evaluate e-learning content was suggested [28]. In the first phase (analysis), both learner and tasks are analysed. Different aspects of learner characteristics (academic, personal and social) are determined and the required learning tasks are identified. In the second phase; design, instructional objectives are specified and strategies to accomplish these objectives are defined. The third phase involves producing the initial draft of learning material by creating concept maps, storyboards, and course evaluation questionnaires. In the following, formative evaluation phase is conducted wherein the produced content is reviewed by experts and tested by sample learners, individually and grouped. However, the last phase; production, includes releasing the first version of the e-learning material.

The technologies used in e-learning involve hardware and software. In its early days, e-learning was based on floppy disks meaning that the incorporated information was limited and lacks of media such as sound and graphics [11]. Later, CD-ROMs were utilised to deliver the learning and training content in what has been called Computer-Based Training Systems (CBT) developed using different authoring software such as Authorware, Macromedia Director and Toolbook [20]. Although CBT systems were innovative and effective, it was difficult and expensive to update; furthermore, tracking learner's performance was not available through these systems [49]. These drawbacks were resolved by the next generation of e-learning known as Learning Management Systems (LMS) making use of computer networks technology [20]. Blackboard and e-College are two examples of LMS which facilitate administrative and tracking features in terms of learners' activities and achievements as well as course management, and provide collaboration environment for chatting, discussion and synchronous learning [20, 43]. In addition to LMS, another four types of software technology can be used either on their own or combined to develop e-learning solutions. These include programming languages, authoring packages, content management systems (CMS), and learning content management systems (LCMS) [43]. Programming language are used for developing online learning but instructors need to have sufficient programming skills to develop the content; however, authoring tools can enable the developer to design an e-learning interface without worrying about the code because it is performed in the background. CMS are dedicated to facilitate managing and manipulating the content of e-learning courses and provide a huge database of material that can be searched and used by instructors in developing their courses. LCMS combines the capabilities of the above mentioned software technologies. In other words, it provides enterprise solutions that fulfil all what is needed to create comprehensive e-learning software [43].

The main focus of this thesis is the empirical investigation of usability issues of using multimodal interaction metaphors in e-learning interfaces and not the development of the e-learning tool itself. In other words, comparing the software technologies that can be used to develop e-learning interfaces is beyond the scope of this research. The programming language Visual Basic 2005 (VB) from Microsoft Corporation was used to develop the experimental e-learning platforms evaluated in this research. VB offers powerful features in graphical user interfaces and object-oriented programming through a large library of built-in components that facilitate the development of computer applications. Additionally, VB provides suitable functionalities to integrate audio and video files in these applications [215].

According to Hamilton et al. [50], the technologies incorporated in e-learning can be classified into two main types; scheduled delivery platforms and on-demand delivery platforms. Scheduled delivery platforms such as video broadcasting, remote library, and virtual classrooms imitate real learning environments but with time and place limitations. This technology was enhanced by the on-demand delivery platforms that facilitate any time and any where learning in the form of interactive CD-ROMs and web-based training [50].

Due to the increasing use of mobile devices and wireless communication technology, a new more flexible educational method called mobile learning (m-learning) was evolved [51]. It is a new stage of computer-aided multimedia and interactive based learning in which the mobile devices (Palms, digital cell phones, PDAs) are used to deliver e-learning [52]. M-learning is aimed at presenting a specific kind of knowledge depending on location, situation, devices, and learner, and could be available at any time and any place [29]. As well, "Mobile learning can be viewed as any form of teaching or studying that happens when the user is interacting through a mobile device" [53]. Another term in the e-learning paradigm is wireless learning (w-learning) which can be used with m-learning interchangeably [51].

2.2.5 Styles of Learning

Individuals differ in the style they use to learn. The way by which each learner acquires, retains and retrieves information is referred to as learning style [54, 55]. It is distinctive and affected by personal characteristics [56]. Several learning style theories were developed. Based on their learning styles, Kolb’s model [57] classifies learners into four types which are: *divergers, assimilators, convergers* and *accommodators*. People in the first category tend to view things from different angles and excel in generating meaningful ideas from group discussions. Assimilators enjoy listening to lectures and are distinguished by their ability to analyse and organize separate information and absorb it as a whole. In this category, learners prefer the well-organised presentation of the learning material. However, convergers who perform better in problem solving, decision making and implementation of ideas, do not like lectures and group work, and prefer to learn by doing. On the other hand, accommodators are intuitive, prefer exploration and tend to adapt what have been learnt for their needs in a creative manner [58].

Kolb's learning style model was modified by Honey and Mumford who developed a learning style questionnaire to be used in identifying learning styles based on learners' strengths and weaknesses [59]. Accordingly, the four learning styles in this model are *activists*, *reflectors*, *theorists* and *pragmatist*. Felder-Silverman learning style model (FSLSM) [60] is another approach that could suit the design and development of e-learning systems [61, 62]. In comparison with Kolb's and Honey and Mumford's models, FSLSM gives more details about individual learning styles [63] and characterizes each learner according to four dimensions; information processing (active/reflective), information perception (sensing/intuitive), input mode (visual/verbal) and understanding (sequential/global) [60]. *Active* learners prefer to learn by experimenting and group working whilst *reflective* learners tend to learn by thinking first and prefer to work individually. *Sensing* learners prefer to learn concrete material and are practical whilst *intuitive* learners like abstraction and more creative. *Visual* learners remember best by seeing things in pictorial format whilst *verbal* learners prefer to learn from words either written or spoken. *Sequential* learners prefer to learn step by step following linear and logical progress whilst *global* learners learn in large leaps [64]. Most individuals prefer one or two of these styles whereas effective learners are those who are able to make use of them all [14].

In addition to experience with computer and gender, learning style preferences of learners could affect their perception and acceptance of ICT-mediated learning [65]. A study by Shaw and Marlow [14] demonstrated that students were not satisfied with the use of ICT and experience a lack of personal interaction. Therefore, it is important to consider different learning styles of learners in instructional design and teaching strategies as it could result in enhancing their motivation and performance [58]. Additionally, enhancing learners' accessibility and experience regarding ICT could be the key factor in obtaining better learning outcomes [14].

2.2.6 Human Memory and Learning

Human memory enables storing and retrieval of information and can be classified into three types; sensory memory, short-term memory and long-term memory [66]. Sensory memory is the initial storage of information captured through sensory channels and lasts for up to a few seconds whereas short-term memory (also referred to as working memory) can hold a small a mount of information in an active state for a short time period (up to 20-30 seconds) before being forgotten or passed on to long-term memory [67]. In other words, sensory memory and short-term memory are restricted in terms of capacity and duration, while stored information is available for a strictly limited time. In comparison, long-term memory offers a much larger storage of information over long time [66] .

For the learning process, working memory capacity is important for the remembrance and retention of the presented information [68]. This capacity is included in the cognitive traits model which profiles learners according to their cognitive abilities [69]. When a large amount of information is presented at a time, the learner will be cognitively overloaded because working memory will be overwhelmed with what is being presented [70] and as a result part of that information could be lost [71]. Dual coding theory postulates that the working memory has two subsystems; verbal to process spoken or auditory information and non-verbal to process visual information [72]. When both systems are utilised simultaneously, the capacity of working memory could be extended [73], cognitive load will be reduced and more processing resources become available providing more effective learning to users [74].

2.2.7 Usability evaluation in e-learning interfaces

Usability is one of the principal factors to evaluate Human-Computer Interaction [75] and software quality [76]. It could be defined as the "*extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction*" [77]. Effectiveness is defined as the "*accuracy and completeness with which users achieve specified goals*" whereas efficiency is the "*resources expended in relation to the accuracy and completeness with which users achieve goals*" and satisfaction is the "*freedom from discomfort, and positive attitudes towards the user of the product*" [77]. These usability attributes could have more specific meaning within e-learning context. In this regard, effectiveness represents learners' ability to accomplish the learning objectives, efficiency is evaluated by the time cost to reach these objectives, and satisfaction is assessed by users' interest toward the presented content and their tendency to continue to learn [78]. Therefore, successful completion of the required tasks can be used as a measure of effectiveness and the time spent to perform these tasks as a measure of efficiency [75]. However, users' responses to questionnaire comprising statements in relation to the tested interface can indicate their satisfaction level as well as their experience due to the interaction with that system. In addition, other usability attributes such as momorability can be considered and measured by users' ability to remember the features of the system and its functionalities [75].

2.3 Multimodal Interaction

Human senses provide different information using different channels in every day interaction. Most of the computer systems use the sight sense in their interfaces and this might result in overloading human visual channel during the interaction [2] causing in some cases loss of information being communicated [3]. Multimodal computer interfaces are those that involve more than one interaction modality to incorporate different human senses in the interaction process [79]. These metaphors can be categorised into visual (e.g. text and graphics), auditory (e.g. speech and non-speech sounds) and audio-visual such as avatars with facial expressions and body gestures. Multimodal metaphors can improve the Human-Computer Interaction by involving more than one channel to convey different information [6], thus reducing information overload [4, 80]. It makes the human computer interaction closer to natural human to human and human to environment interaction [81] and could overcome the lack of face-to-face communication in computer user interfaces [82, 83]. Also, multimodal interaction facilitates the presentation of similar information using different channels [6] enabling the users to deal with the computer application using the most suitable type of interaction to their abilities, preferences and needs [84]. In addition, the usefulness of multimodal interaction in enhancing the usability of user interfaces was proven by several studies [11, 85-93]. Furthermore, multimodal metaphors can be used to provide interface for the disabled [79, 81].

2.3.1 Visual Metaphors

Visual metaphors are used in computer interfaces to represent information in textual and graphical format and communicated to users using their visual channel [94]. These metaphors have been used in early eighties by Xerox Star system [95] and successfully adopted by Apple Macintosh operating system [96]. User interfaces in Macintosh computer systems enabled the users to use the mouse in treating iconic and pictorial representations of files and folders. For example, delete a file by dragging its icon to trash folder.

The use of visual metaphors has been shown to have positive influence on computer systems' usability in terms of offering simpler and easier interaction between the user and the system; however crowding the interfaces with overwhelming graphical and textual information might confuse the users and scatter their concentration [97]. The addition of auditory metaphors could contribute to reduce the load on the visual channel in receiving the communicated information [98]. With visual interaction, users need to keep directing their sight toward the output device. On the other hand, auditory information can be captured from all sides regardless of head and body direction allowing different information to be obtained by other channels (e.g. visual) [6]. For example, non-speech sounds could be used to capture user attention to specific event while user's visual channel is involved in performing different task [99].

2.3.2 Speech Metaphors

The use of speech in Human-Computer Interaction began a long time ago and can be considered as the most suitable metaphor to communicate textual information using the human auditory channel [100]. In addition to communicating auditory feedback related to the current state of the system [101], speech output has been shown to be useful to provide the users with the information they needed in different applications such as help disk [102], e-banking, e-news, and email [103] in addition to search engines [104], note-taking [105], and talking agents in e-commerce [106, 107]. Furthermore, it is widely utilised as assistive technology for visually impaired users [100, 108-111].

Speech sounds can be categorised into natural speech and synthesised speech [109]. Natural speech is a human spoken speech recorded using digital technology [109]. This type of speech is characterised by its naturalness and its ability to provide a human like interaction with computer systems [103]. However, it needs to be pre-recorded, manipulated and saved as sound files prior use. With an increasing volume of these files, large storage space is needed thus recorded speech is not widely applied in those systems that involve a large vocabulary [112] and is restricted to the communication of limited short spoken messages which can not be generated automatically during the interaction process [107]. On the other hand, synthetic speech is a simulation of human speech generated by speech synthesizers based on two different techniques, concatenation or synthesis by role [101]. In the first one, speech messages are produced by concatenating pre-recorded segments of human voices after being stored in a database system. However, the second technique, also referred to as formant speech, is based on creating the speech sounds artificially using phoneme generation rules and thus can be used to produce speech in run-time [113]. Compared to concatenated speech, the formant speech is of poorer quality [114]. Although the speech synthesizer technology is a faster and more flexible solution for the production of high quality speech sounds, the created speech still sounds computer generated and therefore natural recorded speech is recommended [103] and could be comprehended better [115].

2.3.3 Non-Speech Metaphors

Non-speech sound is another multimodal interaction metaphor that has been involved in Human-Computer Interaction to incorporate the auditory channel in the interaction process. It has been shown that the use of non-speech sounds can contribute to enhance users’ performance as well as the usability of interfaces (see the following two subsections). Non-speech sounds, compared to speech sounds, are language-independent, provide faster communication and can be understood quicker with the presence of sufficient training [116]. It can be segregated into two types: earcons and auditory icons.

2.3.3.1 Earcons

Earcons are short sounds of musical nature used in Human-Computer Interaction for the communication of information about objects, operations and interaction in computer interfaces [117]. In other words, earcons are defined as "abstract, synthetic tones that can be used in structured combinations to create sound messages to represent parts of the interface" [98, 118]. These non-speech sounds are constructed from short sequences of musical notes [119] that can be combined to convey more complex information [99]. According to Blattner et al. [117], earcons can be one-element (simple) or a compound. A single note and a single pitch are examples of the one-element earcons. However, compound earcons can be different combinations of simple earcons. In order to discriminate different earcons within these combinations, sound attributes such as timbre, register, pitch, rhythm, duration, tempo, intensity and spatial location can be used [118]. The use of earcons in user interfaces is based on the linkage between the incorporated earcons and the information to be communicated meaning that the user had to rely only on his/her memory to interpret the delivered auditory message [106].

Earcons have been evaluated in different problem domains and demonstrated that it could be effectively utilised to communicate information in sound [99]. It has been employed to enhance users' interaction with graphical components used in user interfaces such as scrollbars [120], buttons [121], menus [122, 123], progress bars [124], and tool palettes [125]. The auditory feedback provided to the users by earcons assisted in resolving usability problems associated with the use of these graphical widgets and contributed to reduce task completion time, error rate, error recovery time and mental workload without annoying or frustrating the users [1]. Examples of these problems include 'kangarooing' with the thumb wheel of scroll bar and 'slipping of' with buttons. In addition, earcons have been shown to be beneficial to enhance users' interaction with mobile devices where the inclusion of structured musical sounds helped the users to overcome the lack of visual feedback due to small screen size in these devices [116, 126, 127]. Furthermore, the software development process has made use of earcons to communicate auditory messages related to program coding, execution and debugging [128-130] in terms of variable values, compilation errors, their types and locations in the code.

Earcons are also utilised as an assistive technology for visually impaired users to access graphical representations [131-135], spreadsheets [136] and numerical data tables [137] and to enable them to draw line graphs of two dimensions as well after its data points are being communicated by musical notes [138]. AudioGraph [134, 139, 140] is an experimental platform by which earcons have been successfully utilised to convey graphical information to users with visual impairments. In this system, coordinate locations and graphical shapes such as lines, squares, rectangles and circles are all communicated by musical sounds. The potential of usability enhancement due to the incorporation of earcons in multimodal user interfaces has been also demonstrated in other application domains such as stock control systems [91, 141-143], knowledge management systems [86, 144], email browsing [145-149] and search engines [92, 104].

2.3.3.2 Auditory Icons

Auditory icons are non-speech sounds from the surrounding every day life used to communicate different objects and actions in computer interfaces [150] based on the mapping between these sounds and the information to be presented [151]. For example, a noise can be introduced as glass breaking sound. SonicFinder [152], SharedARK [153] and ARKola [154] are examples of systems in which auditory icons have been developed and used. In SonicFinder, environmental sounds are recorded and utilised to represent interface objects, operations and attributes along with visual feedback. When selecting a file, for instance, the icon of that file is highlighted and the sound of hitting (selection) wood (file) is played with the file size being communicated by the frequency of the sound. The ARKola, however is a simulation system in which the auditory icons are communicated to monitor a nine-machine bottling factory. The system attaches each machine with specific sound to indicate its status and at the same time all sounds are played together to communicate the overall ongoing processes in the factory. Auditory icons also can be successfully combined with other multimodal metaphors such as speech and earcons to communicate information for mobile telephony users [119].

The implementation of environmental sounds in user interfaces demonstrates that it could be effectively employed to convey both simple and complex information. An important advantage of auditory icons is its ability to convey different information using single sound [155]. For example, in a messaging system, a weighty sound can be played to indicate both the arrival and the size of the received message [150]. In addition, these sounds are well known to users and can provide natural mapping with the delivered data; therefore they can be easily learned and remembered [155]. However, these mappings are sometimes difficult to establish [156]. In Gaver's SonicFinder [152], for example, copying had no equivalent environmental sound. Consequently, it was presented by pouring a liquid auditory icon. Confusing the user is another possible disadvantage of auditory icons particularly when it is derived from the same sound source such as hammering with walking [119].

In comparison, earcons are more flexible as it can be used to represent any object, operation or interaction in computer interfaces [156] and can be designed in structured combinations to represent hierarchical information (such as menus and its components [157]) that could be differentiated by pitch, timbre and other sound attributes [117]. On the other hand, earcons are more abstract sounds that do not have direct meaningful association with the data it represents [156]. Therefore, this association should be learned from scratch so that the users can easily remember its representation [79, 117]. In summary, each of earcons and auditory icons has advantages and disadvantages. According to Brewster [155], combining both of them in a multimodal interface could be the best choice and this has been demonstrated by some experimental studies [86, 119, 158].

2.3.4 Avatars

An avatar is another multimodal interaction metaphor that could involve both visual and auditory human senses. It is a computer-based character that has been utilised to virtually represent one party in an interactive context [159, 160] with the ability to communicate verbal and non-verbal information [161, 162]. Verbal communication refers to the use of speech and written messages whereas nonverbal one can be attained by facial expressions and body gestures [161]. In general, avatars can be classified as abstract, realistic and naturalistic [163]. Abstract avatars are cartoon-like interactive characters with limited animation [107]. The help avatar embodied in Microsoft's office application is an apparent example of these avatars, designed to provide the users with helpful information during the preparation of their documents [162]. Realistic avatars offer real representation of humans being generated based on captured static or video images and are used in several applications such as games, movies and teleconferences [164]. The drawback here is the cost associated with the hardware needed to implement this technology [163]. However, the naturalistic avatars are humanoid in its appearance and widely utilised in collaborative virtual environments to represent the interacting users [165].

The employment of avatars in virtual environments allows users in physically-separated locations to interact with each other [166, 167] in a virtual world wherein everyday human expressions can be used to express users' feelings and emotions [160] and this could provide them with a sense of presence and involvement in social computer-mediated activities [165, 168]. This could enhance the interaction between users who are communicating in these environments. User's avatar can reflect his/her actions, attention and interactive behaviour to the others, thus providing a high level of mutual awareness [165, 169]. Virtual environments are implemented in web-based applications such as entertainment, edutainment, e-learning, simulation and e-commerce [168].

It has been argued by Fabri et al. [170, 171] that facial expressions with simple features can be displayed effectively and efficiently by avatars in user interfaces. They found that the six universal facial expressions (as regarded by Ekman et al. [172]): happiness, surprise, anger, fear, sadness and disgust in addition to neutral, can be correctly recognised by users even when communicated with limited facial features [173]. Another study conducted by Fabri et al. [174, 175] demonstrated that the addition of facially expressive avatars in the interface of instant messaging tool improved users' involvement in the communication tasks and created a more enjoyable experience, providing them with higher senses of presence and togetherness with another person in the chatting process. Facial expressions were also explored as a therapeutic technology for autistic users. This category of users was found capable of understanding and using the facial expression shown by their avatars [174, 176]; however, in this case, different users need different treatments particularly those with severe autism due to significant differences in their social abilities [177].

When speech metaphor is integrated with expressive avatars, a more realistic and intelligible audio-visual interaction could be introduced by which both verbal and non-verbal information are communicated using spoken messages in company with relative facial expressions and body gestures [161, 178]. In order to attain this integration, facial movements in terms of jaw, lips, teeth and tongue need to be synchronized in a normal manner so that the produced speech is correctly articulated [161, 179].

In addition to facial expressions, body gestures are used by humans to communicate non-verbally in a wordless manner where the movements of body, head and hands can be used as an illustration tool to supplement our speech when we feel that it is unable to express what we would like to say [107, 180]. Although different people have different cultures and traditions, most of the human body gestures have common interpretation over the world. As examples, shaking head from side to side denotes negation whilst nodding indicates agreement or confirmation. However, some gestures (such as the thumb-up) have different meanings in different countries. According to Pease [181], it is widely agreed that facial expressions and body gestures are mainly used to convey attitudes during interpersonal communication and in some case it could replace spoken and written messages.

Different studies were devoted to examine the effect of specific facial expressions and body gesture as well as to evaluate users' perception towards these modalities when used by speaking avatars in the interface. For example, Gazepidis and Rigas argued that incorporating talking virtual salesman with facial expressions and body gestures in e-commerce interfaces are more appealing to users compared to the textual presentation of products [82, 83]. Based on further empirical investigation [107, 182], it has been proposed that some facial expressions are more preferred than others and the same is true for body gestures. A set of 13 expressions and 9 gestures were evaluated in both the absence and presence of interactive context; among them the expressions: happy, interested, amazed, neutral, positively surprised and thinking, and the gestures: open palms, head up, chin stroking, hands clenching and hand steepling were found to be the most positively viewed by users. Furthermore, these expressions and gestures resulted in enhancing users' attitude and their ability to remember the delivered information more accurately [107]. These results have been supported later by other experimental studies where the inclusion of positive (happy, amazed and positively surprised), negative (sad, tired and disgusted) and neutral (neutral and thinking) facial expressions significantly contributed to enhance the satisfaction of users as well their understanding and remembrance of the presented knowledge achieving a higher level of usability [106, 144, 183]. Users' perceptions of avatars could be improved when human-like expressions and body gestures are embedded. A study performed by Cowell and Stanny [184] demonstrated that the facial expressions could promote users' feeling of credibility and trust towards interface agents. In addition, animating an avatar's body in a way resembling human gestures could make it more friendly to users [185].

Additionally, even the presence of simple facial animation such as happiness and eye gaze could have a positive influence on the users. In particular, the happy expression was found to be useful in enhancing users' attitude, intentions and experience [186] in addition to making them more pleasant, confident and responsive to the required tasks [187]. Garau experimentally investigated how important it is to use the eye gazes by the avatar when utilised to represent users while speaking to each other and found that it has the potential to enrich the quality of conversation as communication process [188].

In order to maximise the benefits of multimodal interaction metaphors in Human-Computer Interaction, guidelines for the design and implementation of such metaphors were empirically derived as a result of a series of experimental studies. Part of these guidelines was dedicated to help interface designers in the creation and implementation of earcons [1, 118, 134, 189-191] and avatars [107, 180, 185, 192] whilst other guidelines were concerned with the effective use of different combinations of multimodal metaphors such as speech with avatars [107], speech with earcons and auditory icons [97, 119, 193], and earcons with speech [142, 149, 194]. Other guidelines however were introduced to provide general guidance for the design of multimodal user interfaces [6, 84, 195].

2.4 Multimodal E-learning

Multimodal learning can be defined as the learning process that involves more than one modality in the representation of the learning material [196]. More specifically, the content knowledge can be presented to the learners using verbal communication means (such as text and speech) in company with related non-verbal illustrations such as photos, graphics, videos and animations [197]. Multimodal approach could have positive implications on the learning environments. When only one sensory channel (mostly visual) is used to communicate the presented information, overloading cognitive capacity of the learner is more likely to happen as opposed to the use of multimodality where both visual and auditory channels can be involved to acquire the knowledge contained in the learning material [198, 199]. In addition, learners' ability to understand that material could be improved by the multimodal presentation. Therefore, multimodal learning environments are considered to be more effective as compared to that which incorporates single modality [198, 200-202]. In fact, this has been demonstrated by a series of experimental studies where the research on the impact of multimodal interaction metaphors on learning performance as well as the usability of e-learning interfaces have gained a considerably interest.

The use of speech and non-speech sounds to enhance users' learning has been investigated in different domains. For example, the experimental work performed by

Bonebright et al. [9] tested the effectiveness of earcons to communicate data sets represented in two types of graphs from subjects studied by students during the academic coursework. Flute sound was used for bivariate graphs whereas both flute and bassoon sounds were used for multivariate graphs. For bivariate graphs, monophonic sound was presented to both ears while stereo sound for multivariate graphs was distinguished by different timbres, one variable to the right ear and the other variable to the left ear. While a sound was played, four visual graphs were displayed on the screen and the student has to click the graph that matches the communicated sounds. It was found that students were able to successfully match the auditorily conveyed graph with its corresponding visual representation on screen [13].

A learning-training simulator developed by Li et al. [203] incorporated audio and visual feedback for students while they were trained for lathe operations. The inclusion of environmental sounds to communicate warning messages in addition to different basic operations in lathe machine was found to be useful in providing safer, efficient and satisfactory technical training for students.

Speech technology in combination with earcons, text, graphics and animation was investigated in a multimedia online-learning system by Rigas and Hopwood [11] to examine its use with learning topics of increasing complexity and to evaluate as well its effects on users' performance in different complexity learning tasks related to these topics. In addition to text, recorded speech was utilised to communicate instruction to users whereas earcons, animated arrow and graphical transparent icons with animated text were used as navigational cues to communicate the part of the interfaces to be clicked during the interaction. Users' of the multimedia learning interface outperformed their counterparts who have been exposed to the same content with only the text with graphics in terms of responding to the learning tasks more accurately, particularly those related to intermediate and advanced learning topics. Speech and earcons were also investigated in note-taking applications for learning purposes. Specifically, recorded speech was combined with textual and graphical metaphors for taking notes about the displayed content and shown to be contributing to reduce the note-taking time in addition to improving the accomplishment rate and users' satisfaction in such activities compared to visual-based note-taking in e-learning applications [105]. These findings are supported by another study [89, 204] which explored the use of speech and earcons to support students in recording their own notes about the learning content and to make use of these notes in reviewing that content. It was concluded that the usability of note-taking could be substantially promoted by multimodality where users' learning performance and experience in terms of question answering time and correctness of answers as well as users' attitude have been enhanced [205].

The use of avatars in e-learning environment to serve educational purposes has been explored by several studies which demonstrated the benefits that could be gained due to the inclusion of such metaphors in e-learning interfaces. The presence of lifelike avatars could enhance learners' motivation and engagement in the learning activities [206, 207]. When they interact with a lifelike agent, learners may find their learning more entertaining [208] and taking place in an actual learning environment [7]. In fact, the inclusion of avatar as a learning agent could reinforce the social nature of the learning process [209, 210]. For example, Robertson et al. [211] carried out an experimental investigation on the effect of animated avatars on students' attitude when these avatars were incorporated to help them in story writing, and found that students who used the avatar-based interface expressed stronger tendency to use the interface as compared to those who interacted with a traditional graphical interface without avatars. Also, a study by Baylor [8] investigated three different roles of avatars as pedagogical agents and concluded that the one who combines both expertise and motivation features was evaluated by learners to be more engaging and facilitative of learning in comparison with those that played as either motivator or expert agents.

It is believed that the motivating effects of avatar-based interaction are most likely to result in enhancing learners' understanding and their learning outcomes [207, 212]. A study by Moreno et al. [201] compared the learning performance between two groups of students who learned about botanical physiology. The first group received the learning material by speaking full-body animated agent whereas the other group communicated the same material by an on-screen text in the absence of an agent. They found that the group who had interacted with agent-enhanced interface performed better than the other group in applying what they had learned to solve new problems; however the addition of the avatar did not make any difference in learners' performance in retention questions. On the other hand, contradictory results were observed in another study [213] which found that the use of a speaking avatar with gaze and pointing in explaining human cardiovascular system had outperformed the use of speech or text in retention tasks only and no difference among the three metaphors was found in transfer questions.

However, a study on the influences of facial expressions: neutral, happy, sad, scared, surprised, angry and disgusted on students' motivation and their learning outcomes was carried out by Theonas et al. [12, 214] who investigated the use of facially expressive virtual lecturer against another one without any facial features in conducting a set of four virtual lectures of two complexity levels: easy and difficult. The results from their experimental work indicated that the students were more motivated, attentive and excited when they attend the lectures presented by the facially animated lecturer. Furthermore, they noted that student' learning performance was improved with the presence of facial expression particularly in difficult lectures. In a further experiment, Theonas and his colleagues explored the impact of a smiling expression when depicted by the virtual lecturer and found that this expression could attract the students and improve their performance if used appropriately.

2.5 Summary

To summarise, the inclusion of multimodal interaction metaphors has shown to be useful in e-learning interfaces. Although most of the current e-learning interfaces provide a simple and an efficient interaction by means of text and graphics, there is a potential to cause overloading to users' visual channel [1, 2] particularly when the interface becomes crowded with more textual descriptions and graphical illustrations. In this case, users' retention of the delivered learning information will be in question as some important information being communicated could be missed [3]. Also, users are not always satisfied with the computer-based learning [14]. The reviewed literature highlighted the need to address the following issues related to the design of multimodal e-learning interfaces.

- **Lack of face-to-face interaction:** Previous research demonstrated that using speech and non-speech (earcons and auditory icons) sounds could indeed contribute to reduce visual overload by conveying part of the presented information through the auditory channel and consequently allowing a large volume of information to be communicated using different channels. However, the use of these multimodal interaction metaphors in e-learning interfaces could not provide the social interaction that learners are use to have it in traditional learning. Most likely, the users will feel the lack of inter-personal face-to-face interaction when the text with graphics presentation is enhanced only by auditory metaphors.

- **Users' evaluation of facial expressions and body gestures:** The audio-visual inclusion of avatars with facial expressions and body gestures could benefit e-learning interfaces in terms of enhancing users' motivation, engagement and satisfaction as well as their learning performance. Nevertheless, users' views regarding the use of specific facial expressions and body gestures need to be captured in order to obtain an overall feedback for their perception about these metaphors when used within e-learning context. This could contribute to identify which facial expressions and body gestures are more pleasant to learners and consequently to generate more attractive virtual lecturer to them.

- **Investigating different modes of avatars:** The reviewed literature focused mainly on the comparison between the absence and presence of avatars in e-learning interfaces and did not shed the light sufficiently on comparing the different modes through which avatars can be involved as virtual lecturers in these interfaces.

- **Further research on multimodal e-learning:** The e-learning literature highlighted also the need for additional research to integrate multimodal metaphors in e-learning applications where there is a potential for usability and learning enhancement due to the incorporation of these metaphors on their own or when combined with each other.

Therefore, the starting point for this research was initiated by the motivation to investigate whether a combination of recorded speech, earcons and speaking facially expressive avatars could improve the usability and learning performance in the interface of e-learning systems. In addition, a strong encouragement has been established to evaluate three different modes for the inclusion of avatars as virtual lecturers as well as to explore users' opinions in regard to the facial expressions and body gestures demonstrated by these avatars during the presentation of the learning material. The investigation undertaken in this research could provide an additional insight into the usefulness of multimodal interaction metaphors in different computer applications including e-learning.

Chapter 3

Experimental Phase I: An Empirical Investigation into the Use of Multimodal E-Learning Interfaces

3.1 Introduction

This chapter describes an empirical exploration that has been carried out to investigate the usability aspects of e-learning interface that incorporates a combination of typical text with graphic metaphors and multimodal metaphors such as speech sounds (recorded), non-speech sounds (earcons) alongside avatars with facial expressions to be used in the delivery of learning information. The primary question is whether the inclusion of these metaphors can enhance the usability and users' learning performance in e-learning interfaces. The secondary question is related to the contributing role that each of these multimodal metaphors could play in the expected enhancement. An e-learning experimental platform, with two interface versions (a text with graphics and a multimodal) was developed to serve as a basis for this investigation. The e-learning topic was class diagram notation that is often used in the design of software systems. The study involved two groups of users (one group for each interface version) in which the usability performance of the two groups in terms of efficiency, effectiveness, and user satisfaction was compared.

3.2 Aims

This main aim of this experiment was to examine the impact of combining recorded natural speech, earcons and speaking facially expressive avatars on the usability of e-learning interfaces. It is also aimed at evaluating the extent to which the addition of these multimodal metaphors could affect users' learning performance. More specifically, this experiment is aimed at testing the efficiency, effectiveness and user satisfaction of a multimodal e-learning interface as opposed to a typical text with graphics one. An additional aim was to explore these usability factors with different task complexities (i.e. easy, moderate and difficult) and task types (i.e. recall and recognition). In general, this experiment is aimed at investigating the usability aspects and learning performance of e-learning interface that combines recorded speech, earcons and avatar with facial expressions to communicate the learning material. In other words, this study is aimed at exploring if the addition of the aforementioned multimodal metaphors would result in a significant enhancement in terms of efficiency, effectiveness and satisfaction of e-learning interfaces.

3.3 Objectives

In order to fulfil the aims mentioned in the previous section, the following objectives had to be considered:

1. Formulating the experimental hypotheses.
2. Creating two different versions of an experimental e-learning platform to be used in carrying out this empirical investigation. The first version, visual only e-learning platform (VOELP) was based on a text with graphic metaphors in the presentation of learning information about class diagram notation. However, the second one, multimodal e-learning platform (MMELP) offered a multimodal delivery of the same

learning material by the use of natural recorded speech, earcons and facially expressive avatars.

3. Testing the two experimental e-learning platforms independently by two different groups of users.
4. Measuring the efficiency of the tested platforms by the time users spent in completing the required tasks.
5. Measuring the effectiveness of the tested platforms by calculating the percentage of tasks correctly completed by users. This measure was also used for users' learning performance.
6. Measuring users' satisfaction by their ratings for different aspects and learning experience with the tested platforms.

3.4 Hypotheses

It was expected that the usability of e-learning interfaces and the users' learning performance would be influenced by the addition of recorded natural speech, earcons and talking head of a facially expressive avatar as multimodal-based interaction metaphors in an e-learning interface. Accordingly, the following hypotheses have been derived:

H1: The MMELP will be more efficient than the VOELP in terms of the time spent by users to complete the required tasks.

H2: The MMELP will be more efficient than the VOELP as the task complexity increased.

H3: The MMELP will be more efficient than the VOELP for performing both recall and recognition tasks.

H4: The MMELP will be more effective than the VOELP in terms of the percentage of tasks successfully completed by users.

H5: The MMELP will be more effective than the VOELP as the task complexity increased.

H6: Users of the MMELP will outperform VOELP users in terms of the successfully completed recall and recognition tasks.

H7: Users of the MMELP will be more satisfied than the VOELP users.

3.5 Experimental E-learning Platform

An e-learning platform was developed specially to be used in conducting this empirical investigation. The platform provided two different interface versions; a text with graphics interface version, and a multimodal one. Both interface versions of the experimental platform were designed to deliver the same information about class diagram representation of a given problem statement. The presented material, in the form of three common examples, included explanations about classes, associations among classes and the multiplicity of a given class in the diagram. The complexity of these examples was gradually increased, and each of them was given in a separate screen display. Therefore, the graphic metaphor was commonly used in both interface versions to show class diagram representations.

Table 1 shows mapping between the presented learning material and the interaction metaphors incorporated in both versions of the experimental e-learning platform. It can be noticed that the VOELP use text only in communicating all types of the delivered information. On the other hand, the presentation of the learning information in the MMELP was based on a multimodal approach in which different interaction metaphors were used to support the delivery of different learning materials.

Communicated information	VOELP	MMELP		
	Text	Earcons	Recorded speech	Avatars
Classes	√			√
Attributes & behaviour	√			√
Associations	√		√	
Association type	√		√	
Directed	√		√	
Composition	√		√	
Aggregation	√		√	
Inheritance	√		√	
Association label	√		√	
Association name	√		√	
Role name	√		√	
Multiplicity	√	√		

Table 1: Mapping between presented information and interaction metaphors used in both VOELP and MMELP

In sum, the VOELP involved visual only metaphors (text and graphics) whereas the MMELP made use of visual (graphics), auditory (recorded speech and earcons) and audio-visual (speaking and facially expressive avatar) interaction metaphors.

3.5.1 Learning Material

The learning topic demonstrated in this experiment was the class diagram notation usually used in designing software systems [216]. Three different examples of class diagrams were used in the implementation of the experimental platforms. Each example represents a specific statement of problem. The first example illustrates the elevator controller system [217] whereas the second example represents how to compose a document [218]. In the third example, the bank system is illustrated [216].

These examples were gradual in its complexity in terms of number of classes, number and type of the associations and number of multiplicities. Accordingly, the first and second examples were of low and medium complexity respectively. In comparison, the third example was of higher complexity.

		Example 1	Example 2	Example 3
Represented problem		Elevator controller system	Creation of document	Bank system
Classes				
	Number of classes	4	7	10
Associations				
	Number of associations	3	8	12
	Association type			
	Simple	1		4
	Directed	2		
	Composition		2	
	Aggregation		6	
	Inheritance			4
	Recursive			1
	Association label			
	Association name	3		1
	Role name			3
Multiplicity				
	Frequency of multiplicity			
	0..1		1	3
	1..*	2	2	1
	1..2			1
	2..*	1		
	1	3	5	11
	*		5	6
Communicated files				
	Audio-video (Avatars)	4	7	10
	Audio (Recorded speech)	5	5	6
	Audio (Earcons)	6	13	22
	Total	15	25	38

Table 2: Summary of class diagram examples used in the forst experiment

Table 2 briefs the three examples. It can be observed that the third example (which has been supposed to be the most complicated one) has the highest count of classes, associations and multiplicities compared to the other two examples. Therefore, the volume of the presented information was the largest for example 3 followed by examples 2 and 1 respectively. These class diagram examples can be found in Appendix A2.

3.5.2 Text with Graphics E-learning Platform (VOELP)

Figure 2 shows an example screenshot of the text with graphics e-learning interface in which the required information was delivered in a textual approach and could be

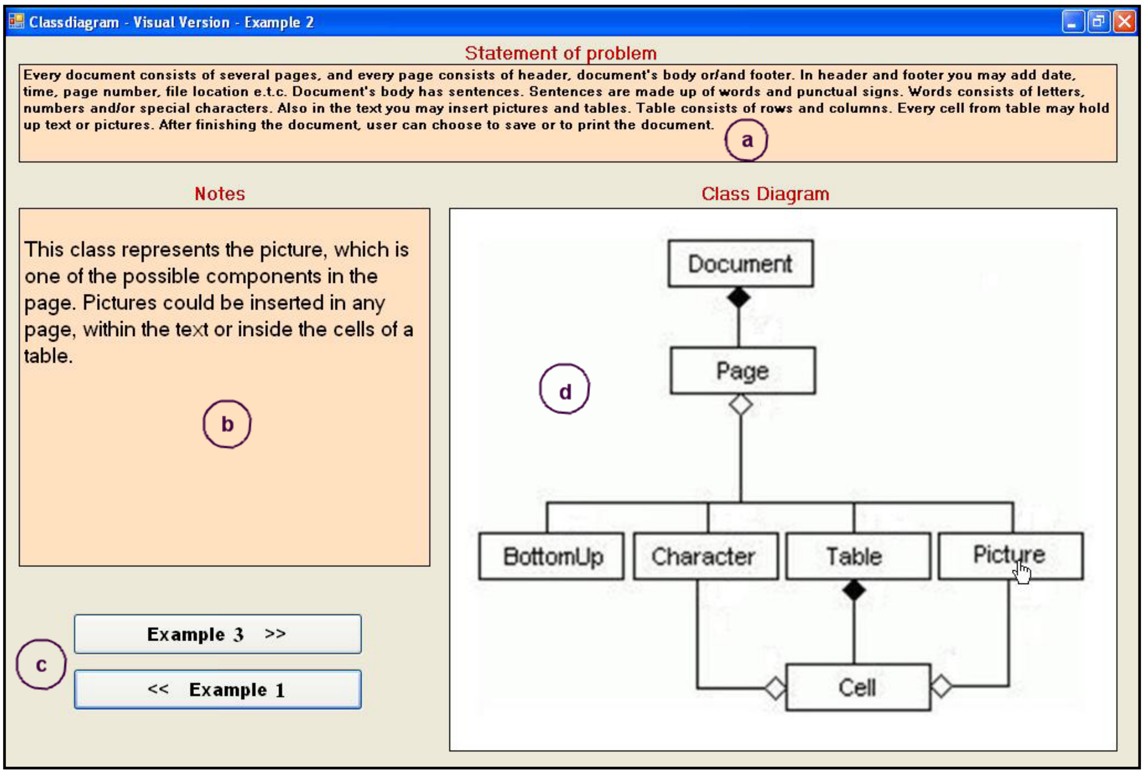


Figure 2: An example screenshot of the text with graphics interface in VOELP

communicated only by the visual channel without making use of any other human senses in the interaction process. This interface was designed to include the following components: (a) text box in the top part of the interface to present the statement of problem related to the given example, (b) notes text box located in the left-hand side of the screen, (c) two command buttons in the bottom right-hand side to enable transition from one example to another, and (d) a still image of class diagram illustration of the given problem, placed in the right-hand side of the interface. When the mouse cursor is placed over a given notation, a textual explanation about that notation is displayed in the notes text box. Therefore, users of this version need to read this explanation as well as to have a look at the diagram in order to understand the communicated information.

3.5.3 Multi-Modal E-learning Platform (MMELP)

Figure 3 shows an example screenshot of the multimodal e-learning interface. Creation

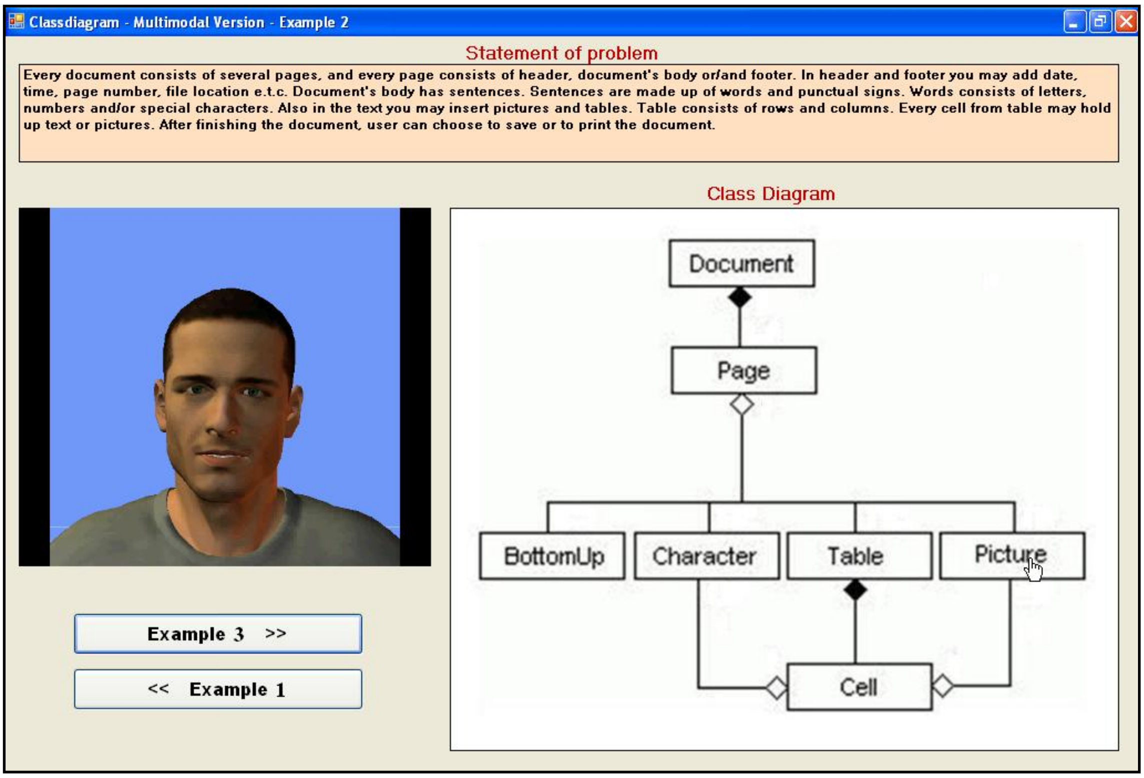


Figure 3: An example screenshot of the multimodal interface in MMELP

of the involved multimodal metaphors was prominently based on the connection between these interaction metaphors and the information being delivered. This connection also considered the previous studies that demonstrated the usefulness of multimodal interaction (refer to Section 2.3). Moreover, guidelines for the design of multimodal information presentation [6] and multimodal user interface [195] were followed. For example, the multimodal output was used to widen the bandwidth of information transfer [5, 6]. Also, graphical displays, speech and non-speech sounds were combined to obtain an effective presentation [195] where speech can be used to transmit short messages [219] and non-speech to supplement other interaction modalities [198].

Apart from notes text box, the same components used in the text with graphics e-learning interface were replicated in the multimodal one. The notes text box was removed and replaced with a combination of recorded speech, earcons, and speaking avatar with facial expressions. When placing the mouse cursor on specific notation in the diagram, related information is introduced. Facially expressive life-like avatar was included to speak the explanations about classes with prosody. Multiplicity of a given class was communicated by earcons whereas the associations among classes were explained by the recorded speech sounds. This way, a different approach for information presentation was offered wherein the user can keep looking at the displayed class diagram whilst listening to the delivered auditory messages.

3.5.3.1 Implementation of Earcons

Different technologies were utilised for the production of earcons used in this study such as music synthesis [220] and sound recording software [221]. Previous research [9, 13, 97, 106, 119, 129, and 139] demonstrated that earcons can be used successfully to communicate numerical information. Therefore, earcons were employed in the multimodal interface (MMELP) to convey the multiplicity. However, the use of auditory icons requires the existence of natural mapping between these sounds and the communicated information and this was not available in the communicated material. Therefore, auditory icons were not considered in this experiment. The used earcons were designed based on the suggested guidelines [118, 134]. Musical notes (starting at middle C in the chromatic scale) were used to create six earcons each of which has been utilized to communicate one of the six different types of multiplicity found in the three class diagram examples. Table 3 shows the design structure of these earcons. It can be seen that the first four earcons were composed of two parts separated by a short pause (0.6 second) in between so that the user can distinguish where the first part finishes and the second starts [118]. Each of these earcons was dedicated to communicate one of the multiplicities: zero or 1 (0..1), one or more (1..*), two or more (2..*), and one or two (1..2).

		Multiplicity					
		0..1	1..*	1..2	2..*	1	*
Duration (second)		1.8	2.2	1.9	2.4	1	1.2
Part 1	Timbre	Seashore	Piano	Piano	Piano	Piano	Piano
	Rhythm	Single note	Single note	Single note	2 serial notes	Single note	4 rising pitch notes
Part 2	Timbre	Piano	Piano	Piano	Piano	-	-
	Rhythm	Single note	4 rising pitch notes	2 serial notes	4 rising pitch notes	-	-

Table 3: Structure of earcons used to communicate multiplicity in MMELP

The remaining two earcons had only one part and had been used to represent the multiplicities: one (1) and many (*). These single-meaning auditory messages were kept as short and simple as possible to enable recognising its meaning [134]. In order to create the required earcons, there was a need to illustrate the values 0, 1, 2, and * musically. For this purpose, different numbers of piano notes were used as follows: one musical note to communicate 1, two rising notes to communicate 2 and four rising pitch notes to communicate many (*). However, to distinguish zero and to represent it in the multiplicity 0..1, only one note of seashore sound was used. Using a different number of notes could be effectively employed to differentiate rhythms of the earcons [118].

3.5.3.2 Implementation of recorded speech

The recorded speech sounds were used in this study to explain the associations among classes. In comparison with synthesised speech, natural human speech is more comprehensible by users [115]. The delivered information explained different aspects of the relationships among classes such as the type of association (i.e. directed, composition, aggregation, inheritance, recursive) and association label (i.e. association name and role name). These sounds were recorded and manipulated using digital audio editor software called Audacity [222].

Neutral Happy Sad Smile

Figure 4: Facial expressions used in MMELP

3.5.3.3 Implementation of avatars

Figure 4 shows the facial expressions used during the creation of avatar files. These expressions were selected as examples of the expressions usually used in everyday life to express human feelings and emotions [171]. In order to create avatar presentations, the following tools were utilised:

- Audacity [222], was used to record and process the natural speech sound and produce it in WAV (Waveform) data format.
- Mimic [223], was used to automate the lip-syncing of a 3D figure for the inserted WAV sound and text files. This software automatically adds the proper mouth movements, eye blinks and head nodes to human-like figures. The output file was then exported to Poser.
- Poser [224] has been used for the generation of facially expressive humanoid 3D figures based on the imported Mimic file. In addition to built-in collection, Poser enables the creation of new figures with customized facial expressions and body gestures. The output file was produced in an Audio Video Interleave (AVI) format.

More details related to the development of the MMELP can be found in Appendix D1.

3.6 Experimental Design

In order to explore the effect of multimodal metaphors and to find out which interface would be better in terms of efficiency, effectiveness, and user satisfaction for the e-learning process, both interface versions of the experimental platform were empirically evaluated by two independent groups of users. One group tested the text with graphics interface to serve as a control and the other tested the multimodal interface in order to serve as an experimental group. This design methodology, *between-subjects* testing, involves the assignment of different users to test different experimental conditions and therefore guarantees controlling the learning effect [225]. In total, 30 users participated individually in the experiment and equally allocated to both groups.

3.6.1 Procedure

In order to keep the consistency throughout the experiment, the same procedure with both groups of users was followed. The experiment was started by reading the introduction to the questionnaire and answering the pre-experimental questions for users' profiling in terms of personal information such as age, gender and education level. Also, at this stage of the experiment, users were required to declare their previous experience in computers, Internet and e-learning applications and to state as well their prior knowledge in object orientation, class diagram notation and avatars. Then, two tutorials were presented; the first tutorial demonstrated the class diagram notation for five minutes and was shown to each user in both groups. The second tutorial had two versions, one for each group to provide an introduction to e-learning interface version that the user was to use. Both these tutorials were run for two minutes. Thereafter, users were instructed to start performing 6 common tasks.

Users	Pre-experimental questionnaire	Tutorial for class diagram notation	Tutorial for the applied interface	Presented example	Evaluation questions	Presented example	Evaluation questions	Presented example	Evaluation questions	Satisfaction questionnaire
1, 7, 13				1		2		3		
2, 8, 14				1		3		2		
3, 9, 15				2		1		3		
4, 10				2		3		1		
5, 11				3		1		2		
6, 12				3		2		1		

Table 4: Procedure followed in conducting the first experiment

After completing all these tasks, users were asked to give their satisfaction ratings about the different aspects of the tested interface version by answering the post-experimental part of the questionnaire. In order to control the learning effect, class diagram examples were presented to users in both groups in a random rotation manner as shown in Table 4. The questionnaire used in conducting this experiment can be seen in Appendix A1.

3.6.2 Tasks

Both groups performed six common tasks. These tasks were evenly associated with the presented class diagram examples and covered all types of presented information such as class attributes and operations, associations among classes, and multiplicities. Previous experimental studies demonstrated that using multimodal metaphors could be affected by tasks type [205, 226] and task complexity [11, 93, 227]. Therefore, the tasks in this experiment were designed to increase in difficulty and they were equally divided into easy, moderate and difficult tasks. Each task comprised a set of requirements each of which asked the user to place the mouse cursor over a specific notation in the displayed class diagram, and to communicate the delivered information related to that notation either visually (in VOELP) or in a multimodal approach (in MMELP).

Task code	Task complexity	Example	NOR	Distribution of requirement		
				Classes	Associations	Multiplicities
ET1	Easy	Example 1	3	1	1	1
ET2			3	1	1	1
MT1	Moderate	Example 2	4	1	1	2
MT2			4	1	1	2
DT1	Difficult	Example 3	6	1	2	3
DT2			5	1	1	3

Table 5: Summary of the required tasks in the second experiment

The complexity of the task depended on two main factors; the number of requirements (NOR) and the nature of delivered information due to the implementation of each requirement. The more complex the task is the more requirements are postulated and thus more information is presented. As a result, difficult tasks involved communicating larger volumes of information as opposed to moderate and easy tasks. Table 5 summarizes the required six common tasks.

Upon completion of each task, each user was requested to answer a memory recall (RL) and recognition (RN) questions. The aim of these questions was to evaluate the learning gained by users due to the information presented by the applied interface. In order to answer recall question correctly, the user had to retrieve part of the presented information from his/her memory. However, the recognition one offered a set of 2 to 4 options and user had to recognize the correct answer among it. In total, each user answered twelve questions consisting of 4 easy, 4 moderate and 4 difficult. Based on question type, these questions were categorised into 6 recall and 6 recognition questions. Refer to the second part of the questionnaire in Appendix A1 for more details about the requirements of the tasks and its relevant evaluation questions. Table 6 shows the multimodal metaphors used to communicate the key information needed by the users of the MMELP to answer the questions successfully. It can be noticed that 42% of the questions were related to information presented by the avatar, all of which were recall (2 easy, 1 moderate and 1 difficult).

	Questions											
	Easy tasks				Moderate tasks				Difficult tasks			
	ET1		ET2		MT1		MT2		DT1		DT2	
	RL	RN	RL	RN	RL	RN	RL	RN	RL	RN	RL	RN
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
Recorded Speech		√		√		√		√				√
Earcons		√		√	√					√		√
Avatar	√		√				√		√		√	

Table 6: Multimodal metaphors used to communicate the key information needed by users in the multimodal interface group to answer the required questions correctly

Another 25% of the questions were based on the information presented by both recorded speech and earcons; 2 easy recall and a single difficult recognition question. The remaining questions, however, asked the users about information communicated by recorded speech and earcons 17% each.

3.6.3 Variables

The variables considered in the experimental design can be classified into three types which are: independent variables, dependent variables and controlled variables [228].

3.6.3.1 Independent Variables

Independent variables represent the factors manipulated in the experiment and assumed to be the cause of the results [229]. These variables include:

IV 1: Presentation mode: the experimental e-learning platform offered two different modes for the presentation of the learning material; text with graphic in VOELP and multimodal in MMELP.

IV 2: Question complexity: this study investigated the usability and learning performance related to three levels of complexity; easy, moderate and difficult.

Variable code	Variable	Levels	Condition 1	Condition 2	Condition 3
IV 1	Presentation mode	2	VOELP	MMELP	
IV 2	Question complexity	3	Easy	Moderate	Difficult
IV 3	Question type	2	Recall	Recognition	

Table 7: Independent variables considered in the first experiment

Variable code	Variable	Measure
DV 1	Question answering time	Efficiency
DV 2	Correctness of answers	Effectiveness and user's learning performance
DV 3	User Satisfaction	Satisfaction

Table 8: Dependent variables considered in the first experiment

IV 3: Question type: this study also investigated the effect of two types of evaluation questions; recall and recognition on the usability of the tested e-learning interfaces as well as on users' learning performance.

These variables are summarised in Table 7.

3.6.3.2 Dependent Variables

These are the variables being measured as a result of manipulating the independent variables [229]. The independent variables regarded in this study are briefed in Table 8 and include the following:

DV 1: Question answering time: this variable was measured by the time taken by users to answer the required questions.

DV 2: Correctness of answers: measured by calculating the percentage of correctly answered questions. In recall questions, partial or total correct answers were considered whilst in the recognition questions, the answer had to be totally correct.

DV 3: User satisfaction: measured by observing users' responses to the satisfaction questionnaire on a 6-point Likert scale. SUS scoring method [230] was used to calculate the satisfaction of each user in regard to overall attitude as well as learning experience with the tested e-learning interface.

3.6.3.3 Controlled variables

These represent the external variables associated with the procedure of the experiment and could affect the obtained results. The controlled variables (known also as confounding variables) should be kept consistent throughout the experiment to avoid their influence on the dependent variables and so ascertain that the independent variables are the only cause of the experimental results [231]. The controlled variables in this experiment were:

CV 1: Required tasks: the same tasks were required from all users.

CV 2: Learning material: the information presented about class diagram examples was similar in both interface versions.

CV 3: Awareness of questions: none of the users were aware of the required questions.

CV 4: Procedure consistency: the experiment has been conducted by the same experimenter on an individual basis with each user. Also, the same procedure was followed during the execution of the experiment including measurement tools and used equipments.

CV 5: Familiarity with the interface: all the users were first-time users of the tested interface with the same level of training.

3.6.4 Users Sampling

A total of 30 volunteer users were involved in this study and were first-time users of the experimental platform. They were equally and randomly assigned (N = 15) to the experimental conditions; text with graphics e-learning interface for the control group, and multimodal interface for the experimental group. Participation of this number of users in both groups could be sufficient to provide the usability evaluation [75]. Also, too many users have not been involved because needed to investigate and carry out an initial test to obtain an overall impression and understanding about the procedure and test criteria and the determination of feasibility as well. The selection of the participants was based on their prior knowledge in the learning topic namely class diagram notation. In this regard, the majority of the users in both groups had no experience indicating that they will rely only on the communicated learning information to answer the required questions.

3.7 Data Collection

The data collection process was based on the experimental observations and questionnaires. Upon completion of each task, each user was required to answer two questions. The time spent to answer each question was observed to help in measuring the efficiency. However, in order to collect the data related to effectiveness, the correctness of user’s answers was checked and the total number of successfully answered questions was counted for each user. The pre-experimental part of the questionnaire was dedicated to gather personal data about users such as age, gender and education. It also helped to obtain data related to users’ prior experience in computers, Internet, object orientation, class diagram notation, e-learning and avatars and facial expressions. Finally, the post-experimental part of the questionnaire was aimed at assessing the users’ satisfaction with the tested e-learning platform. Users’ responses to this questionnaire were used to calculate the satisfaction score for each user in both the control and the experimental groups.

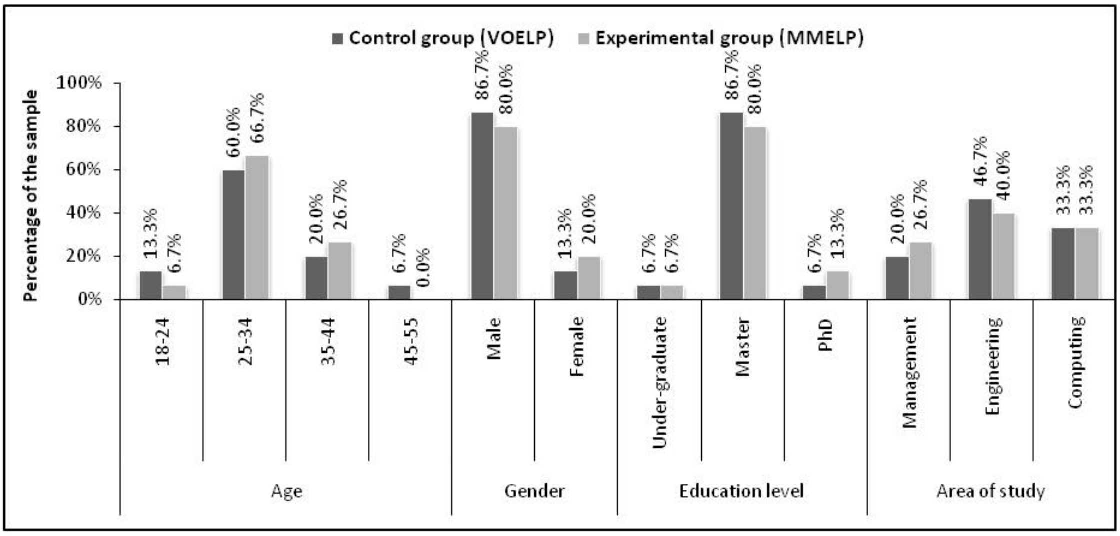


Figure 5: Users' profile in terms of age, gender, education level and area of study in both control and experimental groups

3.8 User Profiling

The data in relation to users' personal and educational information as well as their previous knowledge and experience were collected and analysed on the basis of their responses to the pre-experimental questions (refer to Appendix A3). Figure 5 shows that the age range in the control group was about 60% within 25 – 34, 20% 35 – 44, 13% 18 – 24 and 7% (one user) 45 – 55 years old. In the experimental group, the ages were 67% within 25 – 34, 27% 35 – 44 and one user in the range 18 – 24 years old. The majority of the participants were male by 87% in the control group and 80% in the experimental one. The education level was found to be predominantly postgraduates by 93% in each group while the others were undergraduates. Additionally, the areas of study for users in both groups were computing, engineering and management.

Also, as can be noted in Figure 6, users in both groups were experts on computers and Internet. Eighty percent of the control group use computers for more than ten hours a week compared to 73% in the experimental group. With respect to the weekly use of Internet, the percentage was about 67% in the control group and 80% in the

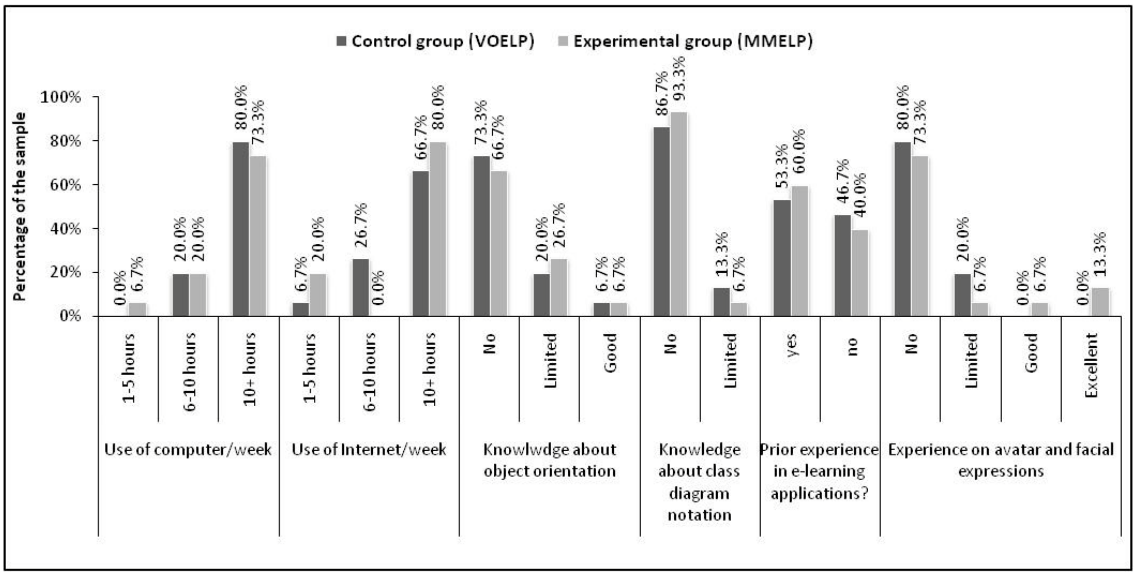


Figure 6: Prior experience for users in both control and experimental groups

experimental group for ten or more hours. What is more important is that most of them in both groups had no prior knowledge about class diagram notation (over 87%) and object-orientation (over 67%). In addition, more than 73% of the sample users were inexperienced in avatars and facial expressions. Moreover, most of the remaining users had limited background in these topics. Finally, Figure 6 revealed that the experimental group was slightly more experienced in e-learning applications in comparison with the control group. Users' profiles shown in Figure 5 and Figure 6 demonstrate that both groups, to a large extent, were equivalent in terms of users' individual characteristics and prior experience. Therefore, any differences between the two experimental conditions in the obtained results could be attributed to the treatment carried out on the participants.

3.9 Results and Analysis

The results of both groups were analysed in terms of efficiency (time users needed to answer the required questions), effectiveness (percentage of correctly answered questions), and user satisfaction (based on a rating scale). For the statistical analysis, the nonparametric Kolmogorov-Smirnov test [232] has been used to test the normal distribution of the obtained results in terms of answering time, questions correctly answered and the satisfaction score. If the data was found to be normally distributed, then the independent t-test was used to evaluate the significance of the difference between the two groups in regard to each of these parameters. This statistical test is applicable when two different experimental conditions are tested independently by two groups of users [231]. Otherwise, Mann-Whitney test was used as a non-parametric equivalent of the independent t-test [233]. Also, Chi-square test was used for statistical analysis of categorical data [234]. These statistical analyses were conducted at $\alpha = .05$ and significant difference was detected if p-value was found to be less than .05.

3.9.1 Efficiency

The time spent to answer the required questions was used as a measure of efficiency. This measure was considered for all questions in total and according to the question type (recall and recognition), question complexity as well as for each question and each user in both groups; control and experimental. Figure 7 shows the mean values of the time taken by the users in both groups to answer all the required questions (A), grouped by the question complexity (B) and question type (C). It can be seen that the consumed answering time was lower in the experimental group for all questions as well as for each complexity level and question type. The raw data for question answering time can be found in Appendix A4.

3.9.1.1 All Questions

Each user had to answer 12 questions in total. As shown in Figure 7A, the mean time consumed to answer these questions in the MMELP condition was lower than that in the VOELP condition. The total time consumed by users of the VOELP in the control group

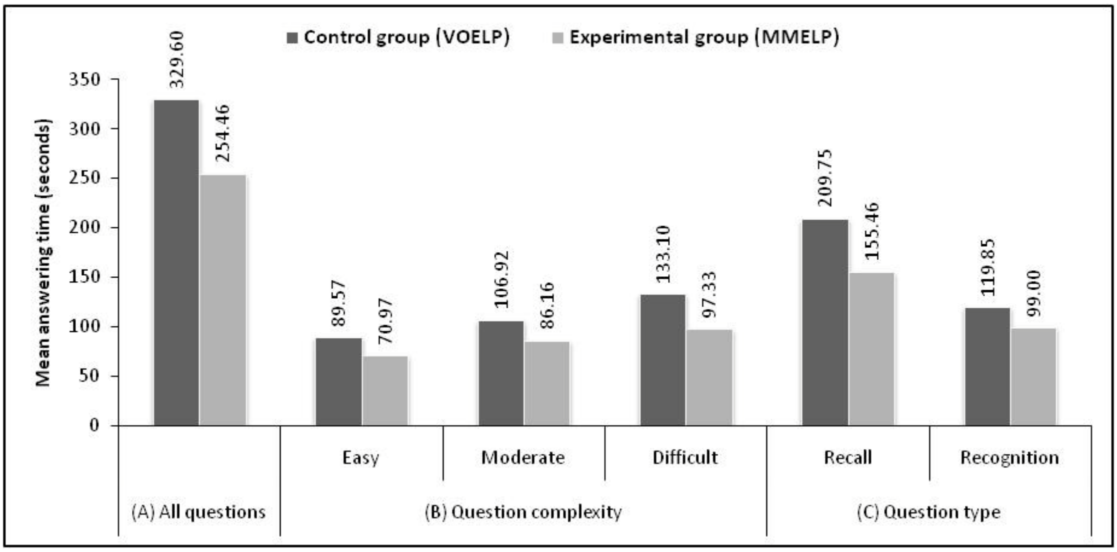


Figure 7: Mean values of time taken by users in both groups to answer all questions (A) and grouped by question complexity (B) and question type (C)

was observed 4943.95 seconds averaging 329.60 seconds per user. In comparison, users of the MMELP in the experimental group spent a total of 3816.78 seconds answering time with 254.46 seconds per user on average. In other words, users of the MMELP were 1127.07 seconds faster than their counterparts who used the VOELP. The t-test calculations showed that the difference in answering time between both groups was significant ($t(28) = 2.69$, $CV = 1.70$, $p < .05$). Experimental observations revealed that users in the control group regularly divided their visual attention between the explanations provided in the notes text box and class diagram representations in order to understand the presented information and in some cases a visual overload might occurred. However, the users in the experimental group maintained their visual attention to the class diagram representations while they were listening to the auditory messages obtaining better concentration on the delivered material. In summary, the users in the experimental group were significantly aided by the addition of the multimodal metaphors in the MMELP which enabled them to spend less time than the users of the VOELP in responding to the required questions. Therefore, it can be said that using speaking facially expressive avatar, earcons and recorded speech was more efficient than using only text with graphic metaphors in presenting the clarifications related to the learning material used in this experiment.

3.9.1.2 Question Complexity

Figure 7B shows the answering time grouped by the complexity of questions. These questions were designed to increase in difficulty and were equally divided into 4 easy, 4 moderate and 4 difficult. On overall, it can be noticed that the answering time in the experimental group was lower for all complexity levels. Also, it can be noticed that the variance in answering time between the two groups increased with an increasing level of question complexity. In easy questions, the mean answering time in MMELP was noted 18.60 seconds less than that in VOELP. The variance between both conditions, however, was slightly larger (20.77 seconds) in responding to moderate questions. In difficult questions, the variance considerably increased to 35.77 seconds in favour of the MMELP. The statistical tests revealed that the users of the MMELP needed significantly less time than the users of the VOELP to answer each of the easy ($t(28) = 2.22$, $CV = 1.70$, $p < .05$), moderate ($U = 62$, $CV = 72$, $p < .05$) and difficult ($t(21) = 2.58$, $CV = 1.72$, $p < .05$) questions. In summary, these results demonstrated that the use of multimodal metaphors had gradually contributed in reducing the answering time consumed by the users when the required evalution questions became more difficult.

3.9.1.3 Question Type

Figure 7C shows the answering time grouped by the question type. The questions were designed to be of two different types; recall and recognition with 6 questions each. On overall, the answering time in the experimental group was lower in both types of questions as opposed to the control group. It can be noticed also that answering the recall questions had taken longer time in comparison with the recognition questions.

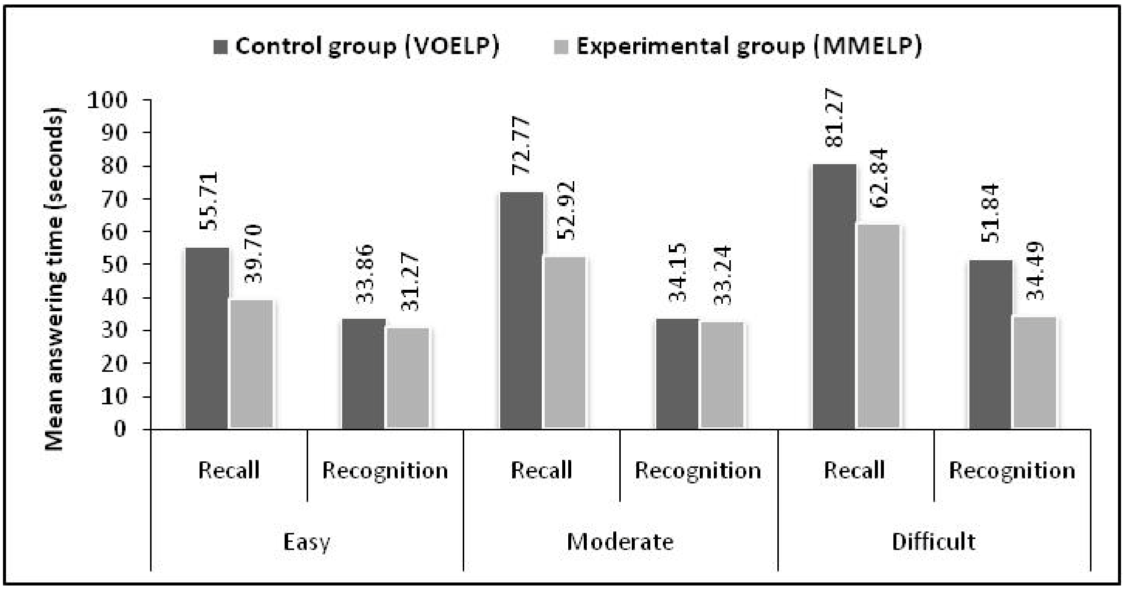


Figure 8: Mean values of time taken by users in both groups to answer the recall and the recognition questions grouped by question complexity

Nevertheless, the variation between the two conditions in answering time was observed larger in the recall questions compared to recognition ones. In responding to the recall questions, users of the MMELP in the experimental group spent 54.28 seconds (on average) less than the users of the VOELP in the control group. However, the variation between the two groups was substantially dropped to 20.86 seconds with respect to answering recognition questions. According to t-test results, The difference between the two groups in answering time was found statistically significant for the recall questions ($t(21) = 3.08$, $CV = 1.72$, $p < .05$) whereas no significant difference has been reached for the recognition questions ($t(28) = 1.49$, $CV = 1.70$, $p > .05$).

A more detailed analysis, shown in Figure 8 revealed that users in the experimental group spent significantly lower time than those in the control group for answering each of the easy recall ($t(28) = 2.56$, $CV = 1.70$, $p < .05$), moderate recall ($t(28) = 3.14$, $CV = 1.70$, $p < .05$) and difficult recall ($t(22) = 1.81$, $CV = 1.72$, $p < .05$) questions. Regarding recognition questions, the statistical results were not significant for the easy ($t(28) = 0.57$, $CV = 1.70$, $p > .05$) and the moderate questions ($U = 89$, $CV = 72$, $p > .05$), but

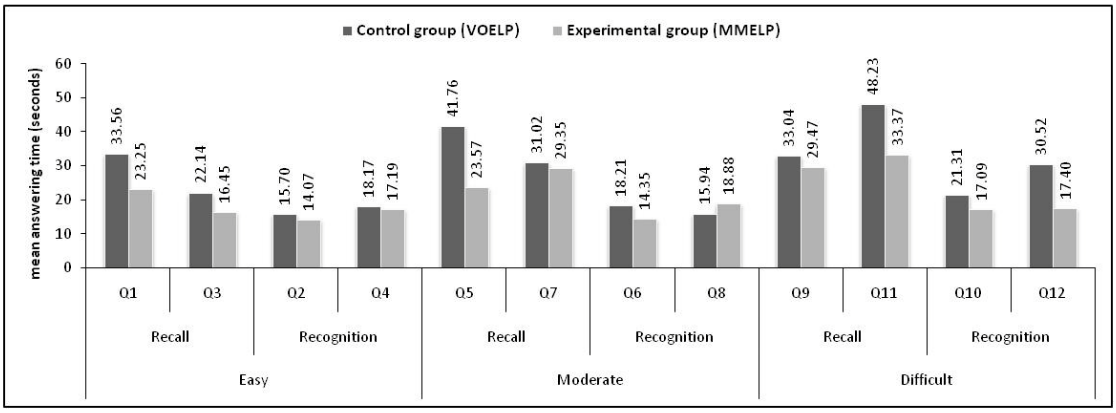


Figure 9: Mean values of time taken by the users in both groups to answer each question

significant difference in favour of the experimental group was observed only for the difficult recognition questions (t(28) = 2.97, CV = 1.70, p < .05) without affecting the overall results. On the whole, these experimental findings indicated that the addition of the multimodal metaphors to the MMELP helped users much more in the recall activities which are more difficult than the recognition ones. In answering recall questions, the users may have taken more time trying to retrieve the required information which is not the case in recognition tasks where users needed only to select the answer among the given options.

3.9.1.4 Each Question

Figure 9 shows the time consumed to answer each question in each group. Apart from the 8^{th} question which needed longer answering time using the MMELP, the experimental group needed shorter time than the control group to answer all the questions. Additionally, the mean time taken to answer a question was 21.20 seconds in the experimental group compared to 27.47 seconds in the control group. It could be noticed that the difference between the two groups in answering times was varied across the twelve questions. These variances could be attributed to the differences in complexity and type of the required questions. Table 9 shows the results of statistical analysis for each question.

Question complexity	Question type	Q#	Statistical test	Statistical result	Significant
Easy	Recall	Q1	t-test	t(28) = 1.88, CV = 1.70, p < .05	Yes
		Q3	Mann-Whitney	U = 66, CV = 72, p < .05	Yes
	Recognition	Q2	t-test	t(28) = 0.67, CV = 1.70, p > .05	No
		Q4	t-test	t(28) = 0.33, CV = 1.70, p > .05	No
Moderate	Recall	Q5	t-test	t(24) = 4.31, CV = 1.71, p < .05	Yes
		Q7	t-test	t(28) = 0.43, CV = 1.70, p > .05	No
	Recognition	Q6	t-test	t(28) = 1.95, CV = 1.70, p < .05	Yes
		Q8	Mann-Whitney	U = 111, CV = 72, p > .05	No
Difficult	Recall	Q9	Mann-Whitney	U = 107, CV = 72, p > .05	No
		Q11	t-test	t(28) = 2.15, CV = 1.70, p < .05	Yes
	Recognition	Q10	t-test	t(28) = 1.18, CV = 1.70, p > .05	No
		Q12	t-test	t(28) = 3.95, CV = 1.70, p < .05	Yes

Table 9: Statistical analysis for answering time taken by the users to answer each question

This table demonstrates that the experimental group spent significantly less answering time to 50% of the questions (Q1, Q3, Q5, Q6, Q11 and Q12). The key information needed to answer these questions was communicated in MMELP using facially expressive speaking avatar (Q1, Q3 and Q11), earcons (Q5), recorded speech (Q6) and both the recorded speech and earcons in Q12. In the remaining questions, no significant differences were obtained. Nevertheless, the obtained results could not be considered as conclusive to clarify the role that each of speech, earcons and avatar played in shortening the answering time when used in the multimodal interface. The reason behind this could be attributed to the design of the required questions. These questions were not designed to explore the individual role of these multimodal metaphors. In few words, the multimodal metaphors applied in the MMELP assisted in reducing the answering time for 92% (11 out of 12) of the required questions.

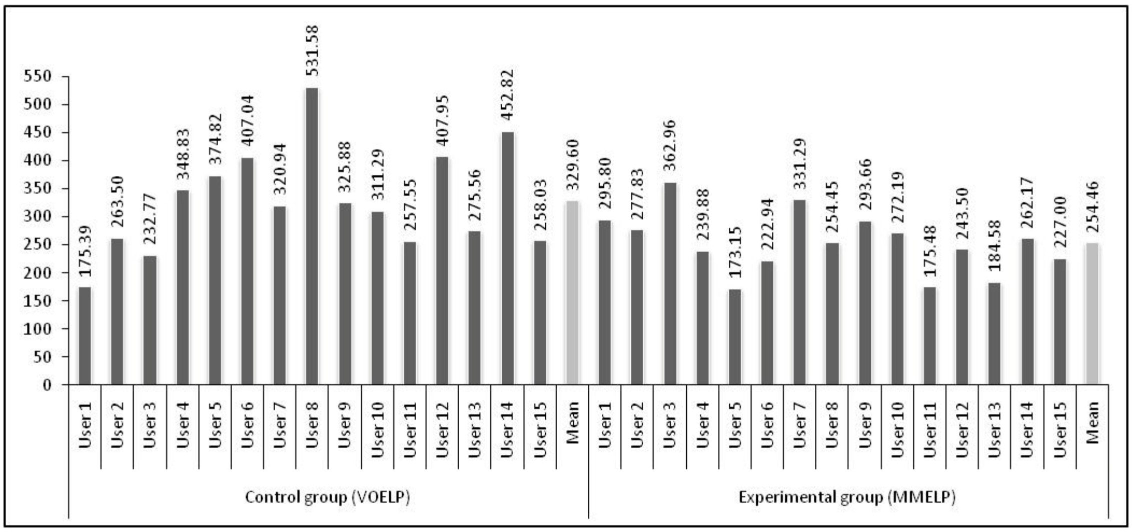


Figure 10: Total time taken by each user in both groups to answer all questions

3.9.1.5 Each user

Figure 10 shows the total time spent by each user in both groups to answer all the 12 questions where a larger consumption of time can be noted for the users of the VOELP compared to the users of the MMELP. The minimum and maximum answering times consumed in the control group were 175.39 seconds (User 1) and 531.58 seconds (User 8) respectively. In the experimental group, the minimum time observed was slightly lower (173.15 seconds by User 5) whereas the maximum time (362.96 seconds by User 3) was 168.62 seconds less than that in the control group. On average, the users of the MMELP were 75.14 seconds faster than their counterparts who used the VOELP.

3.9.2 Effectiveness

The number of correctly answered questions was used as a measure of effectiveness. This measure was considered for all the questions in total, according to the question type (recall and recognition) and question complexity (easy, moderate and difficult) as well as for each question and each user in both control and experimental groups. Figure 11 shows the percentage of correct answers for all questions (A) and according to question complexity (B) and question type (C) in VOELP and MMELP conditions. It

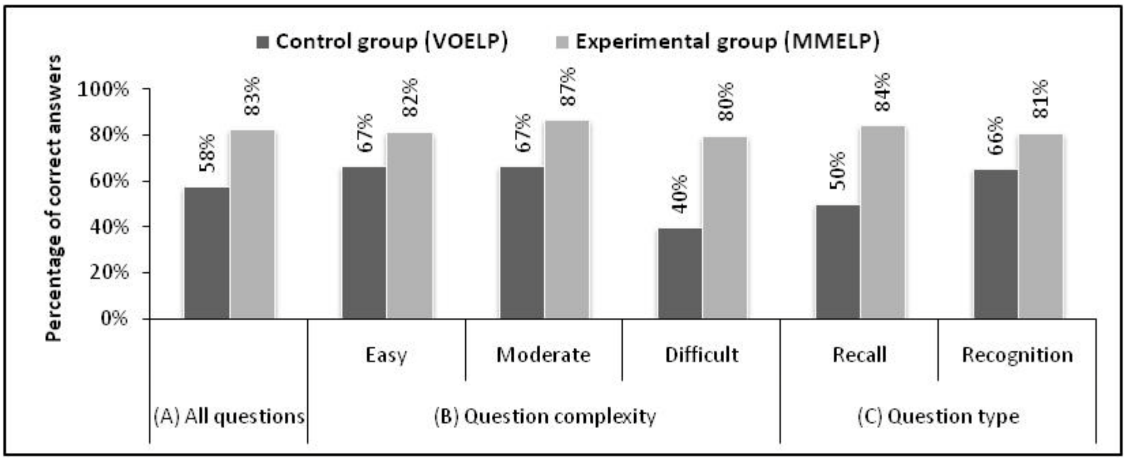


Figure 11: Percentage of correct answers achieved by users in both groups for all the questions (A), grouped by question complexity (B) and question type (C)

can be noticed that users of the MMELP outperformed VOELP users in terms of correctness of answers to all questions as well as to each complexity level and question type. The raw data of the correctness of users' answers can be found in Appendix A5.

3.9.2.1 All Questions

The total number of questions in each group was 180 (15 user * 12 questions per user). In Figure 11A, it can be seen that the users of the MMELP performed better than the users of the VOELP with regard to the correctness rate of all answers. The percentage of correctly answered questions achieved in the MMELP condition (83%) was one-quarter (25%) greater than that attained in the VOELP condition (58%). The total number of correct answers in the experimental group was 149 compared to 104 in the control group. Also, the mean value of correct answers per user calculated for the experimental group (9.93) was 3 correct answers grater than that for the control group (6.93). The t-test results revealed that the difference in correctly answered questions between MMELP and VOELP was significant ($t(28) = 4.75$, $CV = 1.70$, $p < .05$). The incorporation of more than one communication metaphor of different nature in the MMELP helped users in the experimental group to discriminate the different types of information which has been delivered by each of the recorded speech, earcons and speaking avatar enabling them to keep this information for a longer time. As a result, they outperformed the users of the VOELP who received the learning information via visual channel only. In summary, the multimodal interaction metaphors used in the MMELP was more effective in communicating the learning material and considerably assisted the users in the experimental group to achieve a higher effectiveness rate as opposed to the control group users.

3.9.2.2 Question Complexity

Figure 11B shows the percentage of correctly answered easy, moderate and difficult questions in both groups. The total number in each complexity level was 60 questions. It can be noted that the experimental group outperformed the control group in all levels of complexity particularly in answering the difficult questions. What is more, the difference in users' performance increased in favour of the experimental group as the complexity increased. In easy questions, the users of the MMELP scored 15% more correct answers than those of the VOELP. However, the difference was observed larger (20%) with respect to moderate questions and the largest difference (40%) was noted in users' responses to difficult questions where the users in the experimental group achieved double to what has been achieved by the users in the control group. Using the MMELP, the users in the experimental group correctly answered 82%, 87% and 80% of easy, moderate and difficult questions respectively. On the other hand, the users of the VOELP in the control group successfully responded to 67% of easy questions, 67% of moderate questions, and 40% of difficult questions. The results of Man-Whitney test showed that the difference in correct answers between MMELP and VOELP did not reach a statistical significance in easy questions ($U = 77.5$, $CV = 72$, $p > .05$) while it was found significant in moderate ($U = 54$, $CV = 72$, $p < .05$) and difficult questions ($U = 28$, $CV = 72$, $p < .05$). In brief, it can be said that both groups of users accomplished equivalent levels of accuracy of their answers to easy questions. However, the contribution of multimodal metaphors in users' performance was more obvious in their responses to higher complexity questions.

3.9.2.3 Question Type

Figure 11C shows the percentage of correct answers to recall and recognition questions in both control and experimental groups. The total number of questions in each type was 90. It can be noted that users of the MMELP performed better than those of the VOELP in both recall and recognition questions but the difference between the two groups was smaller in the latter type. In recall questions, the percentage of correctly answered questions in the experimental group was 34% higher than that in the control group. However, the percentage of correctly answered recognition questions in the experimental group was 16% higher than that in the control group. Using the MMELP, users in the experimental group gained a correctness rate of 84% and 81% in recall and recognition questions respectively. On the other hand, the users of the VOELP in the control group acquired 50% correctness rate in answering recall questions and 66% in answering recognition questions. The results of Man-Whitney test showed a significant difference in correct answers between MMELP and VOELP conditions for both types of questions; recall ($U = 23$, $CV = 72$, $p < .05$) and recognition ($U = 61.5$, $CV = 72$, $p < .05$). A further analysis (see Figure 12) indicated that the experimental group significantly outperformed the control group in answering both moderate recall ($U = 50.5$, $CV = 72$, $p < .05$) and difficult recall ($U = 42.5$, $CV = 72$, $p < .05$) questions. However the difference in answering easy recall questions was not significant ($U = 75.50$, $CV = 72$, $p > .05$). On the other hand, the experimental group significantly outperformed the control group in only difficult recognition questions ($U = 57$, $CV = 72$, $p < .05$) while the difference in correct answers was not sufficient to reach

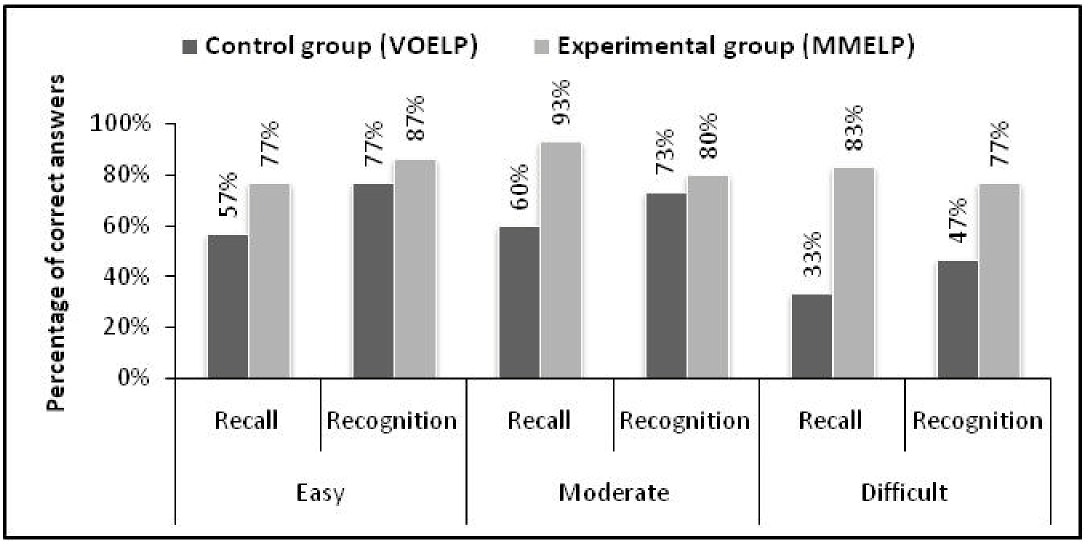


Figure 12: Percentage of correct answers achieved by the users in both groups for recall and recognition questions grouped by complexity level

significant levels in both easy (U = 97, CV = 72, p > .05) and moderate recognition (U = 90.5, CV = 72, p > .05) questions.

In summary, the contribution of multimodal metaphors was more apparent in users' answers to recall activities compared to that in recognition ones. Nevertheless, the experimental group performed significantly better than the control group in overall results for both types of questions.

3.9.2.4 Each Question

The percentage of users' correct answers to each question in each group is shown in Figure 13. It can be seen that the users of the MMELP performed better than VOELP users in 92% (11 out of 12) of the required questions. Only in the 8th question a higher performance was observed for VOELP users. Chi-square results shown in Table 10 demonstrated a significant difference between the control and experimental groups in terms of correct answers to 25% of questions (Q5, Q9 and Q10). The key information required to answer these questions in MMELP (refer to Table 6) was delivered by facially expressive avatar (Q9) and earcons (Q5 and Q10). However, no significant

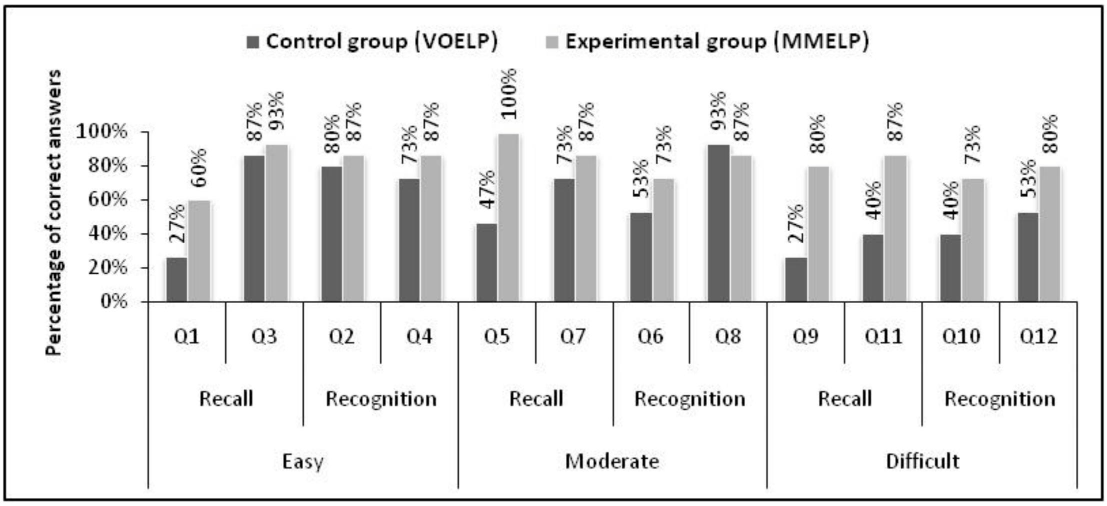


Figure 13: Percentage of correct answers achieved by the users of VOELP and MMELP for each question

differences were obtained in the remaining questions. On the whole, it can be said that the multimodal metaphors used in the MMELP contributed to users' performance in most of the required questions but, to a large extent, the design of the required questions did not permit clear impression about the role played by each of these metaphors in assisting MMELP users.

Question complexity	Question type	Q#	X^2 value	p-value	Significant
Easy	Recall	Q1	3.39	> .05	No
		Q3	.24	> .05	No
	Recognition	Q2	.37	> .05	No
		Q4	.83	> .05	No
Moderate	Recall	Q5	10.91	< .05	Yes
		Q7	1.29	> .05	No
	Recognition	Q6	.83	> .05	No
		Q8	.37	> .05	No
Difficult	Recall	Q9	8.57	< .05	Yes
		Q11	3.39	> .05	No
	Recognition	Q10	7.03	< .05	Yes
		Q12	2.40	> .05	No

Table 10: Chi-square results for the correctness of users' answers to each question in both groups (df = 1, CV = 3.84)

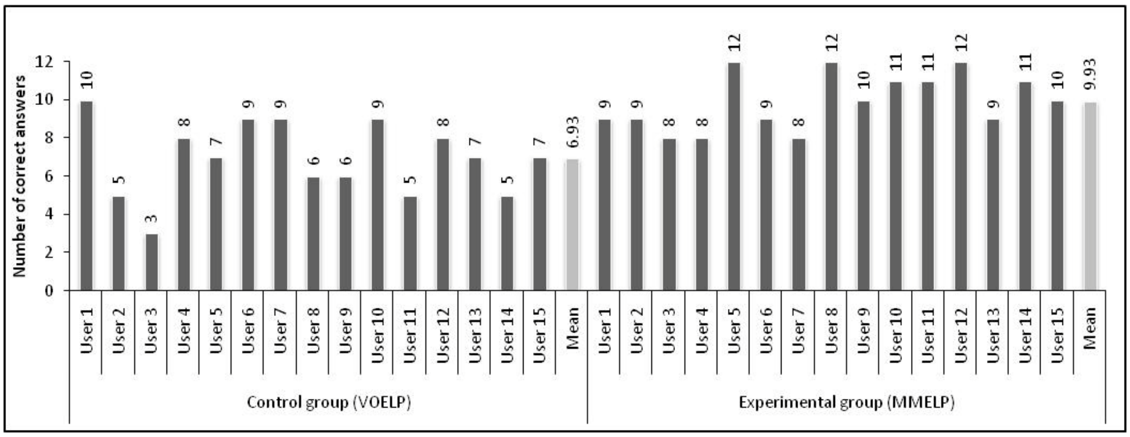


Figure 14: Total number of correct answers achieved by each user in both groups

3.9.2.5 Each User

Figure 14 shows the total number of correct answers achieved by each user in both control and experimental groups. It is worthy to note that 3 users (labelled 5, 8 and 12) of the MMELP correctly answered all the 12 questions and another three users (10, 11 and 14) achieved 11 correct answers. On the other hand, none of the VOELP users was able to reach a similar performance level where the maximum achievement observed was 10 correct answers by User 1. Also, the weakest user in the experimental group (User 7) scored 5 correct answers greater than that in the control group (User 3). On average, the number of correct answers per user in the experimental group was 9.93 compared to 6.93 in the control group. In short, using multimodal metaphors in communicating the learning material enabled the users in the experimental group to outperform their counterparts in the control group in answering the required questions correctly.

3.9.3 User Satisfaction

User satisfaction in regard to different aspects of the applied e-learning platform was measured in both groups by users' answers to the post-experimental questionnaire which consisted of 8 statements related to ease of use, confusion, nervousness, ease of

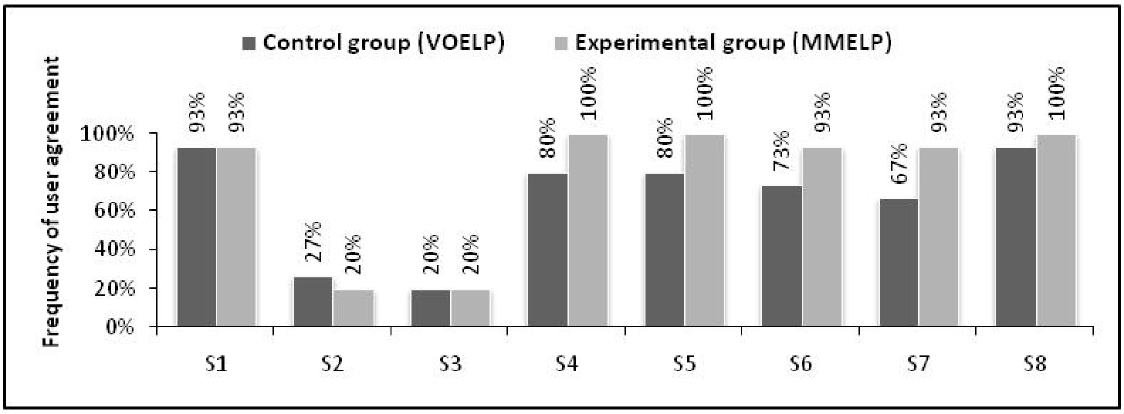


Figure 15: Frequency of users' agreement to each satisfaction statement in both VOELP and MMELP conditions

learning, ease of identification of the presented information related to classes, multiplicity and associations, and overall satisfaction. The six-point Likert scale ranging from 1, the value of strong disagreement, to 6, the value of strong agreement was used for each statement. The overall satisfaction score for each user was calculated using the SUS (System Usability Scale) method [230]. The mean satisfaction score for the users in the experimental groups was 84% compared to 68% for the users in the control group. Statistically, the Man-Whitney test demonstrated that the difference in users' satisfaction between both groups was significant (U = 50, CV = 72, $p < .05$). In other words, the MMELP was more satisfactory than the VOELP.

Figure 15 shows the frequency of user agreement to each statement in the satisfaction questionnaire. Refer to Appendix A6 for users' responses to the satisfaction questionnaire. It can be seen that similar levels of agreement were expressed by the users in both groups for *ease of use* (S1) and *nervousness* level (S3). However, the MMELP was less *confusing* (S2) and could be *easier to learn* (S4) as opposed to the VOELP. In the first statement S1, 93% of users in both group agreed that the tested e-learning interfaces were easy to use. The second statement (S2) asked the users whether they found the interface confusing. In this regard, users of the MMELP expressed a higher level of disagreement 80% than the users of the VOELP 73%. Similar to S1, the same level of agreement was observed for S3 where only 20% of the users in each group *felt nervous* during the interaction with the tested interface. With respect to S4, all of the MMELP users believed that *most people will learn the use of this tool quickly* compared to 80% in the VOELP.

Additionally, it can be noticed that users' agreement in the experimental group was obviously higher as opposed to the control group in terms of aspects related to the learning process (S5, S6 and S7). In other words, in using the MMELP, all users found *it was easy to identify the communicated information about classes* (S5) compared to 80% using the VOELP. Also, 93% of the MMELP users agreed that *it was easy to identify the multiplicity* (S6) *and the communicated information about associations among classes* (S7) compared to 73% and 67% of the VOELP users respectively. On overall, all users in the experimental group were *satisfied with the tested interface* (S8) whereas less percentage 93% was observed for users in the control group. In brief, using the multimodal metaphors to convey the learning material resulted in generating positive views of users. Therefore, the multimodal e-learning interface can be considered more satisfactory than the text-based one.

3.10 Discussion

The present empirical study investigated the usability and learning performance of multimodal e-learning interface as opposed to text with graphics one. The obtained results have been used to compare the two interfaces in terms of efficiency, effectiveness and user satisfaction. The present study also focused on the factors that could affect the role of multimodal interaction metaphors such as the complexity level (easy, moderate and difficult) and the type (recall and recognition) of the required learning activities. Therefore, these results are discussed from the following three angles to get an insight into what contribution has been made by the multimodal metaphors in users' efficiency, effectiveness and satisfaction.

1. Time taken to answer the required learning question in total and in terms of complexity and type (efficiency),
2. Correctness of users' answers to the required questions in total and in terms of complexity and type (effectiveness and learning performance), and
3. User satisfaction and experience with both of the tested e-learning interfaces.

Although the text with graphics interface offered simpler typical interaction, the obtained results showed that the use of multimodal metaphors (recorded speech, earcons, and avatars) was significantly more efficient and effective as well as more satisfactory than using the text with graphics in communicating the learning material in e-learning interfaces.

3.10.1 Question Answering Time

The first hypothesis assumed that the multimodal e-learning interface will be more efficient than the text with graphics one regarding the efficiency of users in answering the required questions. The experimental results, as shown in Figure 7A, demonstrated that using the multimodal interaction metaphors resulted in a significant reduction in the time needed by users in the experimental group to respond to the evaluation questions. Experimental observations revealed that users of VOELP in the control group regularly switched their visual attention between the textual explanations provided in the notes text box and class diagram representations in order to understand the presented information which may have overloaded their visual channel. On the other side, users of the MMELP in the experimental group were able to maintain their visual attention to the class diagram representations while they were listening to the auditory messages delivered by speaking avatar, recorded speech and earcons. The inclusion of different multimodal communication metaphors in the MMELP helped the users to concentrate better on the presented information through the auditory channel while at the same time using the visual channel to understand this information [45]. Therefore, they were significantly aided by the addition of these metaphors in the MMELP in terms of spending lower answering time than users of the VOELP. These results suggested that using the speaking facially expressive avatar, earcons and recorded speech could be significantly more efficient than using only the text with graphic metaphors in presenting clarifications related to the learning material used in this experiment, thus accepting what has been hypothesized in H1.

With respect to question complexity, it was expected, as stated in H2, that the MMELP will be more efficient than the VOELP with an increasing level of complexity. The results of this experiment (refer to Figure 7) showed an increasing difference in answering time, in favour of the experimental group, when the required questions become more difficult. In other words, the more complex the presented learning material, the larger the benefit of using multimodal metaphors. In examples 1 and 2 (easy level), the presented material was simple and limited resources are needed for cognitive processing of that material; however, with increasing complexity, more information is delivered and less cognitive resources become available for processing [197]. In this case, using multimodal metaphors could benefit in extending the capacity of working memory to enable processing of both verbal (auditory) and non-verbal (visual) information [235]. Therefore, the experimental results indicated the gradual contribution of the multimodal metaphors in users' efficiency where users of the MMELP respond significantly faster to the required easy, moderate and difficult evaluation questions, as a result supporting H2. In brief, it was experimentally evidenced that the efficiency of the multimodal metaphors could be influenced by the complexity level of the communicated learning information.

In regard to the question type, the third hypothesis predicted the MMELP to be more efficient for both recall and recognition questions. On the whole, the experimental findings indicated that the addition of the multimodal metaphors as applied in the MMELP particularly contributed to memory recall activities regardless of its complexity. In recall questions, users needed to retrieve the presented information from their memory and this may have taken time depending on the complexity of the task. On the other hand, answering the recognition questions needed only to select the correct option among the given alternatives and this might result in reducing the time needed by users in both groups to answer this type of questions. Therefore, the multimodal metaphors did contribute in recognition questions as much as in recall ones where users in the experimental group consumed significantly less time to complete easy, moderate and difficult recall tasks and difficult only recognition tasks. In other words, no significant difference between the two groups was observed for easy and moderate recognition tasks. On overall, H3 was rejected but could be partially accepted for recall questions. Therefore, it can be said that the effect of the tested multimodal metaphors on answering time is limited to memory recall activities regardless of its difficulty, and it could be beneficial only in high complexity recognition tasks.

3.10.2 Correctness of Answers

It was expected that users of the MMELP will outperform VOELP users in terms of correctly answered question. As shown in Figure 11, the MMELP was superior to the VOELP in enhancing users' learning achievements. It seems that using more than one communication metaphor of a different nature in the MMELP attracted the users and captured their attention. It also assisted users to distinguish among the different types of information provided by each of these metaphors and enabled them to remember this information for longer time. This effect can be returned to multimedia principle [236]; involving other human senses than the visual channel in the interaction process could assist in extending the capacity of working memory and, as a result, the users' ability to perceive and understand the presented information could be enhanced. The fact that users in the experimental group retained the communicated information for longer time (compared to the control group) enabled them to attain significantly a higher number of correct answers than their counterparts in the control group. These findings confirmed the assumptions made in H4.

In terms of question complexity, it was hypothesized that the MMELP will be more effective than the VOELP with an increasing difficulty of the required questions. In this regard, the obtained results (refer to Figure 11B) were similar to those observed for efficiency and therefore supporting H5. Although the MMELP condition outperformed the VOELP condition in answering easy questions, the influence of the multimodal metaphors did not reach a significant level. As mentioned before, these questions were simple and the users in both conditions were able to easily getting the answer. However, a larger contribution of the multimodal metaphors was observed when a higher level of mental processing was needed where users in the experimental group achieved significantly higher correct answers than the control group in moderate and difficult questions. These findings confirm the gradual effect of multimodal metaphors with an increasing complexity tasks and demonstrates that users' learning performance can be improved by the incorporation of these metaphors in e-learning interfaces. In other words, the complexity level of the presented learning content can influence the effectiveness and the efficiency of the tested multimodality in e-learning interfaces.

Considering the question type, the experimental results, as expected in H6, showed that users of the MMELP accomplished substantially a larger number of correct answers than the VOELP users in both recall and recognition questions. In order to successfully answer the recall question, users had to correctly retrieve from their memory part of the communicated learning content. Information in the MMELP was presented in a teacher like scenario in which the avatar simulated a teacher with natural head movement, facial expressions and natural speech while other aspects of the learning materials were presented using earcons and recorded speech. The results of this experiment indicated that user learning experience as formed by the combined multimodal metaphors enabled users to learn better without distracting their attention away from the presented content. This is particularly demonstrated in the moderate and difficult recall activities which are more difficult to be completed than the easy recall ones (refer to Figure 12). The low correctness rate of recall questions in the VOELP condition (50% compared to 84% in the MMELP condition) demonstrates that users' memory in the text with graphics interface was not aided as much as in the multimodal interface. To answer the recognition questions successfully, users had to choose the correct option from the given options. There is always a possibility that this answer could be chosen by the user due to chance, which is far more difficult to happen in responding to recall questions. The successful completion was 66% and 81% in control and experimental groups respectively.

The difference, although smaller than the one in the recall questions, still indicates that users performed better when their e-leaning has taken place in the presence of multimodal metaphors especially in answering difficult recognition questions (refer to Figure 12). In brief, the results suggest the use of multimodal metaphors as combined in the MMELP to enhance users' learning achievements in both recall and recognition activities.

3.10.3 User Satisfaction

On the whole, it was expected that users of the MMELP would be more satisfied than the users of the VOELP. Consistent with this assumption, the multimodal presentation of the learning material in the MMELP has shown to be significantly more satisfactory than the text with graphics in the VOELP. It seems that using the facially expressive avatar in a human-like approach in addition to recorded speech and earcons was interesting and attractive for users in the experimental group. Therefore, they expressed a more positive attitude towards the audio-visual communication of the learning material. Although both of the tested e-learning interfaces were easy to use and learn, neither was confusing nor nervous, the obtained results did not demonstrate a remarkable difference between both groups of users regarding these satisfaction features (refer to S1 to S4 in Figure 15). A larger difference however was observed on specific statements related to learning (refer to S5 to S7 in Figure 15). These results derived from two independent groups and users within those two groups were not presented with both interface versions in order to make an informed comparison. However, users in the experimental group may have had prior experience to typical learning interfaces (refer to Figure 6) and this probably served as a comparison point. Typically, users in the experimental group thought that their learning was better aided by the multimodal metaphors. It was easier for them to identify the learning information about classes, associations, and multiplicity, which have been communicated by avatar, speech, and earcons respectively. This result on its own is not conclusive as it is based on subjective rating of users and the typical mean difference is not large enough (although a statistical significance for the overall satisfaction results was reached). However, when user satisfaction, efficiency and effectiveness results are combined with each other, the argument that users in the experimental group were helped by the multimodal metaphors becomes much stronger. It can therefore be extrapolated that the multimodal aided e-learning is more likely to result in an enjoyable and satisfying experience for the user. This experience is linked with the ability to complete learning tasks correctly and quickly. In sum, the overall results of this experimental study suggest the importance of the tested multimodal interaction metaphors in enhancing user learning performance and the usability of e-learning interfaces in terms of efficiency, effectiveness and user satisfaction.

Nevertheless, the obtained results did not bring out the individual role played by each of the investigated multimodal metaphors in enhancing usability as will as learning performance in the MMELP condition. To some extent, the experimental results (see Figure 9, Figure 13, Table 9 and Table 10) provided a little in this regard when related to Table 6; however it could not be considered as sufficient to determine how each of the speaking avatars, recorded speech and earcons contributed in the obtained results because the design of the required questions was not dedicated for this purpose. Previous studies [8, 10, 12, 197] evidenced that the using a human-like avatar as a virtual pedagogical agent could facilitate the learning process and enhance users' learning performance. Though, these studies did not investigate users' attitude towards facial expressions and body gestures that could be incorporated in avatars when employed as virtual lecturer. Therefore, the next experiment (as described in Chapter 4) has been prepared to explore the usability (in terms of efficiency, effectiveness and user satisfaction) and learning performance of three different modes of utilising avatars as virtual lecturers in e-learning interfaces as well as to obtain users' feedback in relation to the use of specific facial expressions and body gestures in these interfaces.

3.11 Summary

This chapter investigated the influence of multimodal interaction metaphors on usability (in terms of efficiency, effectiveness and user satisfaction) as well as learning performance in e-learning interfaces. This investigation has been carried out by developing two different versions of the experimental e-learning platform. The first version was based on text with graphics in presenting the learning content about class diagram notation. However, the second version involved a novel combination of multimodal metaphors (recorded speech sounds, earcons and avatar with facial expressions) to deliver the same learning material. Both e-learning platforms were then empirically evaluated by two independent groups of users. The first group (control) tested the text with graphics interface and the second one (experimental) tested the multimodal interface in performing common tasks and answering a set of learning evaluation questions.

The results obtained from this experiment confirmed that the multimodal metaphors could indeed help to improve the usability of e-learning interfaces by reducing the time needed in responding to the required learning activities and enabling the users to perform these activities more accurately as well as making the interface more satisfactory. In other words, it can be concluded that the tested multimodal metaphors could significantly contribute in enhancing users' learning performance and the usability of e-learning interfaces in terms of efficiency, effectiveness and user satisfaction. Therefore, the inclusion of multimodal metaphors is suggested and could be taken into consideration when designing user interfaces of e-learning applications.

Chapter 4

Experimental Phase II: Investigating the Role of Avatars in the Multimodal E-Learning Interfaces

4.1 Introduction

Experimental results obtained from the first experiment demonstrated the importance of recorded speech, earcons and speaking facially expressive avatars in enhancing usability and learning performance in e-learning interfaces. These results however did not clearly identify the contributing role of each of these multimodal metaphors in the obtained enhancement.

This chapter describes the second experiment that has been conducted to explore and compare the role of avatars when incorporated as virtual lecturers in e-learning interfaces to present three different lessons about class diagram notation [216]. In addition to textual and graphical communication metaphors, animated speaking avatars were employed in three different modes of presentation which are: virtual lecturer with facial expressions and natural recorded speech, virtual lecturer with facial expressions, body gestures and natural recorded speech, and two virtual lecturers; male and female, with facial expressions and natural recorded speech. The following sections provide a detailed description of the research aims and objectives, hypotheses, experimental platforms, design of the experiment, results and discussion.

4.2 Aims

The aim of this experiment was to obtain an overall feedback from users in regard to their evaluation of facial expressions and body gestures when being used by virtual lecturer in e-learning interface in both the presence and absence of interactive context. It is aimed also to specify which facial expressions and body gestures are more desirable to users. Moreover, the experiment is aimed at testing the usability aspects and users' learning performance of e-learning interfaces that use avatars as virtual lecturers. More specifically, this experiment is aimed at exploring if there are significant differences among the tested e-learning platforms in terms of usability and learning performance and to identify users' most preferred platform as well among the following:

1. Virtual Lecturer with Facial expressions platform (VLFE)
2. Virtual Lecturer with Body Gestures platform (VLBG)
3. Two Virtual Lecturers with Facial expressions platform (TVLFE)

4.3 Objectives

In order to meet the aims stated in section 4.2, the following objectives were considered:

1. Formulation of hypotheses.
2. Development of three experimental platforms that present the learning information about the class diagram notation in three different presentation modes; speaking avatar with facial expressions, speaking avatar with facial expressions and body gestures, and two speaking avatars with facial expressions.
3. Carrying out an experimental investigation for the role of expressive avatars in e-learning interfaces using the developed experimental platforms by one group of users.

4. Obtaining the users’ views for facial expressions and body gestures used by the virtual lecturer in both the absence and presence of interactive context.
5. Measuring the efficiency of each presentation mode by the time users spent to complete the required tasks.
6. Measuring the effectiveness of each presentation mode as well as learning performance by calculating the percentage of tasks successfully completed by users.
7. Measuring the users’ satisfaction in regard to each presentation mode by their ratings for different aspects and learning experience with the applied platform.

4.4 Hypotheses

Based on the assumption that the inclusion of animated human-like virtual lecturers with facial expressions and body gestures would affect the usability and learning performance in e-learning interfaces, the following hypothesis were formulated:

H1: Positive facial expressions will also be rated positively and liked by users when used by an expressive avatar in a non-interactive context to communicate information in e-learning interfaces.

H2: Positive facial expressions of an expressive avatar will be rated by users more positively when used in interactive e-learning interfaces for the communication of learning information.

H3: Positive body gestures will also be rated positively and liked by users when used by an expressive avatar in a non-interactive context to communicate information in e-learning interfaces.

H4: Positive body gestures used by an expressive avatar will be rated by users more positively when used in interactive e-learning interfaces for the communication of learning information.

H5: There will be a difference in efficiency of the experimental platforms VLFE, VLBG, and TVLFE in terms of time taken by users to complete the required tasks.

H6: There will be a difference in effectiveness of the experimental platforms VLFE, VLBG, and TVLFE in terms of tasks successfully completed by users.

H7: The efficiency and effectiveness of the experimental platforms will differ in terms of recall tasks.

H8: The efficiency and effectiveness of the experimental platforms will differ in terms of recognition tasks.

H9: There will be a difference between the experimental platforms VLFE, VLBG, and TVLFE in terms of user's satisfaction.

4.5 Experimental E-Learning Platforms

To serve as a basis for this experiment, three different e-learning platforms were built from scratch and designed to utilize speaking avatars with human-like facial expressions and body gestures as well as natural recorded speech with prosody in order to offer audio-visual presentation of the learning material. The presentation methods provided by these platforms were: (1) speaking avatar with facial expressions in VLFE platform, (2) speaking avatar with facial expressions and body gestures in VLBG platform, and (3) two speaking avatars with facial expressions in TVLFE platform. It is believed that using avatars in this manner imitates to a large extent the traditional face-to-face interaction that typically takes place between the lecturer and the learners.

Lesson	Title	Duration (minutes)	Words	Paragraphs	Lines
Lesson 1	Classes and objects	3.24	342	3	20
Lesson 2	Class naming and drawing	3.28	368	3	22
Lesson 3	Associations and multiplicities	5.90	564	4	33

Table 11: Summary of the learning lessons presented in the second experiment

4.5.1 Learning Material

Three different lessons about class diagram notation were communicated to the users with the aforementioned three ways of presentation. These lessons are summarised in Table 11. The first lesson presents general concepts about classes and objects and the difference between both. In the second lesson, guidelines and rules in relation to class naming and drawing are provided. Lastly, the third lesson explains what is meant by association and multiplicity and how these concepts are implemented in class diagrams. The presentation of these lessons was supported by brief textual notes and graphical illustration in the form of PowerPoint slides. The content of these lessons was adapted from [216] and can be found in Appendix B2. Although the presentation of these lessons varied among the tested platforms, the content and the format was the same.

4.5.2 Virtual Lecturer with Facial Expressions Platform (VLFE)

This platform, as can be seen in Figure 16 uses an expressive avatar with facial expressions as virtual lecturer. The interface provides command buttons (denoted by (a) in Figure 16) to enable the selection of the lesson to be presented. It also provided two separate components for the presentation process namely speaking expressive avatar (g) on the right-hand side of the interface and the PowerPoint presentation (b) on the left-hand side. When the user clicks the button of a given lesson, this button is highlighted with red colour to indicate the current lesson, and the virtual lecturer starts presenting the lesson supported by brief textual notes and graphical illustrations displayed in the

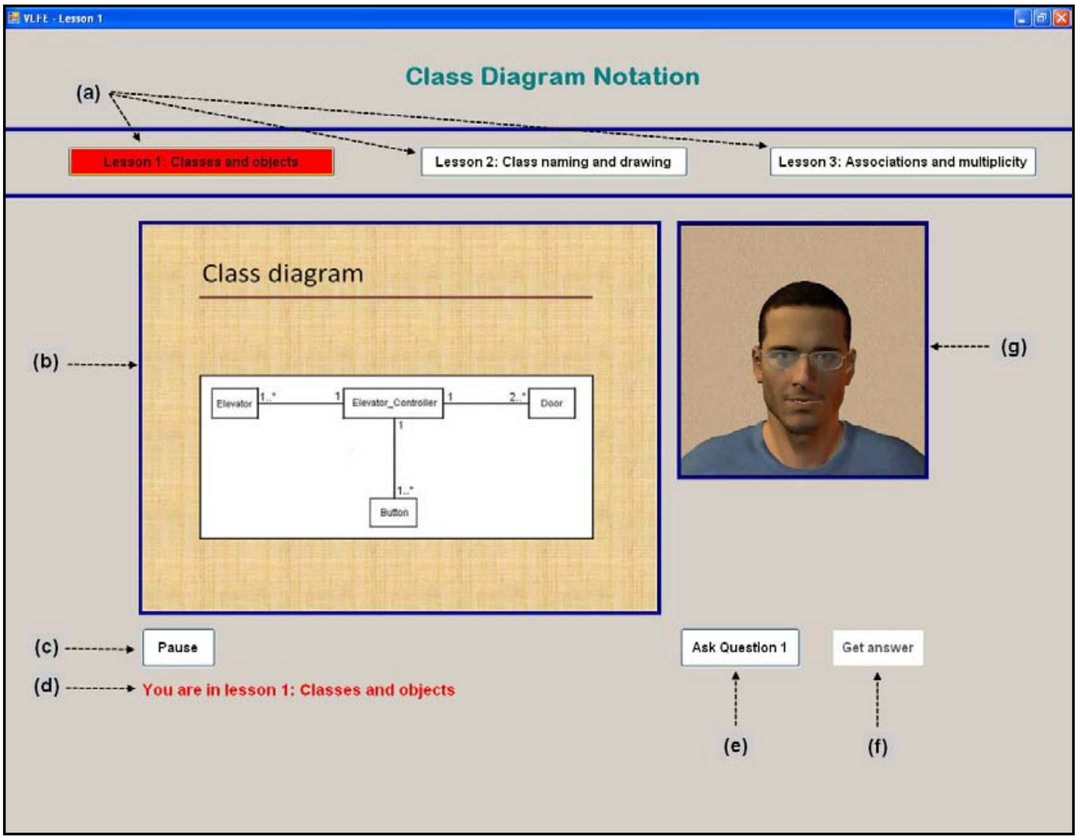


Figure 16: An example screenshot of VLFE platform

PowerPoint slides. The interface also offers pause/play functionalities (c) to enable the user to control his/her learning at any point of time. Upon completion of each lesson, two questions related to the delivered information can be asked to obtain further explanations from the virtual lecturer. When clicking the ask question button (e), the first question is displayed textually in the bottom of that button and the user had to click the get answer button (f) in order to obtain the answer which at that moment is provided by the virtual lecturer with relative explanations shown in the presentation part (b). The same procedure has to be followed in asking and answering the second question. In order to insure the consistency and the confirmation of controlling the experiment, questions related to each lesson were the same for all users. Furthermore, this platform provided a text box (d) in the bottom of left-hand part of its interface to inform which lesson is currently presented.

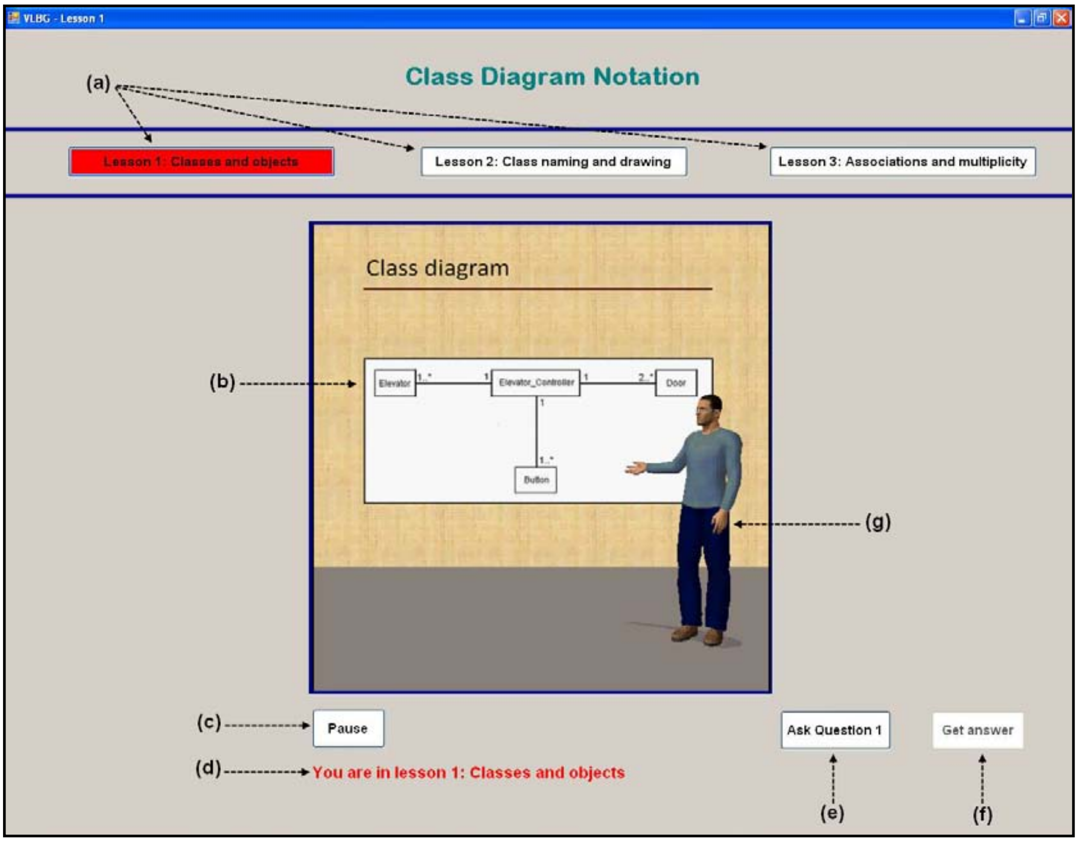


Figure 17: An example screenshot of VLBG platform

4.5.3 Virtual Lecturer with Body Gestures Platform (VLBG)

This platform employed the speaking and expressive avatar with full body gestures to virtually lecture the experimental learning lessons. As can be seen in Figure 17, the PowerPoint presentation (denoted (b)) and the virtual lecturer with full body animation (g) were combined in one scene and shown in the same component placed in the middle part of the interface. This approach could be considered as the most relative to the real class-based learning situation because the virtual lecturer was designed to simulate the same body movements usually performed by the human lecturer in a traditional class room. Similar to the VLFE, interface features for lesson selection (a), pause/play (c), asking and answering questions (e and f respectively) and current lesson highlighting (d) were provided by the interface of VLBG platform. Also, the same procedure for lesson presentation and asking and answering questions was followed.

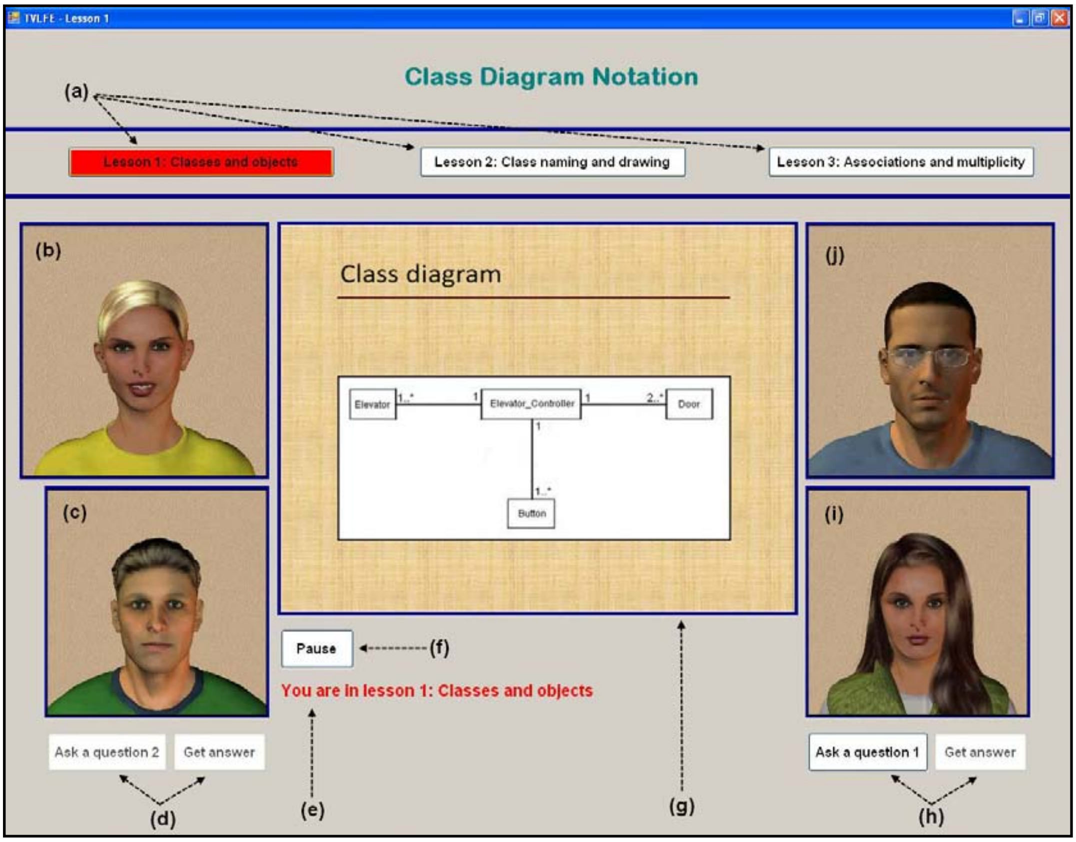


Figure 18: An example screenshot of TVLFE platform

4.5.4 Two Virtual Lecturers with Facial Expressions Platform (TVLFE)

As shown in Figure 18, the aspect that differentiates this platform from the VLFE and VLBG is the incorporation of female (denoted by (b) in Figure 18) and male (j) facially expressive virtual lecturers to share the presentation of each lesson supported by the PowerPoint presentation (g) placed in the middle part of the interface. The learning content was equally distributed between the two lecturers. Additionally, the interface of TVLFE included two more avatars to represent the female (i) and the male (c) students. In contrast to the VLFE and VLBG platforms, the role of the latter tow avatars was to ask the questions related to each lesson vocally. The first question was spoken by the female virtual student by clicking the ask question button placed nearby. This question was related to the learning content communicated by the male lecturer. Therefore, the answer in this case was provided by the male virtual lecturer when clicking the get answer button placed next to the ask question button (h). However, the second question asked by the male student was related to the lesson part that had been presented by the female lecturer and hence she provided the answer in this case. So, two additional buttons to ask and answer the questions (d) were provided in the interface and placed below the male virtual student. The remaining interface features (a, e and f in Figure 18) were similar to those provided by the VLFE and VLBG platforms.

4.5.5 Implementation of Avatars and PowerPoint Presentations

In addition to Mimic [223] and Poser [224] used for the creation of the experimental platform in the first experiment (see Section 3.5.3.3), the following tools have been also utilised to develop the components incorporated later on in the experimental platforms:

1. Microsoft Office PowerPoint 2007, used to create the textual and graphical presentation files related to the communicated lessons.
2. Camtasia Studio by TechSmith Corporation [237] used mainly for recording female and male speech sounds while the PowerPoint presentation of the learning material was running, and then producing the video file of AVI format (Audio Visual Interleave). The output file was a visual presentation of the learning content accompanied by the spoken explanations by the lecturers.

The main problem that has been encountered during the development of the experimental platforms was in generating the AVI files for Poser files. The rendering process was time-consuming particularly when the number of frames becomes larger; sometimes, the machine was suspended. In order to resolve this problem, each lesson was divided into 3 or 4 parts each of which was separately processed by the development tools as mentioned before. Thereafter, the AVI files for these parts were

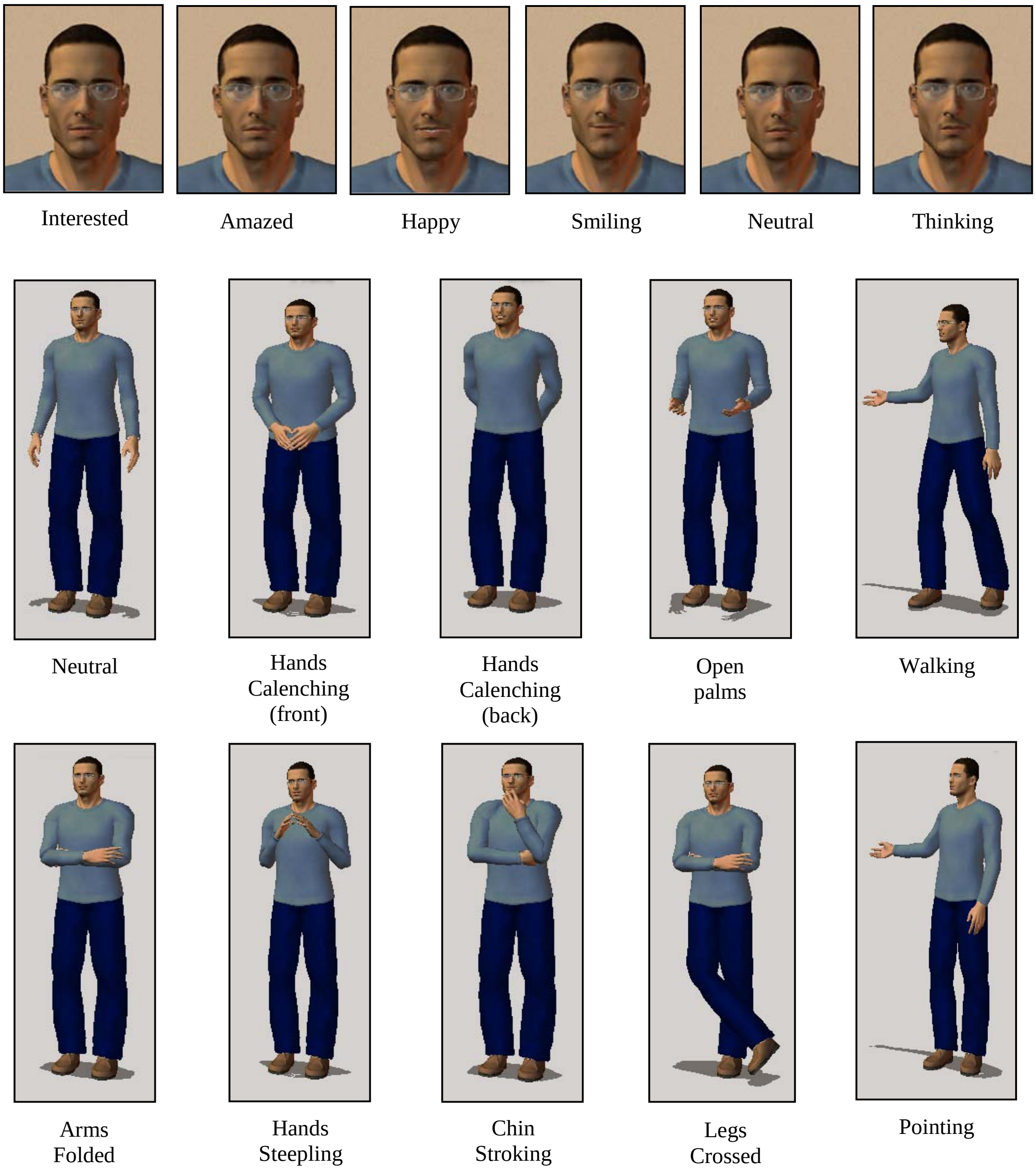


Figure 19: Facial expressions and body gestures used in the experimental e-learning platforms

combined together using the proper functionality provided by Camtasia to produce the final presentation and avatar files for each lesson. Figure 19 shows examples of facial expressions and body gestures used in the experiment.

Facial Expressions	Interested, Amazed, Happy, Smiling	Positive
	Neutral, Thinking	Neutral
Body Gestures	Hands Clenching-front, Hands Clenching-back, Open Palms, Pointing, Chin Stroking, Hands Steepling, Walking	Positive
	Neutral	Neutral
	Arms Folded, Legs Crossed	Negative

Table 12: Facial expressions and body gestures used in the second experiment classified according to [238] and [181] respectively

Six facial expressions were commonly used in both VLFE and TVLFE, whereas 10 body gestures were used in VLBG platform. These expressions and gestures are typically used by people in every day life. As can be seen in Table 12, facial expressions were classified into two groups; positive and neutral [238] while body gestures were categorized into positive, neutral, and negative [181]. More technical details about the development of the experimental e-learning platforms are available in Appendix D2.

4.6 Experimental Design

The *within-subjects* design methodology was followed in carrying out this experimental investigation. This design guarantees the participation of each user in testing all the systems being evaluated; therefore, it brings down the effect of any other external factors that might influence user performance from one treatment to another [239]. Therefore, one group of users was involved to test the experimental e-learning platforms: VLFE, VLBG, and TVLFE. A total of 48 users have taken part in the experiment on an individual basis. This experiment was composed of four main parts. The first part was the pre-experimental questions for users' profiling and to obtain their viewpoints in regard to the use of virtual lecturers and speech output in e-learning interfaces. The second part investigated the users' evaluation (positive or negative) of facial expressions and body gestures used in the experiment when presented to them randomly in the absence of interactive context. Each of these expressions and gestures was shown to users as still image on the screen. In the third part of the experiment, the experimental platforms were demonstrated to users and then used to present the experimental lessons interactively (as shown in section 4.5). This part was aimed at getting the users' perceptions of the same expressions and gestures when communicated in an interactive context, and comparing the experimental e-learning platforms in terms of efficiency and effectiveness as well as users' satisfaction and learning performance. The last part was directed to obtain an overall feedback from users related to the usefulness of the implemented multimodal metaphors, their preferred experimental platform, and any comment or suggestions.

4.6.1 Experimental Procedure and Tasks

A total of 48 users have taken part in the experiment individually and three lessons about class diagram notation were communicated using the experimental platforms. The procedure followed in performing the experiment with each user is shown in Table 13 and detailed in the following subsections. More details can be found in Appendix B1.

4.6.1.1 Pre-Experimental Questions

The experiment was clearly explained to each user before it started by answering the pre-experimental questionns for users' profiling and obtaining their viewpoints in regard to the use of virtual lecturers and speech output in e-learning interfaces. In this questionnaire, users were asked to:

1. Provide personal and educational information.
2. Record previous knowledge about each of class diagram notation, avatars and e-learning.

Users	Pre-experimental questionnaire	Pre-experimental tasks	Demonstration of the experimental platforms	Platform used for lesson 1	Post-conditional tasks	Platform used for lesson 2	Post-conditional tasks	Platform used for lesson 3	Post-conditional tasks	Post-experimental tasks
1, 7, 13, 19, 25, 31, 37, 43				VLFE		VLBG		TVLFE		
2, 8, 14, 20, 26, 32, 38, 44				VLFE		TVLFE		VLBG		
3, 9, 15, 21, 27, 33, 39, 45				VLBG		VLFE		TVLFE		
4, 10, 16, 22, 28, 34, 40, 46				VLBG		TVLFE		VLFE		
5, 11, 17, 23, 29, 35, 41, 47				TVLFE		VLFE		VLBG		
6, 12, 18, 24, 30, 36, 42, 48				TVLFE		VLBG		VLFE		

Table 13: Procedure followed in conducting the second experiment

3. Provide their opinions about the use of virtual lecturers and speech output in e-learning interfaces.

4.6.1.2 Pre-Experimental Tasks

This part of the experiment investigated how users would evaluate the use of each facial expression and body gesture incorporated in the experimental platforms when presented to them randomly in the absence of any interactive learning context. Each of these expressions and gestures was individually shown to users as still image along with its title in the screen (refer to Figure 19), and the users had to:

1. Rate the use of each facial expression positively or negatively.
2. Indicate two expressions most liked and two did not like.

3. Rate the use of each body gesture positively or negatively.
4. Indicate two gestures most liked and two most disliked.

4.6.1.3 Demonstration of the Experimental Platforms and Lessons

At the end of pre-experimental tasks, the experimental platforms were introduced by showing a 2-minute video recording that describes the components integrated in the interface of each platform, and thus provided all the users with a consistent demonstration of these platforms. However, users were allowed to ask questions if more explanations were needed. Thereafter, three learning lessons about the class diagram notation were presented interactively using the experimental platforms. These lessons were dependent on each other. Therefore, the order of presentation was the same for all users (i.e. lesson 1 then lesson 2 then lesson 3). However, each platform was used once with each user presenting one of these lessons. In order to ensure that all experimental platforms had been equally used for each lesson, these platforms were assigned to the three lessons on a systematic random rotation basis (refer to Table 13).

4.6.1.4 Post-Conditional Tasks

These tasks were required to be performed upon completion of each lesson and were aimed at evaluating the users' perceptions of the facial expressions and body gestures after being used in the presence of interactive learning context. It is also aimed at comparing the usability (in terms of efficiency, effectiveness, and user satisfaction) as well as users' learning performance levels across the tested platforms. Each user was asked to answer 4 questions related to the delivered learning material. These questions were divided equally into recall and recognition. Recall questions asked the user to retrieve the required information and write the answer. However, the recognition questions asked the user to select the answer among the given 2 or 4 options. In addition, the user was required to rate positively or negatively each facial expression or body gesture being used by the virtual lecturer(s). These expressions and gestures were similar to those presented in the pre-experimental tasks. Furthermore, the user had to respond to the satisfaction questionnaire. More specifically, this questionnaire was composed of 18 statements each of which had a 5-point Likert scale where one denoted strongly disagree and five denoted strongly agree. The first 10 statements were based on SUS questionnaire [230] to measure users' attitude towards different aspects of the applied platform. SUS questionnaire offers a *"reliable, low-cost usability scale that can be used for global assessment of systems usability"* [230]. However, the remaining statements were added to obtain users' feedback about their learning experience with each experimental platform.

4.6.1.5 Post-Experimental Tasks

The last part of the experiment was dedicated to obtain users' views regarding how useful they found each of the multimodal metaphors used in the experimental platforms. Each user was required to rate each metaphor on the usefulness scale ranging from one, the value of least useful to five, the value of most useful. Also, users were asked to select one experimental platform which they most preferred and finally to provide their suggestions or comments (if any).

4.6.2 Variables

Three types of variables were considered in this experiment which were: independent, dependent and controlled variables. The variables controlled in this experiment were similar to those considered in the first experiment (refer to section 3.6.3.3).

4.6.2.1 Independent Variables

IV 1: Presentation mode: this variable has three conditions; virtual lecturer with facial expressions, virtual lecturer with facial expressions and body gestures, and two virtual lecturers with facial expressions.

IV 2: Facial expressions and body gestures: were presented in both the presence and absence of interactive e-learning context.

IV 3: Question type: the experiment examined the influence of recall and recognition questions on the usability level and users' learning outcomes in each of the tested platforms.

4.6.2.2 Dependent Variables

As shown in Table 14, seven dependent variables were observed.

DV 1: Question answering time: measured by the time consumed by users in answering the required questions.

DV 2: Correctness of answers: measured by the percentage of correct answers achieved by users in response to the required questions.

DV 3: User satisfaction: measured by users' ratings in response to the satisfaction questionnaire comprising statements related to different aspects of the tested platform as well as learning experience gained from the interaction process.

DV 4: Preferred e-learning platform: attained by calculating the percentage of users who preferred each of the tested platforms.

Variable code	Variable	Measure
DV 1	Question answering time	Efficiency
DV 2	Correctness of answers	Effectiveness and user's learning performance
DV 3	User Satisfaction	Satisfaction
DV 4	Preferred platform	
DV 5	Users' views of the tested facial expressions	
DV 6	Users' views of the tested body gestures	
DV 7	Users' selection of the tested expressions and gestures	

Table 14: Dependent variables considered in the second experiment

DV 5: Users' views of facial expressions: estimated by the percentage of positive and negative users' views in regard to each of these expressions in both the absence and presence of interactive e-learning context.

DV 6: Users' views of body gestures: estimated by the percentage of positive and negative users' views in regard to each of these gestures when presented both interactively or non-interactively.

DV 7: Users' selection of facial expressions and body gestures: measured by the percentage of users' liking and disliking of each expression and gesture.

4.6.3 Users Sampling

Computing the factorial of 3; the number of experimental platforms, resulted in 6 possible substitutions for using these platforms in the presentation of the learning lessons. As a result, the number of users was determined to be a multiple of 6. Despite the fact that five test users could provide system usability evaluation, the participation of a larger number of users can offer more adequate usability results [75]. Therefore, the test sample consisted of 48 users participated in the experiment individually. All of them were volunteers and used the experimental platforms for the first time. They were selected to be inexperienced in class diagram notation; the learning topic which presented by the experimental platforms. The involvement of expert users in this topic would affect the experimental results because most probably they would rely on their previous knowledge in answering the required questions and consequently the effect of the tested experimental e-learning platforms on the users' performance will be avoided.

4.7 Data Collection

The collection of experimental data was mainly based on two resources; observations and questionnaires. For example, users' answers to both pre and post experimental questionnaires helped in gathering the data needed for obtaining an overall feedback about the characteristics of the users and their opinions related to the used multimodal e-learning platforms. Additionally, users' responses to the post-conditional tasks contributed to the evaluation of the usability parameters (i.e. efficiency, effectiveness and user satisfaction) and users' learning achievements attached to the implementation of each experimental platform.

4.8 Users' Profiling

Users' responses to the pre-experimental questionnaire were analyzed in order to identify their personal and educational information as well as their prior experience and views in regard to the use of avatars and speech output in e-learning interfaces. Figure 20 shows that half of the participants had an average age between 25 and 34, and another 42% of 35 – 44, 6% within 18 – 24 and 2% (only one user) 45 – 55 years old. In total, there were 13 (27%) female and 35 (73%) male users. The majority of them were overseas residents by 92%, and English was observed as a second language at 94%. In regard to the educational level, 75% of the sample users were master degree holders

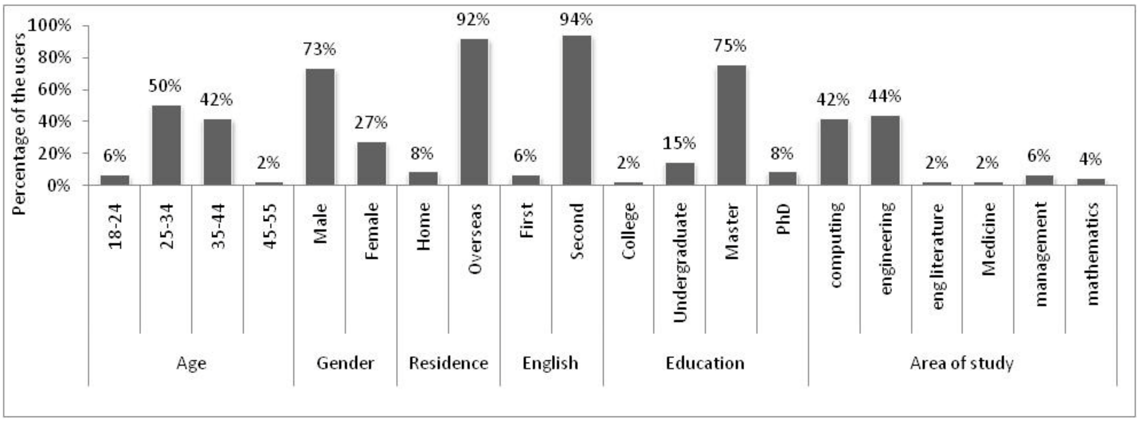


Figure 20: Users' profile in terms of age, gender, residence, English, educational level and area of study

followed by 15% bachelor and 8% doctorate. Only one user (2%) was a college student. Also, the majority of them come from an engineering and computing background with 44% and 42% respectively. As can be seen in Figure 21, and based on users' weekly use of computers and Internet, the participants were regarded as expert users as 83% of them use the computer for more than ten hours and another 15% use it for six to ten hours a week. Additionally, the frequency use of the Internet was found to be 81% ten or more hours and 10% 6 – 10 hours a week. When users were asked about their knowledge in class diagram notation, 85% had no prior knowledge and 8% limited, which demonstrate that a significant 93% were inexperienced in the learning material used in the study. Users were also inexperienced in avatars, facial expressions and body

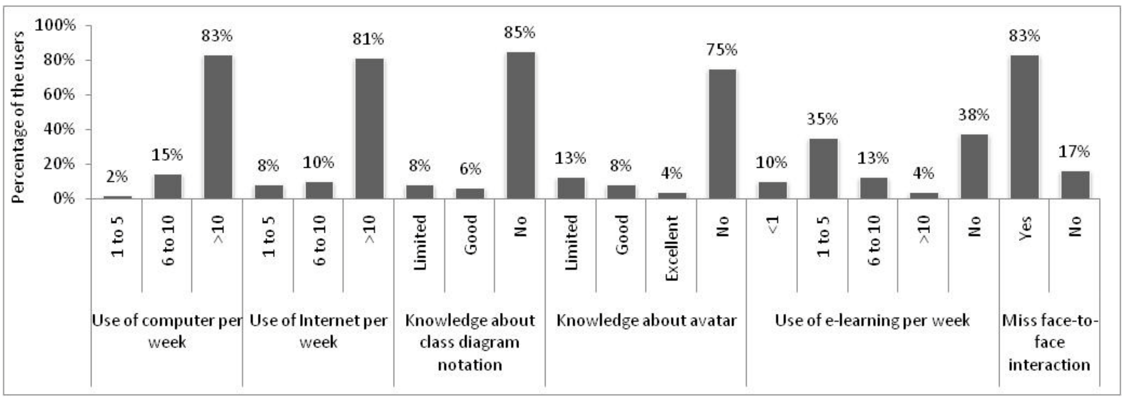


Figure 21: Prior experience of users in the second experiment

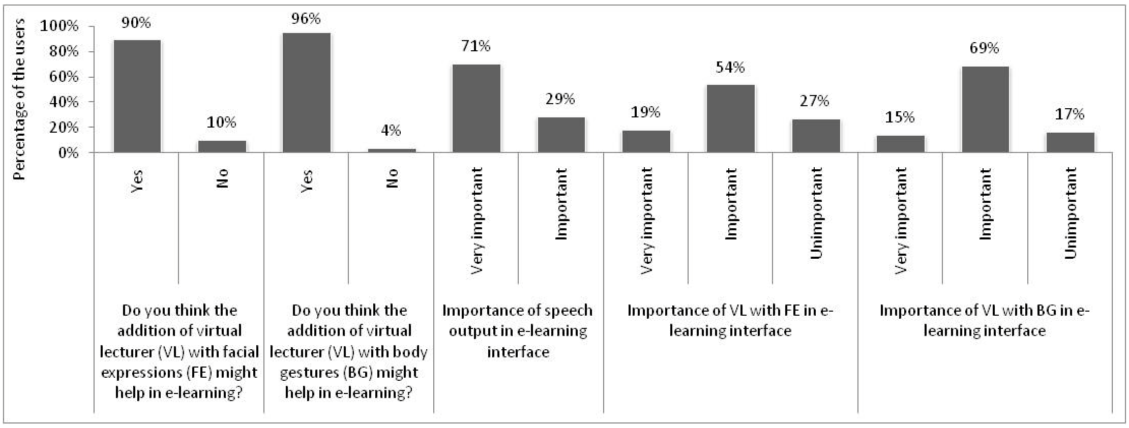


Figure 22: Users' views in relation to the use of virtual lecturers and speech output in e-learning interfaces

gestures. In this respect, 13% regarded their knowledge as limited whilst 75% had no knowledge at all. In regard to the users' involvement in e-learning activities, about 62% usually use e-learning websites or software on a weekly basis but at different periods of time, 83% of them expressed their loss of face-to-face interaction with the lecturer. Most likely, this experience could enable the users to provide a precise feedback based on comparing the tested platforms with other e-learning systems.

Figure 22 shows the users' views related to the use of avatars and speech output in e-learning interfaces. Approximately 90% of the users believe that the addition of facially expressive virtual lecturers could assist the e-learning process particularly in the presence of body language by 96%. In total, the importance of the inclusion of speech output in e-learning interfaces was observed at about 100% importance. Neither unimportant nor very unimportant rates were recorded for this metaphor. Also, according to the users' opinions, body gestures seem to be more important than facial expressions when used by virtual lecturer in e-learning interfaces. About 70% and another 15% of users rate the use of body gestures as important and very important respectively. However, the use of facial expressions was rated by 54% of users as important and another 19% as very important. On the whole, importance rating for body gestures was 84% compared to 73% for facial expressions. See Appendix B3 for more details about users' responses to the pre-experimental questionnaire.

4.9 Results and Analysis

The obtained results were analysed in terms of the users' evaluation of facial expressions and body gestures used in the experimental platforms, answering time (efficiency), correctly answered questions (effectiveness and users' learning performance) and user satisfaction. For statistical analysis, the nonparametric Chi-square test was used to examine the significance of differences in terms of categorical data such as users' views [234]. The Kolmogorov-Smirnov test showed that the remaining data was not normally distributed and hence Friedman's ANOVA was used. This test can be used to test the differences between experimental conditions in within-subjects design when the assumption of normal distribution of the data is violated [233]. Also, Wilcoxon Signed-rank test was used as the non-parametric equivalent of dependent t-test [233] to carry out follow-up pair-wise comparisons across the experimental conditions in this experiment. The significance level used in these statistical tests was $\alpha = 0.05$ indicating the existence of significant difference if p-value was less than that value.

4.9.1 Users' Evaluation of Facial Expressions

Figure 23 shows how users evaluated, positively or negatively, each facial expression used in this experiment, in the absence of interactive context and prior to experimenting any of the experimental platforms (see also Appendix B4). Apart from the *neutral* expression, it can be noticed that all the expressions were positively viewed by the users. More specifically, more than 65% of the participants believed that the positive

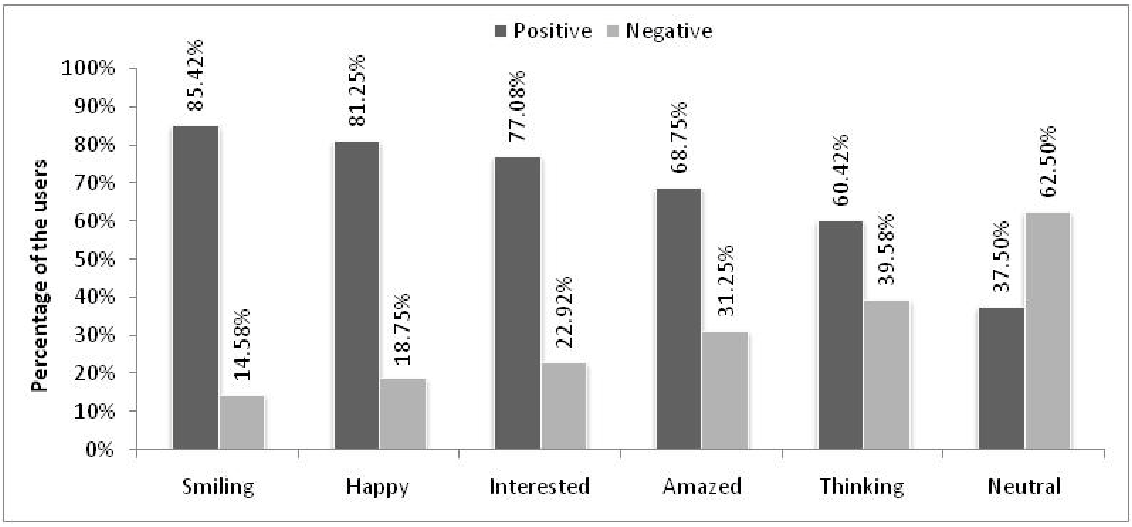


Figure 23: Users' evaluation of facial expressions presented in a non-interactive e-learning context prior to the experiment

expressions such as *smiling*, *happy*, *interested* and *amazed* could be used positively by virtual lecturer. The percentage for *smiling* expression reached about 85% and dropped down for the *happy*, *interested* and *amazed* expressions to about 81%, 77% and 68% respectively. For *thinking* expression, the results were less significant with 60% positive users' views. On contrast, the *neutral* expression had about 37 % of users' positive views which means that they had a negative impression about it. The significance of the difference between positive and negative views was examined by the Chi-square (χ^2) test. Table 15 shows the χ^2 values for all expressions. Positive expressions such as *interested*, *amazed*, *happy*, and *smiling* obtained positive significant results whereas the *neutral* expressions (i. e. *neutral* and *thinking*) did not show any significance.

Facial Expressions	χ^2 value	p-value	Significant
Neutral	3.00	> .05	No
Interested	14.08	< .05	Yes
Amazed	6.75	< .05	Yes
Happy	18.75	< .05	Yes
Smiling	24.08	< .05	Yes
Thinking	2.08	> .05	No

Table 15: Chi-square results for users' evaluation of facial expression used in the absence of interactive learning context (df = 1, CV = 3.84)

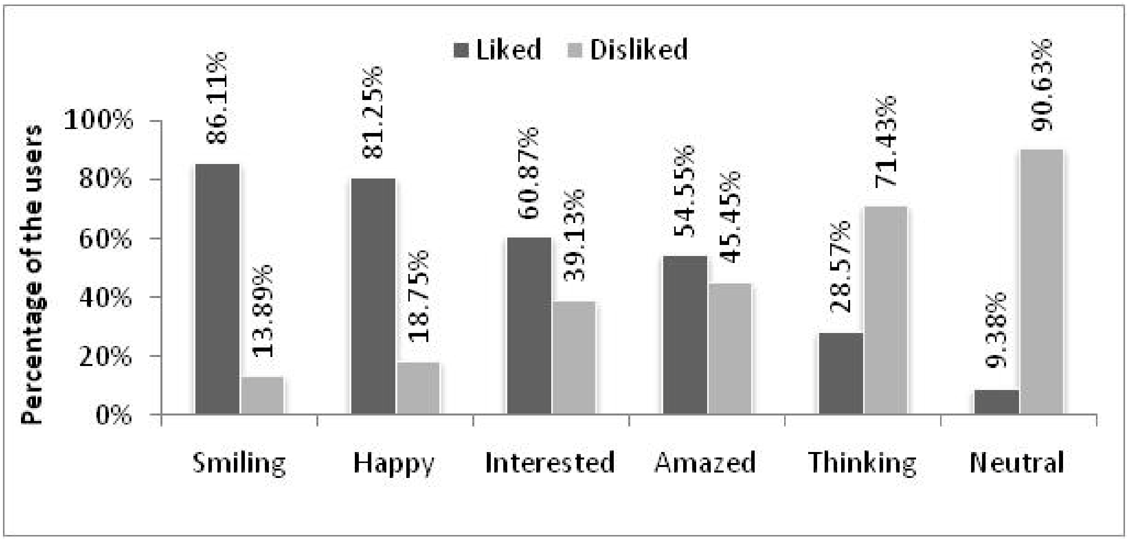


Figure 24: Users' selections of the facial expressions presented in a non-interactive e-learning context prior to the experiment

Facial Expressions	χ^2 value	p-value	Significant
Neutral	21.13	< .05	Yes
Interested	1.09	> .05	No
Amazed	0.18	> .05	No
Happy	12.50	< .05	Yes
Smiling	18.78	< .05	Yes
Thinking	6.43	< .05	Yes

Table 16: Chi-square results for users' selection of facial expression presented in the absence of interactive e-learning context (df = 1, CV = 3.84)

Prior to the experiment, users were also requested to select two facial expressions they liked and another two they did not like (see also Appendix B5). It can be observed from Figure 24 that positive expressions were liked by more than 55% of those users who selected it especially the *smiling* with 86% and *happy* with 81%.

On the other hand, *neutral* and *thinking* expressions were selected to be strongly disliked by 90% and 71% respectively. Although users expressed their liking of *interested* (60%) and *amazed* (55%), the χ^2 values shown in Table 16 indicate non-significant results for these expressions. The other expressions achieved significant results whether liking or disliking. In other words, users were satisfied only with positive facial expressions.

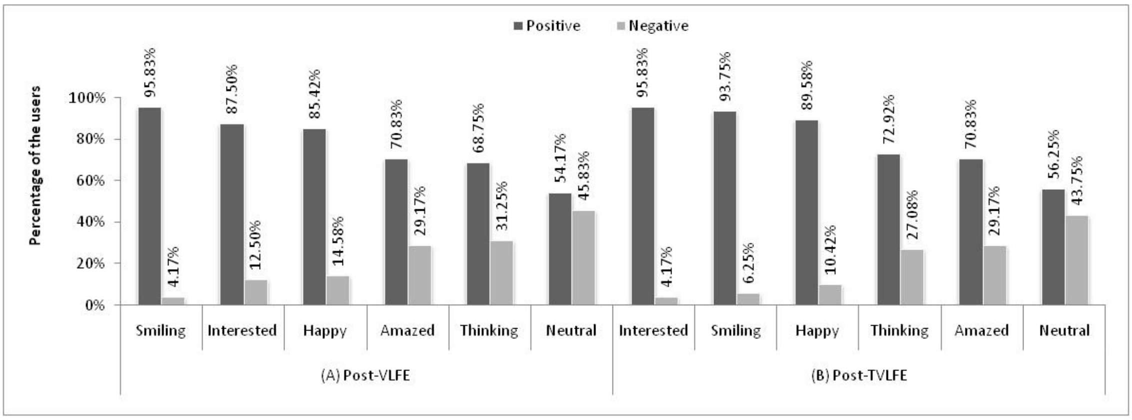


Figure 25: Users' evaluation of facial expressions after being presented interactively in VLFE (A) and TVLFE (B) experimental platforms

Figure 25 shows how the same set of facial expressions were evaluated by the users after they have had the opportunity to interact with both VLFE (A) and TVLFE (B) platforms (see also Appendix B6). It can be seen that users' positive feeling was improved in regard to most of these expressions when being used by virtual lecturers in interactive e-learning interfaces provided in both VLFE and TVLFE platforms. This, in particular, was noticeable for *neutral* expression where the percentage of positive views improved from 37% (see Figure 23) to about 55% in both VLFE and TVLFE conditions. All other expressions were positively rated by higher percentages. Compared to post-VLFE, *interested*, *thinking* and *happy* expressions obtained better results in post-TVLFE with 18%, 12% and 8% increments in positive rating respectively. On the other hand, *smiling* scored the highest positive percentage in post-VLFE. Lastly, users' rating was the same for *amazed* expression after experimentation of both VLFE and TVLFE. The χ^2 results for all these expressions after being used in an interactive e-learning context are shown in Table 17. In spite of the large shift in users' positive rating of *neutral* expression, its χ^2 value did not reach a positive significant level whereas users' evaluation of *thinking* expression became significantly more positive. However, the remaining expressions preserved their positive significant results.

Facial Expressions	χ^2 value		p-value	Significant
	Post-VLFE	Post-TVLFE		
Neutral	0.33	0.75	> .05	No
Interested	27.00	40.33	< .05	Yes
Amazed	8.33	8.33	< .05	Yes
Happy	24.08	30.08	< .05	Yes
Smiling	40.33	36.75	< .05	Yes
Thinking	6.75	10.08	< .05	Yes

Table 17: Chi-square results for users' evaluation of facial expression used in the presence of interactive e-learning context (df = 1, CV = 3.84)

The change of users' impression from pre-experimental phase to post-VLFE conditional phase about each facial expression is shown in Table 18 and the same information but from pre-experimental phase to post-TVLFE conditional phase is presented in Table 19.

Pre-Experimental	Post-VLFE	Neutral	Interested	Amazed	Happy	Smiling	Thinking
Positive	Positive	15	34	28	35	39	23
Positive	Negative	3	3	5	4	2	6
Negative	Positive	11	8	6	6	7	10
Negative	Negative	19	3	9	3	0	9

Table 18: Change of users' impression from pre-experimental phase to post-VLFE conditional phase about each facial expression used in the experiment

Pre-Experimental	Post-VLFE	Neutral	Interested	Amazed	Happy	Smiling	Thinking
Positive	Positive	13	36	26	37	39	23
Positive	Negative	5	1	7	2	2	6
Negative	Positive	14	10	8	6	6	12
Negative	Negative	16	1	7	3	1	7

Table 19: Change of users' impression from pre-experimental phase to post-TVLFE conditional phase about each facial expression used in the experiment

Facial Expression	Pre-Experiment Positive rating	Pre-Experiment Liking	Post-VLFE Positive rating	Post-TVLFE Positive rating	Overall Mean
Smiling	85.42	86.11	95.83	93.75	90.28
Happy	81.25	81.25	85.42	89.58	84.38
Interested	77.08	60.87	87.5	95.83	80.32
Amazed	68.75	54.55	70.83	70.83	66.24
Thinking	60.42	28.57	68.75	72.92	57.66
Neutral	37.50	9.38	54.17	56.25	39.32

Table 20: Overall mean values of users' ratings obtained for each facial expression in both the absence and presence of interactive e-learning context sorted in descending order

Table 20 shows the overall mean values of users' ratings for each expression in both pre-experimental phase and the post-conditional phase. On the whole, positive expressions such as *smiling*, *happy*, *interested*, and *amazed* were shown to be the best rated expressions. On the other hand, neutral expressions (i.e. *neutral* and *thinking*) were found the least regarded by users.

4.9.2 Users' Evaluation of Body Gestures

Figure 26 (and also Appendix B7) shows users' evaluation of the body gestures when presented to them individually in the absence of interactive e-learning context. For positive group (refer to Table 12), it can be seen that these body animations were evaluated positively. The *pointing* gesture obtained 93% positive score, followed by 89% for *open palms*, 85% for *hands steepling* and 81% for *walking*. A lower positive score was observed for *chin stroking* (64%) and *front clenching of the hands* (62%). However, *back clenching of hands* which has been supposed to be a positive gesture was perceived negatively by 58% of users. For the negative gestures, about 75% of the users showed a negative feeling regarding *legs crossed*. This percentage dropped down to about 56% for *folding the arms*. What is more, 60% of the users believed that *neutral* situation gives a negative impression.

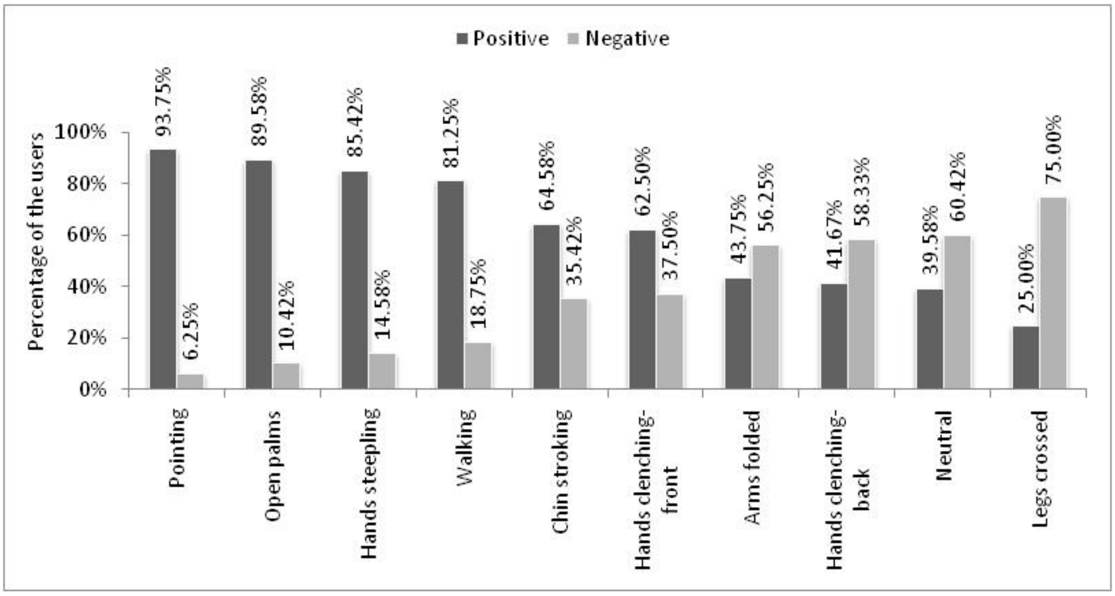


Figure 26: Users' evaluation of body gestures presented in a non-interactive e-learning context prior to the experiment

As can be seen in Table 21, each of *open palms*, *pointing*, *hands steepling*, *chin stroking*, and *walking* obtained positive significant ratings. As far as the negative gestures are concerned, *legs crossed* gesture obtained negative significant results. However, the remaining gestures did show significance neither positively nor negatively.

Body Gestures	χ^2 value	p-value	significant
Neutral	2.08	> .05	No
Hands clenching-front	3.00	> .05	No
Hands clenching-back	1.33	> .05	No
Open palms	30.08	< .05	Yes
Arms folded	0.75	> .05	No
Pointing	36.75	< .05	Yes
Hands steepling	24.08	< .05	Yes
Chin stroking	4.08	< .05	Yes
Legs crossed	12.00	< .05	Yes
Walking	18.75	< .05	Yes

Table 21: Chi-square results for users' evaluation of body gestures used in the absence of interactive e-learning context (df = 1, CV = 3.84)

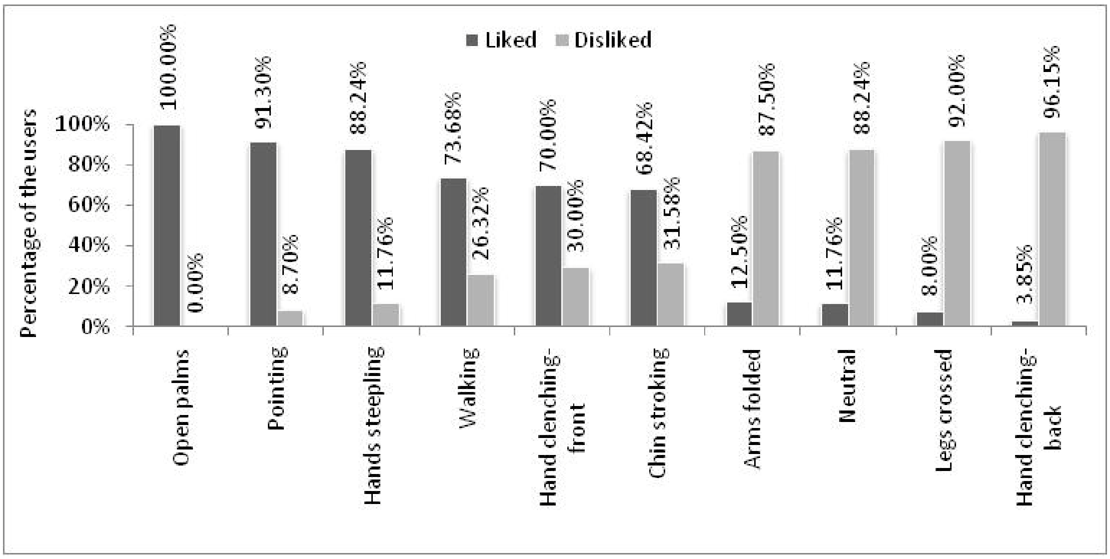


Figure 27: Users' selections of body gestures presented in a non-interactive e-learning context prior to the experiment

Figure 27 shows users' choices of 2 gestures they liked and 2 gestures they didn't like when presented prior to experiment (see Appendix B8). It can be noted that the *open palms* gesture was liked by all users who selected it, followed by the *pointing* with 91%. This percentage ranged between 68% and 88% for the remaining positive gestures excluding the *back clenching of the hands*. In spite the fact that this gesture (i.e. *hands clenching-back*) has been regarded as a positive one (refer to Table 12), the obtained results showed that it was the most hated among all gestures. In addition, the gestures of *legs crossed*, *neutral* and *arms folded* were not satisfactory because these gestures were strongly disliked by users. Table 22 shows the χ^2 results for users' selection. Users who selected *arms folded*, *neutral*, *back clenching of hands* and *legs crossed* gestures significantly expressed their antipathy in regard to these gestures. The *front clenching of hands* as well as *chin stroking* were selected to be liked by users; however its χ^2 values were not significant. The remaining gestures were significantly liked by users who selected it.

Body Gestures	χ^2	p-value	Significant
Neutral	9.94	< .05	Yes
Hands clenching-front	1.60	> .05	No
Hands clenching-back	22.15	< .05	Yes
Open palms	20.00	< .05	Yes
Arms folded	9.00	< .05	Yes
Pointing	15.70	< .05	Yes
Hands steepling	9.94	< .05	Yes
Chin stroking	2.58	> .05	No
Legs crossed	17.64	< .05	Yes
Walking	4.26	< .05	Yes

Table 22: Chi-square results of user' selections of body gesture used in the absence of interactive e-learning context (df = 1, CV = 3.84)

Figure 28 revealed that the inclusion of specific body gestures in interactive e-learning interfaces could be attractive for users. In comparison with pre-experimental results shown in Figure 26, users' positive feeling was improved in regard to all positive gestures as well as *neutral* when these gestures have been used by the virtual lecturer interactively. In particular, positive scores for *neutral* and *back clenching of hands* increased by around 17% and 11% respectively. Also, *pointing* was positively rated by all users and 93% to approximately 98% of them found *hands steepling*, *open palms* and *walking* positive. For other gestures in the positive group, users' positive rating was 75% for *front hands clenching* and 68% for *chin stroking*. Concerning negative gestures, participants in the experiment confirmed their evaluation of both *arms folded* and *legs crossed* where the negative score for these gestures increased approximately 10%. More details about users' evaluation of the body gestures being demonstrated by the virtual lecture interactively can be found in Appendix B9. The χ^2 values for users' views about all gestures after being used in the experimental VLBG platform (see Table 23) demonstrate that positive gestures (i. e. *front clenching of hands*, *open palms*, *pointing*, *hands steepling*, *walking*, and *chin stroking*) obtained significant positive results. However, the negative gestures such as *arms folding* and *legs crossing* showed

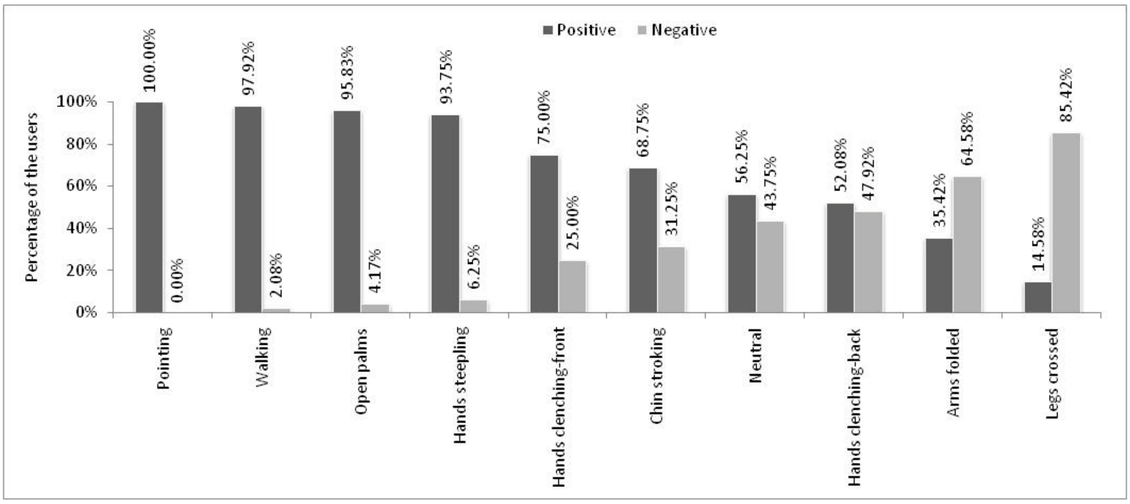


Figure 28: Users' evaluation of body gestures presented in an interactive context after experimenting VLBG platform

significant negative results. Although users largely changed their perceptions of *neutral* and *back clenching of hands*, the results for these gestures did not reach significant levels.

Body Gestures	χ^2 value	p-value	Significant
Neutral	0.75	> .05	No
Hands clenching-front	12.00	< .05	Yes
Hands clenching-back	0.08	> .05	No
Open palms	40.33	< .05	Yes
Arms folded	4.08	< .05	Yes
Pointing	48.00	< .05	Yes
Hands steepling	36.75	< .05	Yes
Chin stroking	6.75	< .05	Yes
Legs crossed	24.08	< .05	Yes
Walking	44.08	< .05	Yes

Table 23: Chi-square values for users' evaluation of body gesture used in the presence of interactive e-learning context

The change of users' perception from pre-experimental phase to post-VLBG conditional phase about each body gesture used in the experiment can be seen in Table 24. However, Table 25 demonstrates the overall mean ratings for each gesture attained by users' responses in both phases. Apart from *back clenching of hands*, positive gestures were all the best rated and negative gestures were the least rated.

Pre - Experimental	Post-VLFE	Neutral	Hands clenching-front	Hands clenching-back	Open palms	Arms folded	Pointing	Hands steepling	Chin stroking	Legs crossed	Walking
+	+	17	23	14	41	17	45	39	24	5	38
+	-	2	7	6	2	4	0	2	7	7	1
-	+	10	13	11	5	0	3	6	9	2	9
-	-	19	5	17	0	27	0	1	8	34	0

Table 24: Change of users' perception from pre-experimental phase to post-VLBG conditional phase about each body gesture used in the experiment

	Pre-Experiment		Post-VLBG	
Body Gestures	Positive rating	Liking	Positive rating	Overall Mean
Open palms	89.58	100.00	95.83	95.14
Pointing	93.75	91.30	100	95.02
Hands steepling	85.42	88.24	93.75	89.14
Walking	81.25	73.68	97.92	84.28
Hands clenching-front	62.50	70.00	75	69.17
Chin stroking	64.58	68.42	68.75	67.25
Neutral	39.58	11.76	56.25	35.86
Hands clenching-back	41.67	3.85	52.08	32.53
Arms folded	43.75	12.50	35.42	30.56
Legs crossed	25.00	8.00	14.58	15.86

Table 25: Overall mean values of users' ratings obtained for each body gesture in both the absence and presence of interactive e-learning context sorted in descending order

4.9.3 Efficiency

Efficiency of each experimental platform was measured using the time taken by users to answer questions related to the learning material when presented by that platform. This measure was considered for all questions and according to question type, recall and recognition. Figure 29 shows the mean answering time for all questions (A) and grouped by question type (B) for each experimental condition. It can be noticed that the answering time was the least when VLBG platform was used. Refer to Appendix B10 for the raw data of question answering time.

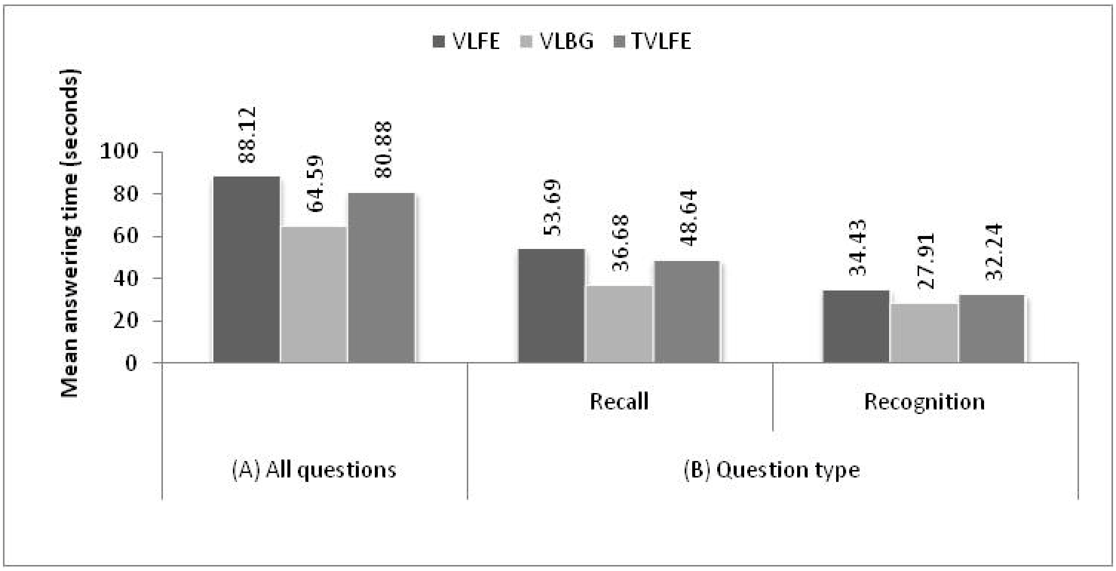


Figure 29: Mean values of time taken by users to answer all questions (A) and grouped by question type (B) in each of experimental condition

4.9.3.1 All Questions

A total of 12 questions were required to be answered by each user and equally distributed over the three lessons. On overall, each lesson was presented 16 times by each platform (refer to Table 13) and so, the time to answer each question was observed 16 times with each platform. The total time taken to answer all the questions was 4229.97 seconds in VLFE compared to 3100.40 seconds in VLBG and 3882.02 seconds in TVLFE. To some extent, the shortest answering time spent by a single user in VLBG (1.70 second) was close to that in TVLFE (1.53 second), and larger in VLFE with 4.06 seconds. On the other hand, the longest time value in VLBG was 62.91 seconds, nearly 19% smaller than that in VLFE (77.79 seconds) and 22% smaller than that in TVLFE (80.15 seconds). It can be observed from Figure 29A that the VLBG was the most efficient platform averaging 64.59 seconds answering time per user, followed by TVLFE 80.88 seconds and VLFE which was found the least efficient platform 88.12 seconds. Statistical calculations performed by Friedman's ANOVA test demonstrated the existence of significant differences in answering time among the three experimental platforms. In other words, the time users spent in answering the required questions was significantly affected by the presentation mode ($\chi^2(2) = 10.04$, CV = 5.99, $p < .05$). However, this finding did not reveal which presentation modes differed from each other. Accordingly, the Wilcoxon signed rank test was applied in the pair-wise comparisons to compare all different combinations of the experimental platforms (VLFE vs. VLBG, VLFE vs. TVLFE and VLBG vs. TVLFE) and significant difference was identified if the absolute value of z was found to be greater than the critical value (CV) [233]. Results of these follow-up comparisons indicated that the time consumed to answer the required questions was significantly lower in the VLBG condition compared to the VLFE ($z = -4.44$, CV = 1.96, $p < .05$) and the TVLFE ($z = -2.02$, CV = 1.96, $p < .05$) conditions. However, no significant difference in question answering time was attained between VLFE and TVLFE ($z = -1.48$, CV = 1.96, $p > .05$).

4.9.3.2 Question Type

Figure 29B shows a breakdown of the mean answering time results according to the question type where the time taken by users to answer recall questions (Q1 and Q2) and recognition questions (Q3 and Q4) was obtained for each of the three experimental conditions. In recall questions, it could be noted that the least time (36.68 seconds) was spent by users when the learning material has been delivered by virtual lecturer with full body animation followed by TVLFE (48.64 seconds) and VLFE (53.69 seconds) platforms respectively. According to Friedman's ANOVA, the difference in answering time across the three conditions was significant ($\chi^2(2) = 10.29$, CV = 5.99, $p < .05$). Results of pair-wise comparisons revealed the existence of a significant decline in time users spent in answering the recall questions from the VLFE to VLBG ($z = -4.08$, CV = 1.96, $p < .05$), and from TVLFE to VLBG ($z = -2.65$, CV = 1.96, $p < .05$). Comparing VLFE and TVLFE, no significant difference was found ($z = -1.40$, CV = 1.96, $p > .05$).

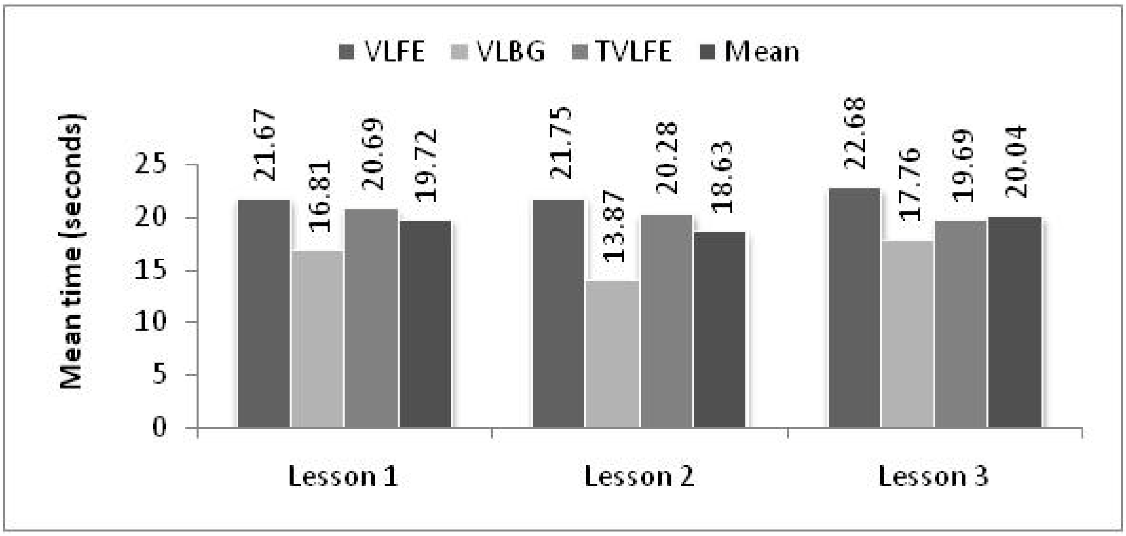


Figure 30: Mean answering time for questions related to each lesson across the experimental e-learning platforms

Similarly, the VLBG scored the minimum time for answering the recognition questions (27.91 seconds) and longer times were observed when facial expressions has been incorporated either with one virtual lecturer (32.24 seconds) or two (34.43 seconds) in the presentation of the learning material. The difference between the three conditions in answering time for this type of questions was smaller than that for the recall questions; however, no significance has been reached ($\chi^2(2) = 1.79$, CV = 5.99, $p > .05$).

4.9.3.3 Each Lesson

Figure 30 shows the time taken by users to answer the questions of each lesson grouped according to the experimental platforms. In each lesson, it can be noted that users were the quickest in providing the answers when virtual lecturer with body gestures was used and the slowest when virtual lecturer with facial expressions was used. Nevertheless, the average answering time was approximately similar at 19.72 seconds for lesson 1, 18.63 seconds for lesson 2 and 20.04 seconds for lesson 3.

Figure 31 illustrates the time results in each platform grouped according to the learning lessons. It can be seen that the most efficient presentation mode was VLBG with an

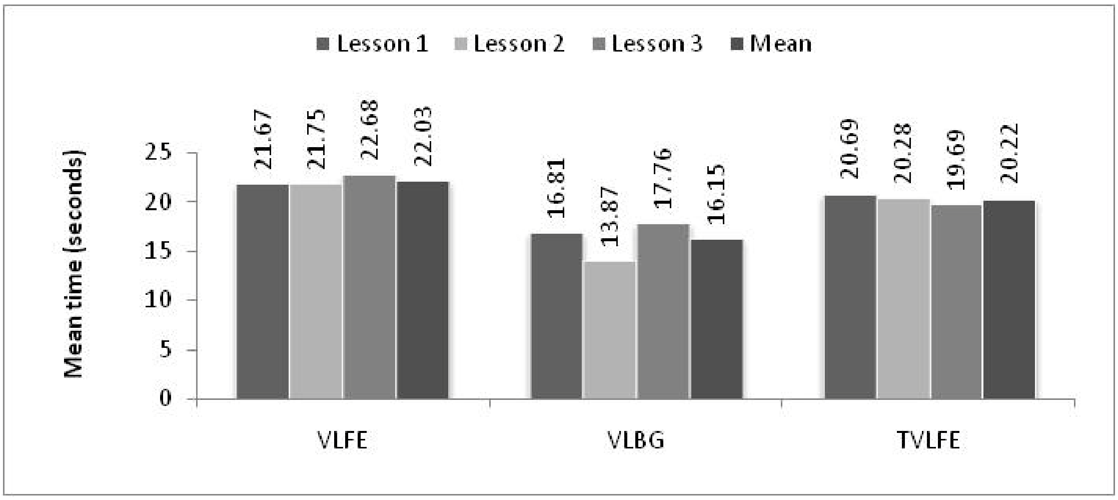


Figure 31: Mean answering time for the experimental platforms across the learning lessons

average answering time of 13.87 seconds for lesson 2, 16.81 seconds for lesson 1, and 17.76 seconds for lesson 3. When these lessons were introduced by the TVLFE platform, users consumed the same answering time across the three lessons averaging 19.69, 20.28, and 20.69 seconds to answer the questions of lesson 3, lesson 2 and lesson 1 respectively. However, the average answering time for all lessons in TVLFE was found larger compared with the VLBG platform. Likewise, the time taken to answer the questions was (more or less) equal in the three lessons in VLFE condition with the highest average time of 22.03 seconds.

4.9.3.4 Each Question

Figure 32 shows the mean time taken to answer each question in each interface. The shortest answering times were found for Q3 related to lesson 2 with 11.44 seconds in VLBG (minimum value), 12.49 seconds in TVLFE and 13.06 seconds in VLFE. On the other hand, the longest answering times were observed for the first question in the first lesson with 22.96 seconds in VLBG, 31.91 seconds in TVLFE, and 35.61 seconds in VLFE (maximum value). It can be noticed also that in 92% of questions (11 out of 12) the shortest time was taken when lessons have been delivered by the VLBG platform.

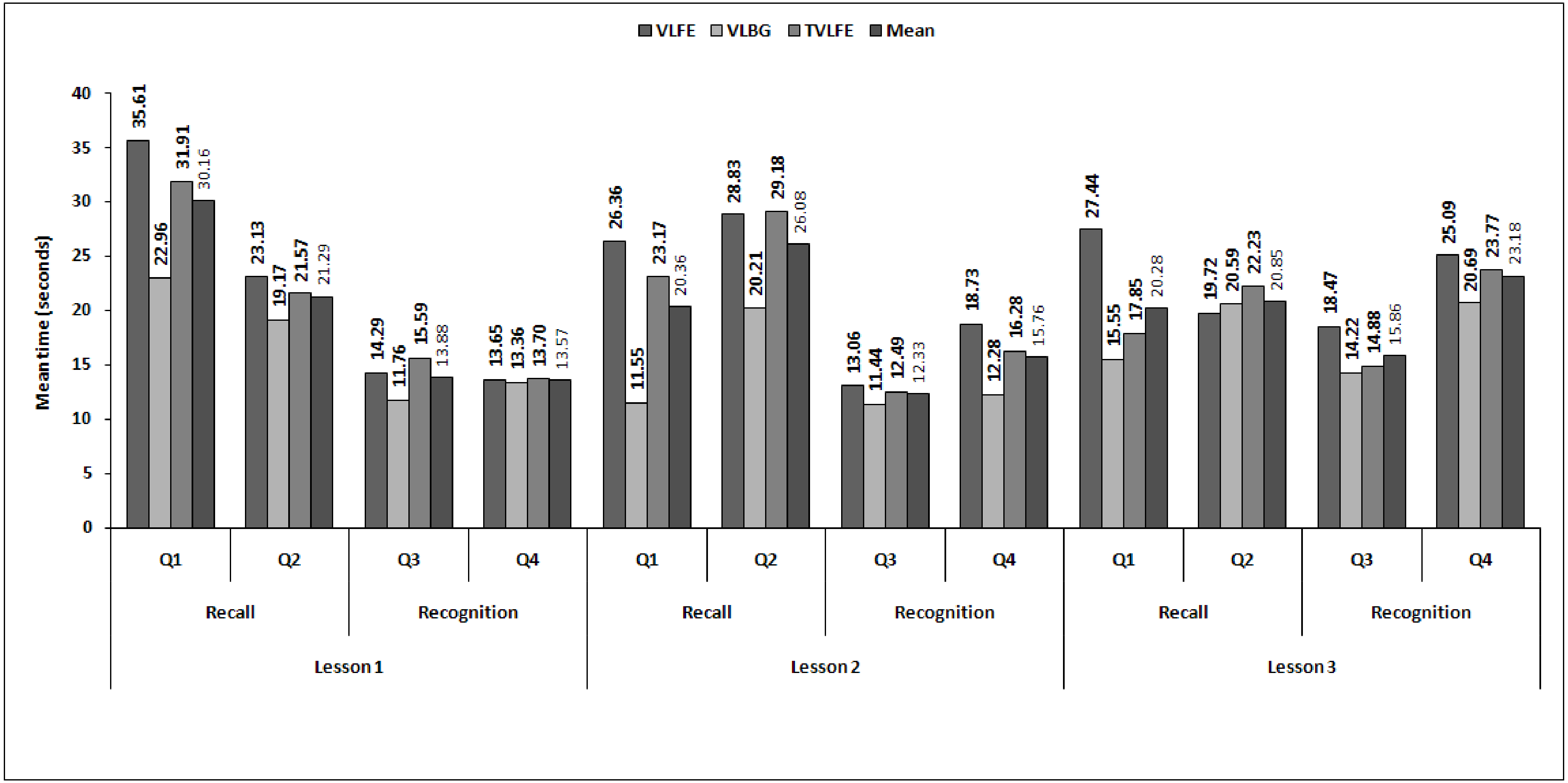


Figure 32: Mean answering time taken for each question in each experimental interface

4.9.3.5 Each User

Figure 33 shows the total time consumed by each user to answer the required questions. In total, the maximum time taken by a single user to answer all questions was 389.38 seconds by user 20, whereas the minimum time was 81.77 seconds by user 37 who also was the quickest user in all conditions. As far as the experimental platforms are concerned, users were more efficient with the use of VLBG platform averaging 64.59 seconds per user as opposed to the use of TVLFE (80.88 seconds) and VLFE (88.12 seconds).

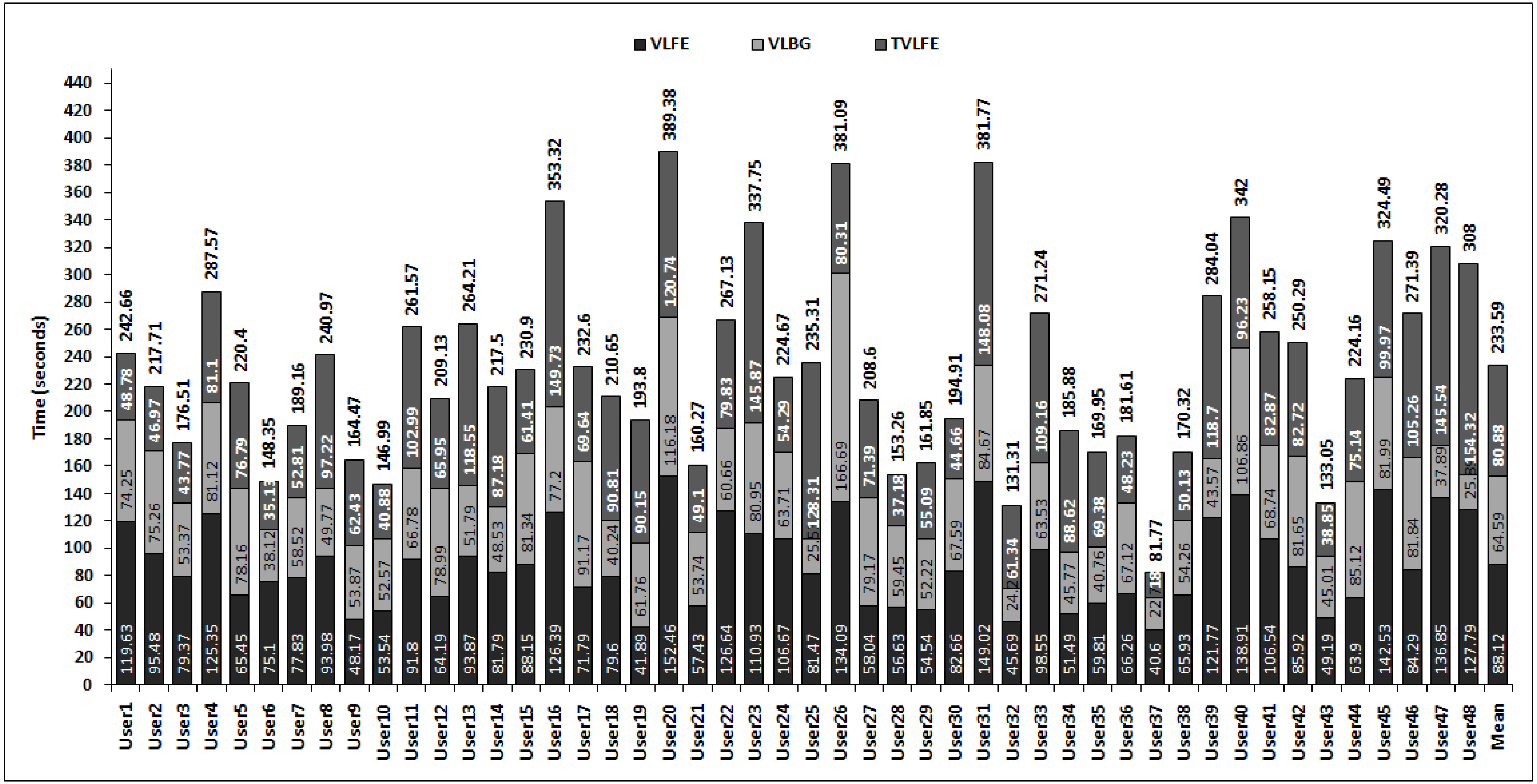


Figure 33: Answering time spent by each user in each experimental interface

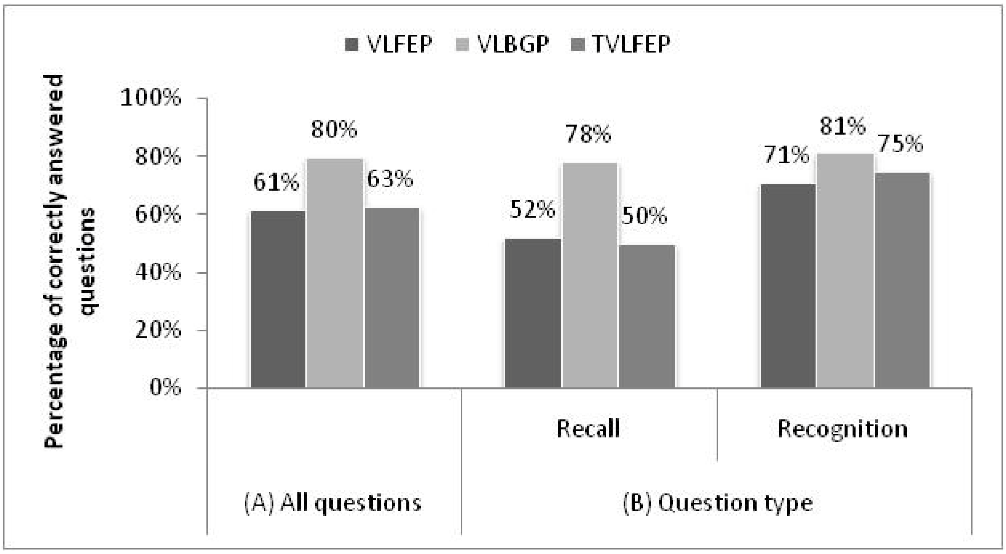


Figure 34: Percentage of correct answers achieved by users for all questions (A) and grouped by question type (B) in each experimental condition

4.9.4 Effectiveness

The measure of effectiveness of the experimental platforms was specified according to the number of correct answers achieved by users when each of these platforms has delivered the learning material. Figure 34 shows the percentage of correct answers for all questions (A) and grouped by question type (B) in each experimental condition. It can be seen that users' performance was the highest with the implementation of the virtual lecturer with full body animation in VLBG. The raw data of the correctness of users' answers can be found in Appendix B11.

4.9.4.1 All Questions

There were 12 questions for each user equally distributed on the experimental platforms at 4 questions related to each lesson. Each platform was used 16 times to introduce each lesson (refer to Table 13), therefore, the maximum number of correct answers that can be accomplished by the users in each experimental condition was 192 (16 * 4 questions per lesson * 3 lessons).

As can be seen in Figure 34A, the VLBG outperformed the other two conditions, VLFE and TVLFE. Using the VLBG, users correctly answered 153 questions out of 192 achieving 80% effectiveness rate compared to 120 (63%) using TVLFE and 118 (61%) using VLFE. Friedman's ANOVA revealed that the difference in users' performance among the experimental conditions was significant ($\chi^2(2) = 17.54$, CV = 5.99, $p < .05$) indicating that the presentation mode significantly influenced users' ability to answer the required questions correctly. More specifically, the results of pair-wise comparisons using Wilcoxon signed rank test (ignore the minus sign of z value [233]) showed that users' achievement using virtual lecturer with body gestures in VLBG platform was significantly better than that using two virtual lecturers with facial expressions in TVLFE platform ($z = -3.61$, CV = 1.96, $p < .05$). The VLBG condition also outperformed the VLFE condition ($z = -3.84$, CV = 1.96, $p < .05$). However, the difference between the VLFE and the TVLFE was not significant ($z = -.45$, CV = 1.96, $p > .05$).

4.9.4.2 Question Type

For each lesson, the four questions were of two types; recall (Q1 and Q2) and recognition (Q3 and Q4). Each of these questions has been asked 16 times with each platform. As a result, the total number of questions in each type was 96 (16 * 2 questions per type * 3 lessons). In recall questions, Figure 34B demonstrates that users' performance was better when using VLBG compared to the other presentation modes. Using VLBG, the total number of users' correct answers to recall questions was 75 giving 78% correctness rate whereas a smaller number of correct answers to the same type of questions was observed when using VLFE, where users correctly answered 50 questions, slightly higher than the half (52%). When the two virtual lecturers shared the delivery of the lessons, users' achievement dropped further to 48 (i.e. 50%) correct answers to recall questions. Based on Friedman's ANOVA calculations, users performed significantly differently amongst the three platforms in regard to recall questions ($\chi^2(2) = 20.17$, CV = 5.99, $p < .05$). When each pair of the experimental conditions has been compared, use of virtual lecturer with body gestures was significantly more effective than the use of the other modes of presentation; virtual lecturer with facial expressions ($z = -3.79$, CV = 1.96, $p < .05$) and two virtual lecturers with facial expressions ($z = -3.91$, CV = 1.96, $p < .05$). However, no significant difference was found between the latter two conditions; VLFE and TVLFE ($z = -.32$, CV = 1.96, $p > .05$).

Although users' performance was better in the recognition questions, the presentation mode did not show overall significant differences among the experimental conditions in answering this type of questions ($\chi^2(2) = 3.05$, CV = 5.99, $p > .05$). Therefore, it can be said the users' performance in answering the recognition questions was not significantly affected by the presentation modes offered in the experimental interfaces. Nevertheless, VLBG scored the highest percentage of users' correct answers (81%) compared to TVLFE (75%) and the VLFE (71%).

4.9.4.3 Each Lesson

A comparison between users' performance in the three lessons is shown in Figure 35. It can be observed that users performed better when each of these lessons were presented by the virtual lecturer with full body gestures (VLBG) where the percentage of correctly answered questions fluctuated between 77% and 83%. In the first and third lessons, this percentage ranged from 63% to 70% when both lessons were presented by the VLFE and TVLFE; however higher percentages were noted for TVLFE in both lesson 1 and lesson 3.

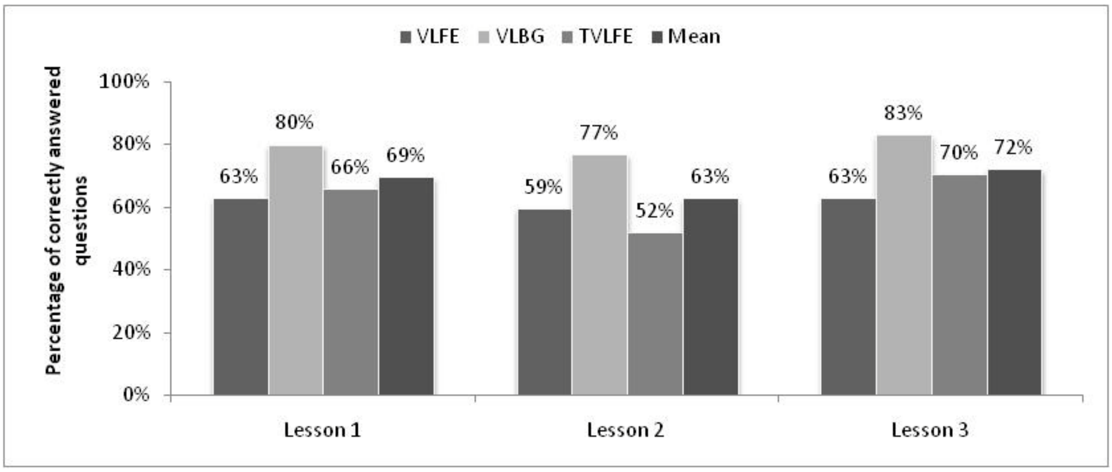


Figure 35: Percentage of correct answers for each lesson across the experimental conditions

In converse, VLFE condition scored 59% correctness percentage in the second lesson outperforming the TVLFE with 52%. Calculating the mean, users achieved 63% correct answers in lesson 2. This performance was higher in lesson 1 with 69%, whilst the highest users' performance was found 72% in lesson 3.

The effectiveness results in each presentation mode according to the communicated lessons are shown in Figure 36. It can be observed that whenever the VLBG platform has been used, the correctness of users' answers was the highest with 77% in lesson 2,

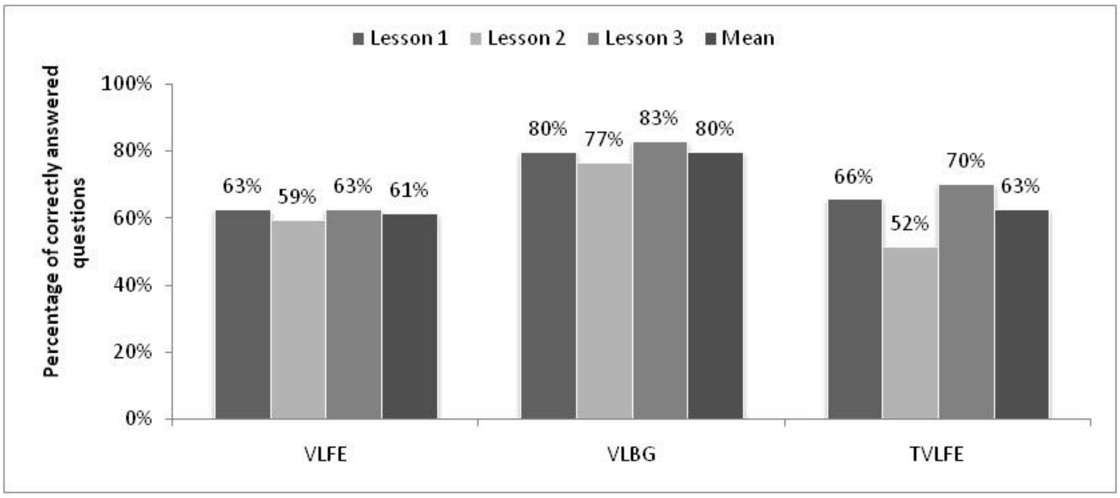


Figure 36: Percentage of correct answers for the experimental conditions across the learning lessons

80% in lesson 1, and 83% in lesson 3 deriving an overall average of 80%. In the second place, users' correct answers was weaker and dropped to 63% on average with the use of two virtual lecturers with facial expressions where the percentages were 52% in lesson 2, 66% in lesson 1 and 70% in lesson 3. However, when these lessons have been presented using single facially expressive virtual lecturer, users' performance was observed the weakest with 59% in lesson 2 and 63% in lessons 1 and 3 attaining approximately similar users' performance with an overall average 61%.

4.9.4.4 Each Question

The percentages of correct answers that users achieved for each question when each platform has been used in the presentation of each lesson are displayed in Figure 37. It can be observed that in 83% (10 out of 12) of the required questions, the highest percentage of correct answers was obtained with the use of VLBG platform while in the remaining questions (Q4 lesson 2 and Q3 lesson3) this presentation mode came in the second place.

On overall, the average percentages for correct answers across the three lessons using VLBG were 83% for Q3, 81% for Q1, 79% for Q4 and 75% for Q2. In regard to the use of the VLFE and TVLFE, the overall percentage dropped dramatically particularly for Q1 and Q2 (i.e. recall questions). The users' achievement in TVLFE was 46% and 54% for the Q2 and Q1 respectively whilst the percentage was 52% for both questions in VLFE. For the recognition questions (i.e. Q3 and Q4), the overall percentage was higher where 69% of users' answers to Q3 were correct in VLFE compared to 79% in TVLFE. Lastly, Q4 achieved relatively close percentages 71% and 73% in TVLFE and VLFE respectively.

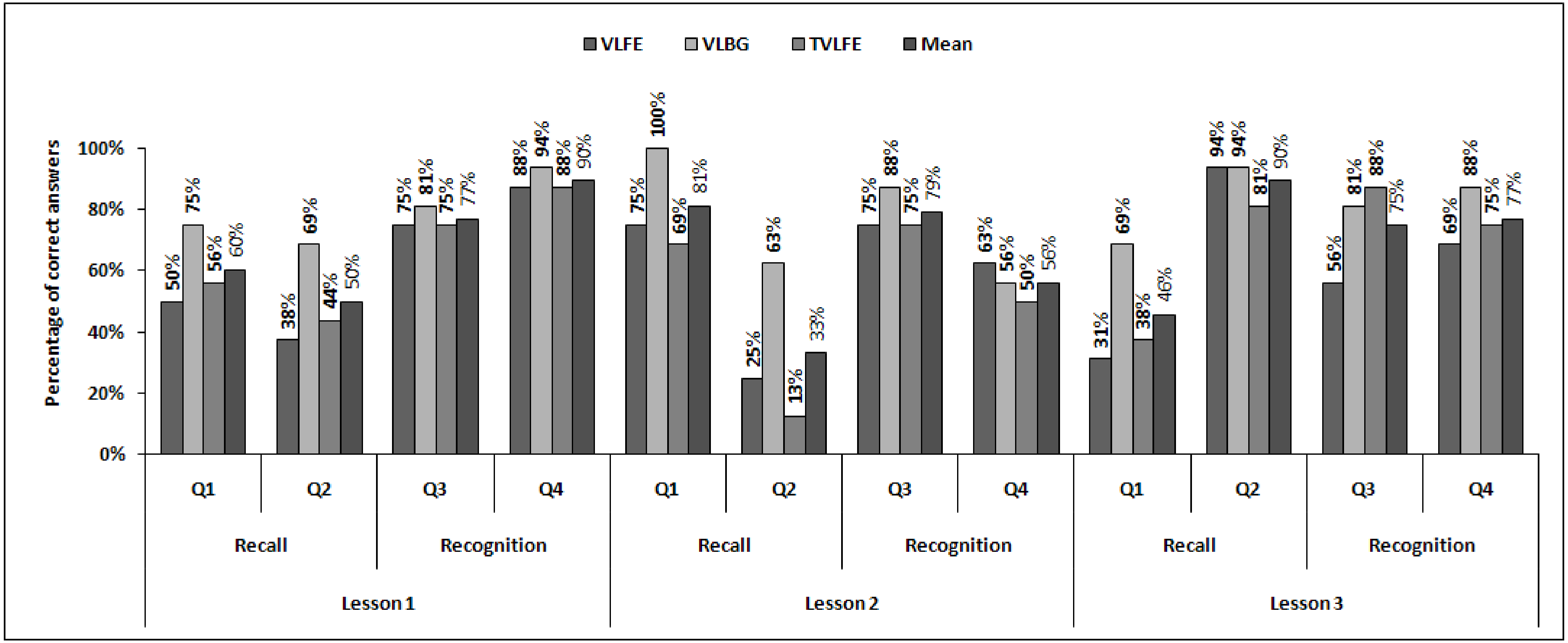


Figure 37: Percentage of correct answers for each question in each experimental condition

4.9.4.5 Each User

Figure 38 shows the number of correct answers achieved by each user. In one hand, only two users (user 9 and user 21) were able to answer all questions correctly. On the other hand, user 46 obtained only 2 correct answers which is the lowest value. Comparing the mean values of users' achievements across the three platforms, the VLBG condition attained the higher average (3.19) of correct answers per user than the TVLFE (2.50) and the VLFE (2.46) conditions.

4.9.5 User Satisfaction

Users' responses to SUS questionnaire (10 statements) was used to measure their attitude after they have had the opportunity to use each experimental platform. Also, users were required to respond to additional eight statements related to interface components and learning experience. Each of the 18 statement was based on a five-point Likert scale where 1 represents strongly disagree and 5 represents strongly agree. SUS scoring method [230] was used for the first ten statements to calculate the satisfaction score for each user in each interface, whereas frequency of users agreement for each statement was used to attain their level of attitude towards different aspects and learning experience of the tested platforms. Findings demonstrated that the VLBG scored the highest satisfaction rate compared to VLFE and TVLFE. The mean SUS score calculated for the VLBG found 85.05 compared to 79.45 and 77.97 for TVLFE and VLFE respectively. Statistical calculations using Friedman's ANOVA showed an overall significance in terms of differences in users' attitudes towards different presentation modes ($\chi^2(2) = 9.59$, CV = 5.99, $p < .05$).

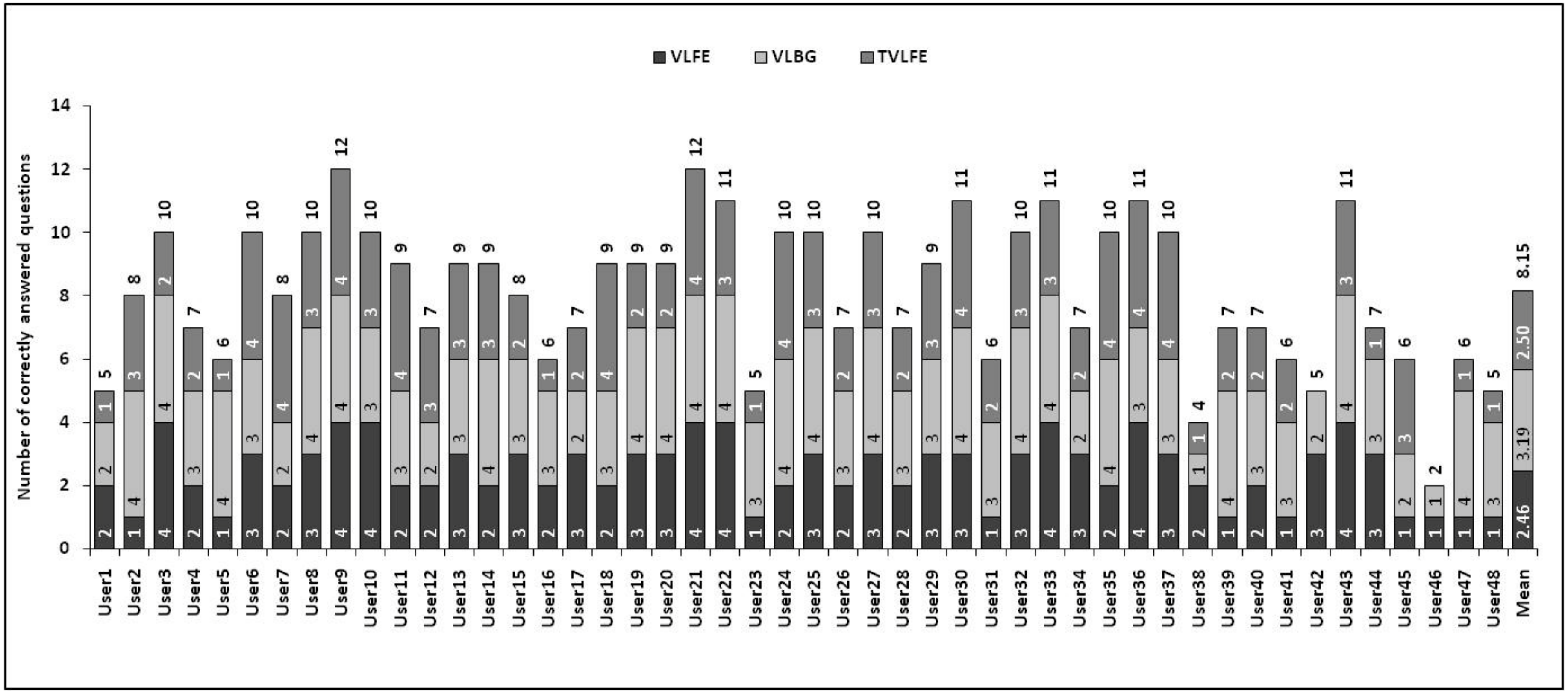


Figure 38: The number of correctly answered questions for each user in each experimental interface

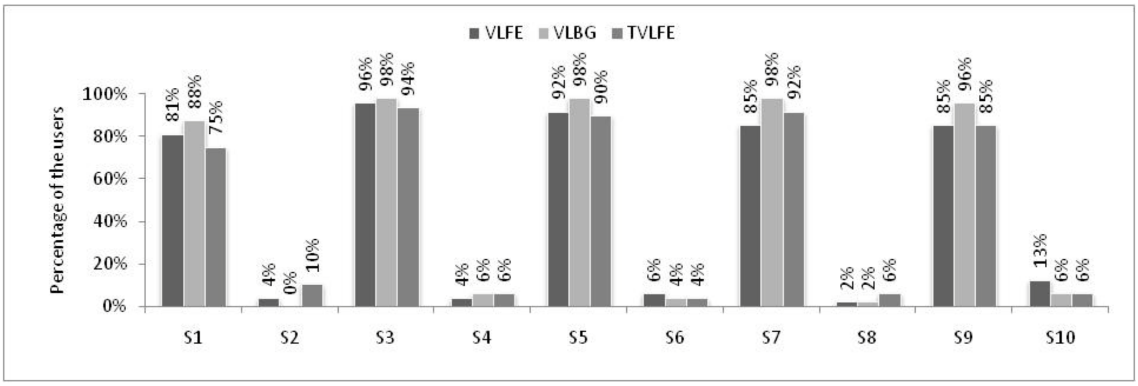


Figure 39: Frequency of users' agreement for each SUS statement in the experimental conditions

Ignoring the minus sign of z value [233], the results of follow up pair-wise comparisons using Wilcoxon signed rank test found the VLBG significantly more satisfactory than TVLFE ($z = -2.61$, $CV = 1.96$, $p < .05$) and VLFE ($z = -3.61$, $CV = 1.96$, $p < .05$). On the other hand, the satisfaction of users did not change significantly between VLFE and TVLFE conditions ($z = -1.34$, $CV = 1.96$, $p > .05$).

Figure 39 shows agreement frequency demonstrated by users for each of the SUS (system usability scale) statements in each experimental condition. As can be seen, users of the experiment expressed higher level of agreement in regard to the positive aspects of the tested interface when they interact with the full body animation of the virtual lecturer (VLBG). In other words, the participants found the VLBG *easier to use* (S3), *well integrated* (S5) and *could be learnt quicker* (S7). Also, they *felt confident* (S9) while using it. Therefore, they express stronger interest to *use* VLBG *frequently* (S1). Contrarily, users show stronger disagreement towards the negative aspects of VLBG platform. More specifically, none of the users agreed that VLBG was *unnecessarily complex* (S2) whereas similar agreement was expressed across the three conditions in terms of the *need for technical support* (S4) and the extent of *inconsistency* in the tested platform (S6). Additionally, only 2% of the users agreed that VLBG and VLFE were *cumbersome to use* (S8) compared to 6% for the TVLFE. The percentage however was slightly higher for S10; *I needed to learn a lot of thing before I could get going with the system*, with 6% for both VLBG and TVLFE and much higher 13% for VLFE. In summary, it can be said that using full body animation of the virtual lecturer in e-learning interfaces is more satisfactory than facial expressions only.

In addition to the system usability scale (SUS) statements, additional statements S11 to S18 were included to obtain users views about their learning experience and interface components as well as the incorporated multimodal features. In other words, these statements investigated users' *excitement and interest about the presented lesson* (S11), whether *the asking and answering feature helped to improve their understanding* (S12), and their l*evel of control over learning* (S13). Also, statements S14 – S16 asked users to rate the role of virtual lecturer's facial expressions (or body gestures) in terms of *increasing* their *attention* and *enjoyment* (S14), *encouraging* them *to keep in* e-*learning with virtual lecturers* (S15), and *easing the process* of following up and understanding the presented lessons (S16). The last two statements were aimed at evaluating an *overall*, users' *satisfaction* (S17) and their *learning experience with the applied interfaces* (S18).

Users' agreement levels for these statements are illustrated in Figure 40. It can be seen that when the virtual lecturer with full body animations has been experienced, users showed a stronger agreement with respect to most of the added statements. In other words, users using the VLBG felt more excited and interested about the presented lessons and the way of asking and answering questions simulated in this platform enhanced their understanding further. Furthermore, animation of virtual lecturer's body made them enjoy more and encouraged them to pay more attention in addition to the ease in pursuing the presented learning material.

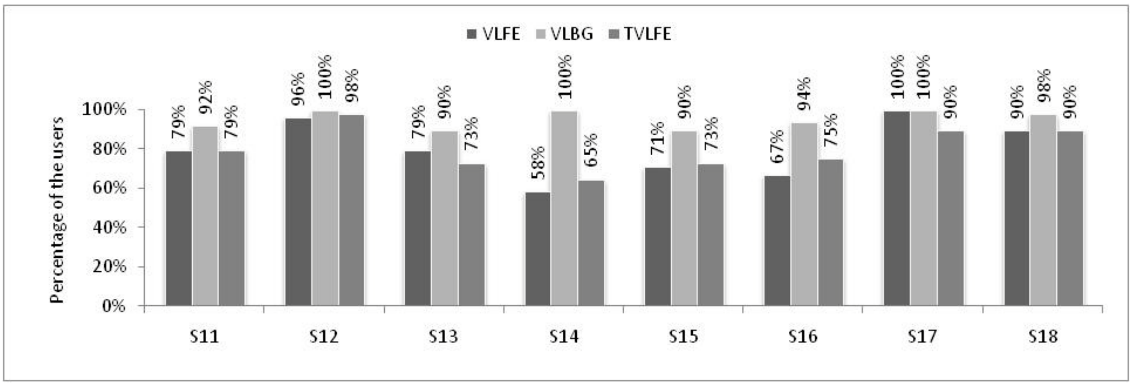


Figure 40: Frequency of users' agreement for the additional 8 satisfaction statements in the experimental conditions

On overall, users were more satisfied and gained more enriching learning experience with the implementation of the VLBG presentation mode. The raw data of users' responses to the satisfaction questionnaire can be found in Appendix B12.

4.9.6 Post-Experimental Users' Views

At the end of the experiment, users were required to rate the usefulness of each of the multimodal metaphors used in the experimental e-learning platforms on a 5-point Likert scale with 1, the value of least useful and 5, the value of most useful. Also, they had to indicate one of these platforms they mostly preferred. Finally, they were asked whether they have any comments or suggestions.

According to users' views shown in Figure 41, incorporating two virtual lecturers with facial expressions (TVLFE) was found to be more impressive than using one (VLFE) as the observed most usefulness rate was about 6% for VLFE and 25% for TVLFE. In comparison, employing body gestures by facially expressive virtual lecturer were found to be the most useful for users where slightly more than half of them (52%) consider their learning to be substantially assisted by this mode of presentation. In addition to speech output, the textual brief notes and graphical illustrations displayed in the PowerPoint presentation part have been used in the same manner in the experimental

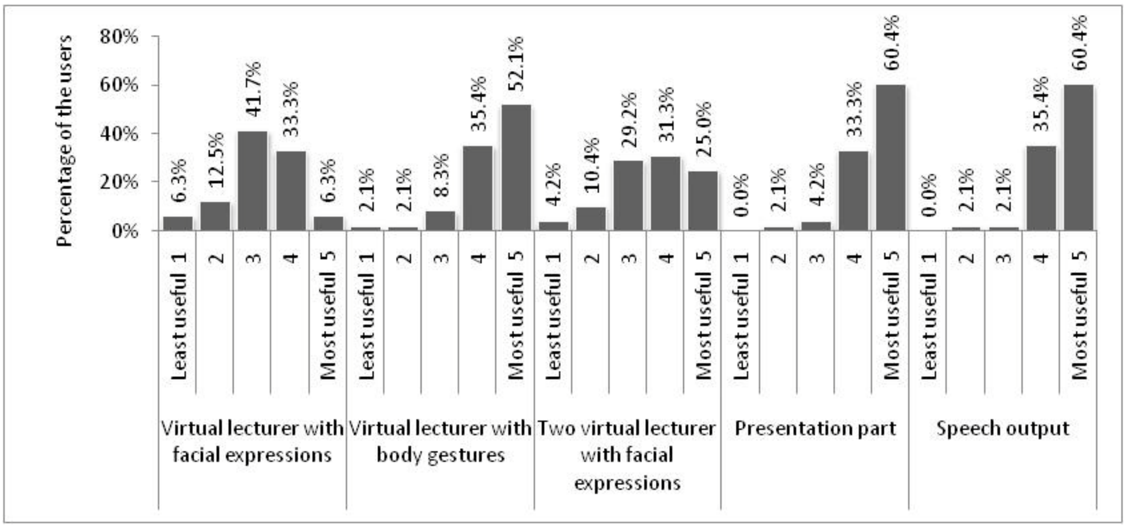


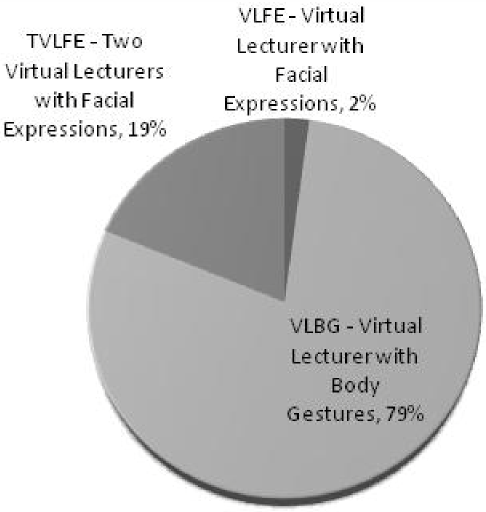
Figure 41: Users' ratings for usefulness of multimodal metaphors used in the experimental platforms

platforms and seem to contribute beneficially in users' learning as both modalities attained 60% most usefulness rate. These results were found to be consistent with users' views obtained prior to the experiment (refer to Figure 22).

As can be seen in Figure 42, the most preferred e-learning platform was VLBG obtaining 79% preference rate among users. This percentage dropped dramatically to 19% for TVLFE and 2% for VLFE. The raw data of users' responses to the post-experimental part of the questionnaire can be found in Appendix B13.

In regard to users' comments and suggestions, it could be summarized in the following points:

1. Allowing users to attend the presented lesson more than once could provide more flexibility. This functionality was neglected in the design of the experimental interfaces as it will bias the results where one user could attend the presented learning material many times while another may prefer to attend it once.
2. Adding more body gestures and facial expressions as it may result in a more interesting and attractive virtual lecturer.



	VLBG	TVLFE	VLFE
Efficiency (Time)	65	81	88
Effectiveness (Correctness)	80%	63%	79%
User Satisfaction	85%	79%	78%

Figure 42: Percentage of users who preferred each of the experimental platforms

3. Directing the questions to the virtual lecturer by voice.
4. Using two lecturers in the TVLFE presentation is distracting and could scatter user’s concentration.
5. Drawing user’s attention to the most important parts of the presentation by means such as flashing the textual notes or adding accompanying sounds.
6. Incorporating additional female lecturer in the VLBG platform with full body animation might attract users more.

4.10 Discussion

This chapter explored users’ views of point towards a set of facial expressions and body gestures when being used by virtual lecturers in both the presence and absence of interactive e-learning context. In total, 6 facial expressions and 10 body gestures have been evaluated.

The results demonstrated in Figure 23 showed that facial expressions which have been regarded as positive ones such as happy, smiling, interested and amazed were the best rated expressions when presented as static pictures without any interactive context. The percentage of positive ratings for these expressions ranged between about 69% for the amazed to 85% for the smiling. On the other hand, the remaining expressions were found to be the most negatively viewed by the users with about 63% for neutral and 40% for thinking. Chi-square test results shown in Table 15 confirmed what has been hypothesized in H1 where only the positive expressions attained significant levels of users acceptance. When users were given the opportunity to select their most desirable two expressions and another two they did not like, positive expressions were selected to be the most liked expressions (refer to Figure 24). Nevertheless, significant results were achieved only for two of them; happy and smiling, as can be seen in Table 16. This could be attributed to the fact that only two expressions were allowed to be selected as liked and users were not able to select more expressions even though they preferred it such as interested and amazed giving that the remaining expressions were significantly hated by users. These findings provide additional support to H1.

When the tested expressions have been used interactively by virtual lectures in both VLFE and TVLFE platforms, a substantial change in users' evaluation for these expressions was observed in favour of positive views (see Figure 25). Integrating these expressions in an interactive e-learning context impressed the users and enabled them to evaluate it in a more realistic situation. Therefore, they appreciated the role that could play in e-learning interfaces. This has been evidenced by users' views which became significantly positive with respect to all of the expressions except the neutral (see Table 17), although this expression scored the highest shift in users' views. Consequently, the second hypothesis of the experiment has been accepted.

Concerning the third hypothesis (H3), the obtained results shown in Figure 26 exhibited that positively regarded gestures were positively perceived by users when introduced in the absence of interactive context. Table 21 showed that these gestures reached a significant level of users' positive views except for the front and back clenching of hands. The *back clenching of hands* attained an unexpected result as this gesture was considered as positive; however about 58% of users (28 out of 48) had a negative impression about it, yet did not reach a significant level. This was explained later where 25 users out of 26 (96%) selected it as disliked. In addition to *back clenching of hands*, the most negatively rated gestures were *neutral*, *arms folded* and *legs crossed*; however a significant result was noted only for the last one. As shown in Table 22, these four gestures were significantly hated by users and this could clarify why users perceived it negatively. In contrary, most of positive gestures were significantly preferred by users. On overall, H3 has been supported by these results.

As predicted in the 4th hypothesis, Figure 28 shows a higher positive rating for each positive gesture when it has been used by the virtual lecturer in interactive e-learning context. In particular, users' thought about the *back clenching of hands* were changed but not significantly. However, the remaining positive gestures scored significant positive ratings (refer to Table 23). Users felt that these body animations could attract them, enhance their concentration about the communicated learning material, and put them in a scenario similar to what usually happen in the actual class-based situation. Therefore, the obtained results expressed users' belief that these body animations have the potential to benefit e-learning interfaces. On the other hand, users' opinions indicated that some body gestures such as *crossing the legs* and *folding the arms* should be avoided by the lecturers as these gestures were significantly negatively perceived. These results confirmed the classification of these body gestures into positive and negative categories [181] and were found to be consistent with the findings of other related studies [107, 182]. On the whole, the experimental results as demonstrated in

Table 20 and Table 25 suggests that specific facial expressions and body gestures could be more appealing and attractive for learners whereas some other expressions and gestures could be not satisfactory for them.

This chapter also investigated three different modes of employing avatars as virtual lecturers in the presentation of the learning material. Namely, these modes were facially expressive virtual lecturer (VLFE), virtual lecturer with full body gestures (VLBG), and tow virtual lecturers with facial expressions (TVLFE). The obtained results were used to compare these experimental conditions in terms of efficiency, effectiveness and user satisfaction.

The difference among the experimental platforms; VLFE, VLBG and TVLFE with respect to each of these usability parameter has been predicted in the hypotheses H5, H6, and H9 respectively. In answering the required learning questions, the participants of the experiment spent different times using different presentation modes offered by the experimental interfaces. Also, the number of correctly answered questions and the satisfaction of users differed across these interfaces. This difference was found to be significant by Friedman's ANOVA calculations. As a result of multiple comparisons, the VLBG was found to be the most efficient and most effective as well as the most satisfactory presentation mode. The way used in VLBG platform to present the learning material enabled the users to be engaged in learning environment similar to real lecturer-to-learner human interaction which takes place in traditional class rooms. In addition, when the full body of speaking virtual lecturer was animated, users were more attracted, more excited and interested about the presentation. Furthermore, presenting the learning material in the background of the virtual lecturer within the same interface component in VLBG platform helped users to watch both at the same time. Consequently, they were more attentive and better concentrated to what is being presented. Also, they kept involved in cognitive processing of the delivered learning information and got better understanding of it. This situation enabled the users to preserve the communicated information and accordingly, the time they spent in responding to the required questions was significantly shortened in comparison with using facially expressive talking head of virtual lecturer either in VLFE or TVLFE. Additionally, VLBG significantly outperformed the other two experimental conditions in terms of correct answers and user satisfaction. Moreover, the experimental results shown in Figure 41 and Figure 42 demonstrated how users rated the VLBG to be the most useful and preferred interface respectively. Hence, supporting what has been hypothesised in H5 and H6.

The comparison between VLFE and TVLFE platforms revealed that usability levels were equivalent and no significant differences have been noted at all. Even though, using two facially expressive virtual lecturers performed better than employing only one and their usefulness was rated higher as shown in Figure 41. During the experiment, it was observed that users' concentration has been spread out with the use of TVLFE platform. It seems that the existence of two facially expressive virtual lecturers and additional two virtual students in different interface components distracted the users away from the delivered information and split their attention as mentioned by some of the users in the post experimental short interview. However, incorporating talking head of single virtual lecturer with facial expressions in an interface component different from that used to present the textual notes and graphics (as applied in VLFE) did not attract users as much as the VLBG.

According to Figure 29B and Figure 34B, efficiency and effectiveness of the experimental platforms were varied in regard to the question type (recall and recognition). There was a difference across these conditions in the time users spent in answering both recall and recognition questions. Similar difference was also noticed in the number of correct answers to both types of questions. These differences were found to be significant for the recall questions only and not for the recognition ones.

On overall, it was observed that the users consumed less time in answering the recognition question than the time they needed in responding to the recall questions (refer to Figure 29). In recall questions, users may have taken more time in trying to retrieve the required information whereas in the recognition tasks, users were required to select the answer among the provided set of options which may contribute in shortening the time they needed to answer. Also, the percentage of correctly answered recognition questions was noted larger in comparison with the recall questions as shown in Figure 34. In recognition questions, users had to successfully choose the correct answer from the given alternatives and this could be done by chance which make it easier for them to answer. On the other hand, in recall questions, no answering options were provided and this might make it more difficult to answer. In this case, users had to rely only on their memory to find out the correct answer which is far to occur due to chance.

Nevertheless, the statistical calculations showed the existence of significant difference across the experimental conditions in recall questions result in favour of the VLBG platform either in terms of answering time or correctness of answers, hence accepting H7. However, no significant difference has been revealed in the recognition questions result regarding both measures; time and correctness, therefore rejecting H8. These results demonstrated the effect of including virtual lecturer with body animations, as applied in the VLBG platform, on users' performance in responding to recall activities faster with higher accuracy. At the same time, this effect did not extend to users' performance in recognition questions. Comparing the results of both VLFE and

TVLFE, no significant difference has been achieved in both types of questions indicating equal effect of including one or two facially expressive virtual lecturers either on efficiency or effectiveness.

With respect to the last experimental hypothesis H9, users' attitude towards each experimental platform was found significantly different confirming what has been hypothesized. In accordance with the post hoc statistical tests applied on the SUS satisfaction questionnaire, the VLBG interface was significantly more satisfactory to users comparable to the VLFE and the TVLFE interfaces. Also, the satisfaction results shown in Figure 40, offered additional support to H9. The design of this experiment involved recruiting one group of users to evaluate all the experimental conditions. In other words, each user had the opportunity to interact with each of the tested experimental platform. Users were pleased and satisfied with different aspects of the VLBG interface as well as to the learning experience they gained using this interface. When the talking head of facially expressive virtual lecturer has been used, users expressed similar levels of satisfaction with both interfaces; VLFE and TVLFE.

4.11 Summary

This chapter documented the experimental work conducted to innovatively explore users' opinions in regards to a specific set of facial expressions and body gestures when used in the absence and the presence of interactive e-learning context. Also, it investigated in a novel approach the usability aspects and learning performance of e-learning interfaces that employed avatars as virtual lecturers through three deferent e-learning interfaces in the presentation of learning information. The first interface incorporated talking head of single facially expressive avatar while the second interface made use of full body animated avatar. In the third interface, talking heads of two facially expressive avatars; male and female, were included to share the presentation. The assessed usability measures included efficiency (in terms of task completion time), effectiveness (in terms of tasks correctly completed) and user satisfaction.

The obtained results demonstrated that facial expressions and body gesture usually considered as positive were also regarded positively by the users. These finding suggests the adoption of these expressions and gestures in the design of avatars in order to play a pleasant and attractive role as virtual lecturers. Also, the results of this experiment provided empirical evidence that using full body animation of speaking virtual lecturer combined with the learning material in the same interface constituent is more efficient, more effective and more satisfactory as opposed to the other two investigated e-learning interfaces. Using talking head of facially expressive avatar as virtual lecturer was shown to be as efficient, effective and satisfactory as the use of two talking heads of facially expressive virtual lecturers. In regard to specific types of the experimental tasks (i.e. recall and recognition), the multimodal audio-visual presentation of the learning material as applied in the VLBG platform, contributed particularly in memory recall activities much more than the recognition ones. However, the results invoke additional questions such as: would the addition of non-speech sound such earcons and auditory icons to the VLBG interface enhance the usability and learning further? In sum, this study recommends some empirically derived guidelines for incorporating an expressive and full body animated avatars in e-learning interfaces. The description and discussion of these guidelines are introduced in chapter 6.

Chapter 5

Experimental Phase III: The Role of Non-Speech Auditory Technologies

5.1 Introduction

Chapter 4 demonstrated how the use of full body animated virtual lecturer outperformed the use of facially expressive ones in terms of enhancing the usability of e-learning interfaces and improving users' learning performance. However, the role of body gestures was found to be limited to a specific type of learning activities; recall, and did not extend to recognition questions. This chapter investigates whether the addition of non-speech sounds such as earcons and auditory icons could contribute in supporting the influence of avatars' body gestures and strengthen it to comprehend both types of learning evaluation questions; recall and recognition particularly in complex activities. This investigation could help in revealing the role that non-speech auditory technology could play in multimodal e-learning interfaces.

5.2 Aims

This experiment was aimed at examining the usability (in terms of effectiveness, memorability and user satisfaction) of e-learning interfaces that incorporate the use of non-speech sounds along with full body animated avatars in the presentation of the learning material. More specifically, it aimed at examining the effectiveness of earcons and auditory icons in delivering supportive auditory messages related to the learning material presented by full-body animated avatar. It is also aimed at assessing the memorability of these non-speech sounds in terms of users' remembrance of its meaning and use. Furthermore, this experiment targeted measuring the user' satisfaction in relation to the applied e-learning interface. Additionally, this experiment is aimed at evaluating the users' performance in responding to the required experimental learning activities.

5.3 Objectives

In order to accomplish the aforementioned aims, the following objectives were needed to be achieved:

1. Implementation of an experimental e-learning platform that employs avatars with full body gestures in a similar way to that applied in the previous experiment (refer to section 4.5.3) but with the addition of earcons and auditory icons as non-speech auditory messages to communicate specific features of presented learning material. This platform has been referred to as Auditory-enhanced Virtual Lecturer with Body Gestures Platform (AVLBG).
2. Empirical evaluation of the AVLBG by one group of users.
3. Measuring the effectiveness (as well as user' learning performance) by calculating the percentage of questions successfully answered by users in order to measure the effectiveness of the tested e-learning platform, and users' learning performance.
4. Measuring the memorability of tested non-speech metaphors by users' ability to remember its meaning and use.
5. Measuring the satisfaction of users by their responses to questionnaire dedicated to assess users' attitudes in relation to the applied e-learning platform.

5.4 Hypotheses

It was assumed that the addition of earcons and auditory icons in AVLBG would influence the usability level and users' learning achievement of the AVLBG e-learning platform. Based on this assumption, the following hypotheses were derived:

H1: The addition of earcons and auditory icons will result in enhancing the effectiveness of the AVLBG in terms of tasks correctly completed of both types recall and recognition.

H2: Users of the AVLBG will express positive views towards the use of earcons and auditory icons in terms of annoyance, frustration, helpfulness and concentration.

H3: Users of the AVLBG will successfully remember the key features of learning material when communicated by earcons and auditory icons.

H4: Users of the AVLBG will correctly recognise the non-speech sounds used to communicate the key features of the presented learning material.

H5: On overall, users will be satisfied with the AVLBG.

5.5 Experimental platform

The VLBG platform used in previous experimental work demonstrated better performance compared to VLFE and TVLFE regarding both usability and users' achievement levels. This was noticeable particularly in recall questions. However, VLBG was found to be as usable as VLFE and TVLFE with respect to both efficiency and effectiveness in recognition questions. These experimental outcomes established the need for further enhancements in the VLBG platform to investigate if the addition of non-speech auditory stimuli could enhance users' performance in recognition as well as recall tasks. Previous experimental studies showed the potential of earcons in improving the usability [120, 124, 129, 139] and learning performance [9, 11, 13]. Also, auditory icons as environmental sounds were successfully used to communicate information in user interfaces [150, 152, 154]. Therefore, the experimental e-learning platform (AVLBG) recruited to carry out this investigation replicated and extended the VLBG by involving earcons and auditory icons to capture the users' attention towards the most important parts of the learning information when delivered by full-body animated virtual lecturer. In other words, the use of these non-speech metaphors (i.e. earcons and auditory icons) was the only feature that distinguished AVLBG from VLBG. Table 26 shows how earcons and auditory icons were used in the AVLBG platform to capture users' attention towards the key parts of the learning content while being communicated by the facially expressive full body animated virtual lecturer. It can be seen that these parts were grouped into 6 different types which are: the beginning and end of an important statement, the beginning and end of an important definition, and the importance level of a specific keyword. Therefore, three types of multimodal interaction metaphors were incorporated in this platform: visual-only metaphors (text and graphics), audio-visual metaphors (speaking avatar with full body gestures) and auditory ones (earcons and auditory icons).

	Statement		Definition		Keyword importance		
	Start	End	Start	End	High	Medium	Low
Earcons					√	√	√
Auditory icons	√	√	√	√			

Table 26: Mapping between key parts of learning material and non-speech sound used in AVLBG

5.5.1 Learning Material

One lesson about class diagram notation was used in this experiment. The learning lessons used in the second experiment were dependent on each other. Therefore, the first lesson (classes and objects in Appendix B2) was selected to be delivered by the AVLBG. The knowledge contained in this lesson presented introductory information about class diagrams and its notations. It also explained in detail what is meant by class and object, and how to differentiate between them.

5.5.2 Implementation of Non-speech Auditory Metaphors

Earcons used in this experiment were utilised to communicate the importance of specific keyword in the lesson when spoken by the virtual lecturer. These keywords were grouped in three categories in terms of its importance; high, medium and low. Each of these levels was represented by a rank as follows: 1 for low, 2 for medium and 3 for high. Due to the potential of utilising earcons to communicate information of numerical nature, these musical metaphors have been used in this experiment and three different single-meaning earcons were designed, each of which was dedicated to communicate, in simple and meaningful format, single importance level at a time. The design of these musical stimuli was based on the guidelines for the creation of earcons [118, 134] where the voice of drum instrument was selected to play a different number of notes to communicate the required auditory messages. The structure of these earcons is shown in Table 27. It can be seen that the first earcon was composed of only one note to communicate low importance whereas the second earcon consisted of two rising notes to indicate medium importance.

		Timbre	Rhythm	Notes	Duration (seconds)
Importance of keyword					
	Low	Drum	Single note	1	0.22
	Medium		Sequenced notes	2	0.44
	High		Rising pitch notes	3	0.65

Table 27: Structure of earcons used in AVLBG to communicate importance level of keywords in the presented learning content

		Sound	Frequency	Duration (seconds)
Important statement				
	Start	Door opening	1	1.42
	End	Door closing	1	0.55
Important definition				
	Start	Bottle opening	1	0.60
	End	Can dropping	1	0.57

Table 28: Auditory icons used in AVLBG to communicate the start and the end of important statements and definitions in the presented learning content

However, in the third earcon, a sequence of three rising pitch notes was used to communicate high importance. This way, the rhythm of these earcons could be differentiated [118]. Also, these earcons were short and simple to facilitate the interpretation of the delivered auditory message [134]. This experiment also utilised auditory icons (or environmental sounds) [150, 152] to communicate other key aspects of the presented learning content such as start and end of important statement or definition were the representation of these aspects by auditory icons could provide natural mapping to help the users to remember and interpret it accurately. As shown in Table 28, the sound of "opening a door" communicated that an important statement will start, and "closing a door" sound communicated that this statement had finished. Also, the sound of "opening a bottle" was used to indicate that an important definition is going to start whereas the end of that definition was communicated by the sound of "dropping a can". These sounds were selected due to the potential of establishing a natural mapping with the communicated information. Both earcons and auditory icons were played during the presentation in pause intervals so that it does not interfere with the speech of the virtual lecturer. Appendix D3 provides more technical details related to the AVLBG platform.

5.6 Experimental Design

Usability and users' learning performance of VLBG e-learning platform was tested in the second experiment reported in Chapter 4. Therefore, only one group of users was involved in this experiment to evaluate the addition of non-speech auditory sound in AVLBG. Although different tasks were designed in this empirical investigation, it was believed that the obtained results could serve as a comparison point to explore if the addition of earcons and auditory icons in AVLBG resulted in enhancing the VLBG in terms of usability and users' learning performance. In total, 24 users participated in the experiment individually.

5.6.1 Procedure

Throughout the experiment, the same procedure was followed with each user. At the beginning of the experiment, each user was requested to read the introductory message of the questionnaire and to provide personal data in relation to age, gender, educational level and course. In addition, each user had to tell about prior experience in Computers, Internet, class diagram notation and e-learning. Then, a brief demonstration video (53 seconds) about the tested platform was presented. In the following, a short training for 90 seconds was provided in which each user had the opportunity to listen to the implemented non-speech sounds. The aim of this training was to insure users' ability to understand and interpret each of these sounds. Upon completion of the training period, each user had to express views regarding the use of these sounds in e-learning interfaces in the absence of any interactive context. Thereafter, the learning lesson about class diagram notation was presented using the experimental e-learning platform AVLBG. Afterwards, the user was instructed to perform the required tasks. Subsequently, the last part of the experiment was devoted to obtain user's opinion in regard to the implemented non-speech sounds as well as to provide any comments or suggestions.

5.6.2 Tasks

The required tasks were grouped into three categories; learning performance (or effectiveness) tasks, memorability tasks and satisfaction tasks. In the first group, each user had to answer four questions related to the presented learning content in order to measure the effectiveness of the interface as well as the learning gained by users from presented material. These questions were of two types; recall and recognition with two questions each. The second group consisted of two tasks and was aimed to test users' ability to remember the key features communicated by earcons and auditory icons, and to identify the sound used to communicate each of these features. In the first task in this group, two paragraphs of the presented learning material were shown to the user and he/she had to highlight where the non-speech sounds were played to indicate the required key features.

However, in the second memorability task, three sounds were played and the user had to recognise which one has been used to communicate each of the given important parts of the lesson. Finally, the last task was aimed to obtain the users' attitude towards the tested platform by responding to the satisfaction questionnaire consisting of 15 statements on a 5-point Likert scale. The first 10 statements were adopted from SUS questionnaire while the other 5 statements were related to learning experience with the tested platform. Table 29 summarises all the required tasks. More details about all the required tasks can be found in Appendix C1.

Table 30 shows how the implemented non-speech sounds were utilised to indicate the key information needed to answer the required 4 questions correctly. Earcons and auditory icons were not used to communicate the content itself but to inform the user that important parts of this content is about to be communicated by the virtual lecturer.

Task category	Task description
Learning performance tasks (effectiveness)	Answer 2 recall questions related to the communicated learning content
	Answer 2 recognition questions related to the communicated learning content
Memorability tasks	Highlight in the given text where the non-speech sounds were used to indicate (a) start of statement (b) end of statement, and (c) high important key word Highlight in the given text where the non-speech sounds were used to indicate (a) start of definition (b) end of definition (c) low important key word and (d) medium important key words
	Identify which of the played sounds has been used to communicate each of the key features in the presented lesson
Satisfaction tasks	Respond to satisfaction questionnaire

Table 29: Summary of the required tasks in the third experiment

	Recall questions		Recall questions	
	Q1	Q2	Q3	Q4
Earcons	√			√
Auditory icons		√	√	
Communicated answer related to	High important key word	Definition	Statement	Medium important key word

Table 30: Non-speech sounds used to indicate the key information needed to answer the learning tasks correctly

It can be noticed that the first and fourth questions were related to information about the key words of high and medium importance respectively. However, the remaining two questions were related to important definition (Q2) and statement (Q3) indicated during the interaction with AVLBG by auditory icons.

5.6.3 Variables

The variables considered in this experiment were independent, dependent and controlled. The controlled variables were similar to those mentioned earlier in Section 3.6.3.3.

Variable code	Variable	Measure
DV 1	Correctness of users’ answers	Effectiveness and user’s learning performance
DV 2	Correct identification of the key features in the learning content	Memorability
DV 3	Correct recognition of earcons and auditory icons	Memorability
DV 4	User Satisfaction	Satisfaction

Table 31: Dependent variables used in the third experiment

5.6.3.1 Independent Variables

IV 1: Presentation mode: this experiment investigated the effect of non-speech metaphors (i.e. earcons and auditory icons) when incorporated with full-body animated virtual lecturer in the interface of e-learning platform.

IV 2: Task type: two different types of questions, recall and recognition, were used to evaluate the users’ learning achievement attained from the knowledge presented by the tested e-learning platform.

5.6.3.2 Dependent Variables

Four dependent variables were considered in this experiment and briefed in Table 31.

DV 1: Correctness of users’ answers to the evaluation questions: measured by calculating the number and percentage of correctly answered questions. In recall questions, the accuracy of the answer was assessed partially or totally. In recognition questions, the correct option must be selected to indicate the successful answer.

DV 2: Identification of key features in the presented learning material: measured by the number and percentage of users who correctly highlighted these features after being communicated by non-speech sounds.

DV 3: Users' recognition of earcons and auditory icons used in the experiment: measured by the number and percentage of users successfully recognised these non-speech sounds after being used in the experimental e-learning platform.

DV 4: User satisfaction: measured by users' responses to satisfaction questionnaire.

5.7 Data Collection

Two main resources were utilised in collecting the obtained data; observations and questionnaires. Users' responses to the pre-experimental part of the questionnaire helped in gathering the data in relation to the individual characteristics of the participants in terms of personal information and previous experience as well as their opinions regarding the use of earcons and auditory icons in e-learning interfaces. However, users' answers to the required tasks were evaluated to attain the data related to effectiveness and learning performance, memorability, and users' satisfaction. Furthermore, the post-experimental part of the questionnaire was devoted to obtain users' feedback with respect to the implemented non-speech sounds after they have had the opportunity to experience it in an interactive learning context.

5.8 Users Profiling

The test sample consisted of 24 users participated in the experiment on an individual basis. All of them were volunteers and first-time users of the experimental platform. Figure 43 shows users profiling in terms of personal data (A) and experience (B). As shown in Figure 43A, most of the participants (71%) were from the age range 25-34, a quarter of them were between 35 and 44 and the remaining (one user) was over 45 years old. The users' gender was observed as 67% (16 users) male and 33% (8 users) female. Also, the largest percentage (79%) of the users was enrolled in a doctorate course whilst

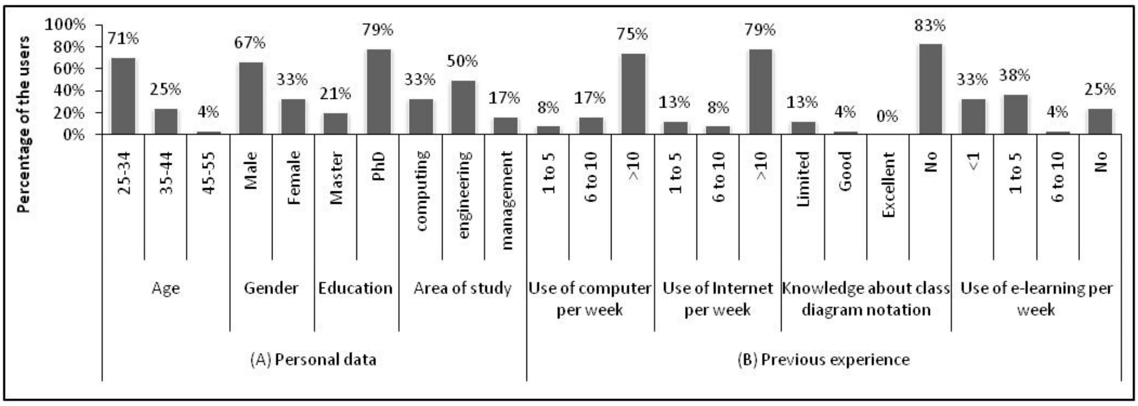


Figure 43: Profile of users in terms of personal data and previous experience

21% were students in the Master degree course. The areas of study were engineering, computing and management with 50%, 33% and 17% respectively.

In terms of prior experience, Figure 43B demonstrates that users were experts in computers (75%) and Internet (79%). When users were asked about their knowledge in class diagram notation, the learning material used in this experiment, 83% (20 users) of them declared that they have no idea, another 13% (3 users) have limited experience and only 4% (one user) had good knowledge on this topic. In other words, an overwhelming 96% of the users were inexperienced in this regard. In total, 75% of the users used e-learning applications with different time intervals per week whereas the remaining quarter did not experience it at all. Appendix C2 provides more details about the characteristics of the users who participated in the third experiment.

5.9 Results and Analysis

The obtained experimental results were analysed in terms of different parameters including users' views regarding the non-speech sounds accompanied the virtual lecturer voice in both the absence and presence of interactive e-learning context. Also, these parameters involved the effectiveness and memorability (in terms of correct and incorrect users' answers) in addition to user satisfaction. The existence of significant

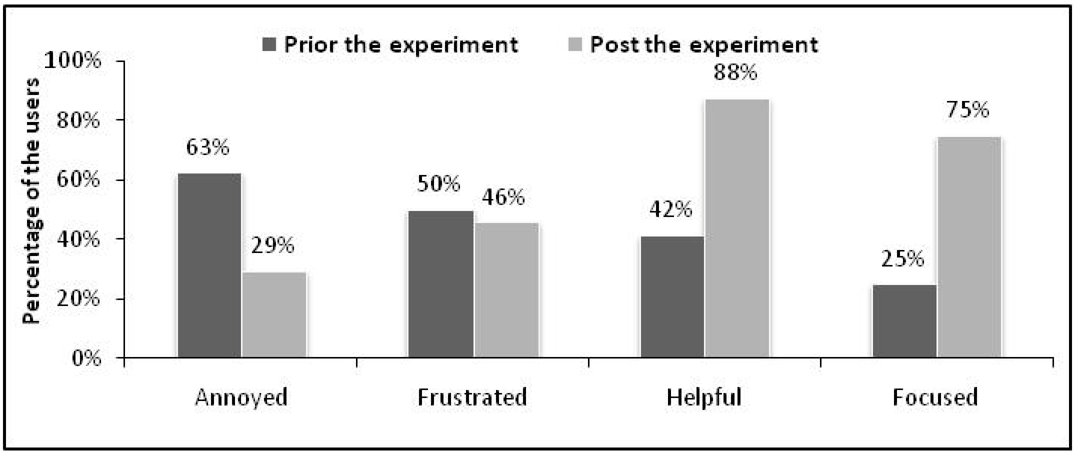


Figure 44: Users' views about earcons and auditory icons accompanied virtual lecturer voice when used in both the absence and presence of interactive e-learning context

difference in users' responses was examined by the nonparametric Chi-square statistical test at $\alpha = .05$ indicating significant difference when p-value was found less than .05.

5.9.1 Users' Evaluation of Earcons and Auditory Icons

Prior to the experiment, users were asked to express their feelings (Yes or No) in terms of annoyance, frustration, helpfulness and focus towards the use of earcons and auditory icons accompanying the voice of virtual lecturer in an e-learning interface. The same question was repeated at the end of the experiment. Users' responses are shown in Figure 44 (see also appendix C3 for raw data). It can be seen that users' feeling was more positive when earcons and auditory icons were used interactively in the AVLBG. About 63% of the users felt annoyed when they heard the tested sounds prior to the experiment. This percentage dropped to 29% after the earcons and auditory icons were being used in the applied interface. Although the level of frustration was approximately similar in both prior (50%) and post (46%) experimentation, users felt less frustrated when earcons and auditory icon incorporated interactively in the interface. In terms of helpfulness, it can be seen that users' opinion has been considerably changed after they experimentally tested the AVLBG. Prior to the experiment, 42%, of the participants thought that involving the tested non-speech sounds could help in enhancing their learning whereas a considerable 88% found these sounds helpful after they have had the opportunity to experience it in the applied platform.

Furthermore, it seems that these sounds did not substantially split users' concentration on the presented material; 75% found themselves focusing during the interaction with AVLBG compared to only 25% who thought that these sounds could enhance their attention when it has been introduced prior to experimentation. In summary, the addition of earcons and auditory icons to the experimental e-learning platform was helpful, not distracting, neither frustrating nor annoying.

5.9.2 Effectiveness

The number of correct and incorrect users' answers to the required learning evaluation questions was used to assess users' learning performance as well as the effectiveness of the AVLBG in presenting the learning material. Each user was required to answer 4 questions of two types; recall (Q1 and Q2) and recognition (Q3 and Q4). Therefore, the total number of questions was 96 (24 user * 4 questions per user) equally distributed over the two types. Figure 45 shows the percentage of correct and incorrect answers achieved by users for all questions (A), grouped by question type (B) and for each question (C). It can be seen, as shown in Figure 45A that the overall percentage of correct answers was 81% compared to 19% incorrect. In other words, 78 out of 96 questions were correctly answered. These results were significant ($\chi^2(1) = 38.00$, CV = 3.84, $p < .05$). In terms of question type, it can be noticed in Figure 45B that the percentage of successfully answered recognition questions was higher than that for the recall ones. In response to 48 questions in each type, the number of correct answers was 43 (90%) and 35 (73%) in recognition and recall questions respectively. Although users

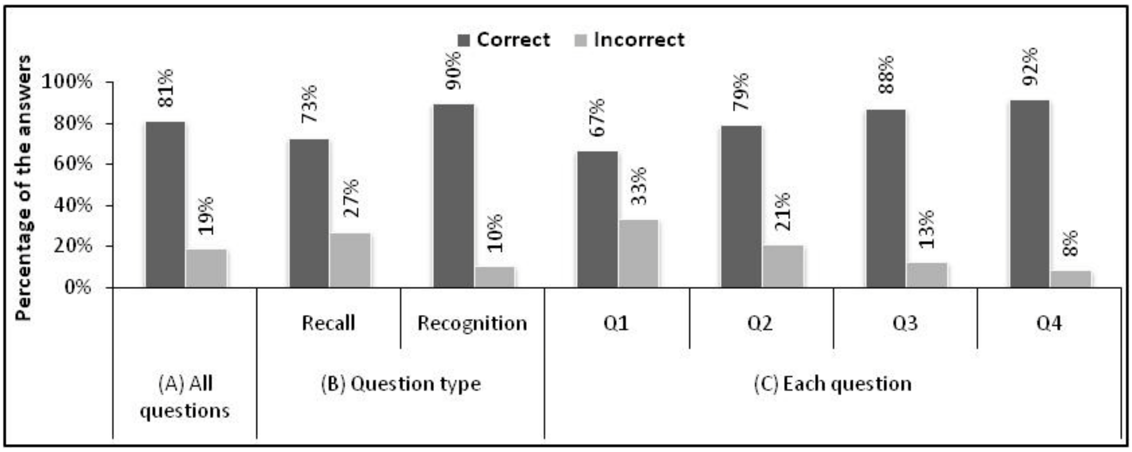


Figure 45: Percentage of correct and incorrect answers achieved by users for all questions (A) grouped by question type (B) and for each question (C)

performed better in the recognition tasks, the difference between correct and incorrect answers was found significant in both types recall ($\chi^2(1) = 10.08$, CV = 3.84, $p < .05$) and recognition ($\chi^2(1) = 30.08$, CV = 3.84, $p < .05$).

Figure 45C shows the correct answers attained by users for each question related to the communicated learning content. It can be seen that users' performance was varied across these questions. More specifically, the percentage of users who correctly answered Q4 and Q3 was 92% (22 users) and 88% (21 users) respectively. However, it seems that the remaining two questions were more difficult to answer. The percentage of correct answers declined to 79% for question 2 and 67% for question 1. These results were significant in terms of the difference between correct and incorrect answers for Q2 ($\chi^2(1) = 8.17$, CV = 3.84, $p < .05$), Q3 ($\chi^2(1) = 13.5$, CV = 3.84, $p < .05$) and Q4 ($\chi^2(1) = 16.67$, CV = 3.84, $p < .05$) whilst no significance has been obtained for Q1 ($\chi^2(1) = 2.67$, CV = 3.84, $p > .05$).

Figure 46 shows the number of correct answers provided by each user. It can be observed that 8 users answered all questions successfully whereas another 11 users accomplished 3 correct answers. However, the remaining users provided accurate

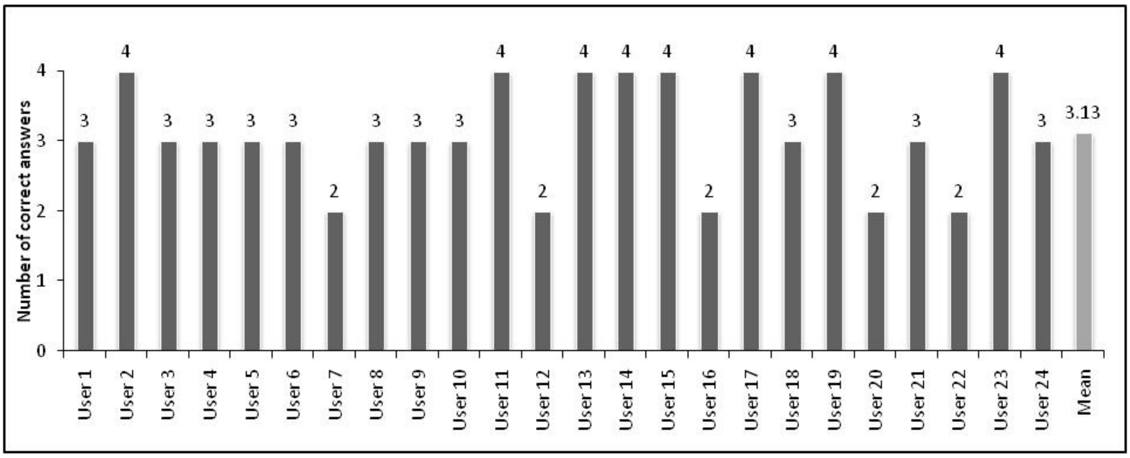


Figure 46: Total number of correct answers attained by each user in the third experiment

responses to only half of the required questions. In other words, 79% of users correctly performed 3 or more questions, which could be regarded as high performance rate. To summarize, it can be said that the incorporation of well known environmental sounds and short musical stimuli along with the virtual lecturer was found to be beneficial in communicating the learning material in e-learning interfaces. In other words, using these auditory messages can complement the role of virtual lecturer and it is more likely to result in capturing the users' attention to key parts of the delivered learning content. As a result, it could help in enhancing learners' performance in responding to different evaluation questions. More details about the correctness of users' answers to the learning evaluation questions can be found in Appendix C4.

5.9.3 Memorability

Upon completion of the learning performance tasks, users were asked to perform two memorability tasks. In the first one, users were presented with two paragraphs of the delivered lesson and they were requested to highlight where each of the incorporated non-speech sounds were used to indicate the key features in these paragraphs (see appendix C1). In this regard, the total number of questions was 168 (24 user * 7 questions per user).

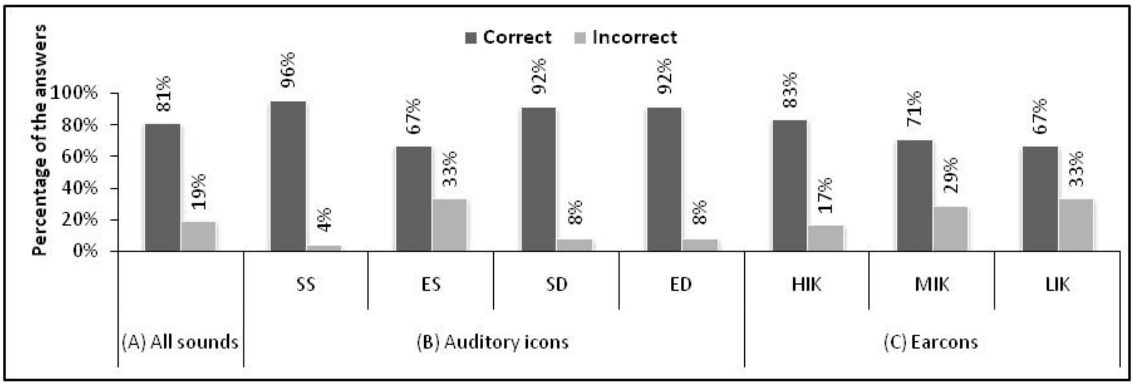


Figure 47: Percentage of users' correct identification of key features communicated by non-speech metaphors for all sounds (A) auditory icons (B) and earcons (C)

Figure 47 shows the correctness rate of users' responses to this task in regard to all sounds (A), auditory icons (B) and earcons (C). More details are presented in Appendix C5. On overall, it can be seen in Figure 47A that 81% (136 out of 168) of key features communicated by earcons and auditory icons were correctly identified by users. Statistically, this result was significant ($\chi^2(1) = 64.38$, CV = 3.84, $p < .05$). Also, Figure 47B demonstrates that most of users identified correctly the key features communicated by auditory icons. More specifically, 96% (23 users) highlighted *start of statement (SS)* correctly whereas 92% (22 users) accurately determined both *start (SD) and end of definition (ED)*. This percentage was dropped to 67% (16 users) for *end of statement (ES)* feature. When users were asked to highlight three words communicated by earcons as high, medium and low importance key words, Figure 47C shows that 83% (20 users) responded correctly for *high (HIK)* compared to 71% (17 users) for *medium (MIK)* and 67% (16 users) for *low importance (LIK)* words. The Chi-square results (see Table 32) revealed significant levels for the correctness of users' answers except for *end of statement* and *low importance* key words.

In the second memorability task, three sounds were played for each of the seven key features and the user had to recognize which sound was used to communicate each feature.

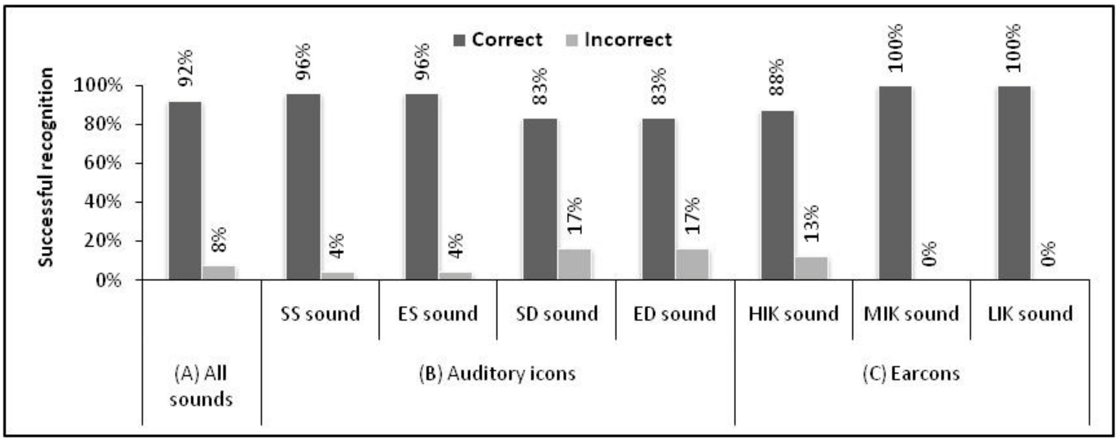


Figure 48: Users' successful recognition of the tested non-speech metaphors for all sounds (A) auditory icons (B) and earcons (C)

Communicated key feature	Non-speech sound	χ^2 value	p-value	Significant
Start of statement	Auditory icons	20	< .05	Yes
End of statement		3	> .05	No
Start of definition		17	< .05	Yes
End of definition		17	< .05	Yes
High importance key word	Earcons	11	< .05	Yes
Medium importance key word		4	< .05	Yes
Low importance key word		3	> .05	No

Table 32: Chi-square results for the correctness of users' identification of key features communicated by auditory icons and earcons (df = 1, CV = 3.84)

The obtained results can be seen in Figure 48 (see also Appendix C6) for all sounds (A), auditory icons (B) and earcons (C). On the whole, 92% of the tested sounds were correctly recognized by users (see Figure 48A), and this result was highly significant ($\chi^2(1) = 120$, CV = 3.84, $p < .05$). In Figure 48B, it can be seen that 96% (23 users) correctly recognised the auditory icons "opening a door" (SS sound) and "closing a door" (ES sound).

However, this percentage was 83% (20 users) for "opening a bottle" (SD sound) and "dropping a can" (ED sound). With respect to earcons, Figure 48C shows that the sounds used for medium (MIK sound) and low (LIK sound) importance were correctly recognised by all users whereas the sound of high importance (HIK sound) was recognised by a lower percentage 88% (21 users) of the users.

Communicated sound	Represented feature	χ^2 value	p-value	Significant
Door opening	Start of statement	20	< .05	Yes
Door closing	End of statement	20	< .05	Yes
Bottle opening	Start of definition	11	< .05	Yes
Can dropping	End of definition	11	< .05	Yes
Drum – 3 notes	High importance key word	14	< .05	Yes
Drum – 2 notes	Medium importance key word	24	< .05	Yes
Drum – 1 note	Low importance key word	24	< .05	Yes

Table 33: Chi-square results for correctness of users' recognition of the implemented non-speech sounds (df = 1, CV = 3.84)

The Chi-square results, as shown in Table 33 demonstrated significant difference between correct and incorrect recognition for all the tested sounds. In brief, the obtained results suggests that the tested auditory icons and earcons could be successfully interpreted and easily remembered by users when utilised in e-learning interfaces to signal the importance of specific content delivered by the virtual lecturer.

5.9.4 User satisfaction

Upon finishing memorability tasks, users were required to respond to the satisfaction questionnaire composed of 15 statements each of which had a 5-point Likert scale with 1 representing strong disagreement and 5 representing strong agreement. The first 10 statements were adopted from SUS questionnaire [230] to obtain users' attitude towards the different aspects of the AVLBG; however, the remaining 5 statements were included to obtain feedback from users regarding their learning experience attained during the interaction with the tested e-learning platform. On average, user satisfaction score calculated using the SUS approach was 80% indicating a high positive attitude. The frequency of users' different responses (i.e. agree, disagree, undecided) to each statement in the satisfaction questionnaire is illustrated in Figure 49 (see also Appendix C7 for more details).

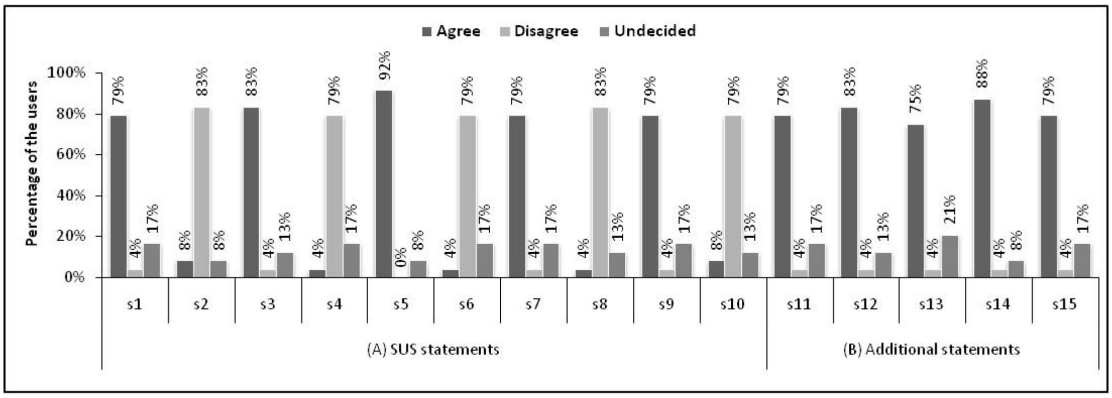


Figure 49: Frequencies of users' responses to SUS statements (A) and other statements (B) included in the satisfaction questionnaire

It can be noticed in Figure 49A that the positive statements (S1, S3, S5, S7, and S9) in SUS questionnaire attained high levels of users' agreement (between 79% and 92%). More specifically, 79% of the users agreed that *they would like to use the AVLBG frequently* (S1) and that *most people will learn how to use it very quickly* (S7). The same percentage of users *felt confident* (S9) during the interaction with AVLBG. However, this percentage rose to 83% and 92% for *ease of use* (S3) and *well integration of functions* (S5) respectively. On the other hand, users' disagreement regarding the negative statements (S2, S4, S6, S8, and S10) was observed high and fluctuated between 79% and 83%. According to most of the users (83%), the interface was neither *complex* (S2) nor *cumbersome to use* (S8). However, a slightly lower percentage of them (79%) did not agree that *using the tested interface requires the need for technical support* (S4) and a similar reaction was observed for S6; *I think that there is too much inconsistency in the interface*.

The additional satisfaction statements explored the users' views with respect to their learning experience in terms of *excitement and interest* about the presented lesson (S11), ease of *identification of the most important parts* of that lesson (S12) and the willingness to *attend similar e-learning activities* (S13). The last two statements also investigated the *overall users' satisfaction* with the interface (S14) and *learning* *experience* (S15) with the tested e-learning platform. The results can be seen in Figure 49B where more than three quarters of the users expressed positive impressions. In more details, users were excited and interested about the presented material (79%), and it was easy for them to identify the key features in this material (83%). Therefore, they would like to participate in e-learning if presented in similar approach (75%). On the whole, the interface and learning experience was satisfactory for users with 88% and 79% agreement rate respectively.

5.10 Discussion

During the experiment, it was observed that the users were concentrated on the delivered learning information. The reason could be attributed to the inclusion of interaction metaphors of different modalities in the tested interface. The textual and graphical metaphors incorporated in the presentation part of the interface along with body gestures of the virtual lecturer contributed to capturing user's visual attention towards the presented information. At the same time, further auditory explanations about this information were provided by the voice of the virtual lecturer. What is more important, using the non-speech sounds provided users with a mechanism to realise when important learning information is about to be presented and when it has been delivered. Also, it helped them to determine the most important key words in the communicated material. The obtained result showed that these sounds did not annoy, frustrate or distract the users and they found it helpful (refer to Figure 44). Also, users were able to remember the features communicated by these sounds (see Figure 47). Therefore, when users were asked about the presented learning material, they were able to provide the correct answer as shown in Figure 45. As a result, they were satisfied with the tested e-learning interface.

The first experimental hypothesis (H1) examined the effect of the added non-speech sounds on users' learning performance. Findings of this experiment demonstrated that users' performance was significantly assisted by the addition of earcons and auditory icon and helped in extending the contribution of full body animated virtual lecturer to comprehend both types of the required evaluation activities, recall and recognition (see Figure 45B). However, it is worthy to note that the contribution of these non-speech auditory stimuli (earcons and auditory icons) was varied across the required learning questions. For example, auditory icons significantly aided users' performance in both recall and recognition questions (Q2 and Q3) where the required information to answer these questions was attached with well known sounds from every day life. On the other hand, earcons contributed particularly in recognition questions (Q4) much more than in recall ones (Q1). In other words, the earcons used in this experiment were less beneficial compared to auditory icons. Nevertheless, the results on overall users' performance were significant in both recall and recognition activities supporting what have been hypothesised in H1.

The results of the experiment indicated that the users were satisfied, to a large extent with the inclusion of auditory icon and earcons in the evaluated e-learning platform (see Figure 49). The majority of users stated that these sounds were neither annoying nor frustrating, helped their learning and did not split their attention away. Despite the short training (90 seconds) provided prior to the experiment, users also were able to correctly interpret and easily remember these sounds (see Figure 47 and Figure 48). This could be referred to the design of the incorporated earcons where it was created using simple and short notes. Also, the used auditory icons were selected to help in making natural mapping between the communicated information and known sounds from the surrounding environment. Furthermore, each of these sounds conveyed only one meaning at a time and used consistently throughout the AVLBG. These aspects are important particularly when it is incorporated in parallel with other auditory and visual metaphors [146, 193, 240]. These aspects, in addition, led to generate positive users' feelings with respect to the tested e-learning platform. In sum, these results supported all the assumptions made by the experimental hypotheses H3, H4, H5 and H6. On the whole, the obtained results suggest that utilising non- speech sounds alongside full body gestures of speaking virtual lecturer could be useful in enhancing the usability and learning performance in e-learning interfaces.

5.11 Summary

The experiment reported in this chapter investigated the effectiveness and memorability of earcons and auditory icons when used in an innovative manner as complementary auditory signals to point out the most important features in the learning content when presented by a virtual lecturer in e-learning interfaces. The experiment also investigated users' satisfaction and their learning performance due to the use of this type of non-speech multimodal metaphors. A total of 24 users have taken part in the experiment to evaluate the experimental e-learning platform that extended the one tested in the second experiment by the addition of auditory icons and earcons. The obtained results demonstrated the effectiveness of these sounds to capture users' attention to important features of the delivered content, and contributed positively to enhance users' performance in different learning activities. Additionally, these sounds have shown to be easy to remember and understand, and was satisfactory to users. Therefore, the use of these metaphors was found to be helpful to enhance the usability of e-learning.

Chapter 6

Final Conclusions and Empirically Derived Guidelines

6.1 Introduction

This chapter presents a brief review of the experimental studies carried out in this research programme to explore the effect of different multimodal interaction metaphors on the usability and learning performance in e-learning interfaces. The chapter also summarises the main conclusions and limitations drawn from the obtained results. A set of empirically derived guidelines for the incorporation of multimodal metaphors in interface of e-learning applications are also included and discussed. These guidelines could contribute to the design of more usable e-learning interfaces to enable better learning performance. In the final part, the chapter concludes with a discussion of recommended future work.

6.2 Review of the Experimental Work

The research programme reported in this thesis was aimed at investigating the usability aspects of e-learning interfaces that utilise multimodal metaphors of a different nature (i.e. visual, audio and audio-visual) in the presentation of learning material about class diagram notation. The tested usability parameters included efficiency, effectiveness, memorability and user satisfaction. The research was aimed also at evaluating the users' learning performance due to the implementation of the experimental multimodal e-learning platforms. In order to fulfil these aims, a total of three experimental studies have been conducted where the obtained results helped in answering the research questions stated in Section 1.1.

The first experiment was aimed at investigating the usability (in terms of efficiency, effectiveness and user satisfaction) of earcons, recorded speech and speaking avatar with animated facial expressions when combined in a multimodal e-learning interface. Two different versions of the experimental e-learning platform were developed to serve in conducting this investigation. The first version named VOELP was based on a textual approach in providing explanations about three different examples of class diagram representations whereas the other version named MMELP involved a combination of earcons, recorded speech and talking head of facially expressive virtual lecturer to deliver the same learning material. The experimental e-learning platforms were evaluated by two independent groups of users each of which consisted of 15 users. Both groups were required to perform six common tasks of increasing complexity. Each task was based on placing the mouse cursor over notations in the displayed class diagram example and communicating the presented explanations visually in VOELP or audio-visually in MMELP. Each task was also followed by two evaluation questions; recall and recognition. Efficiency was measured by the time the users spent in answering the required questions. Effectiveness as well as learning performance was measured by users' correct answers. However, satisfaction of users was assessed by their responses to satisfaction questionnaire. More details about the first experiment can be found in Chapter 3.

The second experiment was dedicated to investigate the usability and learning performance of three experimental designs of e-learning interfaces each of which utilised different modes of avatars in the presentation of three lessons about class diagram notation. The focus was on exploring the role that avatars could play in improving the usability and users' learning achievements. Two of these interfaces incorporated a single facially expressive avatar; one with only a talking head (VLFE) and the other with full body animation (VLBG). However, in the third platform two talking heads of facially expressive avatars (TVLFE), male and female were included to share the presentation of each lesson. This experiment was also aimed to obtain the users' views in regard to a specific set of facial expression and body gestures when used by the virtual lecturers in both the absence and the presence of interactive e-learning context. The experimental interfaces were empirically tested by one group of users consisting of 48 users. Each user was instructed to evaluate positively or negatively a total of 6 facial expressions and 10 body gestures used in the experimental interfaces and to answer evaluation questions related to the presented learning content. The parameters for usability and learning performance were similar to those measured in the first experiment. The second experiment and its results are detailed in Chapter 4.

The third experimental work in this research programme investigated the effectiveness, memorability and user satisfaction of non-speech sounds (earcons and auditory icons) in providing supplementary auditory signals to indicate the key aspects of the learning content when presented by a single virtual lecturer in the presence of full body gestures. The experiment was also aimed to obtain users' views with respect to the added non-speech sounds. Apart from the addition of earcons and auditory icons, the experimental platform tested in this experiment was similar to that (i.e. VLBG) used in the second experiment. In total, 24 users participated in this experiment after being provided with a short training to insure their understanding of the tested sounds. Users were required to answer 4 recall and recognition evaluation questions in relation to the communicated content, and to remember the tested sounds as well as to interpret how these sounds were used during the interaction process. In addition, they were requested to express their attitude towards the evaluated e-learning interface. Chapter 5 described the third experiment and its findings in detail.

6.3 Main Conclusions

This section presents the main conclusions and limitations drawn from the experiments carried out in this research programme.

The results obtained from the first experiment showed that the multimodal metaphors were significantly more usable than the text with graphic metaphors in the presentation of the learning material. Using a combination of earcons, recorded speech and facially expressive avatar was more efficient in terms of reducing the time needed by users to answer the required evaluation questions (refer to Section 3.9.1 and Figure 7). These multimodal metaphors were also found more effective and significantly helped users to respond correctly to a higher number of questions particularly when these questions were of higher complexity (refer to Section 3.9.2 and Figure 11). Additionally, users of the multimodal e-learning interface were significantly more satisfied than their counterparts who used the text with graphics e-learning interface (see Section 3.9.3). These findings, however, did not sufficiently shed the light upon the individual contribution of these multimodal metaphors in the improvements observed on usability and users' learning performance. Therefore, the next experiment was designed to examine the role of avatars with specific facial expressions and body gestures as virtual lecturers in e-learning interfaces.

In the second experiment, the obtained results demonstrated that utilising a full body animation of facially expressive avatar is more usable (in terms of efficiency, effectiveness and satisfaction) than using facial expressions only either by one or two avatars in the design of e-learning interfaces (refer to Sections 4.9.3, 4.9.4 and 4.9.5 respectively). Also, designing the interface of e-learning in a way that combines both full body animated virtual lecturer and the presented learning material in one interface component was shown to be more attentive and attractive (see section 4.9.5) as well as more useful and preferable (see section 4.9.6) to users as their learning achievement was found the best when the experimental learning lessons were presented using this design. However, no difference in terms of usability and learning performance was attained between using one or two facially expressive avatars when incorporated in separate components within e-learning interfaces. Additionally, the results from the second experiment helped in determining both the best and the worst rated facial expressions and body gestures used in the tested e-learning designs. Nevertheless, these results explained that the contribution of the full body animated virtual lecturer in e-learning interfaces was restricted to recall learning activities and did not extend to the recognition ones. As a result, further investigation was needed to explore if the addition of non-speech sounds could support the influence of full-body animated virtual lecturer in both recall and recognition activities.

The third experiment provided empirical evidence that the addition of earcons and auditory icons could indeed help in capturing users' attention to key features of the learning material when delivered by the voice and body gestures of virtual lecturer. These sounds could be effective as supportive auditory messages to strengthen the contribution of full body animated virtual lecturer and hence to enable the users to perform well in different types (i.e. recall and recognition) of learning evaluation tasks (see Section 5.9.2). Results of this experiment also demonstrated that these non-speech sounds could be easily remembered and interpreted by users (see Section 5.9.3) and were shown to be satisfactory for them as shown in Section 5.9.4.

6.4 Empirically Derived Guidelines

The main findings and conclusions of the reported experiments assisted in producing a set of empirically derived guidelines for the design of more usable e-learning interfaces that could help learners in enhancing their learning performance in regard to the learning material used in this research. These guidelines could contribute to the current literature in both areas; e-learning and multimodal interaction. This section presents an overall discussion of the guidelines derived from this research.

6.4.1 Use of Recoded Speech Sounds

The recorded speech sounds were intensively used in this research programme and primarily utilised as the voice of virtual lecturer. These sounds have shown to be a fundamental component in interactive multimodal e-learning interfaces when these interfaces incorporate the use of human-like speaking avatars in communicating the learning content. The obtained results demonstrated the significant contribution of recorded speech in enhancing the interaction process particularly in terms of delivering clear and understandable spoken auditory messages. Most of the participants in all experiments (see Sections 3.9.3, 4.9.5, 4.9.6 and 5.9.4) express positive attitudes towards the tested e-learning platforms which implicitly mean that they were satisfied with the use of recorded natural speech sounds. These results support the findings of previous research (refer to Section 2.3.2) which confirmed that the recorded natural speech is advantageous over that generated by the speech synthesisers. Contrary to synthesised speech, recorded speech can be prepared to fit the needs of e-learning. For example, different tones or pitch could be used to stress users' attention to specific key words or statements in the delivered learning content. Therefore, when recording speech sounds to be used as a voice of virtual lecturer, care should be taken to use different notes and pitch in order to help in attaching proper facial expressions and body gestures. Also, it is recommended to leave short pause intervals among the speech of the virtual lecturer as it could attract the learners and possibly could be used later on to insert supportive auditory sounds of non-speech nature to capture users' attention to specific important parts of the presented material as demonstrated in the third experiment (see Section 5.5.2).

Furthermore, using recorded speech sounds is suggested to prevent splitting users' attention away from that material where users can keep looking at graphical representation and at the same time listening to spoken auditory explanations. This will result in reducing working memory load and offering more resources for cognitive processing of the presented learning material [45]. In brief, using recorded natural speech is recommended when designing e-learning interfaces.

6.4.2 Use of Avatar's Facial Expressions

The seconds experiment investigated users' views in regard to 6 facial expressions in both the absence and presence of interactive e-learning context. Based on the obtained results (see Section 4.9.1), designers of avatars for e-learning should bear in mind to incorporate positive facial expressions such as *smiling*, *happy*, *interested* and *amazed*. These expressions were found to be the most liked and best rated by the users (see Figure 24, Figure 25 and Table 20). The implementation of these expressions by virtual lecturers during the presentation of the learning content could make the e-learning environment more interesting and enjoyable to learners.

Although the low positive feeling expressed by users towards both *neutral* and *thinking* expressions (see Figure 24, Figure 25 and Table 20), still there is a need to use these expressions by avatars in e-learning interfaces. These expressions could be used to change the rhythm of the presentation and to attract users to think in the presented learning information.

6.4.3 Use of Avatar’s Body Gestures

The second experiment also investigated users’ opinions with respect to 10 body gestures when used by the virtual lecturer during the presentation of the learning material in both the absence and presence of interactive e-learning context. based on the experimental results (see Figure 27 and Figure 28), some of these gestures such as *pointing, walking, open palms, hands steepling, front clenching of hands* and *chin stroking* are suggested to be used by virtual lecturers in e-learning interfaces. These gestures were preferred by users and could be used in e-learning applications to attract learners and to enhance their interaction with the delivered material. In particular, *open palms, pointing, hands steepling* and *walking* were the best rated gestures (refer to Table 25) and could be performed by the virtual lecturer to support the presentation of learning content in e-learning interfaces.

On the other hand, body gestures such as *legs crossed* and *arms folded* are not recommended and suggested to be avoided by virtual lecturer during the presentation of the learning material in e-learning interfaces. These gestures were the most negatively regarded and least rated by users (see Figure 28 and Table 25). According to users’ comments, *legs crossed* indicate lack of respect towards the recipients which in turn could negatively influence their attitude and involvement in e-learning activities. However, the remaining gestures (i.e. *neutral* and *back clenching of hands*) can be carefully considered because it did not obtain significant positive users’ views (see Table 23).

6.4.4 Use of Extra Facially Expressive Avatars

The experiments performed in this thesis (see Chapters 3, 4 and 5) demonstrated that when designing e-learning interfaces to make use of facially expressive avatars as virtual lecturers, there is no need to incorporate more than one avatar. The addition of extra avatar does not make any difference in terms of usability and users' learning performance. For example, the results of the second experiment (see Sections 4.9.3, 4.9.4 and 4.9.5) showed that the inclusion of a single facially expressive virtual lecturer in the VLFE was found as usable as the inclusion of two facially expressive avatars in the TVLFE. The inclusion of more than one virtual lecturer with facial expressions could result in more distractive e-learning environment. Users' comments revealed that they felt distracted between the male and female virtual lecturers when both of them shared the presentation of the learning content. In other words, e-learning application can gain more benefits from avatars when incorporated in the form of only one virtual lecturer in the presence of facial expressions and full body animation. However, further experimental work is needed to investigate the inclusion of two full-body animated facially expressive virtual lecturers as compared to one similar to that implemented in the VLBG platform (see Section 4.5.3).

6.4.5 Integration of Virtual Lecturer in E-learning Interface

Another guideline for the use of human-like avatars in e-learning is related to interface component in which this avatar should be placed. The results from the second experiment in this research program (see Sections 4.9.3, 4.9.4, 4.9.5 and 4.9.6) provided an empirical basis for the necessity of combining full body animation of the virtual lecturer and the learning material in the same interface constituent. Placing the learning content (textual, graphical or both) in the background of the virtual lecturer with full body animation in the same scene, as applied in the VLBG (refer to Section 4.5.3) is suggested to be adopted in the design of e-learning interfaces in order to maximise the benefit of body gestures (such as *walking* and *pointing*) particularly in directing learners' visual attention to the related displayed learning information. Also, all the resources needed to integrate and comprehend the delivered information will be available to the learner in the same place without the need to move visual attention elsewhere in the interface. On the other hand, incorporating talking head of facially expressive virtual lecturer and presentation of learning content in separate interface components could result in overloading users' working memory by spending more mental effort in searching for the information related to the spoken material. This guideline is consistent with the results of other experiments in the literature [45] which confirmed the importance of integrating different information elements in one place in the interface.

6.4.6 Use of Non-speech Auditory Sounds

The use of non-speech sounds along with speech sounds has shown to be beneficial in enhancing Human-Computer Interaction in different domains (see Section 2.3.2 and Section 2.3.3). Earcons and auditory icons, as demonstrated by the results of the third experiment in this thesis (see Sections 5.9.1, 5.9.2, 5.9.3 and 5.9.4) can also be added to support and complement the role of virtual lecturer with full body animation in e-learning interfaces to communicate some key aspects of the learning content without annoying or confusing the learner. For example, well-known environmental sounds such as *door opening* and *bottle opening* can be used to inform the learner that an important statement or definition is about to be explained by the virtual lecturer whereas *door closing* and *can dropping* sounds can indicate the end of that information. Also, different numbers of musical tones can be used to convey simple and short auditory signals related to the importance level of specific key words in the presented learning discipline. These sounds could convey single meaning and could be used consistently throughout the interface in order to avoid distracting or confusing the users. Also, it is recommended to add these sounds in the pause intervals in virtual lecturer's speech so that its duration suits these pauses. In other words, these sounds should be communicated in a way which does not overlap with the virtual lecturer speech to enable the learner to remember and interpret it before continuation of the virtual lecturer's speech. Lastly, sufficient training could be provided so that users can easily and quickly remember the features communicated by these sounds.

6.4.7 Complexity and Type of Learning Evaluation Tasks

The experiments performed in this research programme demonstrated that the contribution of multimodal interaction metaphors in e-learning interfaces can be affected by the type and complexity of the required learning evaluation tasks. In the first experiment, the combination of earcons, recorded speech and avatars contributed to substantially enhance learners' efficiency in the three task complexity levels (see Section 3.9.1.2). This was apparent for recall activities whereas in the recognition task the efficiency difference between the two conditions was considerable only in difficult level of complexity (see Section 3.9.1.3). With respect to effectiveness, the results of the first experiment (see sections 3.9.2.2 and 3.9.2.3) showed that the combination of earcons, auditory icons and avatars can particularly benefit in improving learners' performance in moderate to difficult learning evaluation tasks of both recall and recognition types.

However, the second experiment demonstrated that the use of virtual lecturer with full body animation in communicating the learning material can help to enhance both efficiency and effectiveness of learners in recall activities much more than that in recognition tasks (see Sections 4.9.3.2 and 4.9.4.2). This research also revealed the importance role that the addition of non-speech sounds could play in communicating supportive information during the presentation of the learning content by the virtual lecturer. The results of the third experiment (see Section 5.9.2) demonstrated that this addition could result in extending the effect of full-body animated virtual lecturer on learners' performance to include both recall and recognition learning activities. On overall, this thesis recommends the use of multimodal interaction metaphors to enhance the usability of e-learning interfaces, particularly with the increasing complexity of learning activities of both recall and recognition types.

6.5 Future Work

This section proposes ideas for experimental work that can be carried out in the future as a continuation of this research.

6.5.1 More Facial Expressions and Body Gestures

The second experiment in this research investigated 6 facial expressions and 10 body gestures when used by the virtual lecturer during the presentation of learning material. Further experiment can be undertaken to examine additional facial expressions and body gestures in both the absence and presence of interactive e-learning context. The best and least rated among these expressions and gestures can also be evaluated. The expected outcomes could contribute in producing wider and broader guidelines for the use of facial expressions and body gestures in e-learning interfaces.

6.5.2 Interactive Virtual Lecturer in Mobile Learning

Currently, mobile devices and wireless technology are widely used and could offer flexible and convenient mobile learning [51]. These portable devices are continuously developed particularly in terms of screen size and resolution as well as other multimedia features. Therefore, there is potential to explore the usefulness of incorporating avatars with facial expressions and body gestures as virtual pedagogical agents in mobile learning.

6.5.3 Intelligent Virtual Lecturer

The virtual lecturers investigated throughout this thesis were prominently used in the presentation of the learning content. Although users in the second and third experiment were able to textually ask questions and got the answers by the virtual lecturers, these features were programmed in advance to suit research necessities. Therefore, some of the participants expressed the desire of more interactivity in their interaction with the virtual lecturer. For example, speech recognition technology can be involved to enable orally-directed queries by users. In this case, the virtual lecturer could have intelligent capabilities such as retrieval of the required explanations and automatic generation of relevant verbal and non-verbal responses.

6.5.4 Theatre Metaphor

The results obtained from the second experiment in this research program demonstrated that the use of full body animation of facially expressive virtual lecturer outperformed the use of only the talking head of facially expressive one. It would be worthy to conduct an experimental study to explore the usability and learning performance of two facially expressive virtual lecturers with full body animation when used in a theatrical style to share the presentation of the learning material displayed in the background of the same interface component. This exploration could involve the gender of both virtual lecturers (i.e. which is better to use? two males, two females or mixed?).

6.5.5 Personalised Virtual Lecturer

Apart from facial expressions and body gestures, the current thesis did not consider the influence of the virtual lecturer's external characteristics such as gender, voice, ethnicity and age. These individual parameters of the virtual lecturer might affect users' learning performance. However, more research is still needed in this regard. For example, virtual lecturer could be used to serve as a personalised tutor. It would be interested to examine users' learning outcomes in both the absence and presence of virtual lecturer personalization.

6.5.6 Virtual Lecturers for Deaf-Mute Users

Another interesting research idea is to investigate the use of virtual lecturer with full body animation to teach the sign language for deaf-mute users. A large variety of body gestures can be programmed on-demand to demonstrate the components of this language. More advanced features can also be explored such as capturing body gestures of the user and automatically generate the virtual lecturer's body animation in response to the user.

6.6 Epilogue

This thesis has investigated the usability aspects of e-learning interfaces that utilise multimodal interaction metaphors in the presentation of the learning content. The thesis has also explored the effect of these metaphors on users' learning outcomes. The results obtained from three experiments within this research programme have provided empirical evidence that earcons, auditory icons, recorded speech along with avatars with facial expressions and body gestures could indeed help in improving the usability as well as users' learning performance in e-learning interfaces when utilised to communicate the incorporated learning material. The experimental findings as well as the empirically derived guidelines for the design of more usable e-learning applications contribute to the research literature in both the multimodal interaction and the e-learning fields. However, further research highlighted earlier in this chapter could be carried out to reinforce the potential of multimodal metaphors in enhancing Human-Computer Interaction in e-learning domain.

REFERENCES

[1] S. A. Brewster, "The design of sonically-enhanced widgets," Interacting with Computers, vol. 11, pp. 211-235, 1998.

[2] S. A. Brewster, "Using non-speech sound to overcome information overload," Displays, vol. 17, pp. 179-189, 1997.

[3] I. Oakley, M. R. McGee, S. Brewster, and P. Gray, "Putting the feel in'look and feel ‘," Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 415-422, 2000.

[4] M. L. Brown, S. L. Newsome, and E. P. Glinert, "An experiment into the use of auditory cues to reduce visual workload," Proceedings of the SIGCHI conference on Human factors in computing systems: Wings for the mind, pp. 339-346, 1989.

[5] S. Oviatt, "Multimodal Interfaces," The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, 2003.

[6] N. B. Sarter, "Multimodal information presentation: Design guidance and research challenges," International Journal of Industrial Ergonomics, vol. 36, pp. 439-445, 2006.

[7] L. A. Annetta and S. Holmes, "Creating Presence and Community in a Synchronous Virtual Learning Environment Using Avatars," International journal of instructional technology and distance learning, vol. 3, pp. 27-43, 2006.

[8] A. L. Baylor, "The impact of three pedagogical agent roles," Proceedings of Autonomous Agents & Multi-Agent Systems Conference, pp. 928-929, 2003.

[9] T. L. Bonebright, M. A. Nees, T. T. Connerley, and G. R. McCain, "Testing the effectiveness of sonified graphs for education: A programmatic research project," International Conference on Auditory Display, Espoo, Finland, 2001.

[10] J. Holmes, "Designing agents to support learning by explaining," Computers & Education, vol. 48, pp. 523-547, 2007.

[11] D. Rigas and D. Hopwood, "The Role of Multimedia in Interfaces for On-Line Learning," 9th Panhellenic Conference on Informatics (PCI'2003). , Thessaloniki, Greece, 2003.

[12] G. Theonas, D. Hobbs, and D. Rigas, "Employing Virtual Lecturers' Facial Expressions in Virtual Educational Environments," International Journal of Virtual Reality, vol. 7, pp. 31-44, 2008.

[13] R. Upson, "Educational sonification exercises: Pathways for mathematics and musical achievement," International Conference on Auditory Display, Kyoto, Japan, 2002.

[14] G. Shaw and N. Marlow, "The role of student learning styles, gender, attitudes and perceptions on information and communication technology assisted learning," Computers & Education, vol. 33, pp. 223-234, 1999.

[15] D. Steinman, "Educational experiences and the online student," TechTrends, vol. 51, pp. 46-52, 2007.

[16] G. E. Tudor, "Teaching Introductory Statistics Online–Satisfying the Students," Journal of Statistics Education, vol. 14, 2006.

[17] G. Morgan, "Thirteen'must ask' questions about e-learning products and services," The learning organization, vol. 8, pp. 203-211, 2001.

[18] D. Wong, "A Critical Literature Review on e-Learning Limitations," School of Management and Information Technology, UCSI, pp. 55-62, 2007.

[19] IDC, http://www.idc.com. Cited in P. Dunn, "Has e-learning truly arrived in Europe?", Training Journal, Fenman Limited, UK, 2001. Available from: http://e-learningzone.co.uk/feature3.html.

[20] D. Asirvatham, "Borderless Education through E-Learning," presented at MAPCU National Higher Education Conference, Petaling Jaya, LA1238 MNHEC, 2003. Available from: https://mmlscyber.mmu.edu.my/articles/article2.html.

[21] T. Clarke and A. Hermens, "Corporate developments and strategic alliances in e-learning," Education+ Training, vol. 43, pp. 256-267, 2001.

[22] S. Alexander, "E-learning developments and experiences," Education and Training, vol. 43, pp. 240-248, 2001.

[23] U. Bleimann, "Atlantis University: a new pedagogical approach beyond e-learning," Campus-Wide Information Systems, vol. 21, pp. 191-195, 2004.

[24] I. Roffe, "E-learning: engagement, enhancement and execution," Quality Assurance in Education, vol. 10, pp. 40-50, 2002.

[25] S. Sambrook, "E-learning in small organisations," Education+ Training, vol. 45, pp. 506-516, 2003.

[26] T. L. Wentling, C. Waight, J. Gallaher, J. La Fleur, C. Wang, and A. Kanfer, "E-learning-A review of literature," Knowledge and Learning Systems Group, University of Illinois at Urbana-Champaign learning. ncsa. uiuc. edu/papers/elearnlit. pdf, 2000.

[27] R. H. Wild, K. A. Griggs, and T. Downing, "A framework for e-learning as a tool for knowledge management," Industrial Management and Data Systems, vol. 102, pp. 371-380, 2002.

[28] T. Govindasamy, "Successful implementation of e-Learning Pedagogical considerations," The Internet and Higher Education, vol. 4, pp. 287-299, 2001.

[29] F. Mikic and L. Anido, "Towards a Standard for Mobile E-Learning," International Conference on Systems and International Conference on Mobile Communications and Learning Technologies, 2006. ICN/ICONS/MCL 2006. International Conference on Networking, pp. 217-222, 2006.

[30] D. Yu, W. Zhang, and X. Chen, "New Generation of E-Learning Technologies," Proceedings of the First International Multi-Symposiums on Computer and Computational Sciences (IMSCCS'06), 2006.

[31] A. A. Hamid, "e-Learning is it the “e” or the learning that matters?," The Internet and Higher Education, vol. 4, pp. 311-316, 2001.

[32] M. J. Rosenberg, E-learning-strategies for delivering knowledge in the digital age: McGraw-Hill, 2001.

[33] A. P. Correia and P. Dias, "Criteria for evaluating learning web sites: how does this impact the design of e-learning," Actas da II Conferência Internacional de Tecnologias da Informação e Comunicação na Educação: Desafios Challenges, pp. 521-528, 2001.

[34] A. M. Spalter, R. M. Simpson, M. Legrand, and S. Taichi, "Considering a full range of teaching techniques for use in interactive educational software: a practical guide and brainstorming session," Frontiers in Education Conference, 2000. FIE 2000. 30th Annual, vol. 2, 2000.

[35] K. DelVecchio and M. Loughney, "Advantages and Disadvantages of E-Learning," E-Learning Concepts and Techniques, E-Book, pp. 5-6, 2006.

[36] T. A. Urdan and C. C. Weggen, "Corporate e-learning: Exploring a new frontier," San Francisco: WR Hambrecht, 2000.

[37] G. Russell and B. Holkner, "Virtual schools," Futures, vol. 32, pp. 887-897, 2000.

[38] L. Brady, "Fault lines in the terrain of distance education," Computers and Composition, vol. 18, pp. 347-358, 2001.

[39] G. Homan and A. Macpherson, "E-learning in the corporate university," Journal of European Industrial Training, vol. 29, pp. 75-90, 2005.

[40] K. O'Regan, "Emotion and e-learning," Journal of Asynchronous learning networks, vol. 7, pp. 78-92, 2003.

[41] A. P. Rovai and H. M. Jordan, "Blended learning and sense of community: A comparative analysis with traditional and fully online graduate courses," International Review of Research in Open and Distance Learning, vol. 5, pp. 1-13, 2004.

[42] K. A. Meyer, "Face-to-face versus threaded discussions: The role of time and higher-order thinking," Journal of Asynchronous Learning Networks, vol. 7, pp. 55-65, 2003.

[43] K. M. Kapp, "E-learning basics: essay: The e-learning market: it's about the learner, not the instructor!," eLearn, vol. 2003, 2003.

[44] J. R. Evans and I. M. Haase, "Online business education in the twenty-first century: An analysis of potential target markets," Internet Research: Electronic Networking Applications and Policy, vol. 11, pp. 246-260, 2001.

[45] H. Tabbers, L. Kester, H. Hummel, and R. Nadolski, "Interface design for digital courses," pp. In W. Jochems, J. van Merriënboer, & R. Koper (Eds). Integrated e-learning: implications for pedagogy, technology & organisation (pp. 100-111). London: Routledge Farmer, 2005.

[46] T. Mayes and S. de Freitas, "Review of e-learning theories, frameworks and models," Commissioned review report as part of the JISC-funded e-pedagogy desk study on e-learning models Retrieved April 25th, 2007.

[47] M. Frank, G. Kurtz, and N. Levin, "Implications of presenting pre-university courses using the blended e-learning approach," Educational Technology & Society, vol. 5, pp. 1-16, 2002.

[48] P. Henry, "E-learning technology, content and services," Education+ Training, vol. 43, pp. 249-255, 2001.

[49] D. Hoffmaster, "The E-Learning Evolution, E-Learning Concepts and Techniques," Institute for Interactive Technologies , Bloomsburg University of Pennsylvania, USA, 2006, pp. 3-4.

[50] R. Hamilton, C. Richards, and C. Sharp, "An examination of e-learning and e-books," 2001.

[51] T. Georgiev, E. Georgieva, and G. Trajkovski, "Transitioning from e-Learning to m-Learning: Present Issues and Future Challenges," Proceedings of the Seventh ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD'06)-Volume 00, pp. 349-353, 2006.

[52] K. Nyiri, "Towards a philosophy of m-learning," Proceedings of the IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE’02). pp. 121-124, 2002.

[53] A. Trifonova and M. Ronchetti, "A general architecture to support mobility in learning," Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT’04), pp. 26-30, 2004.

[54] L. Curry and C. Adams, "Patterns of learning style across selected medical specialties," Educational Psychology, vol. 11, pp. 247-277, 1991.

[55] R. M. Felder and E. R. Henriques, "Learning and teaching styles in foreign and second language education," Foreign Language Annals, vol. 28, pp. 21-31, 1995.

[56] R. Dunn, "Survey of Research on Learning Styles," Educational Leadership, vol. 46, pp. 50-58, 1989.

[57] D. A. Kolb, "Experimental Learning," Experience as a Source of Learning and Development, 1984.

[58] T. Larkin and D. Budny, "Learning Styles in the Classroom: Approaches to Enhance Student Motivation and Learning," Information Technology Based Higher Education and Training, 2005. ITHET 2005. 6th International Conference on, pp. F4D-1, 2005.

[59] P. Honey and A. Mumford, The Manual of Learning Styles: 10 Linden Avenue, Maidenhead: Peter Honey, 1986.

[60] R. M. Felder and L. K. Silverman, "Learning and Teaching Styles in Engineering Education," Engineering Education, vol. 78, pp. 674-681, 1988, Preceded by a preface in 2002.

[61] C. A. Carver, R. A. Howard, and W. D. Lane, "Enhancing student learning through hypermedia courseware and incorporation of student learning styles," IEEE Transactions on Education, vol. 42, pp. 33-38, 1999.

[62] J. Kuljis and F. Liu, "A comparison of learning style theories on the suitability for elearning," Proceedings of the Conference on Web Technologies, Applications, and Services, pp. 191-197, 2005.

[63] S. Graf and T. Lin, "The relationship between learning styles and cognitive traits–Getting additional information for improving student modelling," Computers in Human Behavior, vol. 24, pp. 122-137, 2008.

[64] R. M. Felder and B. A. Soloman, "Learning styles and strategies," At URL: http://www. engr. ncsu. edu/learningstyles/ilsweb. html, 2000.

[65] D. H. R. Spennemann, "Gender imbalances in computer access among environmental science students," e-Journal of Instructional Science and Technology, vol. 1, 1996.

[66] A. D. Baddeley, Essentials of human memory: Taylor & Francis, 1999.

[67] A. C. Catania, Learning-Remembering, 3rd ed: Prentice-Hall International Editions, 1992.

[68] S. Graf, T. Lin, and Kinshuk, "Analyzing the relationship between learning styles and cognitive traits," Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007), pp. 235-239, 2007.

[69] T. Lin and P. A. Kinshuk, "Cognitive Trait Model-A Supplement to Performance Based Student Models," Proceedings of International Conference on Computers in Education, Hong Kong, pp. 629-632, 2003.

[70] R. Low and J. Sweller, "The Modality Principle in Multimedia Learning," in The Cambridge handbook of multimedia learning, R. Mayer, Ed. New York: Cambridge University Press, 2005, pp. 147–158.

[71] L. J. Chanlin, "The Effects of Verbal Elaboration and Visual Elaboration on Student Learning," International Journal of Instructional Media, vol. 24, pp. 333-39, 1997.

[72] A. Paivio and J. M. Clark, "Dual coding theory and education," Educational Psychology Review, vol. 3, pp. 149-170, 1991.

[73] R. E. Mayer, Multimedia learning. New York: Cambridge University Press, 2001.

[74] M. D. Sankey and M. Nooriafshar, "Multiple representations in multimedia and e-learning materials: an issue of literacy," pp. 149-171, 2005.

[75] J. Nielsen, Usability Engineering: Academic Press Inc., US, 1993.

[76] M. F. Costabile, "Usability in the Software Life Cycle," Handbook of software engineering and knowledge engineering, vol. 1, pp. 179–192, 2001.

[77] ISO, "ISO 9241: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) - Part 11: Guidelines on usability," pp. 2, 1998.

[78] L. L. Lohr, "Designing the instructional interface," Computers in Human Behavior, vol. 16, pp. 161-182, 2000.

[79] A. Dix, G. Abowd, J. Finlay, and R. Beale, Human-Computer Interaction (3rd Edition). Prentice Hall, 2004.

[80] D. Rigas and M. Alsuraihi, "A Toolkit for Multimodal Interface Design: An Empirical Investigation," in Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments, vol. 4552/2007, Lecture Notes in Computer Science, 2007, pp. 196-205.

[81] A. J. Dix, Human-computer interaction. New York; London: Prentice-Hall, 1993.

[82] D. Rigas and N. Gazepidis, "Facial Expressions and Body Gestures as Metaphors for B2C Interfaces: An Empirical Study," Proceedings of the International Conference on e-Business, Setubal, Portugal, 2006.

[83] D. Rigas and N. Gazepidis, "A further investigation of facial expressions and body gestures as metaphors in e-commerce," Proceedings of the 7th Conference on 7th WSEAS International Conference on Applied Informatics and Communications, pp. 148-153, 2007.

[84] S. Buisine and J. C. Martin, "Design principles for cooperation between modalities in bi directional multimodal interfaces," Proceedings of the CHI 2003 workshop on ‘‘Principles for multimodal user interface design’’, 2003.

[85] M. Alotaibi and D. Rigas, "How Effective is it to Communicate Customer Knowledge with the aid of Audio-Visual Metaphors," In Proceedings of the Saudi International Innovation Conference 2008, pp. 29-34, 2008.

[86] M. Alotaibi and D. Rigas, "How efficient is it to communicate customer knowledge with the aid of audio-visual metaphors?," Proceeding of the 2008 conference on Knowledge-Based Software Engineering: Proceedings of the Eighth Joint Conference on Knowledge-Based Software Engineering, pp. 233-240, 2008.

[87] M. Alotaibi and D. Rigas, "How satisfactory is it to communicate customer knowledge with the aid of multimodal interaction metaphors?," in Proceedings of the 8th WSEAS International Conference on Software engineering, parallel and distributed systems. Cambridge, UK: World Scientific and Engineering Academy and Society (WSEAS), 2009.

[88] M. Alotaibi and D. Rigas, "Interactive Customer Knowledge Management Systems: a Comparative Evaluation of Users' Perception of Trust and Level of Knowledge," Proceedings of the 7th WSEAS International Conference on E-ACTIVITIES, pp. 54-59, 2008.

[89] S. Rashid and D. Rigas, "E-Learning: An Experimental Note-Taking Platform," WSEAS Transactions on Computer Rresearch, vol. 1, pp. 254-259, 2006.

[90] D. Rigas and M. Alotaibi, "The Role of Audio-Visual Metaphors in Aiding the Communication of Customer Knowledge: User Satisfaction Perspective," In Proceedings of International Conference on E-Business (ICETE ICE-B), pp. 143-148, 2008

[91] D. Rigas and I. Bahadur, "A Two Platform Empirical Study to Investigate the Use of Audio in a Stock Control Application," WSEAS Transactions on Computers, vol. 5, pp. 21-40., 2006.

[92] D. Rigas and A. Ciuffreda, "An empirical investigation of multimodal interfaces for browsing internet search results," in Proceedings of the 7th Conference on 7th WSEAS International Conference on Applied Informatics and Communications - Volume 7. Vouliagmeni, Athens, Greece: World Scientific and Engineering Academy and Society (WSEAS), 2007.

[93] D. Rigas and A. Ciuffreda, "The use of Multi-Modality in Delivering Web Results: An Empirical Investigation," IADIS International Conference in Computer Graphics and Visualisation, Lisbon, Portugal, 2007.

[94] L. Barfield, Design for new media: Interaction design for multimedia and the web: Addison-Wesley, 2004.

[95] J. Johnson, T. L. Roberts, W. Verplank, D. C. Smith, C. H. Irby, M. Beard, and K. Mackey, "The xerox star: A retrospective," Computer, vol. 22, pp. 11-26, 1989.

[96] G. Williams, "Apple Macintosh computer," Byte, vol. 9, pp. 30-31, 1984.

[97] A. Ciuffreda, "An Empirical Investigation in Using Multi-modal Metaphors to Browse Internet Search Results," in Department of Computing, School of Informatics, vol. PhD. Bradford, UK: University of Bradford, 2008.

[98] S. A. Brewster, P. C. Wright, and A. D. N. Edwards, "A detailed investigation into the effectiveness of earcons," Auditory display, sonification, audification and auditory interfaces. The Proceedings of the First International Conference on Auditory Display, pp. 471-498, 1992.

[99] S. A. Brewster, P. C. Wright, and A. D. N. Edwards, "An evaluation of earcons for use in auditory human-computer interfaces," Proceedings of InterCHI, vol. 93, pp. 222-227, 1993.

[100] J. L. Alty, D. Rigas, and P. Vickers, "Music and speech in auditory interfaces: When is one mode more appropriate than another?," Proc. ICAD2005, 2005.

[101] J. Preece, Y. Rogers, H. Sharp, D. Benyon, S. Holland, and T. Carey, "Human Computer Interaction," Addison-Wesley, New York, 1994.

[102] G. Di Fabbrizio, D. Dutton, N. Gupta, B. Hollister, M. Rahim, G. Riccardi, R. Schapire, and J. Schroeter, "AT&T help desk," presented at The Seventh International Conference on Spoken Language Processing Denver, Colorado, USA, 2002.

[103] B. Duggan and M. Deegan, "Considerations in the usage of text to speech (tts) in the creation of natural sounding voice enabled web systems," presented at Proceedings of the 1st international symposium on Information and communication technologies 2003.

[104] A. Ciuffreda and D. Rigas, "A usability Study of multimodal interfaces for the presentation of Internet Search Results," International Journal of Computers, NAUN, vol. 2, pp. 120-125, 2008.

[105] D. Rigas and M. Sallam, "An Empirical Study on E-learning Note-Taking Platform Using Multimodal Metaphors," International journal of Applied Mathematics and Informatics, University Press, UK., vol. 2, pp. 88-95, 2008.

[106] M. Alotaibi, "Electronic Customer Knowledge Management Systems: a Multimodal Interaction Approach," in Informatics Research Institute, vol. PhD. Bradford, UK: University of Bradford, 2009.

[107] N. Gazepidis, "Facial Expressions and Body Gestures in E-Commerce: an Empirical Investigation," in Department of Computing, School of Informatics, vol. PhD. Bradford, UK: University of Bradford, 2008.

[108] C. Asakawa and T. Itoh, "User interface of a home page reader," 1998.

[109] L. Lines and K. S. Hone, "Older adults' evaluations of speech output," presented at Proceedings of the 5th International ACM SIGCAPH Conference on Assistive Technologies (ASSETS’02), Edinburgh, Scotland, 2002.

[110] R. W. Massof, "Auditory assistive devices for the blind," presented at The 2003 International Conference on Auditory Display, Boston, MA, USA, 2003.

[111] J. Sanchez and F. Aguayo, "Mobile messenger for the blind," Lecture Notes in Computer Science, vol. 4397, pp. 369, 2007.

[112] M. S. Sanders and E. J. McCormick, Human factors in engineering and design: McGraw-Hill Science/Engineering/Math, 1993.

[113] S. Weinschenk and D. T. Barker, Designing effective speech interfaces: John Wiley & Sons, Inc. New York, NY, USA, 2000.

[114] D. Gibbon, I. Mertins, and R. K. Moore, "Handbook of multimodal and spoken dialogue systems: resources, terminology and product evaluation," Computational Linguistics, vol. 27, pp. 149-150, 2001.

[115] E. Janse, "Time-Compressing Natural and Synthetic Speech," presented at Proceedings of 7th International Conference on Spoken Language Processing, Denver, USA, 2002.

[116] S. Brewster, G. Leplâtre, and M. Crease, "Using non-speech sounds in mobile computing devices," 1998.

[117] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, "Earcons and Icons: Their Structure and Common Design Principles," Human-Computer Interaction, vol. 4, pp. 11-44, 1989.

[118] S. A. Brewster, P. C. Wright, and A. D. N. Edwards, "Experimentally derived guidelines for the creation of earcons," Adjunct Proceedings of HCI, vol. 95, pp. 155-159, 1995.

[119] D. I. Rigas, D. Memery, D. Hopwood, and M. A. Rodrigues, "Empirically derived design issues in auditory information processing for mobile telephony," Information Technology: Coding and Computing, 2000. Proceedings. International Conference on, pp. 462-469, 2000.

[120] S. A. Brewster, P. C. Wright, and A. D. N. Edwards, "The design and evaluation of an auditory-enhanced scrollbar," Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence, pp. 173-179, 1994.

[121] S. A. Brewster, P. C. Wright, A. J. Dix, and A. D. N. Edwards, "The sonic enhancement of graphical buttons," Proceedings of Interact, vol. 95, pp. 43-48, 1995.

[122] S. A. Brewster and M. Crease, "Making Menus Musical," Proceedings of IFIP Interact'97, pp. 389-396, 1997.

[123] S. A. Brewster and M. G. Crease, "Correcting menu usability problems with sound," Behaviour & Information Technology, vol. 18, pp. 165-177, 1999.

[124] M. Crease and S. Brewster, "Making Progress With Sounds-The Design and Evaluation Of An Audio Progress Bar," Proceedings of ICAD, vol. 98, 1998.

[125] S. A. Brewster and C. V. Clarke, "The design and evaluation of a sonically enhanced tool palette," ACM Transactions on Applied Perception (TAP), vol. 2, pp. 455-461, 2005.

[126] S. A. Brewster, "Sound in the interface to a mobile computer," Communication, Cooperation, and Application Design., pp. 43, 1999.

[127] A. Walker and S. Brewster, "Spatial audio in small screen device displays," Personal and Ubiquitous Computing, vol. 4, pp. 144-154, 2000.

[128] C. J. DiGiano, R. M. Baecker, and R. N. Owen, "LogoMedia: a sound-enhanced programming environment for monitoring program behavior," Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 301-302, 1993.

[129] D. I. Rigas and J. L. Alty, "Using sound to communicate program execution," Proceedings of the 24th EUROMICRO Conference, vol. 2, pp. 625–632, 1998.

[130] D. I. Rigas, M. A. R. Kirby, and D. O'Connel, "Audiotest: Utilising audio to communicate information in program debugging," the Eighth International Conference on Human-Computer Interaction: Ergonomics and User Interfaces, pp. 1293-1297, 1999.

[131] J. L. Alty, D. Rigas, and P. Vickers, "Using music as a communication medium," Conference on Human Factors in Computing Systems, pp. 30-31, 1997.

[132] J. L. Alty and D. I. Rigas, "Communicating graphical information to blind users using music: the role of context," Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 574-581, 1998.

[133] L. Brown, S. Brewster, R. Ramloll, W. Yu, and B. Riedel, "Browsing modes for exploring sonified line graphs," Proceedings of British HCI, vol. 2, pp. 2–5, 2002.

[134] D. I. Rigas, "Guidelines for Auditory Interface Design: An Empirical Investigation," PhD thesis, Loughborough University of Technology, 1996.

[135] D. I. Rigas and J. L. Alty, "The Use of Music in a Graphical Interface for the Visually Impaired," in Proceedings of the IFIP TC13 Interantional Conference on Human-Computer Interaction: Chapman & amp; Hall, Ltd., 1997.

[136] T. Stockman, "The Design and Evaluation of Auditory Access to Spreadsheets," Proceedings of ICAD, 2004.

[137] J. Kildal and S. A. Brewster, "Explore the Matrix: Browsing Numerical Data Tables Using Sound," Proc. ICAD2005, 2005.

[138] L. Brown and S. A. Brewster, "Drawing by ear: Interpreting sonified line graphs," Proceedings of ICAD 2003, pp. 152–156, 2003.

[139] J. L. Alty and D. Rigas, "Exploring the use of structured musical stimuli to communicate simple diagrams: the role of context," International Journal of Human-Computer Studies, vol. 62, pp. 21-40, 2005.

[140] D. Rigas and J. Alty, "The rising pitch metaphor: An empirical study," International Journal of Human Computer Studies, vol. 62, pp. 1-20, 2005.

[141] D. Rigas and I. Bahadur, "The Use of Audio in Stock Control Application," in Proceedings of the 10th WSEAS International Conference on Computers. Athens, Greece, 2006

[142] D. Rigas, H. Yu, and D. Memery, "Experiments Using Speech, Non-Speech Sound and Stereophony as Communication Metaphors in Information Systems," EUROMICRO, pp. 383-390, 2001.

[143] D. Rigas, H. Yu, D. Memery, and D. Howden, "Combining Speech with Sound to Communicate Information in a Multimedia Stock Control System," 2001.

[144] M. B. Alotaibi and D. I. Rigas, "A usability evaluation of multimodal metaphors for customer knowledge management," International Journal of Computers and Communications, University Press, UK, vol. 2, pp. 59-68, 2008.

[145] S. Alharbi and D. Rigas, "Email Threads: A Comparative Evaluation of Textual, Graphical and Multimodal Approaches," International Journal of Computers, NAUN, vol. 3, pp. 238-250, 2009.

[146] D. Rigas, D. Memery, and K. Klearhou, "A Multimedia E-mail Tool: An Empirical Case of using Auditory Metaphors to Communicate and Browse Data," Applied Informatics 2002, pp. 199–203, 2002.

[147] D. Rigas, H. Yu, K. Klearhou, and S. Mistry, "Designing Information Systems with Audio-Visual Synergy: Empirical Results of Browsing E-Mail Data," in Panhellenic Conference on Human-Computer Interaction: Advances on Human-Computer Interaction. Patras, Greece, 2001.

[148] D. I. Rigas and D. Memery, "Multimedia e-mail data browsing: the synergistic use of various forms of auditory stimuli," Proceedings of the International Conference on Information Technology: Computers and Communications (ITCC'03), pp. 582-586, 2003.

[149] D. I. Rigas and D. Memery, "Utilising audio-visual stimuli in interactive information systems: a two domain investigation on auditory metaphors," Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'02), pp. 190-195, 2002.

[150] W. W. Gaver, "Auditory Icons: Using Sound in Computer Interfaces," Human-Computer Interaction, vol. 2, pp. 167-177, 1986.

[151] W. W. Gaver, "Auditory interfaces," in Handbook of human-computer interaction, 2nd ed, 1997, pp. 1003-1041.

[152] W. W. Gaver, "The SonicFinder: An Interface That Uses Auditory Icons," Human-Computer Interaction, vol. 4, pp. 67-94, 1989.

[153] W. W. Gaver and R. B. Smith, "Auditory Icons in Large-Scale Collaborative Environments," ACM SIGCHI Bulletin, vol. 23, 1990.

[154] W. W. Gaver, R. B. Smith, and T. O'Shea, Effective sounds in complex systems: the ARKOLA simulation: ACM Press New York, NY, USA, 1991.

[155] S. A. Brewster, "Chapter 13: NonSpeech Auditory Output," In Sears, A. and Jacko, J. (Eds.), The Human Computer Interaction Handbook 2nd Edition: Lawrence Erlbaum Associates, USA, 2008, pp. 247-264

[156] S. Garzonis, S. Jones, T. Jay, and E. O'Neill, "Auditory icon and earcon mobile service notifications: intuitiveness, learnability, memorability and preference," in the 27th international conference on Human factors in computing systems: ACM New York, NY, USA, 2009, pp. 1513-1522.

[157] G. Leplâtre and S. A. Brewster, "Designing non-speech sounds to support navigation in mobile phone menus," Proceedings of the International Conference on Auditory Display (ICAD), 2000.

[158] D. I. Rigas, D. Hopwood, and D. Memery, "Communicating spatial information via a multimedia-auditory interface," EUROMICRO Conference, pp. 398-405, 1999.

[159] C. Bartneck, T. Takahashi, and Y. Katagiri, "Cross Cultural Study of Expressive Avatars," Proceedings of the Social Intelligence Design, 2004.

[160] M. D. Dickey, "3D virtual worlds: an emerging technology for traditional and distance learning," Proceedings Ohio Learning Network: The convergence of learning and technology—windows on the future; http://www. oln. org/conferences/OLN2003/papers/Dickey3DVirtualWorlds. pdf, 2003.

[161] J. Beskow, "Animation of Talking Agents," Proceedings of AVSP, vol. 97, pp. 149-152, 1997.

[162] R. Sheth, "Avatar Technology: Giving a Face to the e-Learning Interface," The eLearning Developers' Journal, 2003.

[163] B. Salem and N. Earle, "Designing a non-verbal language for expressive avatars," in Proceedings of the third international conference on Collaborative virtual environments. San Francisco, California, United States: ACM, 2000, pp. 93-101.

[164] M. C. Villa-Uriol, M. Sainz, F. Kuester, and N. Bagherzadeh, "Automatic creation of three-dimensional avatars," in SPIE-IS&T Electronic Imaging: Citeseer, 2003, pp. 14-25.

[165] D. Burford and E. Blake, "Real-time facial animation for avatars in collaborative virtual environments," in South African Telecommunications Networks and Applications Conference, vol. 99, 1999, pp. 178-183.

[166] B. Krenn, B. Neumayr, E. Gstrein, and M. Grice, "Lifelike Agents for the Internet: ACross-Cultural Case Study," Agent Culture: Human-Agent Interaction in a Multicultural World, 2004.

[167] E. Prasolova-Fřland and M. Divitini, "Supporting learning communities with collaborative virtual environments: Different spatial metaphors," International Conference on Advanced Learning Technologies (ICALT) 2002, pp. 259-264, 2002.

[168] D. Thalmann, "Virtual Humans In Virtual Environments: A New View of Mullimedia Applications," in Inrernationol Workshop on Synrhefic-Natural Hybrid Coding and Three Dimensimral Imaging. Rhodes, 1997, pp. 3-7.

[169] M. Fabri, D. J. Moore, and D. J. Hobbs, "The emotional avatar: Non-verbal communication between inhabitants of collaborative virtual environments," Lecture Notes in Computer Science, pp. 269-276, 2000.

[170] M. Fabri, D. Moore, and D. Hobbs, "Mediating the expression of emotion in educational collaborative virtual environments: an experimental study," Virtual Reality, vol. 7, pp. 66-81, 2004.

[171] M. Fabri, D. J. Moore, and D. J. Hobbs, "Face Value: Towards Emotionally Expressive Avatars," In proceeding of AISB Convention on Animated Expressive Characters for Social Interaction, London, UK, 2002.

[172] P. Ekman, W. V. Friesen, and P. C. Ellsworth, Emotion in the human face: Pergamon Press, New York, 1972.

[173] M. Fabri, D. J. Moore, and D. J. Hobbs, "Expressive Agents: Non-verbal Communication in Collaborative Virtual Environments," Proceedings of Autonomous Agents and Multi-Agent Systems (Embodied Conversational Agents), July, 2002.

[174] M. Fabri, S. Y. A. Elzouki, and D. Moore, "Emotionally Expressive Avatars for Chatting, Learning and Therapeutic Intervention," Lecture Notes in Computer Science, vol. 4552, pp. 275, 2007.

[175] M. Fabri and D. J. Moore, "Is empathy the key? Effective communication via instant messaging," Proceedings of the 11th EATA International Conference on Networking Entities, 2005.

[176] M. Fabri and D. Moore, "The use of emotionally expressive avatars in Collaborative Virtual Environments," Proceeding of Symposium on Empathic Interaction with Synthetic Characters, held at Artificial Intelligence and Social Behaviour Convention, 2005.

[177] S. Y. A. Elzouki, M. Fabri, and D. J. Moore, "Teaching severely autistic children to recognise emotions: finding a methodology," Proceedings of The 21st British HCI Group Annual Conference on HCI, vol. 2, pp. 137-140, 2007.

[178] J. Beskow, "Talking heads-communication, articulation and animation," presented at Fonetik 96, Sweedish Phonetics Conference, 1996.

[179] S. Fagel and C. Clemens, "Two Articulation Models for Audiovisual Speech Synthesis-Description and Determination," presented at AVSP 2003-International Conference on Audio-Visual Speech Processing, 2003.

[180] A. J. Cowell and K. M. Stanney, "On manipulating nonverbal interaction style to increase anthropomorphic computer character credibility," Proceedings of AAMAS 2003 workshop: Embodied conversational characters as individuals, 2003.

[181] A. Pease, "Body Language - How to read others' thoughts by their gestures," Sheldon Press, 1981.

[182] N. Gazepidis and D. Rigas, "Evaluation of Facial Expressions and Body Gestures in Interactive Systems," International Journal of Computers, NAUN, vol. 2, pp. 92-97, 2008.

[183] M. Alotaibi and D. Rigas, "The Role of Avatars with Facial Expressions to Communicate Customer Knowledge," International Journal of Computers, NAUN, vol. 3, pp. 1-10, 2009.

[184] A. J. Cowell and K. M. Stanney, "Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility," International journal of human-computer studies, vol. 62, pp. 281-306, 2005.

[185] H. M. McBreen and M. A. Jack, "Evaluating humanoid synthetic agents in e-retail applications," IEEE Transactions on Systems, Man and Cybernetics, Part A, vol. 31, pp. 394-405, 2001.

[186] L. Gong, "Is happy better than sad even if they are both non-adaptive? Effects of emotional expressions of talking-head interface agents," International Journal of Human-Computer Studies, vol. 65, pp. 183-191, 2007.

[187] B. Wild, M. Erb, and M. Bartels, "Are emotions contagious? Evoked emotions while viewing emotionally expressive faces: quality, quantity, time course and gender differences," Psychiatry Research, vol. 102, pp. 109-124, 2001.

[188] M. Garau, M. Slater, S. Bee, and M. A. Sasse, "The impact of eye gaze on communication using humanoid avatars," Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 309-316, 2001.

[189] J. Lumsden, S. Brewster, M. Crease, and P. D. Gray, "Guidelines for audio-enhancement of graphical user interface widgets," in Proceedings of British HCI, vol. 2002, 2002, pp. 6-9.

[190] D. Rigas and J. Alty, "How Can Multimedia Designers Utilise Timbre," in HCI-98, People and Computers XIII. Sheffield, UK: Springer, British Computer Society, 1998.

[191] D. A. Sumikawa, M. M. Blattner, K. I. Joy, and R. M. Greenberg, "Guidelines for the syntactic design of audio cues in computer interfaces," in 19th Annual Hawaii International Conference on System Sciences, 1986.

[192] M. Fabri, "Emotionally Expressive Avatars for Collaborative Virtual Environments," PhD Thesis, Leeds Metropolitan University, 2006.

[193] D. Rigas, "Empirically Derived Multimedia Design Guidelines for Browsing Large Volumes of E-Mail Data," Annals of Mathematics, Computing and Teleinformatics, vol. 1, pp. 77-83, 2003.

[194] L. M. Brown, S. A. Brewster, R. Ramloll, M. Burton, and B. Riedel, "Design guidelines for audio presentation of graphs and tables," in Proceedings of ICAD 2003 Workshop on Auditory Displays In Assistive Technologies. Boston, Massachusetts, USA, 2003.

[195] L. M. Reeves, J. C. Martin, M. McTear, T. V. Raman, K. M. Stanney, H. Su, Q. Y. Wang, J. Lai, J. A. Larson, and S. Oviatt, "Guidelines for multimodal user interface design," Communications of the ACM, vol. 47, pp. 57-59, 2004.

[196] A. Paivio, Mental representations: A dual coding approach. Oxford, UK: Oxford University Press, USA, 1990.

[197] R. Moreno and R. Mayer, "Interactive Multimodal Learning Environments," Educational Psychology Review, vol. 19, pp. 309-326, 2007.

[198] G. Kramer, B. Walker, T. Bonebright, P. Cook, J. Flowers, N. Miner, and J. Neuhoff, "Sonification Report: Status of the Field and Research Agenda.," National Science Foundation, 1999.

[199] R. E. Mayer, "principles for managing essential processing multimedia learning: segmenting, pretraining, and modality principles," in Cambridge handbook of mulrimedial learning, R. Mayer, Ed. New York: Cambridge University Press, 2005, pp. 169-182.

[200] R. E. Mayer, "Introduction to multimedia learning," in Cambridge handbook of multimedia learning, R. Mayer, Ed. New York: Cambridge University Press, 2005, pp. 1–16.

[201] R. Moreno, R. E. Mayer, H. A. Spires, and J. C. Lester, "The case for social agency in computer-based teaching: do students learn more deeply when they

interact with animated pedagogical agents?," Cognition and Instruction, vol. 19, pp. 177-213, 2001.

[202] R. Moreno and R. E. Mayer, "Learning Science in Virtual Reality Multimedia Environments: Role of Methods and Media," Journal of Educational Psychology, vol. 94, pp. 598-610, 2002.

[203] Z. Li, H. Qiu, and Y. Yue, "Development of a Learning-Training Simulator with Virtual Functions for Lathe Operations," Virtual Reality, vol. 6, pp. 96-104, 2002.

[204] S. Rashid and D. Rigas, "An Initial Empirical Study into E-Note Taking," IADIS International Conference Interfaces and Human-Computer Interactions, Lisbon, Portugal, pp. 161-164, 2007.

[205] S. Rashid and D. Rigas, "A two-group evaluation to e-note," International Journal of Education and Information Technologies, NAUN, vol. 2, pp. 7-11, 2008.

[206] C. Elliott, J. Rickel, and J. Lester, "Lifelike pedagogical agents and affective computing: An exploratory synthesis," Artificial Intelligence Today: Recent Trends and Developments, pp. 195-211, 1999.

[207] M. Moundridou and M. Virvou, "Evaluating the persona effect of an interface agent in an intelligent tutoring system," Journal of Computer Assisted Learning, vol. 18, pp. 253-261, 2002.

[208] S. Van Mulken, E. André, and J. Müller, "The persona effect: how substantial is it?," in People and Computers XIII: Proceedings of HCI'98, H. Johnson, L. Nigay, and C. Roast, Eds., 1998, pp. 53-66.

[209] A. Gulz, "Social enrichment by virtual characters-differential benefits," Journal of Computer Assisted Learning, vol. 21, pp. 405-418, 2005.

[210] J. M. Moshell and C. E. Hughes, "Virtual environments as a tool for academic learning," in Handbook of virtual environments, K. A. Stanney, Ed.: Lawrence Erlbaum Ssociate, Inc, 2002, pp. 893-910.

[211] J. Robertson, B. Cross, H. Macleod, and P. Wiemer-Hastings, "Children's interactions with animated agents in an intelligent tutoring system," International Journal of Artificial Intelligence in Education, vol. 14, pp. 335-357, 2004.

[212] W. L. Johnson, S. Kole, E. Shaw, and H. Pain, "Socially intelligent learner-agent interaction tactics," Proceedings of the International Conference on Artificial Intelligence in Education, 2003.

[213] Q. Dunsworth and R. K. Atkinson, "Fostering multimedia learning of science: Exploring the role of an animated agent's image," Computers & Education, vol. 49, pp. 677-690, 2007.

[214] G. Theonas, D. Hobbs, and D. Rigas, "The Effect of Facial Expressions on Students in Virtual Educational Environments," International Journal of Social Sciences, vol. 2, pp. 42-49, 2008.

[215] H. M. Deitel and P. J. Deitel, Visual Basic 2005: how to program: Prentice Hall, 2006.

[216] T. C. Lethbridge and R. Laganiere, Object-oriented software engineering: McGraw-Hill Education, UK, 2001.

[217] http://www.geocities.com/siliconvalley/network/1582/uml-example.htm, May 2007

[218] http://www.hep.wisc.edu/~pinghc/classDiagram.htm, May 2007

[219] P. R. Michaelis and R. H. Wiggins, "A human factors engineer's introduction to speech synthesizers," Directions in Human±Computer Interaction, 1982.

[220] http://www.shitalshah.com/vmusic, September 2007

[221] http://www.mp3do.com/soundrecorder.html, September 2007

[222] http://audacity.sourceforge.net, October 2007

[223] http://www.daz3d.com, October 2007

[224] http://www.smithmicro.com, October 2007

[225] A. P. Field, "Repeated-Measures Designs," Discovering statistics using SPSS, 2nd edition, SAGE Publications, pp. 427-482, 2005.

[226] M. Alotaibi and D. Rigas, "A Usability Evaluation of Multimodal Metaphors for Customer Knowledge Management," International Journal of Computers and Communications, universitypress, UK, vol. 2, 2008.

[227] D. Rigas and A. Ciuffreda, "An empirical Investigation of Multimodal Interfaces for Browsing Internet Search Results," Proceedings of the 7th WSEAS International Conference on Applied Informatics & Communication, pp. 194-199, 2007.

[228] N. J. Salkind, "The Research Process: Coming to Terms," in Exploring research, Sixth ed: Prentice Hall, Upper Saddle River, NJ, 2006, pp. 19-32.

[229] J. H. Kranzler, "Frequency Distributions," Statistics For The Terrified, 3rd edition, Prentice Hall, pp. 36-37, 2003.

[230] J. Brooke, "SUS: a" quick and dirty" usability scale," Usability evaluation in industry, pp. 189-194, 1996.

[231] A. P. Field, "Comparing two means," Discovering statistics using SPSS, 2nd edition, SAGE Publications, pp. 269-308, 2005.

[232] A. P. Field, "Exploring data," Discovering statistics using SPSS, 2nd edition, SAGE Publications, pp. 63-106, 2005.

[233] A. P. Field, "Non-parametric tests," Discovering statistics using SPSS, 2nd edition, SAGE Publications, pp. 521-570, 2005.

[234] J. H. Kranzler, "Nonparametric Statistics: Chi-Square," Statistics For The Terrified, 3rd edition, Prentice Hall, pp. 160-165, 2003.

[235] M. D. Sankey and M. Nooriafshar, "Multiple representations in multimedia and e-learning materials: an issue of literacy," Enhancing learning and teaching: Pedagogy, technology and language, pp. 149-171, 2005.

[236] J. D. Fletcher and S. Tobias, "The multimedia principle," The Cambridge handbook of multimedia learning, pp. 117–133, 2005.

[237] http://www.techsmith.com, July 2008

[238] http://ccoder.com/GainingFace, January 2008

[239] A. P. Field, Discovering statistics using SPSS: Sage Publications Inc, 2005.

[240] D. Rigas, D. Memery, and H. Yu, "Experiments in using structured musical sound, synthesised speech and environmental stimuli to communicate information: is there a case for integration and synergy?," Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing, pp. 465-468, 2001.

[241] A. Brown, "Learning from a distance, " Journal of Property Management, vol. 71, pp. 42-45, 2006.

[242] S. H. Liu, H. L. Liao, and J. A. Pratt, "Impact of media richness and flow on e-learning technology acceptance, " Computers & Education, vol. 52, pp. 599-607, 2009.

[243] D. Kopf, "e-Learning Market to hit $52.6B by 2010, " The Journal, 2007.

APPENDICES

Appendix A – First Experiment (An Empirical Investigation into the Use of Multimodal E-Learning Interfaces)

A1: Questionnaire

I am pleased to present my self to you as one of the postgraduate research students in the School of Informatics in the University of Bradford. I am currently investigating the use of multimodal metaphors in e-learning interfaces, and I would like to obtain your views regarding the use of such multimodal metaphors such as: avatar with facial expressions, recorded speech, and non-speech sounds.

Please follow the following procedure:

- Answer the pre-experiment questions.
- Read of each task carefully.
- Perform the task.
- On completion of the task, answer the required related questions.
- After completion of all tasks, answer the satisfaction questionnaire.

Please answer all the questions as truthfully as possible. It would be grateful if you could fill in the following questionnaire sincerely and provide your views. Your privacy is guaranteed as your name will not be mentioned in any part of the study.

Thank you very much, and I highly appreciate your participation.

Name: ……………………………………………………….

Part 1	Pre-experiment Questions

What is your age?

☐ 18 - 24. ☐ 25 - 34. ☐ 35 - 44. ☐ 45 - 54. ☐ 55 +.

What is you gender?

☐ Male ☐ Female

What is your education level?

☐ High School ☐ College ☐ Under-graduate

☐ Master Degree ☐ Doctorate Degree ☐ Other

Area of study: ………………………………………………………..

How often do you use the computer (average) per week?

☐ Never ☐ Less than 1 hour ☐ 1-5 hours

☐ 6-10 hours ☐ More than 10 hours

How many hours do you use the internet (average) per week?

☐ Never ☐ Less than 1 hour ☐ 1-5 hours

☐ 6-10 hours ☐ More than 10 hours

Do you have knowledge about Object Orientation?

☐ Limited ☐ Good ☐ Excellent ☐ No

Do you have knowledge about Class Diagram notation?

☐ Limited ☐ Good ☐ Excellent ☐ No

Did you practice the use of any e-learning web sites or software?

☐ Yes ☐ No

Do you have knowledge about avatar and facial expressions?

☐ Limited ☐ Good ☐ Excellent ☐ No

Part 2	Tasks and related questions

In example 1, you have a problem statement talks about the elevator controller system which is used to control the use of elevators (lifts) in a building of many floors. The class diagram of this problem appears in the right hand side of the screen. You are asked to answer some questions related to the presented information.

Task 1:

To accomplish this task, perform the following requirements and communicate the presented information:

- Move the mouse cursor over the class named "Door".
- Move the mouse cursor over the multiplicity of “Door” class.
- Move the mouse over the arrow > of the association between “Door” class and the "Elevator_Controller" class.

Now, please answer the following questions:

Q1.1 What is the main attribute of "Door" class?

…………………………………

Q1.2 Which of the following statements is TRUE?

□ multiplicity of "Door" class is 1..*

□ the association between "Door" class and "Elevator_Controller" class is directed

<u>Task 2:</u>

To accomplish this task, perform the following requirements and communicate the presented information:

- Move the mouse cursor over the class named "Elevator".
- Move the mouse cursor over the multiplicity of “Elevator”.
- Move the mouse over the verb "control" between "Elevator" class and "Elevator_Controller" class.

Now, please answer the following two questions:

Q2.1 Write down only one operation that could be performed on the elevator?

…………………………….

Q2.2 Which of the following statements is FALSE?

□ in its association with "Elevator_Controller" class, "Elevator" class is the subject.

□ 1..* is the multiplicity of "Elevator" class.

In example 2, a brief description for writing a text document appears in the "Statement of Problem" box. The class diagram of this description appears in the right hand side of the screen. You are asked to answer some questions related to the presented information.

<u>Task 3:</u>

To accomplish this task, perform the following requirements and communicate the presented information:

- Move the mouse cursor over the "Page" class and read the notes.
- Move the mouse cursor over the multiplicity of "Page" class in its association with "Document" class.
- Move the mouse cursor over the multiplicity of "Page" class in its association with ("BottomUp", "Character", "Table", "Picture") classes.
- Move the mouse cursor over the diamond notation of "Page" class.

Now, please answer the following questions:

Q3.1 What are the multiplicities of "Page" class in its association

with "Document" class? ……………………………

with "Character" class? ……………………………

Q3.2 Which of the following classes is not related in aggregation association with "Page" class?

□ Table □ Picture

□ Document □ BottomUp

<u>Task 4:</u>

To accomplish this task, perform the following requirements and communicate the presented information:

- Move the mouse cursor over the "Table" class and read the notes.
- Move the mouse cursor over the multiplicity of "Table" class in its association with "Page" class.
- Move the mouse cursor over the multiplicity of "Table" class in its association with "Cell" class.
- Move the mouse cursor over the diamond notation of "Table" class.

Then, please answer the following questions:

Q4.1 Write down 2 operations that can be performed on the table?

1- ……………………………

2- ……………………………

Q4.2 The class "Table" is the 'whole' side in the association with one of the following?

□ Page class □ Character class

□ Picture class □ Cell class

In example 3, a Bank System is introduced and represented in a class diagram. You are asked to answer some questions related to the presented information

<u>Task 5:</u>

To accomplish this task, perform the following requirements and communicate the presented information:

- Move the mouse cursor over the "Employee" class and read the notes.
- Move the mouse cursor over the multiplicity of the "Employee" class in its association as a manager with "OrganizationalUnit" class.
- Move the mouse cursor over the multiplicity of the "Employee" class in its other association with "OrganizationalUnit" class.
- Move the mouse cursor over the multiplicity of the "Employee" class in its association with "Customer" class.
- Move the mouse cursor over the phrase "personalBanker".
- Move the mouse cursor over the phrase "manager".

Now, please answer the following questions:

Q5.1 Write down 3 attributes of each employee.

1- …………………..

2- …………………..

3- ………………….

Q5.2 The class "Employee" has 3 multiplicities. Which of the following is not a multiplicity of "Employee"?

☐ 1 ☐ 1..* ☐ * ☐ 0..1

<u>Task 6:</u>

To accomplish this task, perform the following requirements and communicate the presented information:

- Move the mouse cursor over the "Account" class and read the notes.
- Move the mouse cursor over the multiplicity of the "Account" class in its association with "Customer" class.
- Move the mouse cursor over the multiplicity of the "Account" class in its association with "Branch" class.
- Move the mouse cursor over the multiplicity of the "Account" class in its association with "AccountType" class.
- Move the mouse cursor over the triangle symbol of the “Account” class.

Now, please answer the following questions:

Q6.1 Write down 3 operations performed on each account.

1- …………………..

2- …………………..

3- ………………….

Q6.2 Two of the following statements are TRUE? Tick them.

- ☐ "Account" is the super class in its association with "MortgageAccount" class
- ☐ Multiplicity of "Account" class in its association with "Branch" is *
- ☐ "CreditCard" class is a subclass of "Account" class

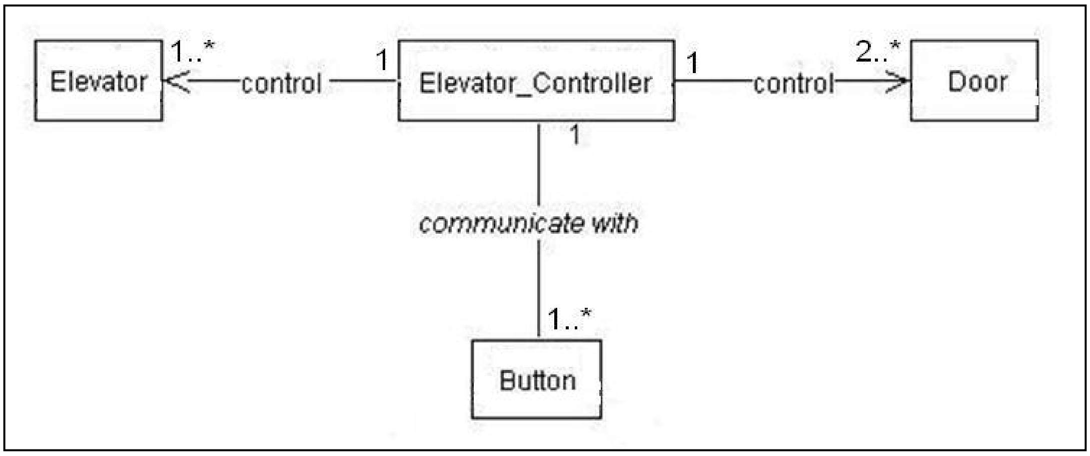
Part 3	Satisfaction

For each statement below, please express your view by placing a tick in the appropriate column.

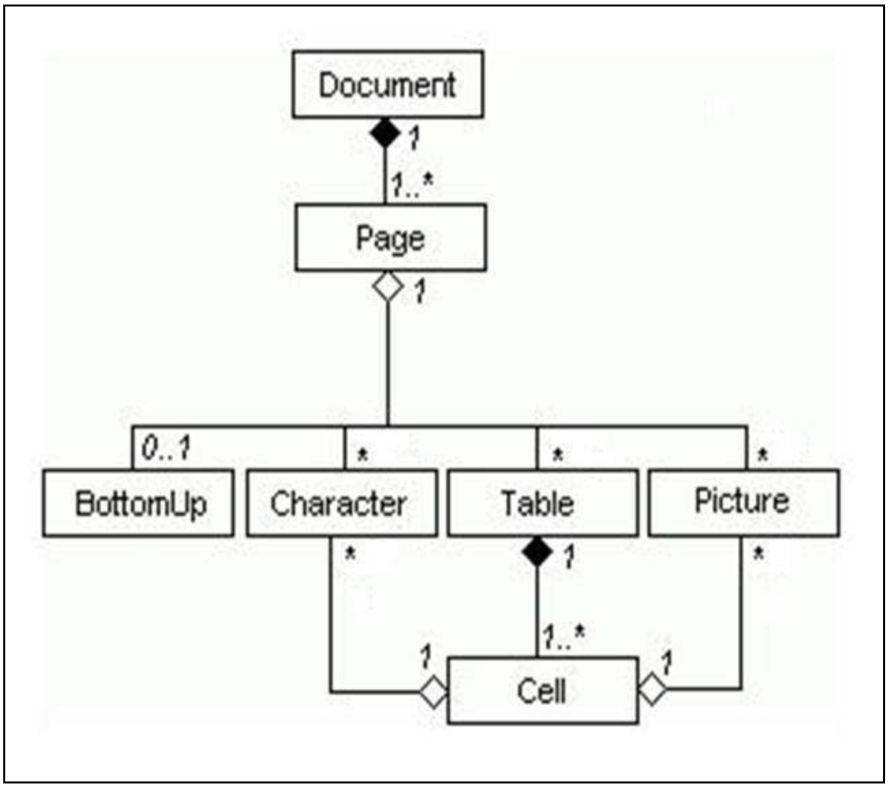
	Statements	Agreement			Disagreement		
		Strongly Agree	Moderately Agree	Slightly Agree	Slightly Disagree	Moderately Disagree	Strongly Disagree
S1	I think the interface was easy to use.	6	5	4	3	2	1
S2	I think the interface is confusing.	6	5	4	3	2	1
S3	There have been times while interacting with the system where I felt nervous.	6	5	4	3	2	1
S4	I think that most people will learn the use of this tool quickly.	6	5	4	3	2	1
S5	It was easy to identify the information about the class. (attributes and operations)	6	5	4	3	2	1
S6	It was easy to identify the multiplicity of the class.	6	5	4	3	2	1
S7	It was easy to identify the information about the association between classes. (aggregation, composition, inheritance, directed, not directed, association name, and role name)	6	5	4	3	2	1
S8	Overall, I am satisfied with the interface.	6	5	4	3	2	1

Thank you very much for your patience and generous help.

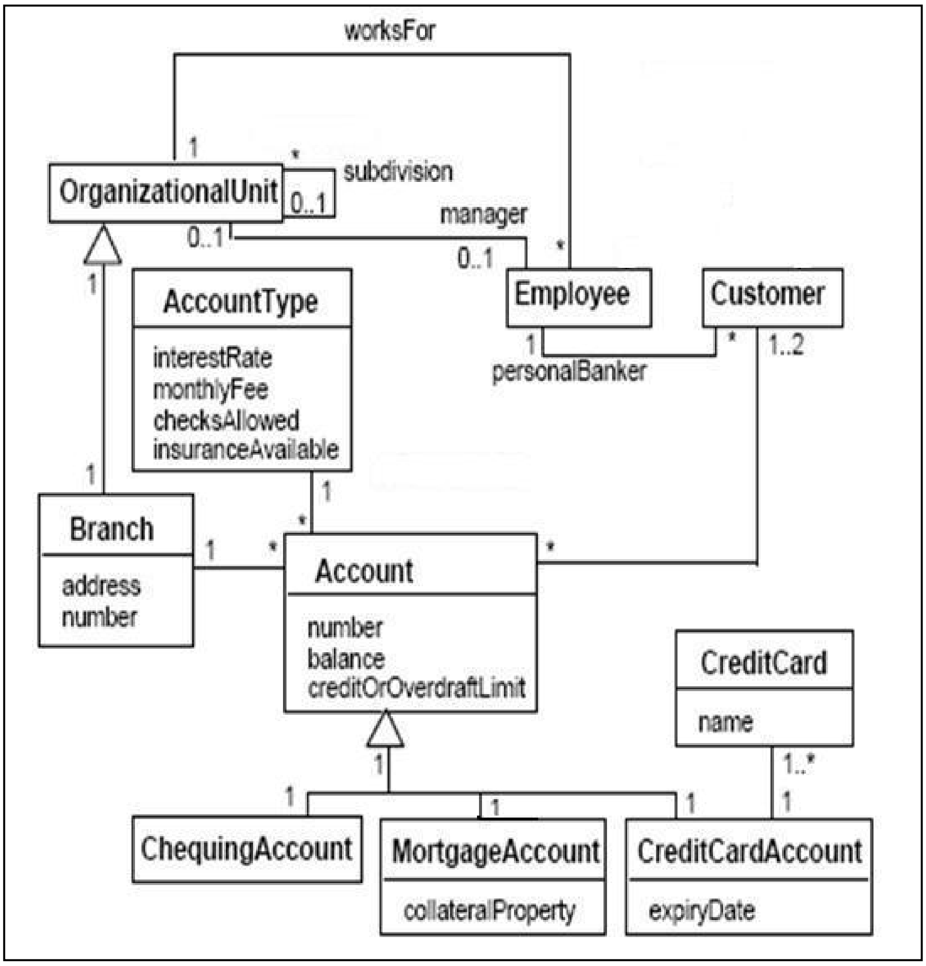
A2: Class Diagram Examples



Example 1: Elevator Controller System



Example 2: Document



Example 3: Bank System

A3: Frequency Table for Users' Characteristics

		Control group (VOELP		Experimental group (MMELP)	
Age	18-24	2	13.3%	1	6.7%
	25-34	9	60.0%	10	66.7%
	35-44	3	20.0%	4	26.7%
	45-55	1	6.7%	0	0.0%
	Total	15	100.0%	15	100.0%
Gender	Male	13	86.7%	12	80.0%
	Female	2	13.3%	3	20.0%
	Total	15	100.0%	15	100.0%
Education level	Under-graduate	1	6.7%	1	6.7%
	Master	13	86.7%	12	80.0%
	PhD	1	6.7%	2	13.3%
	Total	15	100.0%	15	100.0%
Area of study	Management	3	20.0%	4	26.7%
	Engineering	7	46.7%	6	40.0%
	Computing	5	33.3%	5	33.3%
	Total	15	100.0%	15	100.0%
Use of computer/week	1-5 hours	0	0.0%	1	6.7%
	6-10 hours	3	20.0%	3	20.0%
	10+ hours	12	80.0%	11	73.3%
	Total	15	100.0%	15	100.0%
Use of Internet/week	1-5 hours	1	6.7%	3	20.0%
	6-10 hours	4	26.7%	0	0.0%
	10+ hours	10	66.7%	12	80.0%
	Total	15	100.0%	15	100.0%
Knowledge about object orientation	No	11	73.3%	10	66.7%
	Limited	3	20.0%	4	26.7%
	Good	1	6.7%	1	6.7%
	Total	15	100.0%	15	100.0%
Knowledge about class diagram notation	No	13	86.7%	14	93.3%
	Limited	2	13.3%	1	6.7%
	Total	15	100.0%	15	100.0%
Prior experience in e-learning applications?	yes	8	53.3%	9	60.0%
	No	7	46.7%	6	40.0%
	Total	15	100.0%	15	100.0%

Experience on avatars and facial expressions	No	12	80.0%	11	73.3%
	Limited	3	20.0%	1	6.7%
	Good	0	0.0%	1	6.7%
	Excellent	0	0.0%	2	13.3%
	Total	15	100.0%	15	100.0%

A4: Raw Data of Question Answering Time

Raw Data of Question Answering Time for the Control Group (VOELP)

	Easy tasks				Moderate tasks				Difficult tasks				Question Complexity			Question Type		
	Task 1		Task 2		Task 3		Task 4		Task 5		Task 6		Easy	Moderate	Difficult	Recall	Recognition	
Users	Q1.1	Q1.2	Q2.1	Q2.2	Q3.1	Q3.2	Q4.1	Q4.2	Q5.1	Q5.2	Q6.1	Q6.2						Total
U01	15.3	5.4	20.5	10.8	19.1	16.9	17.7	5.5	13.1	8.0	23.9	19.1	52.0	59.2	64.2	109.7	65.7	175.4
U02	30.4	7.7	18.0	8.7	49.1	14.1	30.5	12.2	18.8	15.6	27.0	31.4	64.8	105.8	92.8	173.8	89.7	263.5
U03	11.3	7.2	24.7	9.6	31.2	14.3	34.2	7.3	16.8	11.5	30.3	34.4	52.8	87.0	93.0	148.5	84.3	232.8
U04	51.7	14.6	31.0	21.3	28.6	12.3	25.3	13.0	38.4	15.5	62.1	35.1	118.5	79.2	151.1	237.1	111.8	348.8
U05	49.8	22.5	30.6	19.8	27.3	24.1	26.1	9.8	39.9	31.0	55.0	39.0	122.7	87.3	164.9	228.7	146.1	374.8
U06	24.1	36.0	13.2	16.4	59.1	27.7	34.0	27.3	63.0	20.0	50.9	35.5	89.7	148.0	169.4	244.2	162.8	407.0
U07	23.7	15.5	18.7	40.8	37.1	22.7	30.1	12.4	20.4	12.8	42.8	44.1	98.7	102.3	120.0	172.7	148.2	320.9
U08	40.4	28.2	14.7	32.3	54.2	24.0	56.9	43.5	82.8	48.1	60.9	45.7	115.6	178.6	237.5	309.9	221.7	531.6
U09	39.7	16.1	16.4	20.7	50.9	16.8	28.7	10.9	34.7	25.2	33.8	32.1	92.9	107.2	125.8	204.2	121.7	325.9
U10	18.5	17.6	18.9	15.9	52.5	21.3	14.7	15.6	48.7	26.4	42.1	19.1	70.9	104.1	136.3	195.5	115.8	311.3
U11	23.7	16.9	24.2	11.2	28.7	7.1	23.6	17.4	25.9	15.3	40.9	22.7	76.0	76.8	104.8	166.9	90.7	257.6
U12	36.4	15.5	43.9	16.0	67.2	20.3	27.3	22.3	16.2	30.1	71.3	41.2	111.9	137.2	158.8	262.4	145.6	408.0
U13	19.7	8.3	23.9	23.9	36.8	25.4	37.2	9.4	20.2	20.7	31.3	18.6	75.8	108.9	90.8	169.1	106.4	275.6
U14	63.2	12.0	22.1	16.9	49.6	13.3	54.3	17.8	28.7	29.4	119.3	26.3	114.1	135.0	203.7	337.2	115.6	452.8
U15	55.5	12.0	11.3	8.3	35.1	13.1	24.5	14.6	28.0	10.2	31.9	13.6	87.1	87.3	83.7	186.3	71.7	258.0

Raw Data of Question Answering Time for the Experimental Group (MMELP)

Users	Easy tasks				Moderate tasks				Difficult tasks				Question Complexity			Question Type		Total
	Task 1		Task 2		Task 3		Task 4		Task 5		Task 6		Easy	Moderate	Difficult	Recall	Recognition	
	Q1.1	Q1.2	Q2.1	Q2.2	Q3.1	Q3.2	Q4.1	Q4.2	Q5.1	Q5.2	Q6.1	Q6.2						
U01	44.8	11.8	11.8	20.1	28.0	12.4	30.0	9.3	57.5	7.3	45.3	17.5	88.5	79.7	127.6	217.4	78.4	295.8
U02	50.8	10.4	7.5	19.4	20.4	12.0	50.2	11.3	19.9	14.8	38.5	22.5	88.1	93.9	95.8	187.3	90.5	277.8
U03	11.5	22.5	25.2	28.3	28.7	15.0	41.1	71.8	18.9	36.6	41.6	21.6	87.6	156.6	118.7	167.1	195.9	363.0
U04	9.6	17.4	16.8	22.6	19.4	26.8	22.9	7.2	25.2	15.5	43.1	13.3	66.5	76.3	97.1	137.0	102.8	239.9
U05	35.1	9.7	11.9	8.1	25.1	11.6	9.2	9.3	17.4	8.1	24.5	3.3	64.7	55.2	53.2	123.1	50.0	173.1
U06	28.0	12.8	11.3	12.7	22.6	7.9	23.6	14.2	22.9	20.6	17.6	28.7	64.9	68.4	89.7	126.0	96.9	222.9
U07	31.2	19.2	26.3	23.1	47.2	16.5	25.6	38.1	22.8	30.8	24.3	26.1	99.9	127.4	104.0	177.5	153.8	331.3
U08	13.0	16.2	10.0	12.1	28.5	13.4	30.7	11.2	47.3	13.1	44.4	14.6	51.4	83.7	119.4	173.9	80.6	254.4
U09	35.4	21.4	21.5	24.6	20.3	19.0	22.6	11.3	52.3	26.4	21.3	17.4	102.9	73.3	117.5	173.5	120.1	293.7
U10	25.6	15.6	37.8	23.9	15.5	17.0	35.8	12.3	25.1	15.1	35.9	12.6	102.9	80.6	88.7	175.7	96.5	272.2
U11	12.8	7.9	10.5	9.7	19.2	18.9	30.9	21.8	10.1	8.2	22.3	3.3	40.8	90.7	43.9	105.8	69.7	175.5
U12	9.2	11.9	11.0	8.5	20.0	15.3	40.2	16.3	21.3	23.4	50.4	16.0	40.6	91.8	111.1	152.0	91.5	243.5
U13	21.0	10.3	7.7	11.8	13.5	9.3	21.0	13.4	31.6	6.6	14.9	23.5	50.7	57.2	76.6	109.7	74.9	184.6
U14	14.3	13.9	11.7	22.1	32.3	12.0	26.9	20.8	32.4	11.6	35.6	28.7	62.1	91.9	108.2	153.1	109.1	262.2
U15	6.2	10.2	25.6	11.0	13.1	8.2	29.5	14.7	37.4	18.3	40.9	11.9	53.0	65.5	108.5	152.8	74.2	227.0

A5: Raw Data of Answers' Correctness

Raw Data of Answers' Correctness for the Control Group (VOELP)

1: Correct answer, 0: Incorrect answer

	Easy tasks				Moderate tasks				Difficult tasks				Question Complexity			Question Type		
	Task 1		Task 2		Task 3		Task 4		Task 5		Task 6		Easy	Moderate	Difficult	Recall	Recognition	
Users	Q1.1	Q1.2	Q2.1	Q2.2	Q3.1	Q3.2	Q4.1	Q4.2	Q5.1	Q5.2	Q6.1	Q6.2						Total
U01	1	1	1	1	0	1	1	1	1	1	1	0	4	3	3	5	5	10
U02	0	1	1	1	0	0	0	1	0	1	0	0	3	1	1	1	4	5
U03	0	0	0	0	1	0	0	1	0	0	0	1	0	2	1	1	2	3
U04	0	1	1	1	0	1	1	1	0	0	1	1	3	3	2	3	5	8
U05	1	1	1	0	1	1	1	1	0	0	0	0	3	4	0	4	3	7
U06	1	1	1	1	1	0	1	1	0	1	0	1	4	3	2	4	5	9
U07	0	1	1	1	1	0	1	1	1	0	1	1	3	3	3	5	4	9
U08	0	1	1	1	1	0	0	1	0	0	0	1	3	2	1	2	4	6
U09	0	0	1	1	1	1	1	1	0	0	0	0	2	4	0	3	3	6
U10	1	1	1	1	0	1	1	0	0	1	1	1	4	2	3	4	5	9
U11	0	0	1	0	1	0	1	1	0	0	0	1	1	3	1	3	2	5
U12	0	1	1	1	0	1	1	1	1	1	0	0	3	3	2	3	5	8
U13	0	1	0	1	0	1	0	1	1	0	1	1	2	2	3	2	5	7
U14	0	1	1	1	0	0	1	1	0	0	0	0	3	2	0	2	3	5
U15	0	1	1	0	0	1	1	1	0	1	1	0	2	3	2	3	4	7

Raw Data of Answers' Correctness for the Experimental Group (MMELP)

1: Correct answer, 0: Incorrect answer

	Easy tasks				Moderate tasks				Difficult tasks				Question Complexity			Question Type		
	Task 1		Task 2		Task 3		Task 4		Task 5		Task 6		Easy	Moderate	Difficult	Recall	Recognition	
Users	Q1.1	Q1.2	Q2.1	Q2.2	Q3.1	Q3.2	Q4.1	Q4.2	Q5.1	Q5.2	Q6.1	Q6.2						Total
U01	0	1	1	1	1	1	1	1	0	1	0	1	3	4	2	3	6	9
U02	1	0	1	1	1	1	0	1	1	1	1	0	3	3	3	5	4	9
U03	1	1	1	1	1	0	1	0	1	0	0	1	4	2	2	5	3	8
U04	1	0	1	0	1	0	1	1	1	0	1	1	2	3	3	6	2	8
U05	1	1	1	1	1	1	1	1	1	1	1	1	4	4	4	6	6	12
U06	0	1	1	1	1	1	0	1	1	0	1	1	3	3	3	4	5	9
U07	1	1	1	0	1	0	1	0	1	1	1	0	3	2	3	6	2	8
U08	1	1	1	1	1	1	1	1	1	1	1	1	4	4	4	6	6	12
U09	0	1	1	1	1	1	1	1	0	1	1	1	3	4	3	4	6	10
U10	0	1	1	1	1	1	1	1	1	1	1	1	3	4	4	5	6	11
U11	1	1	1	1	1	1	1	1	1	0	1	1	4	4	3	6	5	11
U12	1	1	1	1	1	1	1	1	1	1	1	1	4	4	4	6	6	12
U13	0	1	1	1	1	0	1	1	1	1	1	0	3	3	3	5	4	9
U14	1	1	0	1	1	1	1	1	1	1	1	1	3	4	4	5	6	11
U15	0	1	1	1	1	1	1	1	0	1	1	1	3	4	3	4	6	10

A6: Raw Data of Users' Responses to the Satisfaction Questionnaire

S1 – S2: Statements of the satisfaction questionnaire shown in Appendix A1

6: Strongly Agree, 5: Moderately Agree, 4: Slightly Agree, 3: Slightly Disagree, 2: Moderately Disagree, 1: Strongly Disagree

	Control Group (VOELP)								
Users	S1	S2	S3	S4	S5	S6	S7	S8	SUS score
U01	5	2	2	5	5	4	5	5	77.5
U02	5	3	3	4	5	5	4	5	70
U03	5	2	3	4	5	5	4	5	72.5
U04	5	1	4	4	5	2	2	5	60
U05	2	5	5	2	2	2	2	2	20
U06	5	2	2	3	2	2	3	4	52.5
U07	6	5	5	5	5	5	6	6	72.5
U08	5	2	3	5	4	5	5	5	75
U09	5	2	2	5	5	5	5	5	80
U10	6	2	2	6	5	4	3	5	77.5
U11	4	4	3	3	4	3	3	4	50
U12	5	4	2	4	5	4	5	5	70
U13	6	2	1	6	6	6	5	6	95
U14	6	3	2	5	3	5	5	5	75
U15	5	2	2	5	4	5	5	5	77.5

	Experimental Group (MMELP)								
Users	S1	S2	S3	S4	S5	S6	S7	S8	SUS score
U01	5	4	4	6	5	3	6	5	70
U02	6	2	1	5	6	6	6	6	95
U03	6	2	1	5	6	5	6	5	90
U04	6	1	1	5	6	6	5	6	95
U05	6	1	2	6	6	6	6	6	97.5
U06	6	2	1	5	6	6	5	6	92.5
U07	4	4	4	5	5	6	5	5	70
U08	5	2	1	5	6	6	6	5	90
U09	6	1	1	6	6	5	6	6	97.5
U10	6	2	2	5	5	5	5	5	82.5
U11	5	3	1	6	5	4	4	4	75
U12	6	1	2	6	5	5	5	6	90
U13	3	4	5	4	4	5	2	4	47.5
U14	5	2	3	5	5	5	5	5	77.5
U15	6	1	1	5	6	6	4	6	92.5

Appendix B – Second Experiment (Investigating the Role of Avatars in the Multimodal E-Learning Interfaces)

B1 – Questionnaire

I am pleased to present my self to you as one of the postgraduate research students in the School of Informatics in the University of Bradford. I am currently investigating the use of multimodality in e-learning human computer interaction, and I would like to obtain your views regarding the use of avatar and natural recorded speech in e-learning interfaces.

Please answer all the questions as truthfully as possible. It would be grateful if you could fill in the following questionnaire sincerely and provide your views. Your privacy is guaranteed as your name will not be mentioned in any part of the study.

Thanks for your co-operation, and your participation is highly appreciated.

Name: ____

Part 1	Pre-experiment questions

Age: ☐ 18 - 24 ☐ 25 - 34 ☐ 35 - 44 ☐ 45 - 54 ☐ 55 +

Gender: ☐ Male ☐ Female

Residence: ☐ Home ☐ Overseas

English is: ☐ First language ☐ Second language

Education level: ☐ High School ☐ College ☐ Under-graduate
☐ Master ☐ PhD ☐ Other…………………………………..

Area of study: ____

How often do you use the computer per week?
☐ Never ☐ Less than 1 hour ☐ 1-5 hours
☐ 6-10 hours ☐ More than 10 hours

How often do you use the internet per week?
☐ Never ☐ Less than 1 hour ☐ 1-5 hours
☐ 6-10 hours ☐ More than 10 hours

Do you have knowledge about Class Diagram notation?

 Yes (Limited Good Excellent)

 No

Do you have knowledge about avatars, facial expressions and body gestures?

 Yes (Limited Good Excellent)

 No

Did you use any e-learning web sites or software?

 Yes

 No

If yes, how often per week? Less than 1 hour 1-5 hours

 6-10 hours More than 10 hours

Did you miss the face-to-face interaction in these e-learning websites or software?

 Yes No

Do you think that the addition of virtual lecturer with facial expressions might help you in e-learning? Yes No

Do you think that the addition of virtual lecturer with body gestures might help you in e-learning? Yes No

Circle as appropriate the importance rate for each of the following elements in e-learning interfaces.

(VI=Very Important, I=Important, U=Unimportant, VU=Very Unimportant)

	VI	I	U	VU
Speech output	1	2	3	4
Virtual lecturer with facial expressions	1	2	3	4
Virtual lecturer with body gestures	1	2	3	4

Part 2	Obtaining users' overall evaluation in regard to FACIAL EXPRESSIONS & BODY GESTURES presented without interactive context/speech

Assume that you are using e-learning interface which employs virtual lecturer with facial expressions. How would you feel with the following facial expressions? Positively or Negatively?

		Rating	
	Facial Expressions	Positive	Negative
1	Neutral		
2	Interested		
3	Amazed		
4	Happy		
5	Smiling		
6	Thinking		

Please indicate two expressions that you most liked/disliked in order of preferences.

Liked: A __________________ B _________________

Disliked: A __________________ B _________________

Assume that you are using e-learning interface which employs virtual lecturer with body gestures. How would you feel with the following body gestures? Positively or Negatively?

		Rating	
	Body Gestures	Positive	Negative
1	Neutral		
2	Hands Clenching - front		
3	Hands Clenching - back		
4	Open Palms		
5	Arms Folded		
6	Pointing		
7	Hands Steepling		
8	Chin Stroking		
9	Legs Crossed		
10	Moving forward & backward		

Please indicate two gestures that you most liked/disliked in order of preferences.

Liked: A __________________ B _________________

Disliked: A __________________ B _________________

Part 3	Presentation of lesson 1 (Classes and objects) VIRTUAL LECTURER with FACIAL EXPRESSIONS (VLFE)

Now, information about classes and objects will be presented in lesson 1. Please pay your attention to the interface and the presented information.

Depending on the presented information, please answer the following questions?

Q1: There are three main symbols shown in class diagram. Write down only one of them.

……………………..………

Q2: What is the name of the software component that represents a set of similar objects?

………………………………………………..…………………………………….

Q3: Which one of the following may change the state of an object?

□ Properties □ Behavior □ Association □ Multiplicity

Q4: Something should be an object if it could be a member of a set defined by a class.

□ True □ False

Please rate each of the following facial expressions used by the virtual lecturer? positively (+) or negatively (-)?

1	2	3	4	5	6
Neutral	Interested	Amazed	Happy	Smiling	Thinking
+	+	+	+	+	+
-	-	-	-	-	-

For each statement below, please indicate your agreement rate using the following rating scale.

1=Strongly Disagree 2=Disagree 3=Undecided 4=Agree 5=Strongly Agree

S1	I think I would like to use this software frequently	1	2	3	4	5
S2	I found the interface unnecessarily complex	1	2	3	4	5
S3	I thought the interface was easy to use	1	2	3	4	5
S4	I think that I would need the support of technical	1	2	3	4	5

	person to be able to use this interface					
S5	I found the various functions in this interface were well integrated	1	2	3	4	5
S6	I think that there is too much inconsistency in this interface	1	2	3	4	5
S7	I would imagine that most people will learn to use this interface very quickly	1	2	3	4	5
S8	I found the system very cumbersome to use	1	2	3	4	5
S9	I felt very confident using the software	1	2	3	4	5
S10	I needed to learn a lot of things before I could get going with this system	1	2	3	4	5
S11	I was excited and interested about the lesson	1	2	3	4	5
S12	Asking and answering questions in the interface helped to improve my understanding of the presented information	1	2	3	4	5
S13	I felt that I have a high level of control over my learning	1	2	3	4	5
S14	The virtual lecturer's facial expressions increased my attention and I enjoyed it	1	2	3	4	5
S15	I would attend lessons with virtual lecturers again	1	2	3	4	5
S16	The virtual lecturer made it easier for me to follow and understand the lesson	1	2	3	4	5
S17	Overall, I am satisfied with the interface	1	2	3	4	5
S18	Overall, I had an enriching experience with the interface	1	2	3	4	5

Part 4	Presentation of lesson 2 (Class naming and drawing) VIRTUAL LECTURER with BODY GESTURES (VLBG)

Now, information about classes and objects will be presented in lesson 1. Please pay your attention to the interface and the presented information.

Depending on the presented information, please answer the following questions?

Q1: Class symbol could be divided into three parts which are: class neme, attributes, and ………………………………………..

Q2: To show a class in full detail, we can include the ………….…… of attributes and the signature of operations.

Q3: In the first stages of software development, it could be enough to identify the name of the class?

□ True □ False

Q4: Which one of the following could be accepted name of a class?

□ passenger □ Routes □ Bank Account □ Game

Please rate each of the following body gestures used by the virtual lecturer? positively (+) or negatively (-)?

1	2	3	4	5	6	7	8	9	10
Neutral	Hands clenching-front	Hands clenching-back	Open palms	Arms folded	Pointing	Hands steepling	Chin stroking	Legs crossed	Moving
+	+	+	+	+	+	+	+	+	+
-	-	-	-	-	-	-	-	-	-

For each statement below, please indicate your agreement rate using the following rating scale.

1=Strongly disagree 2=Disagree 3=Undecided 4=Agree 5=Strongly agree

S1	I think I would like to use this software frequently	1	2	3	4	5
S2	I found the interface unnecessarily complex	1	2	3	4	5
S3	I thought the interface was easy to use	1	2	3	4	5
S4	I think that I would need the support of technical person to be able to use this interface	1	2	3	4	5
S5	I found the various functions in this interface were well integrated	1	2	3	4	5
S6	I think that there is too much inconsistency in this interface	1	2	3	4	5
S7	I would imagine that most people will learn to use this interface very quickly	1	2	3	4	5
S8	I found the system very cumbersome to use	1	2	3	4	5
S9	I felt very confident using the software	1	2	3	4	5
S10	I needed to learn a lot of things before I could get going with this system	1	2	3	4	5
S11	I was excited and interested about the lesson	1	2	3	4	5
S12	Asking and answering questions in the interface helped to improve my understanding of the presented information	1	2	3	4	5
S13	I felt that I have a high level of control over my learning	1	2	3	4	5

S14	The virtual lecturer's body gestures increased my attention and I enjoyed it	1	2	3	4	5
S15	I would attend lessons with virtual lecturers again	1	2	3	4	5
S16	The virtual lecturer made it easier for me to follow and understand the lesson	1	2	3	4	5
S17	Overall, I am satisfied with the interface	1	2	3	4	5
S18	Overall, I had an enriching experience with the interface	1	2	3	4	5

Part 5	Presentation of lesson 3 (Associations and multiplicity) TWO VIRTUAL LECTURERS with FACIAL EXPRESSIONS (TVLFE)

Now, information about classes and objects will be presented in lesson 1. Please pay your attention to the interface and the presented information.

Depending on the presented information, please answer the following questions?

Q1: There are two types of association labels; association name and ……..………….. name.

Q2: What should be added to determine the direction of the association between two classes?

…………………………………………………………………………………………
……………...…

Q3: The symbol * means that the multiplicity is?

☐ 1 or more ☐ only one ☐ zero or more ☐ zero

Q4: Which one of the following is used to indicate how many objects of one class can be related with an object of another class?

☐ attributes ☐ multiplicity ☐ association name ☐ class name

Please rate each of the following facial expressions used by the two virtual lecturers? positively (+) or negatively (-)?

1	2	3	4	5	6
Neutral	Interested	Amazed	Happy	Smiling	Thinking
+	+	+	+	+	+
-	-	-	-	-	-

For each statement below, please indicate your agreement rate using the following rating scale.

1=Strongly disagree 2=disagree 3=Undecided 4=Agree 5=Strongly agree

S1	I think I would like to use this software frequently	1	2	3	4	5
S2	I found the interface unnecessarily complex	1	2	3	4	5
S3	I thought the interface was easy to use	1	2	3	4	5
S4	I think that I would need the support of technical person to be able to use this interface	1	2	3	4	5
S5	I found the various functions in this interface were well integrated	1	2	3	4	5
S6	I think that there is too much inconsistency in this interface	1	2	3	4	5
S7	I would imagine that most people will learn to use this interface very quickly	1	2	3	4	5
S8	I found the system very cumbersome to use	1	2	3	4	5
S9	I felt very confident using the software	1	2	3	4	5
S10	I needed to learn a lot of things before I could get going with this system	1	2	3	4	5
S11	I was excited and interested about the lesson	1	2	3	4	5
S12	Asking and answering questions in the interface helped to improve my understanding of the presented information	1	2	3	4	5
S13	I felt that I have a high level of control over my learning	1	2	3	4	5
S14	The virtual lecturers' facial expressions increased my attention and I enjoyed it	1	2	3	4	5
S15	I would attend lessons with virtual lecturers again	1	2	3	4	5
S16	The virtual lecturer made it easier for me to follow and understand the lesson	1	2	3	4	5
S17	Overall, I am satisfied with the interface	1	2	3	4	5
S18	Overall, I had an enriching experience with the interface	1	2	3	4	5

Part 6	Post-experiment questions

How useful did you find each of the following in the interface? (Circle as appropriate)

Virtual lecturer with facial expressions.	Least useful	1	2	3	4	5	Most useful
Virtual lecturer with body gestures.	Least useful	1	2	3	4	5	Most useful
Two virtual lecturers with facial expressions.	Least useful	1	2	3	4	5	Most useful

The presentation part.	Least useful	1	2	3	4	5	Most useful
Speech output.	Least useful	1	2	3	4	5	Most useful

Select the most preferred interface of the following:
Virtual lecturer with facial expressions
Virtual lecturer with body animations
Two virtual lecturers with facial expressions

Do you have any suggestions or comments?

____
____
____
____

Thank you very much for your patience and generous help.

B2 – Learning Content about Class Diagram Notation

Lesson 1: Classes and objects

Class diagram is one of the main diagram types included in the Unified Modeling Language; UML. It describes classes and their relationships. It also describes the data found in a software system. Here, we have an example of class diagram. The main symbols shown on class diagrams are: classes, associations and multiplicities. Let us start with classes.

A class is a software component that represents and defines a set of similar objects. All the objects with the same properties and behaviour are instances of one class. So, an object is a piece of structured data that can represent anything with which we can associate properties and behaviour. Properties represent the characteristics or attributes of the object. Behaviour is the operations which represent how an object acts and reacts. For example, here we have some of the objects and their properties that might be important to a particular banking system. It could be noticed that the employees Jon and Janet have the same properties. So, they can be considered as objects of a class called Employee. All objects of this class have a name, a date of birth, an address and a position.

The problem is that sometimes it's difficult to decide what should be a class and what should be an object. In general, something should be a class if it could have objects, and something should be an object if it could be a member of the set defined by a class. For example, in the university system, we have many students, each student is considered as an object of a class called student. So, in general, student is a class but student whose name is Jo is an object. Also, course is another class but if we take the computer applications course, it is an object. So, when we draw a class in a class diagram, we are saying that the system will contain a class by that name, and that when the system runs, objects of that class will be created.

Q1: could you explain the difference between properties and behaviour?

Alright, properties are the attributes which describe the current state of the object, while behaviour is the operations which may change the state of the object. For example, width and height are attributes of the rectangle, and its values describe the current dimensions and area of that rectangle. While, resize, is an operation which may change the values of width and height. This will change the state of the rectangle because it will change its dimensions and area.

Q2: could you give another example on the difference between class and object?

Well, country is a class which has many objects such as United Kingdom, Italy, and Greece. So, each of these countries is considered as an object of the class country.

Lesson 2: Class naming and drawing

Now, how can we choose the name of a class? There are important rules that should be followed to name the classes. The name of the class should always be noun, singular, and starts with a capital letter. Using the singular format ensures that an instance or object of the class is a single item, not a list of collection. If we want to give a class a name consisting of more than one word, then delete the spaces and capitalize the first letter of each word.

To draw the class, we have many options because it can be drawn at several different levels of details. The simplest way is to draw a box with the name of the class inside. In this case, we are simply hiding the attributes and operations. Optionally, we may also show the attributes and operations contained in the class. This is done by dividing a class box into three boxes: the top box contains the class name, in the middle box we list the attributes which represent the properties of the class objects, and in the bottom box we list the operations which represent the behaviour of class objects.

As an example, if we have a class called rectangle, we can show only its name, or we can show its name and attributes, here we have tow attributes: height and width; each rectangle has a height and a width. Also, we can show the name of the class and its operations, here each rectangle has tow operations which are compute area and resize. Another option is to show all of the class information; name, attributes and operations. To show a class in a full detail, we can include the type of attributes and the signature of operations. In our example, we can see that both of height and width are integer numbers, and compute area is an operation that compute the area of a rectangle and return it as an integer value. Resize is another operation that modifies the integer values of both height and width. Note that the class name should always appear, the attributes are a singular noun format, and the operations are verbs.

Q1: why we should initialize the name of the class with a capital letter?

The reason is to determine what is a class and what is not. So, the following nouns could be names of classes, while the following is not because it's not initialized with capital letter.

Q2: when drawing a class, how can we decide which level of details should be used?

Well, how much detail we show in the class drawing depends on the phase of software development and on what we want to communicate. So, in the first stages of development we may use the class name, but in the subsequent phases of development we give more details for each class. For example, in the programming stage, we need to give full details for each class.

Lesson 3: Associations and multiplicity

An association is used to show how classes are related to each other. It represents the relationship between classes. It's represented as a line between a pair of classes. For example, there is a relationship or association between the company class and the employee class. The multiplicity indicates how many objects of one class can be associated with an object of the other class. Symbols representing multiplicity are shown at each end of the association. Here, the multiplicity of company class could be one and the multiplicity of employee class could be one to many. This means that there can be only one company associated with each employee and, one or more employees can be associated with a company. Note that many is represented by the star symbol, and that if the multiplicity of a class is one, which is the default case, no need to write it.

For example, in the association between company class and board of directors class, both of the two classes has the multiplicity one which means that each company has one board and each board is assigned to one company. The multiplicity could be optional. In this case the notation zero or one is used. For example, there can be zero or one office per employee. In other words, it is optional to assign an employee to an office because some employees may work at home or in a job that doesn't require an office. Also, we can specify the multiplicity to be an interval or range, which is shown as two dots between the lower and upper bound. In the same example, an office could be allocated to one, two, or three employees. If an interval has no upper bound, then we use the star symbol. Therefore, star means many or zero or more, one two dots star means one or more, or, at least one, and one two dots three means at least one and at most three.

To explain the nature of the association, we can label it. There are two types of label; association name and role name. An association name should be a verb or verb phrase, and is placed next to the middle of the association. As a result, one class becomes the subject and the other class becomes the object. For example, the association between employee and company could be called works for and we can read it as 'an employee works for a company'. Role name is a noun or noun phrase that could be attached to either or both ends of an association. For example, in the association between person and board of directors, board member is a role name that describes the people who are members of the board.

So, we can read the association as 'a board of directors has three to seven persons as board members'. If we delete both the association name and role names, then the association name is simply has. For example, in the association between company and board of directors, we read it as 'a company has a board of directors. In general, associations could be read in both directions but it's possible to limit the direction of an association by adding an arrow next to the association name as shown in this example.

Q1: there is a relationship between student and lecturer, how can we represent it in a simple class diagram?

Ok, let us think about it. Here, we have two classes: student and lecturer. We know that each lecturer teaches one or more student and each student may taught by one or more lecturer. So, the multiplicity of both classes is one to many.

Q2: what labels could be added to this association?

Here, we can add the verb "teach" as an association name. In this case, lecturer is the subject and student is the object of this verb. Also, we can add an arrow to explain that this association is directed from lecturer to student and it could be read as: a lecturer teaches one or more student.

B3 – Frequency Table for Users' Responses to the Pre-experiment Questions

Age	18-24	3	6%
	25-34	24	50%
	35-44	20	42%
	45-55	1	2%
	Total	48	100%
Gender	Male	35	73%
	Female	13	27%
	Total	48	100%
Residence	Home	4	8%
	Overseas	44	92%
	Total	48	100%
English	First	3	6%
	Second	45	94%
	Total	48	100%
Education	College	1	2%
	Undergraduate	7	15%
	Master	36	75%
	PhD	4	8%
	Total	48	100%
Area of study	Computing	20	42%
	Engineering	21	44%
	Eng. literature	1	2%
	Medicine	1	2%
	Management	3	6%
	Mathematics	2	4%
	Total	48	100%
Use of computer per week (hours)	1 to 5	1	2%
	6 to 10	7	15%
	>10	40	83%
	Total	48	100%
Use of Internet per week (hours)	1 to 5	4	8%
	6 to 10	5	10%
	>10	39	81%
	Total	48	100%
Knowledge about class diagram notation	Limited	4	8%
	Good	3	6%

	No	41	85%
	Total	48	100%
Knowledge about avatar	Limited	6	13%
	Good	4	8%
	Excellent	2	4%
	No	36	75%
	Total	48	100%
Use of e-learning per week (hours)	<1	5	10%
	1 to 5	17	35%
	6 to 10	6	13%
	>10	2	4%
	No	18	38%
	Total	48	100%
Miss face-to-face interaction?	Yes	25	83%
	No	5	17%
	Total	30	100%
Do you think the addition of virtual lecturer (VL) with facial expressions (FE) might help in e-learning?	Yes	43	90%
	No	5	10%
	Total	48	100%
Do you think the addition of virtual lecturer (VL) with body gestures (BG) might help in e-learning?	Yes	46	96%
	No	2	4%
	Total	48	100%
Importance of speech output in e-learning interface	Very important	34	71%
	Important	14	29%
	Total	48	100%
Importance of virtual lecturer with facial expressions in e-learning interface	Very important	9	19%
	Important	26	54%
	Unimportant	13	27%
	Total	48	100%
Importance of virtual lecturer with body gestures in e-learning interface	Very important	7	15%
	Important	33	69%
	Unimportant	8	17%
	Total	48	100%

B4 – Users' Evaluation of Facial Expression Presented in the Absence of Interactive Context

P: Positive, N: Negative

Users	Facial Expressions					
	Neutral	Interested	Amazed	Happy	Smiling	Thinking
U01	N	N	N	P	P	N
U02	N	N	P	P	P	P
U03	P	P	P	P	P	N
U04	N	N	N	P	P	N
U05	N	P	P	N	P	P
U06	P	P	N	P	P	P
U07	N	P	N	P	P	P
U08	N	N	P	P	P	P
U09	P	N	P	N	P	P
U10	P	P	N	P	P	P
U11	N	P	N	P	P	P
U12	P	P	P	P	P	P
U13	P	P	P	P	P	P
U14	N	P	P	P	P	N
U15	P	N	P	N	P	P
U16	N	P	P	P	N	N
U17	N	P	N	N	P	P
U18	N	P	P	P	P	N
U19	N	P	P	P	P	P
U20	N	P	N	P	P	N
U21	N	N	N	P	P	P
U22	P	P	P	P	P	P
U23	N	P	P	P	P	P
U24	N	P	P	P	N	P
U25	P	P	P	N	P	N
U26	N	P	N	P	P	P
U27	N	N	P	P	P	P
U28	P	P	P	P	P	P
U29	N	P	P	P	P	P
U30	N	P	P	N	P	P
U31	N	P	N	P	P	N

U32	P	P	N	P	P	P
U33	N	P	P	P	N	P
U34	P	P	P	P	P	N
U35	N	P	P	P	P	N
U36	N	P	P	P	P	N
U37	P	N	N	P	N	N
U38	P	P	P	N	P	P
U39	P	P	P	P	P	P
U40	P	P	N	P	P	N
U41	N	P	P	P	N	P
U42	N	P	P	P	P	P
U43	N	P	P	P	P	N
U44	N	N	P	N	P	N
U45	N	P	N	P	P	N
U46	P	P	P	N	P	P
U47	N	P	P	P	N	N
U48	P	N	P	P	N	N

B5 – Users' Evaluation of Facial Expression Presented in the Absence of Interactive Context

L: Like, D: Dislike

Users	Facial Expressions					
	Neutral	Interested	Amazed	Happy	Smiling	Thinking
U01	D			L	L	D
U02	D			L	L	D
U03		L			L	D
U04	D			L	L	D
U05	D	L		D	L	
U06			D	L	L	
U07	D		D	L		L
U08	D	D		L	L	
U09		D	L	D	L	
U10			D	L		L
U11	D		D	L	L	
U12				L	L	D
U13				L	L	
U14	D		L	L		D
U15	L	D		D		L
U16	D	L	L			D
U17	D	L			L	D
U18	D			L	L	D
U19	D			L		L
U20	D			L	L	D
U21	D	D		L	L	
U22				L	L	D
U23	D	L			L	D
U24	D	L		L	D	
U25	L		L	D		D
U26	D		D		L	L
U27	D	D	L			L
U28				L	L	D
U29	D	L			L	
U30	D		L	D	L	
U31			D	L	L	D

U32			D	L	L	D
U33	D	L		L		D
U34		L	D		L	D
U35	D	L	L			D
U36	D	L		L		D
U37	L	D		L	D	
U38	D		L		L	D
U39			D		L	L
U40		L	D		L	D
U41	D	L			D	L
U42	D	L		L		D
U43	D			L	L	D
U44	D	D	L		L	
U45		D		L	L	D
U46			L	D	L	
U47	D		L		D	L
U48		D	L		D	L

B6 – Users' Evaluation of Facial Expression Presented in the Presence of Interactive Context

P: Positive, N: Negative

Users	Post-VLFE						Post-TVLFE					
	Neutral	Interested	Amazed	Happy	Smiling	Thinking	Neutral	Interested	Amazed	Happy	Smiling	Thinking
U01	P	P	N	P	P	P	P	P	N	P	P	P
U02	N	P	P	P	P	N	N	P	P	P	P	P
U03	N	P	P	P	P	P	N	P	P	P	P	P
U04	P	P	P	P	P	P	P	P	P	P	P	P
U05	N	P	P	P	P	P	N	P	P	P	P	P
U06	P	P	N	P	P	P	P	P	N	P	P	P
U07	P	P	N	P	P	N	P	P	N	P	N	N
U08	N	P	P	N	P	N	N	P	P	N	P	N
U09	P	P	P	N	P	P	P	P	P	P	P	P
U10	P	P	P	P	P	P	P	P	P	P	P	P
U11	N	P	N	P	P	N	P	P	P	P	P	N
U12	P	P	P	P	P	P	N	P	N	P	P	P
U13	P	P	P	P	P	P	P	P	P	P	P	P
U14	N	N	P	P	N	N	N	P	P	P	P	N
U15	P	N	P	P	P	P	P	P	N	P	P	P
U16	P	P	P	P	P	N	P	P	P	P	P	N
U17	P	P	P	P	P	P	P	P	P	P	P	P
U18	P	P	N	P	N	N	P	P	N	P	P	P
U19	N	P	P	P	P	P	N	P	P	P	P	P
U20	P	N	N	P	P	P	N	N	N	P	P	P
U21	N	P	P	N	P	N	N	P	P	P	P	P
U22	P	P	P	P	P	P	P	P	P	P	P	P
U23	N	P	P	P	P	P	N	P	P	P	N	P
U24	P	P	N	P	P	P	P	P	N	P	N	P
U25	P	P	P	P	P	P	P	P	N	P	P	N
U26	N	P	N	P	P	P	N	P	N	P	P	P
U27	N	N	P	P	P	P	N	N	P	P	P	P
U28	P	P	P	P	P	P	P	P	P	P	P	P
U29	N	P	P	P	P	P	N	P	P	P	P	N
U30	N	P	P	N	P	P	N	P	P	N	P	P

U31	P	P	P	P	P	P		P	P	P	P	P	P
U32	P	P	N	P	P	P		P	P	P	P	P	P
U33	N	P	P	P	P	P		P	P	P	P	P	N
U34	N	P	P	P	P	N		P	P	P	P	P	N
U35	N	N	P	N	P	N		N	P	P	P	P	N
U36	N	P	P	P	P	P		N	P	N	N	P	N
U37	P	P	N	N	P	N		N	P	N	P	P	N
U38	P	P	N	P	P	P		P	P	P	P	P	P
U39	P	P	P	P	P	P		N	P	P	P	P	P
U40	P	P	P	P	P	P		P	P	P	P	P	P
U41	N	P	P	P	P	P		N	P	P	P	P	P
U42	N	P	N	P	P	N		N	P	P	P	P	N
U43	N	P	P	P	P	N		P	P	P	P	P	P
U44	P	N	N	P	P	P		P	P	N	N	P	P
U45	N	P	N	P	P	N		P	P	N	P	P	P
U46	P	P	P	N	P	P		P	P	P	N	P	P
U47	P	P	P	P	P	P		P	P	P	P	P	P
U48	N	P	P	P	P	N		N	P	P	P	P	P

B7 – Users' Evaluation of the Body Gestures Presented in the Absence of Interactive Context

P: Positive, N: Negative

Users	Body Gestures									
	Neutral	Hands clenching-front	Hands clenching-back	Open palms	Arms folded	Pointing	Hands steepling	Chin stroking	Legs crossed	Walking
U01	P	P	N	N	N	N	N	P	P	N
U02	P	P	P	N	P	N	N	P	N	N
U03	N	P	P	N	P	P	N	P	N	N
U04	N	P	N	P	N	P	N	N	N	N
U05	N	P	P	P	P	P	P	P	N	P
U06	P	P	P	P	P	P	P	P	N	P
U07	N	P	P	P	N	P	P	P	N	N
U08	N	N	N	P	N	P	P	N	N	P
U09	P	P	N	P	N	P	P	P	N	P
U10	P	P	N	P	P	P	P	P	P	P
U11	N	P	P	P	N	P	P	N	P	N
U12	P	P	P	P	P	P	P	N	N	P
U13	P	P	P	P	P	P	P	P	N	P
U14	N	P	N	P	P	P	P	P	P	P
U15	P	P	N	P	P	P	P	N	P	P
U16	N	N	N	P	N	P	P	N	N	P
U17	P	P	P	P	N	P	P	P	P	P
U18	N	N	N	P	P	P	P	P	N	P
U19	N	N	N	P	N	P	P	N	N	P
U20	N	N	N	P	N	P	P	P	N	P
U21	N	P	N	P	N	P	P	P	N	N
U22	P	N	P	P	N	P	P	N	N	P
U23	N	N	N	P	N	P	P	P	N	P
U24	N	N	N	P	N	P	P	P	P	P
U25	P	N	N	P	N	P	P	N	N	N
U26	N	N	N	P	P	P	P	P	N	P
U27	N	P	P	P	P	P	P	P	P	P
U28	P	P	N	P	N	P	P	N	N	P
U29	N	P	P	P	P	P	P	P	N	P
U30	N	P	P	P	P	P	P	P	P	P

U31	N	N	N	P	N	P	P	N	N	P
U32	P	N	N	N	P	P	N	N	N	P
U33	N	N	N	P	N	P	P	P	N	P
U34	P	P	P	P	P	P	P	P	P	P
U35	N	P	P	P	N	P	P	N	N	P
U36	N	N	P	P	P	P	P	P	N	P
U37	P	P	P	P	P	P	P	P	N	P
U38	P	P	N	P	P	P	N	N	N	P
U39	N	P	P	P	P	P	P	P	P	P
U40	P	P	P	P	P	N	P	P	P	N
U41	N	N	N	P	N	P	P	N	N	P
U42	N	N	N	P	N	P	P	P	N	P
U43	N	P	P	N	N	P	P	N	N	P
U44	P	N	N	P	N	P	P	P	N	P
U45	N	P	N	P	N	P	P	N	N	P
U46	P	N	N	P	N	P	N	P	N	P
U47	N	P	N	P	N	P	P	P	N	P
U48	N	P	N	P	N	P	P	P	N	P

B8 – Users' Evaluation of the Body Gestures Presented in the Absence of Interactive Context

L: Like, D: Dislike

Users	Body Gestures									
	Neutral	Hands clenching-front	Hands clenching-back	Open palms	Arms folded	Pointing	Hands steepling	Chin stroking	Legs crossed	Walking
U01			D		D	L		L		
U02			D	L		L			D	
U03	D	L				L			D	
U04			D		D		L			L
U05	D			L		L			D	
U06	L		D						D	L
U07	D	L			D			L		
U08	D			L					L	D
U09					D			L	D	L
U10			D				L	L		D
U11		L	D		D		L			
U12						L		D	D	L
U13				L		D		L	D	
U14	D		D		L			L		
U15			D	L			L	D		
U16	D			L	D			L		
U17		D		L	D					L
U18			D	L				L	D	
U19			D		D	L	L			
U20			D			L			D	L
U21			D				L	L	D	
U22				L	D	L			D	
U23			D	L				L	D	L
U24	D		D			L				L
U25	L		D	L				D		
U26		D	D		L		L			
U27	D				D				L	L
U28			D			L			D	L
U29				L	D		L		D	
U30	D		D	L						L

U31			D		D	L	L			
U32			D	L			D			L
U33	D			L		L			D	
U34		L	L				D			D
U35	D					L	L	D		
U36	D					L		L	D	
U37							L	L	9	D
U38		L				L		D	D	
U39	D					L	L		D	
U40		L				D	L			D
U41	D		D	L			L			
U42	D		D			L		L		
U43		L					L	D	D	
U44		D				L			D	L
U45			D	L		L			D	
U46			D	L	D					L
U47			D	L		L			D	
U48				L	D	L			D	

B9 – Users' Evaluation of the Body Gestures Presented in the Presence of Interactive Context

P: Positive, N: Negative

Users	Post-VLBG									
	Neutral	Hands clenching-front	Hands clenching-back	Open palms	Arms folded	Pointing	Hands steepling	Chin stroking	Legs crossed	Walking
U01	P	P	N	P	N	P	P	N	N	P
U02	P	P	P	P	P	P	P	P	N	P
U03	N	P	P	P	P	P	P	P	N	P
U04	N	P	P	P	N	P	P	P	N	P
U05	N	P	P	P	P	P	P	P	N	P
U06	P	P	P	P	P	P	P	P	N	P
U07	N	N	N	P	N	P	N	N	N	P
U08	N	N	N	P	N	P	P	N	N	P
U09	P	P	P	P	N	P	P	P	N	P
U10	P	P	N	P	P	P	P	P	N	P
U11	N	P	P	P	N	P	P	N	N	P
U12	P	N	N	P	N	P	P	P	N	P
U13	P	P	P	P	N	P	P	P	N	P
U14	N	P	P	P	P	P	P	P	P	P
U15	P	N	N	P	N	P	P	P	N	P
U16	N	N	N	P	N	P	P	N	N	P
U17	P	N	P	P	N	P	P	P	N	P
U18	N	N	N	P	P	P	P	P	N	P
U19	N	P	P	P	N	P	P	P	P	P
U20	P	P	P	P	N	P	P	P	N	P
U21	N	N	N	P	N	P	P	P	N	P
U22	P	P	P	P	N	P	P	P	P	P
U23	P	P	P	P	N	P	P	P	N	P
U24	P	P	P	P	N	P	P	P	P	N
U25	P	P	N	P	N	P	P	N	N	P
U26	N	P	N	P	P	P	P	P	N	P
U27	N	N	N	P	P	P	P	P	N	P
U28	P	P	N	P	N	P	P	P	N	P
U29	N	P	P	P	P	P	P	N	N	P
U30	N	P	P	P	P	P	P	P	P	P

U31	P	P	P	P	N	P	P	P	N	P
U32	N	P	N	P	P	P	P	P	N	P
U33	P	P	N	P	N	P	P	N	N	P
U34	N	P	N	P	N	P	P	P	N	P
U35	N	N	N	P	N	P	P	N	N	P
U36	N	P	P	P	P	P	P	P	N	P
U37	P	P	P	P	P	P	P	N	N	P
U38	P	P	P	P	P	P	P	N	N	P
U39	P	P	P	P	P	P	P	P	P	P
U40	P	P	P	N	P	P	P	P	P	P
U41	N	N	N	P	N	P	P	N	N	P
U42	N	N	N	P	N	P	P	N	N	P
U43	P	P	N	P	N	P	P	N	N	P
U44	P	P	N	P	N	P	P	N	N	P
U45	P	P	N	P	N	P	N	P	N	P
U46	P	P	N	P	N	P	N	P	N	P
U47	P	P	P	P	N	P	P	P	N	P
U48	P	P	P	N	N	P	P	P	N	P

B10 – Raw Data of Question Answering Time for the Second Experiment

VLFE: Virtual Lecturer with Facial Expressions, VLBG: Virtual Lecturer with Body Gestures, TVLFE: Two Virtual Lecturers with Facial Expressions

Users	Lesson 1				Lesson 2				Lesson 3				All questions/Lesson			Recall questions			Recognition questions			Total
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	VLFE	VLBG	TVLFE	VLFE	VLBG	TVLFE	VLFE	VLBG	TVLFE	
U01	62.6	29.9	11.4	15.7	18.3	28.8	20.3	6.8	10.1	26.0	3.8	8.9	119.6	74.3	48.8	92.5	47.2	36.1	27.1	27.1	12.7	242.7
U02	35.9	28.3	15.2	16.1	10.0	13.8	10.0	13.2	8.3	27.5	14.5	25.0	95.5	75.3	47.0	64.2	35.8	23.8	31.3	39.5	23.2	217.7
U03	22.5	8.4	4.8	17.6	18.0	17.1	5.2	39.1	16.5	10.2	8.8	8.3	79.4	53.4	43.8	35.0	30.9	26.7	44.3	22.4	17.1	176.5
U04	33.8	23.0	8.9	15.4	15.1	31.2	15.4	19.4	37.3	21.9	36.6	29.6	125.4	81.1	81.1	59.2	56.8	46.2	66.1	24.3	34.9	287.6
U05	20.0	18.0	16.6	22.2	24.8	21.8	9.8	9.1	19.0	23.6	21.6	13.9	65.5	78.2	76.8	46.6	42.6	38.0	18.8	35.6	38.8	220.4
U06	11.0	7.8	6.6	9.7	9.5	12.8	7.4	8.5	34.5	14.3	11.0	15.3	75.1	38.1	35.1	48.9	22.3	18.8	26.2	15.8	16.3	148.4
U07	36.2	12.7	14.0	15.0	18.2	18.9	13.1	8.3	17.9	8.7	7.2	18.9	77.8	58.5	52.8	48.9	37.2	26.7	29.0	21.4	26.2	189.2
U08	47.1	21.0	11.0	14.9	9.7	47.3	23.7	16.5	2.8	19.3	7.7	20.1	94.0	49.8	97.2	68.1	22.0	57.0	25.9	27.8	40.2	241.0
U09	30.1	14.5	5.0	4.3	16.8	9.7	8.2	13.6	9.3	12.9	23.3	17.0	48.2	53.9	62.4	26.5	44.6	22.1	21.7	9.3	40.3	164.5
U10	15.6	8.5	12.3	16.2	8.5	15.3	6.1	10.9	12.7	10.1	7.3	23.5	53.5	52.6	40.9	22.8	24.2	23.9	30.8	28.4	17.0	147.0
U11	51.3	19.7	12.1	19.9	19.7	40.5	19.9	11.7	30.2	16.5	5.1	15.0	91.8	66.8	103.0	60.2	46.7	71.0	31.6	20.1	32.0	261.6
U12	33.1	15.8	6.3	10.8	21.1	21.4	14.2	22.4	19.6	16.0	9.9	18.8	64.2	79.0	66.0	35.5	42.5	48.9	28.7	36.5	17.0	209.1
U13	28.4	35.1	24.0	6.4	2.9	27.5	11.0	10.4	8.0	47.5	9.1	54.0	93.9	51.8	118.6	63.5	30.4	55.4	30.4	21.4	63.1	264.2
U14	22.6	16.8	22.0	20.4	14.5	34.0	20.5	18.2	26.7	5.6	6.8	9.5	81.8	48.5	87.2	39.4	32.3	48.5	42.4	16.3	38.7	217.5
U15	30.1	25.1	9.7	16.4	26.4	36.5	14.0	11.2	13.3	30.6	6.6	11.0	88.2	81.3	61.4	63.0	55.3	43.9	25.2	26.1	17.6	230.9
U16	32.3	25.2	10.6	9.1	80.2	27.9	12.9	28.8	50.0	35.9	21.0	19.5	126.4	77.2	149.7	85.9	57.5	108.1	40.5	19.7	41.7	353.3
U17	20.0	19.0	19.3	11.3	12.5	28.6	11.9	18.8	44.7	15.1	14.9	16.5	71.8	91.2	69.6	41.1	59.8	39.0	30.7	31.4	30.6	232.6
U18	29.7	43.9	6.9	10.2	5.8	21.5	8.9	4.1	11.4	9.9	14.9	43.4	79.6	40.2	90.8	21.3	27.2	73.7	58.3	13.0	17.1	210.7
U19	14.6	9.3	8.3	9.7	2.8	24.1	7.9	27.0	25.0	33.4	15.8	15.9	41.9	61.8	90.2	23.8	26.9	58.5	18.1	34.9	31.7	193.8
U20	65.4	44.5	26.2	16.4	16.7	56.4	9.6	38.1	21.0	30.2	24.6	40.4	152.5	116.2	120.7	109.9	51.2	73.1	42.6	65.0	47.6	389.4
U21	28.6	9.4	5.4	10.4	18.4	8.4	9.3	21.4	5.6	5.0	9.5	29.1	57.4	53.7	49.1	26.7	38.0	10.6	30.7	15.7	38.6	160.3

U22	32.3	10.7	7.6	10.0	14.9	48.0	8.0	9.0	17.3	24.4	53.5	31.4	126.6	60.7	79.8	41.7	43.0	62.9	84.9	17.6	17.0	267.1
U23	66.5	23.1	44.3	12.0	24.8	27.7	19.1	39.4	15.7	29.3	3.5	32.5	110.9	81.0	145.9	52.5	45.0	89.6	58.4	35.9	56.3	337.8
U24	27.7	10.4	9.2	7.0	12.9	29.6	9.6	11.5	20.0	29.9	13.5	43.3	106.7	63.7	54.3	49.9	42.5	38.1	56.8	21.2	16.2	224.7
U25	26.7	22.7	17.1	15.0	6.4	6.8	7.9	4.4	50.1	19.4	32.2	26.7	81.5	25.5	128.3	49.4	13.2	69.5	32.1	12.4	58.8	235.3
U26	56.3	46.3	14.8	16.7	7.4	28.3	19.7	25.0	22.4	62.9	50.6	30.8	134.1	166.7	80.3	102.6	85.3	35.6	31.5	81.4	44.7	381.1
U27	15.8	12.4	36.9	14.1	18.3	19.3	10.6	9.9	10.6	21.4	29.2	10.2	58.0	79.2	71.4	37.6	28.2	32.0	20.5	51.0	39.4	208.6
U28	4.3	30.4	10.8	14.0	10.2	10.7	6.8	9.4	13.5	12.1	12.4	18.6	56.6	59.5	37.2	25.6	34.7	20.9	31.0	24.8	16.3	153.3
U29	8.9	22.2	11.2	12.8	10.0	15.8	14.9	13.9	6.1	18.3	16.2	11.7	54.5	52.2	55.1	25.8	24.3	31.1	28.7	27.9	24.0	161.9
U30	13.3	11.2	7.5	12.7	15.5	33.0	9.4	9.7	21.0	17.3	22.8	21.6	82.7	67.6	44.7	38.3	48.5	24.5	44.4	19.1	20.2	194.9
U31	77.8	27.9	12.5	30.8	7.3	36.1	25.7	15.5	30.5	56.8	11.0	49.7	149.0	84.7	148.1	105.7	43.4	87.4	43.3	41.2	60.7	381.8
U32	25.2	7.8	6.7	6.0	20.8	26.1	5.2	9.3	1.7	8.1	4.3	10.2	45.7	24.3	61.3	33.0	9.8	46.8	12.7	14.5	14.5	131.3
U33	6.7	22.4	9.7	24.7	34.5	41.8	8.2	14.1	38.5	19.4	26.6	24.7	98.6	63.5	109.2	76.2	29.2	57.9	22.3	34.4	51.3	271.2
U34	17.1	8.9	4.3	15.5	23.9	29.6	15.7	19.4	24.6	12.2	7.2	7.5	51.5	45.8	88.6	36.8	25.9	53.5	14.7	19.8	35.1	185.9
U35	35.1	13.3	8.6	12.3	15.0	25.0	11.2	8.7	7.9	6.8	9.2	16.9	59.8	40.8	69.4	40.0	14.7	48.4	19.9	26.1	21.0	170.0
U36	18.8	11.3	6.4	11.7	19.1	17.4	6.5	24.2	20.0	11.4	4.4	30.5	66.3	67.1	48.2	31.4	36.4	30.1	34.9	30.7	18.1	181.6
U37	16.3	13.9	6.4	4.1	6.0	8.3	3.7	4.7	1.5	3.5	3.3	10.1	40.6	22.8	18.4	30.1	14.3	5.1	10.5	8.4	13.4	81.8
U38	30.0	17.6	8.3	10.1	17.3	13.9	10.6	8.4	8.0	9.0	11.9	25.4	65.9	54.3	50.1	47.6	17.0	31.2	18.4	37.3	19.0	170.3
U39	8.2	19.8	6.3	9.2	25.5	45.4	18.5	32.4	22.8	20.3	33.4	42.1	121.8	43.6	118.7	70.8	28.0	43.2	51.0	15.6	75.5	284.0
U40	39.1	33.5	17.3	16.9	38.3	32.4	11.7	13.9	52.3	35.7	28.2	22.8	138.9	106.9	96.2	88.0	72.6	70.7	50.9	34.2	25.6	342.0
U41	24.7	25.3	12.9	20.0	43.3	28.7	16.3	18.3	8.6	18.1	20.2	21.8	106.5	68.7	82.9	71.9	26.7	50.0	34.6	42.1	32.9	258.2
U42	37.5	20.3	18.1	6.8	9.6	18.9	25.9	27.2	28.5	15.0	15.4	27.1	85.9	81.7	82.7	43.5	28.6	57.8	42.5	53.1	24.9	250.3
U43	8.8	10.3	17.2	12.9	21.3	8.0	7.1	8.6	9.5	7.1	11.3	10.9	49.2	45.0	38.9	19.1	29.4	16.6	30.1	15.7	22.3	133.1
U44	15.9	25.9	13.8	8.3	24.8	23.9	12.7	13.8	15.1	25.7	11.1	33.2	63.9	85.1	75.1	41.8	40.8	48.6	22.1	44.3	26.5	224.2
U45	27.8	26.9	18.4	8.9	59.0	47.3	14.1	22.1	16.4	33.6	7.0	43.0	142.5	82.0	100.0	106.3	54.7	50.0	36.3	27.3	50.0	324.5
U46	22.9	27.6	20.1	11.2	58.5	28.4	11.1	7.2	25.5	18.3	11.7	28.8	84.3	81.8	105.3	43.8	50.5	86.9	40.5	31.3	18.4	271.4
U47	53.9	39.2	33.5	19.0	54.9	48.1	17.9	16.0	10.7	13.7	5.3	8.2	136.9	37.9	145.5	103.0	24.3	93.1	33.9	13.6	52.4	320.3
U48	59.0	44.5	30.0	20.9	8.1	10.2	4.3	3.2	51.0	31.1	25.9	19.8	127.8	25.9	154.3	82.1	18.3	103.5	45.7	7.6	50.9	308.0

B11 – Raw Data of Answers' Correctness for the Second Experiment

VLFE: Virtual Lecturer with Facial Expressions, VLBG: Virtual Lecturer with Body Gestures, TVLFE: Two Virtual Lecturers with Facial Expressions

1: Correct answer, 0: Incorrect answer

Users	Lesson 1				Lesson 2				Lesson 3				All questions/Lesson			Recall questions			Recognition questions			Total
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	VLFE	VLBG	TVLFE	VLFE	VLBG	TVLFE	VLFE	VLBG	TVLFE	
U01	0	0	1	1	1	0	1	0	0	0	1	0	2	2	1	0	1	0	2	1	1	5
U02	0	0	0	1	1	0	1	1	1	1	1	1	1	4	3	0	2	1	1	2	2	8
U03	1	1	1	1	1	1	1	1	0	1	0	1	4	4	2	2	2	1	2	2	1	10
U04	1	0	1	1	1	0	1	0	1	1	0	0	2	3	2	2	1	1	0	2	1	7
U05	0	0	0	1	0	0	1	0	1	1	1	1	1	4	1	0	2	0	1	2	1	6
U06	1	1	1	1	1	0	1	1	0	1	1	1	3	3	4	1	1	2	2	2	2	10
U07	0	0	1	1	1	0	1	0	1	1	1	1	2	2	4	0	1	2	2	1	2	8
U08	1	1	1	0	1	0	1	1	1	1	1	1	3	4	3	2	2	1	1	2	2	10
U09	1	1	1	1	1	1	1	1	1	1	1	1	4	4	4	2	2	2	2	2	2	12
U10	0	1	1	1	1	0	1	1	1	1	1	1	4	3	3	2	1	1	2	2	2	10
U11	1	1	1	1	1	0	0	1	0	1	1	1	2	3	4	1	1	2	1	2	2	9
U12	1	0	1	1	1	0	1	0	0	0	1	1	2	2	3	0	1	1	2	1	2	7
U13	1	1	0	1	1	1	1	0	0	1	1	1	3	3	3	2	2	1	1	1	2	9
U14	0	1	0	1	1	0	1	1	1	1	1	1	2	4	3	1	2	1	1	2	2	9
U15	1	0	1	1	1	0	1	1	1	0	1	0	3	3	2	1	1	1	2	2	1	8
U16	1	0	1	1	0	0	1	0	0	1	0	1	2	3	1	1	1	0	1	2	1	6
U17	0	0	1	1	1	0	1	1	0	1	0	1	3	2	2	1	1	0	2	1	2	7
U18	1	1	1	1	1	1	1	0	1	1	0	0	2	3	4	2	2	2	0	1	2	9
U19	1	0	1	1	1	1	1	1	0	0	1	1	3	4	2	1	2	0	2	2	2	9
U20	1	0	1	1	1	1	0	0	1	1	1	1	3	4	2	1	2	2	2	2	0	9
U21	1	1	1	1	1	1	1	1	1	1	1	1	4	4	4	2	2	2	2	2	2	12

U22	1	1	1	1	1	0	1	1	1	1	1	1	4	4	3	2	2	1	2	2	2	11
U23	0	0	0	1	1	0	0	0	1	0	1	1	1	3	1	1	1	0	0	2	1	5
U24	1	1	1	1	1	1	1	1	0	1	1	0	2	4	4	1	2	2	1	2	2	10
U25	1	0	1	1	1	1	1	1	0	1	1	1	3	4	3	1	2	1	2	2	2	10
U26	0	0	1	1	1	0	0	1	0	1	1	1	2	3	2	0	1	1	2	2	1	7
U27	1	1	1	1	1	0	1	1	0	1	1	1	3	4	3	1	2	1	2	2	2	10
U28	1	0	1	1	0	1	1	0	0	1	0	1	2	3	2	1	1	1	1	2	1	7
U29	1	0	1	1	1	0	1	1	0	1	1	1	3	3	3	1	1	1	2	2	2	9
U30	1	1	1	1	1	1	1	1	0	1	1	1	3	4	4	1	2	2	2	2	2	11
U31	0	0	0	1	1	1	0	1	0	1	1	0	1	3	2	0	2	1	1	1	1	6
U32	0	1	1	1	1	0	1	1	1	1	1	1	3	4	3	1	2	1	2	2	2	10
U33	1	1	1	1	1	1	1	1	0	1	1	1	4	4	3	2	2	1	2	2	2	11
U34	0	1	1	0	1	0	0	1	0	1	1	1	3	2	2	1	1	1	2	1	1	7
U35	1	1	1	1	1	0	0	1	1	1	1	1	2	4	4	1	2	2	1	2	2	10
U36	1	1	1	1	1	0	1	1	1	1	1	1	4	3	4	2	1	2	2	2	2	11
U37	1	1	1	0	1	1	1	0	1	1	1	1	3	3	4	2	2	2	1	1	2	10
U38	0	0	1	1	0	0	1	0	0	1	0	0	2	1	1	0	1	0	2	0	1	4
U39	1	1	1	1	0	0	1	0	1	1	0	0	1	4	2	0	2	2	1	2	0	7
U40	1	1	0	1	1	0	1	0	0	1	0	1	2	3	2	1	2	1	1	1	1	7
U41	0	0	1	1	0	0	1	0	1	1	1	0	1	3	2	0	2	0	1	1	2	6
U42	0	0	0	0	1	1	0	0	0	1	1	1	3	2	0	1	2	0	2	0	0	5
U43	1	1	1	1	1	1	1	1	0	1	1	1	4	4	3	2	2	1	2	2	2	11
U44	1	0	1	1	0	0	1	0	1	1	0	1	3	3	1	1	2	0	2	1	1	7
U45	0	1	0	1	0	0	1	0	0	1	1	1	1	2	3	0	1	1	1	1	2	6
U46	0	0	0	1	0	0	0	0	0	1	0	0	1	1	0	1	0	0	0	1	0	2
U47	0	0	1	0	1	0	0	0	1	1	1	1	1	4	1	1	2	0	0	2	1	6
U48	0	0	0	1	1	0	1	1	0	1	0	0	1	3	1	1	1	0	0	2	1	5

B12: Raw Data of Users' Responses to the Satisfaction Questionnaire

Raw Data of Users' Satisfaction towards VLFE

S1 – S18: Statements of the satisfaction questionnaire shown in Appendix B1

5: Strongly Agree, 4: Agree, 3: Undecided, 2: Disagree, 1: Strongly Disagree

Users	SUS statements										Additional statements								SUS score
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	
U01	3	4	4	2	3	2	4	2	4	2	3	4	4	4	4	3	4	4	65.0
U02	4	1	5	1	4	1	5	1	4	1	3	5	4	4	4	4	5	5	92.5
U03	4	1	4	1	4	1	4	1	5	1	4	4	4	4	5	5	5	4	90.0
U04	4	3	4	2	4	2	4	2	4	2	4	4	3	4	3	4	4	4	72.5
U05	5	2	3	4	4	2	4	2	4	4	4	4	4	4	5	4	5	4	65.0
U06	4	1	5	1	4	1	4	1	4	1	3	4	4	3	3	3	4	4	90.0
U07	3	1	5	1	5	5	5	3	5	1	3	5	5	4	4	4	4	4	80.0
U08	4	2	5	2	4	3	3	2	4	4	4	5	2	4	5	4	4	4	67.5
U09	4	3	5	1	4	1	4	3	4	1	4	4	4	4	4	5	4	3	80.0
U10	4	2	2	1	5	1	4	1	4	2	4	5	4	4	4	4	4	4	80.0
U11	5	1	5	2	4	1	4	1	4	1	5	5	4	3	5	5	5	5	90.0
U12	4	2	4	2	4	2	4	2	4	2	4	4	4	4	4	4	4	4	75.0
U13	4	2	4	2	4	2	4	2	4	2	4	4	4	4	4	4	4	4	75.0
U14	4	3	4	2	5	1	4	3	2	3	5	5	4	3	5	3	4	5	67.5
U15	4	3	5	2	4	3	4	2	4	4	4	4	3	4	4	4	4	4	67.5

U16	4	2	4	2	4	2	4	2	4	2	4	4	4	4	4	4	4	4	75.0
U17	4	2	3	3	4	2	4	2	3	3	4	3	4	4	4	3	4	4	65.0
U18	4	2	5	1	5	1	5	3	4	1	4	4	4	3	3	4	4	4	87.5
U19	3	2	4	3	4	1	4	1	5	1	5	5	5	3	3	3	5	5	80.0
U20	3	2	4	3	4	2	4	2	4	2	4	4	2	2	2	2	4	3	70.0
U21	4	1	4	1	5	1	4	1	5	2	4	5	4	3	4	3	4	4	90.0
U22	5	1	5	1	5	1	5	1	5	1	5	5	5	5	5	5	5	5	100.0
U23	3	2	4	2	3	2	2	2	4	2	3	4	4	2	3	3	4	4	65.0
U24	4	1	5	1	4	5	5	1	5	1	3	4	4	4	4	4	4	4	85.0
U25	5	1	5	1	4	1	5	1	5	1	5	5	4	5	5	5	5	5	97.5
U26	2	3	4	2	5	2	4	2	3	4	4	3	4	5	4	3	4	4	62.5
U27	4	2	4	2	4	2	4	4	4	2	4	4	4	4	2	4	4	4	70.0
U28	4	1	4	1	4	1	4	1	4	4	4	4	4	4	4	4	5	5	80.0
U29	4	1	5	1	4	1	5	2	5	2	5	4	4	3	4	4	5	4	90.0
U30	5	1	5	1	5	1	4	1	5	1	5	5	5	4	5	5	5	5	97.5
U31	4	1	4	2	5	2	5	1	5	2	5	5	4	5	4	5	5	4	87.5
U32	3	2	3	1	4	1	5	1	4	2	2	4	2	3	3	2	4	3	80.0
U33	5	1	4	1	4	2	4	1	5	1	4	5	4	4	4	5	4	4	90.0
U34	3	2	4	2	4	2	3	3	2	2	4	4	4	3	3	4	4	4	62.5
U35	5	1	5	1	5	1	4	1	5	1	4	5	5	3	3	4	5	5	97.5
U36	4	2	4	1	4	2	2	2	4	2	4	4	3	3	4	3	4	4	72.5
U37	4	1	5	1	5	1	5	1	5	1	3	2	5	3	5	5	5	5	97.5
U38	4	2	4	1	4	2	4	2	4	2	4	4	3	4	4	4	4	4	77.5

U39	4	3	4	2	4	3	3	3	4	3	4	4	3	3	3	2	4	3	62.5
U40	3	2	4	3	3	2	3	2	3	2	4	3	3	2	3	3	4	4	62.5
U41	4	3	4	2	2	2	1	2	4	2	4	4	4	2	4	2	4	4	60.0
U42	4	2	4	1	5	1	5	1	5	1	5	5	4	4	5	5	5	5	92.5
U43	5	1	4	2	4	2	4	2	4	2	5	4	4	2	4	4	4	4	80.0
U44	4	1	5	1	5	2	3	1	3	2	3	4	3	2	2	3	3	4	82.5
U45	2	4	5	1	4	1	5	2	4	2	2	4	5	1	4	1	4	3	75.0
U46	3	1	5	4	4	2	4	2	3	1	4	4	4	4	3	4	4	4	72.5
U47	4	2	4	3	2	3	4	2	3	2	2	4	3	2	3	2	4	3	62.5
U48	4	2	4	2	2	4	3	2	3	4	2	4	3	2	3	2	3	3	55.0

Raw Data of Users' Satisfaction towards VLBG

S1 – S18: Statements of the satisfaction questionnaire shown in Appendix B1

5: Strongly Agree, 4: Agree, 3: Undecided, 2: Disagree, 1: Strongly Disagree

Users	SUS statements										Additional statements								SUS score
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	
U01	4	1	5	2	4	2	4	2	4	2	5	4	4	5	5	4	5	4	80.0
U02	5	1	5	1	5	1	4	1	4	1	4	5	4	5	5	5	5	5	95.0
U03	5	1	5	1	5	1	4	1	5	1	5	5	4	4	5	5	5	4	97.5
U04	4	3	4	3	4	2	4	2	4	2	4	4	4	5	3	4	4	4	70.0
U05	5	2	4	4	5	3	4	2	4	3	4	4	4	5	4	5	5	5	70.0
U06	4	1	5	2	5	2	5	2	5	1	3	4	4	5	3	3	4	5	90.0
U07	5	1	5	1	5	4	4	4	5	1	3	4	4	4	4	4	5	5	82.5
U08	4	2	4	2	4	2	4	1	4	2	4	5	4	4	4	5	4	5	77.5
U09	4	2	4	3	4	2	4	3	4	2	4	5	4	4	4	4	5	4	70.0
U10	3	2	4	1	5	2	4	2	5	2	5	5	4	4	4	4	4	4	80.0
U11	5	1	5	1	5	1	5	1	5	1	5	5	4	4	5	5	5	5	100.0
U12	4	2	4	2	4	2	4	2	4	2	4	5	5	5	4	4	4	4	75.0
U13	4	2	4	2	4	2	4	2	4	4	4	4	4	4	4	4	4	4	70.0
U14	5	1	4	1	4	1	5	1	4	2	5	5	5	5	4	5	5	5	90.0
U15	3	2	5	4	4	4	5	2	4	3	4	4	4	5	4	5	4	4	65.0
U16	4	2	5	1	4	2	4	2	5	2	5	5	4	4	4	4	5	4	82.5
U17	3	2	3	3	4	2	3	2	3	4	4	4	4	4	4	3	4	4	57.5

U18	5	1	5	1	5	1	5	1	5	1	5	5	5	4	5	5	5	5	100.0
U19	3	1	5	1	4	1	5	1	5	4	4	5	5	4	3	5	5	4	85.0
U20	4	2	4	2	5	2	4	2	4	2	4	5	2	5	4	4	5	4	77.5
U21	4	1	5	1	5	1	4	2	5	1	5	5	4	5	4	4	4	4	92.5
U22	5	1	5	1	5	1	5	1	5	1	5	5	5	5	5	5	5	5	100.0
U23	4	2	4	2	4	2	5	2	4	2	4	4	4	5	4	4	4	4	77.5
U24	5	1	5	1	4	2	5	1	5	2	5	5	4	4	4	5	4	4	92.5
U25	5	1	5	1	5	1	5	1	5	1	5	5	5	5	5	5	5	5	100.0
U26	3	2	4	3	4	3	4	3	4	3	4	4	3	4	4	4	4	4	62.5
U27	4	1	4	1	4	1	4	1	4	2	4	4	4	4	4	4	4	4	85.0
U28	5	2	4	1	4	1	4	1	4	1	5	5	5	4	4	5	4	5	87.5
U29	4	1	4	2	4	2	5	2	4	2	5	4	5	4	4	4	5	5	80.0
U30	5	1	5	1	5	1	4	1	5	1	5	5	5	5	5	5	5	5	97.5
U31	4	2	4	2	4	2	5	1	5	2	5	5	4	5	4	4	4	5	82.5
U32	4	1	5	1	4	1	5	1	4	2	4	4	3	4	4	5	4	5	90.0
U33	5	1	5	2	4	2	4	1	4	1	5	4	4	5	4	5	4	4	87.5
U34	4	2	4	2	3	2	4	2	4	2	3	4	3	4	4	3	4	2	72.5
U35	5	1	5	1	5	1	5	1	5	1	5	5	5	5	5	5	5	5	100.0
U36	4	2	4	1	4	2	4	2	4	2	4	5	3	4	4	4	4	4	77.5
U37	5	1	5	1	5	1	5	1	5	1	5	4	5	5	5	5	5	5	100.0
U38	4	2	4	2	4	2	4	3	4	2	4	4	4	4	4	4	4	4	72.5
U39	5	1	5	1	4	1	5	2	5	1	5	4	4	5	5	5	4	5	95.0
U40	4	1	5	1	4	1	4	1	4	1	4	4	4	4	3	4	5	4	90.0

U41	4	2	4	1	5	2	5	1	5	1	5	4	4	5	5	5	5	5	90.0
U42	5	2	4	1	4	2	5	1	5	1	5	5	5	4	5	5	5	4	90.0
U43	4	1	5	2	5	1	5	1	5	1	5	5	5	4	5	4	5	5	95.0
U44	4	1	5	1	4	1	5	1	5	1	4	5	5	5	5	5	4	5	95.0
U45	4	1	5	1	4	1	5	1	5	1	5	5	5	5	5	4	4	4	95.0
U46	2	2	4	5	4	2	4	2	3	2	3	4	4	4	3	5	4	5	60.0
U47	5	1	5	1	5	1	5	1	5	1	5	5	5	5	5	4	5	5	100.0
U48	5	1	5	1	5	1	5	1	5	1	5	5	4	5	5	5	5	5	100.0

Raw Data of Users' Satisfaction towards TVLFE

S1 – S18: Statements of the satisfaction questionnaire shown in Appendix B1

5: Strongly Agree, 4: Agree, 3: Undecided, 2: Disagree, 1: Strongly Disagree

	SUS statements										Additional statements								
Users	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	SUS score
U01	4	1	5	2	4	2	4	1	4	1	4	5	4	4	4	4	5	4	85.0
U02	5	1	5	1	4	1	5	1	5	1	4	5	5	4	4	4	4	4	97.5
U03	5	1	4	1	4	1	4	1	5	2	4	5	4	5	5	5	5	4	90.0
U04	3	2	4	2	4	2	4	2	4	2	4	4	3	4	3	4	4	4	72.5
U05	5	2	4	5	4	2	3	2	4	3	4	4	4	4	5	4	4	4	65.0
U06	5	1	5	3	4	1	4	1	4	2	5	4	3	3	4	4	4	4	85.0
U07	4	4	5	1	4	2	5	1	5	1	4	4	4	4	4	4	5	5	85.0
U08	4	1	4	2	4	2	5	2	4	2	5	5	4	5	5	4	5	4	80.0
U09	4	1	5	1	4	1	4	3	4	1	4	5	4	4	4	4	4	4	85.0
U10	4	3	4	2	5	2	4	2	4	1	5	5	4	2	1	2	3	3	77.5
U11	5	1	5	2	5	2	4	2	4	2	5	5	4	4	5	5	5	5	85.0
U12	4	2	5	2	4	2	5	1	3	2	4	4	4	5	4	4	4	4	80.0
U13	4	2	4	2	4	2	4	2	4	4	4	4	4	4	4	4	4	4	70.0
U14	5	3	5	1	4	1	4	1	4	3	5	5	5	4	5	3	5	5	82.5
U15	3	4	5	2	4	3	4	2	4	3	4	5	3	4	4	4	5	4	65.0
U16	4	2	4	2	4	2	4	4	5	4	4	4	4	4	4	4	5	4	67.5
U17	4	2	4	3	4	2	4	2	3	3	4	4	3	4	4	3	4	4	67.5

U18	5	1	4	1	5	1	5	1	5	1	5	5	5	3	5	5	5	5	97.5
U19	4	1	4	1	5	1	5	1	5	1	4	4	4	3	4	3	5	4	95.0
U20	4	2	4	2	4	2	4	2	4	2	3	4	5	4	3	4	4	4	75.0
U21	3	4	2	2	4	3	2	4	3	2	3	3	2	3	2	2	3	3	47.5
U22	3	2	5	1	5	1	5	1	5	1	5	5	5	5	5	5	5	5	92.5
U23	4	2	5	2	4	2	4	2	4	2	4	4	4	3	3	3	4	4	77.5
U24	4	1	5	1	4	2	5	1	4	5	3	4	4	4	5	5	5	4	80.0
U25	5	1	5	1	5	1	5	1	5	1	5	5	5	5	5	5	5	5	100.0
U26	3	2	4	3	4	2	4	3	4	3	4	4	3	4	4	4	4	4	65.0
U27	4	2	4	2	4	2	4	2	4	2	4	4	4	4	4	4	4	4	75.0
U28	4	1	4	1	4	1	4	1	4	2	4	5	4	4	4	4	5	5	85.0
U29	4	2	5	1	4	2	5	2	5	1	5	4	4	3	4	5	5	4	87.5
U30	4	1	5	1	5	2	4	1	5	1	5	5	5	4	5	4	5	5	92.5
U31	4	2	4	2	3	2	5	1	5	2	5	5	4	5	4	5	5	4	80.0
U32	3	2	4	1	4	4	5	1	4	2	3	4	3	2	3	2	4	4	75.0
U33	5	1	4	1	4	2	4	1	5	1	4	4	4	5	5	4	4	5	90.0
U34	3	2	4	2	4	2	3	4	4	2	4	4	3	3	4	4	4	3	65.0
U35	5	1	4	1	5	1	4	1	4	1	5	5	5	4	5	5	5	5	92.5
U36	4	3	5	1	4	1	4	2	4	2	4	5	3	3	3	4	4	4	80.0
U37	3	5	5	1	1	5	5	1	5	1	1	5	3	1	3	1	3	3	65.0
U38	4	2	3	2	2	2	4	2	4	2	4	4	3	4	4	4	4	4	67.5
U39	4	2	3	3	4	2	3	2	4	2	3	4	3	2	2	2	3	3	67.5
U40	4	1	5	1	5	1	4	1	5	1	4	4	4	5	4	4	4	4	95.0

U41	2	4	5	2	2	2	4	2	4	2	5	4	4	2	2	4	2	4	62.5
U42	5	1	4	1	5	2	5	1	5	2	5	5	5	4	5	4	5	4	92.5
U43	5	1	5	1	5	1	5	1	5	1	5	5	5	3	5	5	5	5	100.0
U44	4	1	4	1	5	2	4	1	2	1	3	4	5	1	3	1	4	5	82.5
U45	2	1	5	1	2	2	5	1	4	1	2	4	4	2	4	2	4	4	80.0
U46	3	2	4	4	4	2	4	2	3	1	3	4	4	4	3	4	4	4	67.5
U47	4	2	4	2	5	2	4	1	3	3	4	4	3	4	3	4	4	4	75.0
U48	3	2	4	4	4	2	4	2	3	1	2	4	5	2	4	2	4	4	67.5

B13: Raw Data of Users' Responses to the Post-experiment Questions

1: Least useful, 5: Most useful

Users	Usefulness of the used multimodal metaphors					Preferred interface
	Virtual lecturer with facial expressions	Virtual lecturer with body gestures	Two virtual lecturers with facial expressions	Presentation part	Speech output	
U01	3	4	5	5	4	VLBG
U02	5	5	4	5	5	VLBG
U03	1	2	3	5	5	TVLFE
U04	3	4	5	4	4	TVLFE
U05	3	4	4	5	5	VLBG
U06	3	3	3	5	5	TVLFE
U07	1	1	1	5	5	TVLFE
U08	3	5	5	4	5	VLBG
U09	4	5	4	4	4	VLBG
U10	4	5	2	5	5	VLBG
U11	5	5	3	5	5	VLBG
U12	4	5	5	5	5	VLBG
U13	4	4	4	4	4	VLBG
U14	4	5	4	5	5	VLBG
U15	3	5	3	5	4	VLBG
U16	4	4	3	4	4	TVLFE
U17	3	5	3	2	2	VLBG
U18	2	5	5	5	5	VLBG
U19	4	4	5	5	5	VLBG
U20	3	5	3	4	4	VLBG
U21	3	4	4	5	5	VLBG
U22	4	5	4	5	5	VLBG
U23	2	4	2	4	4	VLBG
U24	4	5	5	4	3	VLBG
U25	5	4	4	5	5	VLFE
U26	4	5	3	4	5	VLBG
U27	3	5	4	4	4	VLBG
U28	3	4	3	5	5	VLBG
U29	4	4	5	4	4	TVLFE
U30	3	4	5	5	5	VLBG
U31	3	5	4	5	5	VLBG

U32	3	4	2	5	5	VLBG
U33	3	5	3	4	5	VLBG
U34	2	4	3	3	4	VLBG
U35	4	5	3	4	5	VLBG
U36	3	4	4	5	4	VLBG
U37	2	5	1	5	5	VLBG
U38	4	3	5	5	5	VLBG
U39	3	5	4	5	5	VLBG
U40	4	3	5	3	4	TVLFE
U41	3	5	4	5	4	VLBG
U42	3	3	4	5	5	TVLFE
U43	4	4	5	4	4	TVLFE
U44	1	5	2	5	5	VLBG
U45	3	5	2	4	4	VLBG
U46	4	5	3	5	5	VLBG
U47	2	5	4	5	5	VLBG
U48	2	4	3	4	4	VLBG

Appendix C – Third Experiment (Investigating the Role of Non-Speech Auditory Technology)

C1 – Questionnaire

I am pleased to present my self to you as one of the postgraduate research students in the School of Informatics in the University of Bradford. I am currently investigating the use of multimodal interaction metaphors in e-learning interfaces, and I would like to obtain your views regarding the use of avatar, natural recorded speech along with non-speech sounds in this domain.

Please answer all the questions as truthfully as possible. It would be grateful if you could fill in the following questionnaire sincerely and provide your views. Your privacy is guaranteed as your name will not be mentioned in any part of the study.

Thanks for your co-operation, and your participation is highly appreciated.

Name: …………………………………….

Part 1	Pre-experiment Questions

Age: ☐ 18 - 24 ☐ 25 - 34 ☐ 35 - 44 ☐ 45 - 54 ☐ 55 +

Gender: ☐ Male ☐ Female

Education level: ☐ High School ☐ College ☐ Under-graduate
☐ Master ☐ PhD ☐ Other……………….

Area of study : ____

How often do you use the computer per week?
☐ Never ☐ Less than 1 hour ☐ 1-5 hours
☐ 6-10 hours ☐ More than 10 hours

How often do you use the internet per week?
☐ Never ☐ Less than 1 hour ☐ 1-5 hours
☐ 6-10 hours ☐ More than 10 hours

Do you have knowledge about Class Diagram notation?
☐ Yes (☐ Limited ☐ Good ☐ Excellent)
☐ No

Do you use any e-learning web sites or software?

□ Yes (how often/week? □Less than 1 hour □1-5 hours □6-10 hours □+10 hours)

□ No

During e-learning environment, if other sounds such as earcons and auditory icon besides the lecturer voice accompanied learning, would you feel?

Annoyed	□ Yes	□ No
Frustrated	□ Yes	□ No
Helpful	□ Yes	□ No
Focused	□ Yes	□ No

Part 2	Presentation of the lesson: Classes and Objects

Now, information about classes and objects will be presented. Please pay your attention to the interface and the presented information.

Depending on the presented information, please answer the following questions?

<u>Learning evaluation:</u>

Q1. The main symbols (components) shown in class diagram are?

(a)…..……………….. (b) ….……………….. (c) Mulltiplicities

Q2. What does it mean by □ehaviour of object?

……………………………………………………………………………………

……………………………………………………………………………………

Q3. Which one of the following statements is TRUE?

Something should be a class if it could have objects.

Something could be an object if it could have classes.

Something could be a class if it could be a member of an object.

UML stands for United Modelling Library.

Q4. Which of the following indicates how an object acts and reacts?

(a) Class. (b) Attributes. (c) Operations. (d)Properties.

<u>Memorability task 1:</u>

The following two paragraphs have been presented by the interface. Highlight where the sounds were used to indicate each of the following:

- Start of a statement.

- End of a statement.
- Keyword of high importance.

Class diagram is one of the main diagram types included in the Unified Modeling Language; UML. It describes classes and their relationships. It also describes the data found in a software system. Here, we have an example of class diagram. The main symbols shown on class diagrams are: classes, associations and multiplicities. Let us start with classes.

- Start of a definition.
- End of a definition.
- Keyword of medium importance.
- Keyword of low importance.

A class is a software component that represents and defines a set of similar objects. All the objects with the same properties and behaviour are instances of one class. So, an object is a piece of structured data that can represent anything with which we can associate properties and behaviour. Properties represent the characteristics or attributes of the object. Behaviour is the operations which represent how an object acts and reacts. For example, here we have some of the objects and their properties that might be important to a particular banking system.

Memorability task 2:

For each of the following you will hear three different sounds. Choose the correct sound used in the interface. (Please write 1, 2 or 3)

		Sound
(a)	Start of a statement.	
(b)	End of a statement.	
(c)	Start of a definition.	
(d)	End of a definition.	
(e)	Keyword of high importance.	
(f)	Keyword of medium importance.	
(g)	Keyword of low importance.	

Satisfaction:

For each statement below, please indicate your agreement rate using the following rating scale.

1=Strongly Disagree 2=Disagree 3=Undecided 4=Agree 5=Strongly Agree

S1	I think I would like to use this software frequently	1	2	3	4	5
S2	I found the interface complex	1	2	3	4	5
S3	I thought the interface was easy to use	1	2	3	4	5
S4	I think that I would need the support of technical person to be able to use this interface	1	2	3	4	5
S5	I found the various functions in this interface were well integrated	1	2	3	4	5
S6	I think that there is too much inconsistency in this interface	1	2	3	4	5
S7	I would imagine that most people will learn to use this interface very quickly	1	2	3	4	5
S8	I found the system very cumbersome to use	1	2	3	4	5
S9	I felt very confident using the software	1	2	3	4	5
S10	I needed to learn a lot of things before I could get going with this system	1	2	3	4	5
S11	I was excited and interested about the lesson	1	2	3	4	5
S12	It was easy to identify the most important parts of the presented information	1	2	3	4	5
S13	I would like to attend e-learning again if presented this way	1	2	3	4	5
S14	Overall, I am satisfied with the interface	1	2	3	4	5
S15	Overall, I am satisfied with the learning experience using this interface	1	2	3	4	5

Part 4	Post-experiment Questions

How did you find the use of the added sounds in the tested e-learning interface?

Annoyed ☐ Yes ☐ No

Frustrated ☐ Yes ☐ No

Helpful ☐ Yes ☐ No

Focused ☐ Yes ☐ No

Please add any suggestions or comments

....
....
....
....

Thank you very much for your patience and generous help.

C2 – Frequency Table for Users' Characteristics

Age	25-34	17	71%
	35-44	6	25%
	45-55	1	4%
	Total	24	100%
Gender	Male	16	67%
	Female	8	33%
	Total	24	100%
Education	Master	5	21%
	PhD	19	79%
	Total	24	100%
Area of study	Computing	8	33%
	Engineering	12	50%
	Management	4	17%
	Total	24	100%
Use of computer per week	1 to 5	2	8%
	6 to 10	4	17%
	>10	18	75%
	Total	24	100%
Use of Internet per week	1 to 5	3	13%
	6 to 10	2	8%
	>10	19	79%
	Total	24	100%
Knowledge about class diagram notation	Limited	3	13%
	Good	1	4%
	No	20	83%
	Total	24	100%
Use of e-learning per week	<1	8	33%
	1 to 5	9	38%
	6 to 10	1	4%
	No	6	25%
	Total	24	100%

C3 – Raw Data of Users' Evaluation for the Use of Non-speech Sounds Prior and Post to the Experiment

1: Yes, 0: No

Users	Prior to experiment				Post to experiment			
	Annoyed	Frustrated	Helpful	Focused	Annoyed	Frustrated	Helpful	Focused
U01	1	1	1	0	1	1	0	0
U02	0	0	1	1	0	1	1	1
U03	1	1	0	0	1	1	1	0
U04	0	1	0	0	0	0	1	1
U05	0	0	1	1	1	0	1	1
U06	0	0	1	1	0	0	1	1
U07	1	1	0	0	0	1	1	1
U08	0	1	0	0	0	0	1	1
U09	0	1	0	0	0	1	1	1
U10	1	0	1	0	0	0	1	1
U11	1	0	1	1	0	0	1	1
U12	1	1	0	0	0	1	0	1
U13	1	0	1	0	0	1	1	1
U14	0	0	1	1	0	1	1	1
U15	1	0	0	0	0	0	1	1
U16	1	1	0	0	1	0	1	0
U17	1	1	0	0	0	0	1	1
U18	0	0	1	0	1	1	1	0
U19	0	0	1	1	0	0	1	1
U20	1	1	0	0	1	1	0	0
U21	1	1	0	0	0	0	1	1
U22	1	1	0	0	0	0	1	1
U23	1	0	0	0	0	0	1	1
U24	1	0	0	0	1	1	1	0

C4 – Raw Data of Correctness of Users' Answers to the Learning Evaluation Questions in part 2 of the questionnaire

1: Correct answers, 0: Incorrect answer

Users	Q1	Q2	Q3	Q4	Recall (Q1 & Q2)	Recognition (Q3 & Q4)	Total
U01	1	1	1	0	2	1	3
U02	1	1	1	1	2	2	4
U03	0	0	1	1	0	2	2
U04	0	1	1	1	1	2	3
U05	1	1	0	1	2	1	3
U06	1	1	1	1	2	2	4
U07	0	1	1	1	1	2	3
U08	1	0	1	1	1	2	3
U09	0	1	1	1	1	2	3
U10	1	0	1	1	1	2	3
U11	1	1	1	1	2	2	4
U12	1	0	0	1	1	1	2
U13	1	1	1	1	2	2	4
U14	1	1	1	1	2	2	4
U15	1	1	1	1	2	2	4
U16	0	1	0	1	1	1	2
U17	1	1	1	1	2	2	4
U18	0	1	1	1	1	2	3
U19	1	1	1	1	2	2	4
U20	1	0	1	0	1	1	2
U21	1	1	1	1	2	2	4
U22	0	1	1	1	1	2	3
U23	1	1	1	1	2	2	4
U24	0	1	1	1	1	2	3

C5 – Raw Data of Correctness of Users' Answers to the First Memorability Task in part 2 of the questionnaire

SS: Start of Statement, ES: End of Statement, KHI: Keyword of High Importance, SD: Start of Definition, ED: End of Definition, KMI: Keyword of Medium Importance, KLI: Keyword of Low Importance

1: Correct answers, 0: Incorrect answer

Users	SS	ES	KHI	SD	ED	KMI	KLI	Total
U01	1	1	1	1	1	1	1	7
U02	1	0	0	1	1	1	1	5
U03	1	1	1	1	0	0	1	5
U04	1	1	1	1	1	1	1	7
U05	1	0	0	1	1	1	0	4
U06	1	1	1	1	1	1	0	6
U07	1	1	1	1	1	1	0	6
U08	1	0	1	1	1	1	0	5
U09	1	1	1	1	1	1	0	6
U10	1	1	1	1	1	0	1	6
U11	1	1	1	1	1	0	1	6
U12	1	1	0	1	1	1	1	6
U13	1	0	1	1	1	1	1	6
U14	1	1	1	1	0	0	1	5
U15	1	1	1	1	1	0	1	6
U16	1	0	1	0	1	1	0	4
U17	0	1	0	1	1	1	0	4
U18	1	1	1	1	1	1	1	7
U19	1	1	1	0	1	0	0	4
U20	1	0	1	1	1	1	1	6
U21	1	0	1	1	1	1	1	6
U22	1	1	1	1	1	1	1	7
U23	1	0	1	1	1	1	1	6
U24	1	1	1	1	1	0	1	6

C6 – Raw Data of Correctness of Users' Answers to the Second Memorability Task in part 2 of the questionnaire

SS: Start of Statement, ES: End of Statement, KHI: Keyword of High Importance, SD: Start of Definition, ED: End of Definition, KMI: Keyword of Medium Importance, KLI: Keyword of Low Importance

1: Correct answers, 0: Incorrect answer

Users	SS sound	ES sound	SD sound	ED sound	KHI sound	KMI sound	KLI sound	Total
U01	1	1	1	1	1	1	1	7
U02	1	1	1	0	1	1	1	6
U03	1	1	1	1	1	1	1	7
U04	1	1	1	1	1	1	1	7
U05	1	1	1	1	1	1	1	7
U06	1	1	1	1	1	1	1	7
U07	1	1	1	1	1	1	1	7
U08	1	1	1	1	1	1	1	7
U09	1	1	1	1	1	1	1	7
U10	1	1	0	1	0	1	1	5
U11	1	1	1	1	1	1	1	7
U12	1	1	0	0	1	1	1	5
U13	1	1	1	1	1	1	1	7
U14	1	1	1	0	1	1	1	6
U15	1	1	1	1	1	1	1	7
U16	1	1	1	1	1	1	1	7
U17	0	0	1	1	1	1	1	5
U18	1	1	1	1	1	1	1	7
U19	1	1	1	0	1	1	1	6
U20	1	1	1	1	0	1	1	6
U21	1	1	0	1	0	1	1	5
U22	1	1	1	1	1	1	1	7
U23	1	1	0	1	1	1	1	6
U24	1	1	1	1	1	1	1	7

C7: Raw Data of Users' Responses to the Satisfaction Questionnaire

S1 – S15: Statements of the satisfaction questionnaire shown in Appendix C1

5: Strongly Agree, 4: Agree, 3: Undecided, 2: Disagree, 1: Strongly Disagree

	SUS statements										Additional statements					SUS score
Users	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	
U01	4	3	4	2	4	3	5	2	3	2	3	4	4	4	4	70.0
U02	4	3	4	2	5	2	5	3	3	2	4	4	4	4	4	72.5
U03	5	1	5	1	5	1	5	2	5	1	4	5	4	5	5	97.5
U04	5	2	5	1	5	1	5	1	5	2	4	3	4	4	4	95.0
U05	5	2	5	2	5	3	4	2	5	2	4	5	5	4	5	82.5
U06	5	2	4	2	5	1	5	2	4	1	5	4	4	5	4	87.5
U07	5	1	5	2	4	2	5	1	5	2	4	3	5	4	4	90.0
U08	5	2	5	2	5	2	4	2	5	2	4	4	4	5	4	85.0
U09	4	2	4	2	4	3	5	2	3	4	5	5	5	4	4	67.5
U10	5	2	5	1	5	1	5	1	5	1	4	4	4	4	4	97.5
U11	4	2	4	2	4	2	5	2	4	2	4	4	4	4	4	77.5
U12	3	2	4	3	4	3	4	2	4	3	4	4	3	3	3	65.0
U13	5	2	5	2	5	2	5	1	5	2	5	4	5	5	5	90.0
U14	4	2	3	3	4	2	3	3	4	3	4	4	3	4	3	62.5
U15	5	2	5	1	5	2	4	2	5	2	5	4	4	4	5	87.5
U16	3	1	4	1	5	2	3	2	4	2	3	2	3	5	3	77.5
U17	5	1	5	1	5	2	5	2	4	1	5	4	5	5	5	92.5
U18	2	4	2	3	3	2	3	3	2	3	3	4	2	2	2	42.5
U19	4	2	5	3	4	2	5	1	5	1	4	4	4	5	5	85.0
U20	3	4	3	4	3	4	2	4	3	4	2	3	3	3	3	35.0
U21	5	1	5	2	5	1	5	1	5	2	5	4	4	5	5	95.0
U22	5	2	5	1	5	2	4	1	5	1	4	5	5	4	4	92.5
U23	5	2	5	1	5	2	5	2	5	1	4	4	4	5	4	92.5
U24	3	1	3	2	5	1	3	2	4	2	3	4	3	4	4	75.0

Appendix D – Technical Description

D1- Experimental Platform Used in the First Experiment (Chapter 3)

In order to keep the consistency, both versions (VOELP – Figure 2 and MMELP – Figure 3) of the experimental platform have the same design in terms of interface components and background colours.

In order to assure clarity and accessibility of class diagram notations, the largest part of the interface in both versions was dedicated to display the class diagram examples in the form of three JPG images.

In each class diagram, the coordinates of each notation (i.e. class, association and multiplicity) were determined so that the learning information is communicated when the mouse cursor is placed on these notations (see the sample code).

The learning information communicated to the users was related to three different types of class diagram notation: classes, associations and multiplicities.

Earcons:

In total, six different multiplicities were found in the learning examples which were: 0..1, 1..*, 2..*, 1..2, 1 and *. Therefore, different numbers of musical notes were used to represent the values 0, 1, 2 and *. Accordingly, the earcons used in the MMELP were designed as shown in Table 3.

In the design of earcons that have lower and upper bounds (such as those represented the multiplicities 0..1, 1..*, 2..* and 1..2), a short pause intervals were added between the two bounds in order to help the listener to differentiate the end of the lower bound and the start of the upper bound. Also, in order to insure consistent representation, the same number of musical notes (4) was determined to communicate the * value in the multiplicities 1..*, 2..* and *.

All the earcons were created using music synthesiser software and recorded by different software for sound recording and producing it in waveform data files (.wav).

Recorded Speech:

Recorded speech was selected to communicate the learning messages about the association among classes due to its naturalness and clarity compared to the synthesised speech.

The voice of the researcher was used to record the required information and then manipulated and produced in .wav format.

During the recording of the speech, short pause intervals and different rhythms and pitches were used to attract the listeners and to capture their attention.

Avatars:

The facially expressive avatars were selected to communicate longer learning messages compared to those communicated by the recorded speech.

The same background colour was used in all the avatar files in order to avoid distracting users' attention.

In order to create the avatar files, the following procedure was used:

- Preparation of the text to be communicated and save it in .txt file.
- Preparation of the recorded speech related to this text and save it in .wav file.
- Importing both text and speech files to the Mimic software to automate the synchronization between the lips movement of built-in 3d figure and the text. Also, this software automatically adds the eye blinks and head nodes. The produced file is exported then to the Poser software.
- In Poser, a built-in character was used and facial expressions were designed and then added in specific frames (pause intervals in the speech) to attain suitable synchronisation between these expressions and the spoken material.
- Playing the output audio-video file for testing.
- Rendering of the frames in order to produce the final file in AVI format.

Sample code:

```
Private Sub PictureBox1_MouseMove(ByVal sender As Object, ByVal e As
System.Windows.Forms.MouseEventArgs) Handles PictureBox1.MouseMove


        Dim a As Graphics = PictureBox1.CreateGraphics()
        Dim xc As Integer
        Dim yc As Integer
        xc = e.X
        yc = e.Y
        '--------------------------------------------------------------
        'Operating avatar files for classes
        '--------------------------------------------------------------
        If xc >= 17 And xc <= 86 And yc >= 35 And yc <= 124 Then
            If Not AWMP.Ctlenabled Then
                AWMP.Visible = True
                AWMP.Ctlenabled = True
                AWMP.URL = "elevator class.avi"
            End If

        ElseIf xc >= 221 And xc <= 371 And yc >= 38 And yc <= 123 Then
            If Not AWMP.Ctlenabled Then
                AWMP.Visible = True
                AWMP.Ctlenabled = True
                AWMP.URL = "elevator controller class.avi"
            End If

        ElseIf xc >= 261 And xc <= 303 And yc >= 368 And yc <= 453
Then
            If Not AWMP.Ctlenabled Then
                AWMP.Visible = True
                AWMP.Ctlenabled = True
                AWMP.URL = "button class.avi"
            End If

        ElseIf xc >= 505 And xc <= 575 And yc >= 37 And yc <= 123 Then
            If Not AWMP.Ctlenabled Then
                AWMP.Ctlenabled = True
                AWMP.Visible = True
                AWMP.URL = "door class.avi"
            End If
```

```
        '-------------------------------------------------------------
        'Operating recorded speech files for associations
        '-------------------------------------------------------------
        ElseIf xc >= 90 And xc <= 100 And yc >= 69 And yc <= 90 Then
            If Not AWMPsound.Ctlenabled Then
            AWMPsound.Ctlenabled = True
            AWMPsound.URL = "arrow from controller to elevator
ssociation.wav"
            End If

        ElseIf xc >= 132 And xc <= 176 And yc >= 74 And yc <= 92 Then
            If Not AWMPsound.Ctlenabled Then
                AWMPsound.Ctlenabled = True
                AWMPsound.URL = "verb from controller to elevator
association .wav"
            End If

        ElseIf xc >= 418 And xc <= 461 And yc >= 74 And yc <= 92 Then
            If Not AWMPsound.Ctlenabled Then
                AWMPsound.Ctlenabled = True
                AWMPsound.URL = "verb from controller to door
association .wav"
            End If

        ElseIf xc >= 491 And xc <= 501 And yc >= 64 And yc <= 94 Then
            If Not AWMPsound.Ctlenabled Then
                AWMPsound.Ctlenabled = True
                AWMPsound.URL = "arrow from controller to door
association .wav"
            End If

        ElseIf xc >= 230 And xc <= 353 And yc >= 229 And yc <= 249
Then
            If Not AWMPsound.Ctlenabled Then
                AWMPsound.Ctlenabled = True
                AWMPsound.URL = "verb from controller to button
association .wav"
            End If

        '-------------------------------------------------------------
        'Operating earcons files for multiplicities
        '-------------------------------------------------------------
        ElseIf xc >= 249 And xc <= 300 And yc >= 129 And yc <= 149
Then
            If Not AWMPsound.Ctlenabled Then
                AWMPsound.Ctlenabled = True
                AWMPsound.URL = "1.wav"
            End If

        ElseIf xc >= 293 And xc <= 300 And yc >= 344 And yc <= 361
Then
            If Not AWMPsound.Ctlenabled Then
                AWMPsound.Ctlenabled = True
                AWMPsound.URL = "1..many.wav"
            End If

        ElseIf xc >= 104 And xc <= 119 And yc >= 75 And yc <= 85 Then
            If Not AWMPsound.Ctlenabled Then
                AWMPsound.Ctlenabled = True
                AWMPsound.URL = "1..many.wav"
            End If
```

```
        ElseIf xc >= 197 And xc <= 216 And yc >= 77 And yc <= 85 Then
            If Not AWMPsound.Ctlenabled Then
                AWMPsound.Ctlenabled = True
                AWMPsound.URL = "1.wav"
            End If

        ElseIf xc >= 376 And xc <= 392 And yc >= 75 And yc <= 85 Then
            If Not AWMPsound.Ctlenabled Then
                AWMPsound.Ctlenabled = True
                AWMPsound.URL = "1.wav"
            End If

        ElseIf xc >= 477 And xc <= 489 And yc >= 76 And yc <= 84 Then
            If Not AWMPsound.Ctlenabled Then
                AWMPsound.Ctlenabled = True
                AWMPsound.URL = "2..many.wav"
            End If

        Else
            AWMP.Ctlenabled = False
            AWMP.URL = ""
            AWMP.Visible = False
            AWMPsound.Ctlenabled = False
            AWMPsound.URL = ""
            AWMPsound.Visible = False
            PictureBox1.Image =
System.Drawing.Bitmap.FromFile("elevcont-v2.jpg")

        End If
    End Sub
```

D2- Experimental Platforms Used in the Second Experiment (Chapter 4)

Consistent design of the three experimental e-learning platforms (VLFE, VLBG and TVLFE) were used in terms of incorporating similar interface components such as lesson selection buttons, pause/play buttons, and asking/answering the questions in addition to the same background colours.

The largest part of the interface in the three platforms was dedicated for displaying the learning material.

In designing the interface for e-learning applications, it is recommended to inform the learner about the current topic being delivered. For example, highlight its title using different colour. Also, enable the learner to control his learning by pause/play buttons and lesson selection buttons.

The PowerPoint presentation for each lesson was pre-prepared and saved in AVI files.

The same set of facial expressions was used across the three interfaces and was incorporated consistently.

In the VLBG platform, both of the learning material and the virtual lecturer were combined in the same interface component to resemble the traditional class-based situation.

Similar to the facial expressions, the body gestures used in the VLBG were designed in Poser and then included in specific frames.

The same procedure explained in Appendix D1 for the production of avatar files was followed. However, the learning material (i.e. three lessons about class diagram notations) was larger compared to the communicated content in the first experiment. Therefore, the rendering process was found to be time-consuming and could suspend the PC. To resolve this problem, each of the three lessons was decomposed into smaller parts each of which processed independently and then the output files were combined to produce the final AVI avatar file related to each lesson.

Sample code:

```
        '-------------------------------------------------------------
        'Pause button
        '-------------------------------------------------------------

Private Sub pausebtn_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles pausebtn.Click
        AVWMP.Ctlcontrols.pause()
        pwmp.Ctlcontrols.pause()
        pausebtn.Visible = False
        playbtn.Visible = True
    End Sub
        '-------------------------------------------------------------
        'Play button
        '-------------------------------------------------------------

    Private Sub playbtn_Click(ByVal sender As System.Object, ByVal e
As System.EventArgs) Handles playbtn.Click
        AVWMP.Ctlcontrols.play()
        pwmp.Ctlcontrols.play()
        pausebtn.Visible = True
        playbtn.Visible = False
        pwmp.settings.mute = True
    End Sub
```

```
        '------------------------------------------------------------
        'Lesson 1 button
        '------------------------------------------------------------
Private Sub btnl1_Click_1(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles btnl1.Click
        btnl1.BackColor = Color.Red
        btnl2.BackColor = Color.White
        btnl3.BackColor = Color.White
        Label3.Text = "You are in lesson 1: Classes and objects"
        txtask.Text = ""
        btnask.Text = "Ask Question 1"
        AVWMP.Visible = True
        pwmp.Visible = True
        AVWMP.URL = "lm1.avi"
        pwmp.URL = "pm1.avi"
        btnask.Visible = True
        btnask.Enabled = False
        btnanswer.Visible = True
        btnanswer.Enabled = False
        pwmp.settings.mute = True
        pausebtn.Visible = True
        playbtn.Visible = False
    End Sub

        '------------------------------------------------------------
        'Lesson 2 button
        '------------------------------------------------------------
    Private Sub btnl2_Click_1(ByVal sender As System.Object, ByVal e
As System.EventArgs) Handles btnl2.Click
        btnl1.BackColor = Color.White
        btnl2.BackColor = Color.Red
        btnl3.BackColor = Color.White
        AVWMP.URL = "lm2.avi"
        pwmp.URL = "pm2.avi"
        Label3.Text = "You are in lesson 2: Class naming and drawing"
        pwmp.settings.mute = True
        btnask.Text = "Ask Question 1"
        pausebtn.Visible = True
        playbtn.Visible = False
    End Sub

        '------------------------------------------------------------
        'Lesson 3 button
        '------------------------------------------------------------
    Private Sub btnl3_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles btnl3.Click
        btnl1.BackColor = Color.White
        btnl2.BackColor = Color.White
        btnl3.BackColor = Color.Red
        AVWMP.URL = "lm3.avi"
        pwmp.URL = "pm3.avi"
        Label3.Text = "You are in lesson 3: Associations and
multiplicity"
        pwmp.settings.mute = True
        btnask.Text = "Ask Question 1"
        pausebtn.Visible = True
        playbtn.Visible = False
    End Sub
```

D3- Experimental Platform Used in the Third Experiment (Chapter 5)

The experimental e-learning platform used in the third experiment was similar to the VLBG platform used in the second experiment but with the addition of earcons and auditory icons.

Similar to the first experiment, the used earcons were designed using musical notes to establish suitable representation of the values related to the importance level of specific keywords when mentioned by the virtual lecturer (i.e. one note = low, two notes = medium and three notes = high as shown in Table 27).

In addition, auditory icons were utilised to offer natural mapping between environmental sounds and specific important aspects in the presented learning lessons as explained in Table 28.

Earcon and auditory icons used in the third experiment were included in the pause intervals of the virtual lecturer. This is important to avoid overlapping both sounds and to facilitate remembering its meaning before continuing the speech of the virtual lecturer. Therefore, the duration of earcon or auditory icons should suit these pause intervals.

Earcons and auditory icons were in .wav format and incorporated in the background of the virtual lecturer speech. Then, the final presentation of the learning content was produced in AVI file.

Sample code:

```
        '--------------------------------------------------------------
        'Lesson 1 button
        '--------------------------------------------------------------
Private Sub btnl1_Click_1(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles btnl1.Click
        pwmp.Visible = True
        pwmp.URL = "l1-v3.avi"
        pausebtn.Visible = True
        playbtn.Visible = False
    End Sub
```

List of Publications

- **Book Chapter:**

1. Alseid M. and Rigas D. (2009): “An Empirical Investigation into the Use of Multimodal E-Learning Interfaces”, Human-Computer Interaction, IntechWeb, ISBN 978-953-7619-26-8. In press.

- **Journal Publications:**

1. Rigas D. and Alseid M. (2009): “Using Multimodal Metaphors to Communicate Information in E-learning: A Two-Group Empirical Study”, International Journal of Computer Science and Systems (IJCSS), Universal Society of Applied Research, ISSN: 1803-8336, 1(1): 52-59.

2. Alseid M. and Rigas D. (2008): "Empirical Results for the Use of Facial Expressions and Body Gestures in E-Learning Tools", International Journal of Computers and Communications, University Press, UK, ISSN: 2074-1294, 2(3): 87-94

- **Conference Contributions:**

1. Alseid M. and Rigas D. (2009): “Utilising Multimodal Interaction Metaphors in E-learning Applications, An Experimental Study”, In Proceedings of the IEEE 23rd International Conference on Advanced Information Networking and Applications (AINA-09), p.p. 944-949, University of Bradford, Bradford, UK, May 26-29, 2009.

2. Alseid M. and Rigas D. (2009): “Users’ Views of Facial Expressions and Body Gestures in E-Learning Interfaces: An Empirical Evaluation”, In Proceedings of the 8th WSEAS International Conference on Software Engineering, Parallel and Distributed Systems (SEPADS '09), p.p. 121-126, Cambridge, UK, ISBN 978-960-474-052-9.

3. Alseid M. and Rigas D. (2008): "Efficiency of Multimodal Metaphors in the Presentation of Learning Information", In Proceedings of the 22nd Annual Conference of Interaction, People and Computers XXII, HCI 2008, (Omar Abuelmatti and David England eds), p.p. 107-110, British Computer Society, Volume 2, Liverpool, UK, ISBN 978-1-906124-06.

4. Rigas D. and Alseid M. (2008): "Multi-modal Aided Presentation of Learning Information: A Usability Comparative Study", In Proceedings of IADIS International Conference on Interfaces and Human-Computer Interaction, p.p. 234-238, InderScience Publishers. IADIS, Amsterdam, Netherlands, ISBN 978-972-8924-5.