Three different modes of avatars as virtual lecturers in e-learning interfaces: a comparative usability study

This is the Published Version of the final output.

Three Different Modes of Avatars as Virtual Lecturers in E-learning Interfaces: A Comparative Usability Study

Marwan Alseid* and Dimitrios Rigas

Department of Software Engineering, Faculty of Information Technology, Applied Science University, Amman, Jordan

Abstract: Most of recent e-learning applications are predominantly based on textual and graphical metaphors to communicate the learning material in its user interfaces. Previous research demonstrated that incorporating both visual and auditory sensory channels in the interaction with e-learning interfaces could enhance the usability and users' learning performance. Also, the presence of humanoid virtual lecturers in e-learning interfaces was found to be attractive for e-learners. This paper describes an empirical evaluation of three different modes for the employment of speaking avatars as virtual pedagogical agents in a multimodal-based e-learning. This study aimed at comparing the usability aspects (efficiency, effectiveness and user satisfaction) as well as learning performance of three different e-learning platforms recruited to present three common lessons about class diagram notation. The first platform involved talking head of facially expressive avatar while the second platform used a full-body animated one. However, the third tested e-learning platform was based on talking heads of tow facially expressive virtual lectures; male and female who shared the presentation of the learning material. In total, 48 users participated in the experiment to test the three platforms in a within-subject design. Results of the experiment showed that the inclusion of speaking virtual lecturer with full body gestures was the most efficient in terms of question answering time, most effective in terms of correctly answered questions, and the most satisfactory as opposed to the other tow investigated e-learning platforms. Experimental results also revealed that using facially expressive avatars either singularly or coupled with female one scored similar levels of usability and learning performance.

Keywords: Usability, Avatars, E-learning, multimodal interaction.

INTRODUCTION

E-learning is a general term that is used to describe the learning process in which information and communication technology could be utilized [1, 2]. Recently, most of the e-learning interfaces are largely dependent on text and graphics as a mean of information delivery means. Making use of multimodal interaction metaphors such as speech, sounds and avatars with facial expressions and body gestures is still limited and need to be investigated further. The experimental study described in this paper is one of the main experiments in a research program that aimed at exploring the usability aspects of multimodal e-learning systems. Previous experiment [3, 4] in this research showed that the inclusion of avatar, earcons and recorded speech could be beneficial in e-learning interfaces. However, it highlighted the need for further research to explore the contributing role of each of these metaphors. This experiment examined the role of avatar's facial expressions and body gestures in an interactive e-learning interfaces. More specifically, the aim of this experiment was to obtain an overall feedback from users in regards to their evaluation of each facial expression and body gesture when being used in e-learning interface in both the presence and absence of interactive context. It aimed also at specifying which facial expressions and body gestures are

more desirable to the users. Moreover, the experiment aimed at testing the usability aspects and users' learning performance of e-learning interfaces that use avatars as virtual lecturers. In other words, this experiment aimed to explore if there are significant differences among the three applied e-learning platforms in terms of usability and learning performance as well as to identify the most preferred way for presentation of learning information among the three presentation ways implemented in the three experimental platforms: Virtual Lecturer with Facial Expressions Platform (VLFEP), Virtual Lecturer with Body Gestures Platform (VLBGP) and Tow Virtual Lecturers with Facial Expressions Platform (TVLFEP). The following sections present an overview of the relevant work in e-learning and multimodal interaction, the experimental e-learning platforms, the experimental design and methodology, analysis and discussion of the results. It also provides conclusion and directions for future work.

E-LEARNING

The accelerated developments in computer networks and machines resulted in facilitating easier and faster access to a huge amount of educational content. Therefore, research in the field of e-learning as well as the technologies employed in the development of e-learning applications has been increased. Scheduled and on-demand delivery platforms are examples of the technology used in e-learning [5]. Scheduled delivery platforms such as video broadcasting, remote libraries, and virtual classrooms imitate real learning environ-

*Addres correspondence to this author at the Department of Software Engineering, Faculty of Information Technology, Applied Science University, Amman 11931, P.O.Box 926296 Jordan; Tel: 00962 6 5609999; Ext: 1053; E-mail: mnks1967@yahoo.com

ments but with time and place limitations. This technology has been enhanced by the on-demand delivery platforms that facilitate anytime and anywhere learning in the forms of interactive training CD ROMs and web-based training. The Internet technology could be beneficial for the learning process in terms of handling the learning content and monitoring students' progress [6]. It is expected that there will be about five million online learners within the next ten years [7]. In comparison with traditional learning, e-learning offers more flexible learning in terms of time and location and allows better adaptation to individual needs [6]. E-learning also enables online collaborative learning over the Internet [8] and could be used to suit a variety of pedagogical teaching approaches [9]. Additionally, e-learning could increase learners' motivation and interest about the presented material [10]. Nevertheless, technology needed in e-learning is not always accessible [11]. Furthermore, it was found that students felt uncomfortable with computer-based learning and missed traditional face-to-face interaction with teacher. Therefore, users' accessibility and their attitude in regard to e-learning should be enhanced [12].

Pedagogically, it is not always true that every e-learning virtual environment provide high-quality learning and so, fundamental pedagogical principles must be applied to insure successful e-learning solutions [13]. According to Govindasamy [13], development and evaluation of e-learning involves learner and task analysis, defining instructional objectives and strategies, testing the environment with users and producing the initial version of the e-learning tool. Also, e-learning interfaces should be designed to support users' individual differences and enable them to learn independently [14].

MULTIMODAL INTERACTION

Multimodal interaction is a human computer interaction in which more than one human senses is involved. It could be utilised to enhance the usability of user interfaces. Multimodality allows conveying different information using different channels [15]. Also, it enables users to employ the most suitable communication metaphor to their abilities [16]. So, learning experience could be enhanced by the assistance of Information and Communication Technology (ICT) where visual, aural, haptic and other channels could be integrated in a multimodal approach to perceive and learn the communicated disciplines.

Sound and visual output are complement to each other and variety of information could be distributed across both. However, sound is more flexible as it can be heard without paying visual attention to the output device. Speech sounds could be used to communicate the current state of the system through auditory feedback [17]. It also could help users with visual disability [18]. Speech sounds could be categorized into natural speech and synthesized speech. In comparison with the synthesized speech which is created by speech synthesizers, natural speech was found to be more understandable [19]. A study performed by Ciuffreda and Rigas [20] showed that speech sounds could contribute with other multimodal interaction metaphors such as graphics and nonspeech sounds in improving the usability of search engines interfaces in terms of learnability and memorability of users

as well as reducing their errors and the time they spent in completing the required searching tasks. It was found that the incorporation of recorded speech and short musical sounds (earcons) helped users to perform different learning tasks more successfully [21].

Avatars

Avatar is another interface component through which both auditory and visual human senses could be involved. It is a computer-based character that could be utilized to represent human-like or cartoon-like characters [22]. It has been used in interactive computer interfaces to communicate verbal and non-verbal information through facial expressions and body gestures [23]. Facial expressions show human emotions, feelings, and linguistic information by different modalities such as lip synchronization, eye gazing and blinking [24] however, body movements are usually used in everyday life to confirm our speech. According to Gazepidis and Rigas [25], the most popular facial expressions are happy, interested, amazed, and positive surprised. It was found that users' satisfaction and their ability to understand and remember the provided knowledge has been enhanced by the incorporation of speaking avatar with facial expressions [26]. Also, facially expressive avatars were used to improve users' involvement and enjoyment in instant messaging applications [27]. Several studies have been carried out to evaluate the role of avatar as a pedagogical agent in e-learning. Results of these studies showed the positive contribution of avatars in terms of facilitating the learning process [28-30]. Furthermore, avatars could be employed in e-learning environments to enhance users' attitude towards online courses [31]. Fabri *et al.* [32] suggested that facially expressive avatars could be used to teach users with special needs (i.e. autism). A study conducted by Theonas *et al.* [10] demonstrated that the use of facial expressions particularly the smiling resulted in a more interesting and motivating learning experience and improved students' learning performance.

EXPERIMENTAL E-LEARNING PLATFORMS

Three different e-learning platforms were built from scratch to serve as a basis for this empirical study. In addition to textual brief notes and graphics, these platforms has been designed to utilize speaking avatars as virtual lecturers with human-like facial expressions and body gestures as well as natural recorded speech with prosody in order to offer an audio-visual presentation of the learning material. This presentation was provided in three forms: avatar with facial expressions and recorded speech, avatar with full body gestures (and facial expressions) and recorded speech, and tow avatars with facial expressions and recorded speech. It is believed that using avatar in this manner imitates to a large extent the face-to-face interaction that typically takes place between the lecturer and the learners in classrooms. Three different lessons about class diagram notation were communicated to the participants using the aforementioned presentation modes. 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

content of these lessons were adapted from [33] and its duration was 3.24, 3.28 and 5.9 minutes respectively. Although the presentation of these lessons varied across the three platforms, the content and the format was the same.

Virtual Lecturer with Facial Expressions Platform (VLFEP)

This platform, as can be seen in Fig. (**1**) utilized an expressive avatar with facial expressions as virtual lecturer. The interface provided command buttons to enable selecting the lesson to be presented. It also provided two separate components for the presentation process; the speaking expressive avatar on the right-hand side of the interface and the PowerPoint presentation show on the left-hand side. When user clicks the button of a given lesson, this button is highlighted with the red colour to indicate the current lesson, and the virtual lecturer starts presenting the lesson supported by the textual brief notes and graphical illustrations displayed in the PowerPoint slides simultaneously. The interface also offered the pause/play functionalities which enable the user to control his/her learning at any point of time. Upon completion of each lesson, user can ask two questions related to the delivered material in that lesson. When clicking the ask question button, the first question displayed textually below that button and user had to click the get answer button in order to obtain the answer which at that moment provided by the speaking virtual lecturer with relative textual and graphical explanations. The same procedure was followed in asking and answering the second question. In order to ensure consistency and to confirm controlling the experiment, the questions of each lesson were the same for all users. Furthermore, this platform provided text box in the bottom part of its interface to inform the user which lesson is currently presented.

Virtual Lecturer with Body Gestures Platform (VLBGP)

This platform employed speaking and expressive avatar with full body gestures to virtually lecture the experimental learning lessons. Contrary to the VLFEP, both the body-animated virtual lecturer and the PowerPoint presentation of

Fig. (1). An example screenshot of the virtual lecturer with facial expressions platform (VLFEP).

the learning content were combined in the same scene and shown in one component placed in the middle part of the interface. This approach is considered as the most relevant 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 the classroom. Similarly to the VLFEP, interface features for lesson selection, pause/play, asking and answering questions and current lesson highlighting were provided by the interface of VLBGP. Also, the same procedure for lesson presentation, asking and answering questions were followed. A screenshot of VLBGP is shown in Fig. (**2**).

Fig. (2). An example screenshot of the virtual lecturer with full body gestures platform (VLBGP).

Tow Virtual Lecturers with Facial Expressions Platform (TVLFEP)

As shown in Fig. (**3**), the aspect that differentiates this platform from the VLFE and VLBG is that its interfaces involved male and female facially expressive virtual lecturers were both of them shared the presentation of each lesson supported by the PowerPoint show placed in the middle part of the interface. The learning content of the three lessons were equally distributed among both lecturers. Additionally, the interface of TVLFE included two more avatars to represent male and female students. In contrast with the VLFE and VLBG platforms, the role of the latter tow avatars was to ask the tow questions related to each lesson vocally. The first question was spoken by the female virtual student by clicking the ask question button placed directly below her. This question concerned 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. However, the second question was asked by the male student and related to the lesson part that has been presented by the female lecturer hence; she provides the answer in this case. So, two additional ask question/get answer buttons were provided in the interface and placed below the male virtual student. The remaining interface features were similar to those provided by the VLFEP and VLBGP platforms.
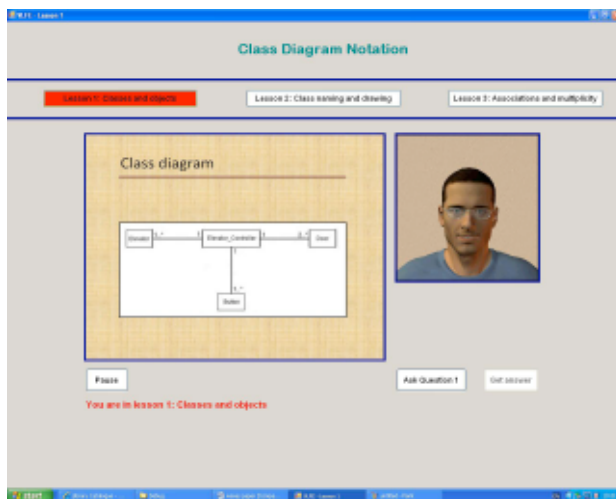
Fig. (**4**) shows examples of the facial expressions and body gestures used in the experiment. Six common facial

Fig. (3). An example screenshot of the two virtual lecturers with facial expressions platform (TVLFEP).

Fig. (4). Facial expressions and body gestures used in the experimental e-learning platforms.

expressions were used in both VLFEP and TVLFEP, and another 10 body gestures were used in the VLBGP platform. These expressions and gestures are typically used in our everyday life. The facial expressions were classified into two groups; positive and neutral [34] while the body gestures were categorized into positive, neutral, and negative [35]. This categorization of expressions and gestures is shown in Table 1.

Table 1.   **The used Facial Expressions and Body Gestures Classified According to [34] and [35] Respectively**

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

EXPERIMENTAL DESIGN AND METHODOLOGY

The within-subjects design methodology was followed in carrying out this experimental investigation. This experimental 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. It also requires less number of participants [36]. Therefore, one group of users was involved to test the three experimental e-learning platforms: VLFEP, VLBGP, and TVLFEP. A total of 48 users have taken part in the experiment in an individual basis. The participation of this number of users can provide satisfying results in terms of system usability [37].

Hypothesis

Based on the assumption that the inclusion of animated human-like virtual lecturers with facial expressions, body gestures and natural speech would affect the usability and learning performance in e-learning interfaces, the main hypothesis stated that there experimental platforms will differ in terms of efficiency, effectiveness and user satisfaction.

Variables

Three types of variables were considered in this experiment which were: independent, dependant, and controlled variables. The independent variables are the manipulated factors therefore, the presentation mode which had three conditions represented by the experimental platforms. The dependant variables are the obtained results due to the experimental manipulation including task completion time, tasks successfully completed and the satisfaction of users. The controlled variables are the external factors associated with the procedure of the experiment. These variables were kept consistent to avoid its influence on the dependant vari-

ables and to insure that the independent variables are the only cause of the experimental results. Therefore, all users performed the same tasks and none of them were aware of these tasks. Also, the same procedure was followed during the execution of the experiment including measurement tools and used equipments.

Users

A total of 48 users participated in the experiment in an individual basis. All of them were volunteers and first-time users of the tested platforms. They were selected to be inexperienced in class diagram notation; the learning material used in this study. Involving expert users in this regard would bias the experimental results because most probably they will 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. The majority of users were male (73%), postgraduate students (83%) coming from a scientific background (82%) and 92% of them were 25-44 years old. Also, they were considered as experts as 81% of them use Computer and Internet more than 10 hours a week. Moreover, 85% of the participants inexperienced in class diagram notation. In regards to e-learning applications, about 62% were found expert, 83% out of them expressed their missing of face-to-face interaction with the lecturer. Most likely, this experience would enable the users to provide precise feedback based on comparing the tested platforms with other e-learning systems.

Procedure and Tasks

The experiment was clearly explained to each user and started by filling the pre-experimental questionnaire for user profiling. Thereafter, a 2-minute video recording was presented demonstrating the experimental platforms. Once this recording had finished, the three lessons about class diagram notation were introduced in an interactive learning context. Due to dependency of lessons on each other, the order of these lessons was constant for all users but each experimental platform had to be used for the presentation of only one lesson at a time. In order to control the learning effect and to make sure that all experimental platforms had been equally used for the presentation of each lesson, these platforms were assigned to the three lessons on a systematic random rotation basis. Upon completion of each lesson, user has been asked to answer 4 evaluation questions related to the delivered learning material. These questions were divided equally into recall and recognition. In order to answer the first type of questions, user had to recall part of the presented learning content, however the recognition questions offered multiple choice alternative answers and user had to recognize the correct one among them. In total, each user answered 12 questions; half of which recall and the other half recognition. Furthermore, user had to respond to the satisfaction questionnaire composed of 18 statements. The first 10 statements were based on the SUS; System Usability Scale [38] to measure users' attitude towards different aspects of the applied platform whereas the remaining statements had been added to obtain users' feedback about their learning experience with each of the three experimental platforms.

Finally, the post-experimental questionnaire asked the users to evaluate the usefulness of the implemented multimodal features as well as to identify the most preferred experimental platform.

RESULTS AND ANALYSIS

Results of the experiment were analysed in terms of efficiency (time spent in answering the required questions), effectiveness (accuracy of users' answers), user satisfaction and the obtained feedback from the post-experimental questionnaire. Based on the random rotation of the three platforms over the three lessons, each platform has been used 16 times in the presentation of each lesson. Therefore, the total number of observations for time and correctness of answers was 192 each; 96 for recall and 96 for recognition. For statistical analysis, the repeated-measures ANOVA were used at alpha .05 to test the existence of significant differences among the three applied platforms with respect to each of the aforementioned parameters. Also, follow-up comparisons were carried out using post-hoc Bonferroni test to find out which presentation modes differed from each other.

Efficiency

Efficiency of each experimental platform was measured by the time users consumed to answer the questions related to the learning material when it has been presented by that platform. This measure was considered for all questions in total and according to the question type; recall and recognition. The total time taken to answer all questions was 3100.40 seconds in VLBGP compared to 3882.02 in VLFEP and 4229.97 in TVLFEP. Fig. (**5**) shows the mean (M) answering time in each condition for all questions as well as for each question type (recall and recognition). It can be seen from Fig. (**5A**) that the VLBGP was the most efficient; $M = 16.18$ (standard deviation $SD = 10.07$) followed by the TVLFEP ($M = 20.22$, $SD = 13.90$) and the VLFEP ($M = 22.03$, $SD = 13.23$) which found the least efficient platform. On the whole, the ANOVA results demonstrated that there was a significant effect of the presentation mode on the time spent by users' to answer ($F(2, 382) = 12.15$, $p<.05$). The pairwise comparisons revealed that the answering time in VLBGP was significantly lower than both the VLFEP ($p=.000$) and the TVLFEP ($p=.002$).
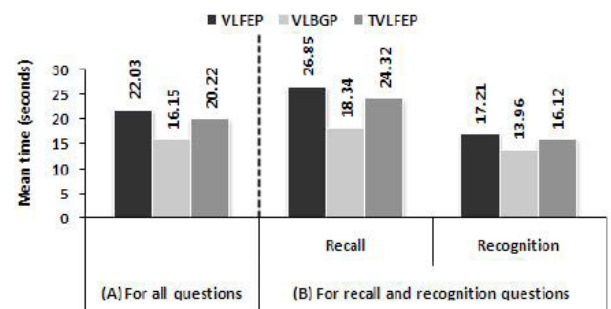
Fig. (5). Mean answering time for all question (A) and grouped by question type (B) in each of the experimental platforms.

Fig. (**5B**) shows a break down of the time results by question type. In recall questions, the least time was spent when the learning lessons have been delivered by the virtual

lecturer with full body animation (M = 18.34, SD = 10.71) followed by TVLFEP (M = 24.32, SD = 15.68) and VLFEP (M = 26.85, SD = 14.72) respectively. According to ANOVA, the time for answering recall questions was significantly affected by the presentation mode (F(2, 190) = 9.05, p<.05). The parwise comparisons revealed significant declines in answering time from the TVLFEP to VLBGP (p=.006) and from VLFEP to VLBGP (p=.000). Comparing VLFEP and TVLFEP, no significant difference was found (p=.821).

Similar to results of recall questions, the VLBGP scored the lowest time for answering the recognition questions (M = 13.96, SD = 8.91) and longer times were observed when facially expressive talking head has been incorporated either by a single virtual lecturer in VLFEP (M = 17.21, SD = 9.41) or by tow ones in TVLFEP (M = 16.12, SD = 10.41). However, no significant difference across the three conditions was reached for recognition questions (F(2, 190) = 2.94, p>.05).

Effectiveness

The measure of effectiveness of the three experimental platforms has been specified according to the number of correct users' answers obtained when each of these platforms has delivered the learning material. Fig. (**6**) shows the percentage of correctly answered questions in each condition for all questions as well as for each question type (recall and recognition). Using the VLBGP, users achieved 80% effectiveness rate (M = 3.19, SD = 0.84) compared to 63% (M = 2.50, SD = 1.15) using the TVLFEP and 61% (M = 2.46, SD = 1.01) using the VLFEP (see Fig. **6A**). The ANOVA showed significant difference in users' performance among the three experimental conditions (F(2, 94) = 12.22, p<.05) indicating that the presentation mode significantly affected users' ability to answer the required questions successfully. More specifically, results of the multiple comparisons revealed that the use of single virtual lecturer with body gesture; VLBGP, performed significantly better than the use of tow facially expressive virtual lecturers; TVLFEP (p=.000). The VLBGP condition also outperformed the use of single facially expressive virtual lecturer, VLFEP (p=.000). The difference between the VLFEP and the TVLFEP was not significant (p=1.000).

In recall questions, Fig. (**6B**) demonstrates that users' performance was better when using the VLBGP compared to

the other tow platforms. Using the VLBGP, the total number of correct answers to recall questions was 75 out of 96 giving 78% performance rate (M = 1.56, SD = 0.54) whereas a smaller number of correct answers to the same type of questions was observed when using VLFEP, where users correctly answered 50 recall questions, slightly higher than the half; 52% (M = 1.04, SD = 0.71). When tow virtual lecturers shared the delivery of the lessons, users' performance dropped further and only 50% of the recall questions have been correctly answered (M = 1.00, SD = 0.74). Based on ANOVA calculations, users performed significantly differently across the tested platforms (F(2, 94) = 13.61, p<.05). When each pair of the experimental conditions has been compared, the use of VLBGP was significantly more effective the use of other presentation modes; VLFEP (p=.000) and TVLFEP (p=.000). However, no significant difference was found between the latter tow conditions (p=1.000).

Although users' performance was better in the recognition questions (see Fig. **6B**), the tested platforms did not differ significantly in terms of the correctness of users' answers to this type of questions (F(2, 94) = 1.48, p>.05). Nevertheless, the VLBGP scored the highest percentage of users' correct answers 81% (M = 1.62, SD = 0.57) as opposed to the TVLFEP 75% (M = 1.50, SD = 0.65) and the VLFEP 71% (M = 1.42, SD = 0.71).

User Satisfaction

The SUS questionnaire has been used to measure the satisfaction of users after they have had the opportunity to use each of the experimental platforms being assessed. Also, users were required to respond to additional eight statements related to the interface components and learning experience. Each of the 18 statement was based on a 5-point Likert scale where 1 represented strongly disagreement and 5 represented strongly agreement. For the analysis of results, the SUS scoring method has been used for the first ten statements, whereas the mode and median were calculated for the remaining statements. Findings showed that the VLBGP scored the highest SUS satisfaction score (M = 85.05, SD = 12.03) compared to TVLFEP (M = 79.45, SD = 11.84) and VLFEP (M = 77.97, SD = 12.10). The ANOVA showed an overall significance in terms of the differences in users' attitude towards the three presentation modes (F(2, 94) = 8.95, p<.05). The results of follow up pairwise comparisons found the VLBGP significantly more satisfactory than TVLFE
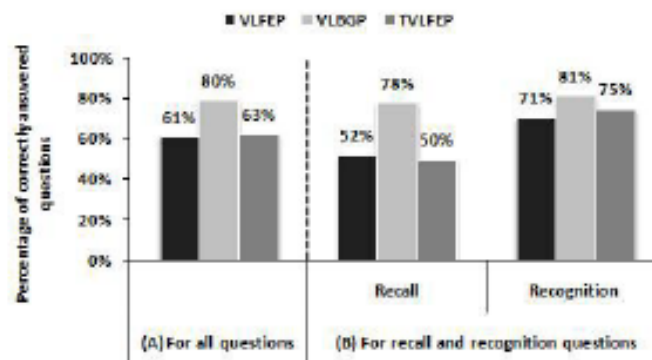

Fig. (6). Percentage of correct answers for all question (A) and grouped by question type (B) in each of the experimental.

(p=.010) and VLFEP (p=.001). However, users' satisfaction did not differ significantly between VLFEP and TVLFEP (p=1.000).

The mode and median scores for each of the SUS statements are shown in Fig. (**7**). In regards to the first statement (S1), the same value 4 was noted for both mode and median in the three platforms indicating that users agreed in *using the experimental platforms frequently*. The second statement asked users whether they found the system *unnecessarily complex*. Users expressed their disagreement in VLFEP with mode noted 2, whereas they strongly disagreed in both VLBGP and TVLFEP with one point lower mode value, however the median values was tow in both VLFEP and TVLFEP compared to one in VLBGP. Wherever the facially expressive virtual lecturers have been used, similar agreement ratings 4 were observed for the *ease of use* (S3) in both VLFEP and TVLFEP whilst strongly agreement was found when VLBGP was used. Consistently, the mode values for the fourth statement; *I think that I would need the support of technical person to be able to use this system*, were 1 on all platforms reporting strongly disagreement among users. The median value was 1 in VLGBP as opposed to 2 in the other tow platforms. The results of the statements S5, S6 and S7 were similar where users equally agreed that they found *the various functions in all the experimental platforms as well integrated* and that these platforms could be *learnt quickly*. Also, users equally disagreed that the tested platforms had *too much inconsistency* (S6). In regards to S8; *I found the system very cumbersome to use*, the level of users' disagreement was stronger in both VLBGP and TVLFEP in comparison with the VLFEP. When virtual lecturer has been applied with full body gestures, users felt higher degree of *confidence* (S9) with mode and median values 5 and 4.5 respectively. However, this confidence was slightly lower when facially expressive virtual lecturer has been used in VLFEP and TVLFEP with mode and median values 4. According to users responses to S10; *I need to learn a lot of things before I could get going with the system*, users expressed stronger disagreement in both VLBGP and TVLFEP in comparison with VLFEP.

In addition to the SUS statements, the statements S11 to S18 were included to obtain users views about their learning experience and interface components as well as the incorporated multimodal features. More specifically, these statements investigated users' excitement and interest about the presented lesson (S11), whether the asking and answering feature helped to improve their understanding of the presented material (S12), and their level of control over learning (S13). Also, the 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. The last tow statements aimed to evaluate, in overall, users' satisfaction (S17) and their learning experience with the applied interface (S18). Users' responses to the additional eight statements are shown in Fig. (**8**). The same level of users' ratings for all these statements could be observed in VLFEP and TVLFEP with mode and median valued four. In this regard, the users expressed their agreement, although they believe that the inclusion of one or tow virtual lecturers with facial expressions in e-learning interfaces does not make any differences regarding the examined features. When the virtual lecturer with full body animations has been experienced, users showed a stronger agreement with respect to most of the added statements where the mode and median ratings were 5 for statements S11 and S12 however the most occurring value and median score were 5 and 4.5 respectively for S14 and S16 - S18. Using the VLBGP, the users felt more excited and interested about the presented lessons, and the way asking and answering questions simulated in this platform enhanced their understanding further. Additionally, the full-body animation of the virtual lecturer's body was more enjoyable for users and encouraged them to pay more attention to the introduced information. Also, it was easier for users' to comprehend this information when presented by the VLBGP. The results of statements S13 and S15 found similar for the three conditions. On overall, users were more satisfied and gained more enriching learning experience with the implementation of VLBGP presentation mode.

Post-Experimental Users' Feedback

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.
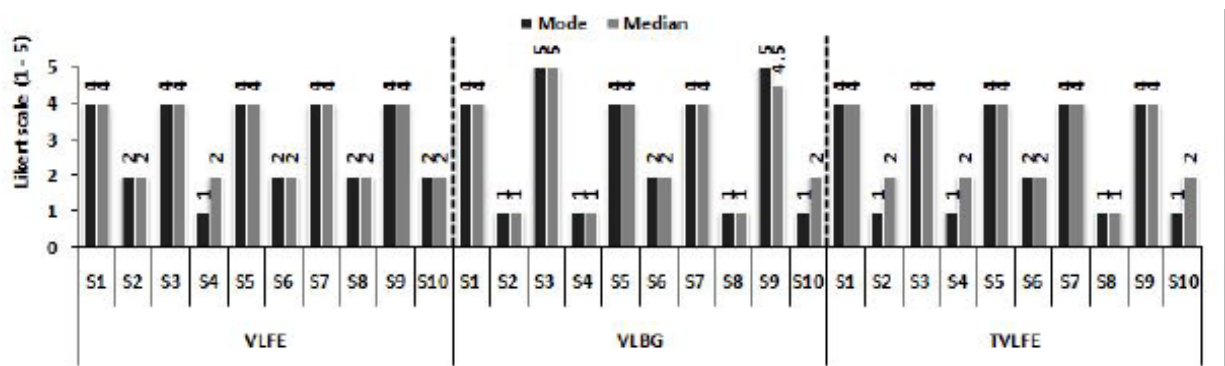

Fig. (7). A comparison of users' rating for the SUS questionnaire statements across the experimental platforms.
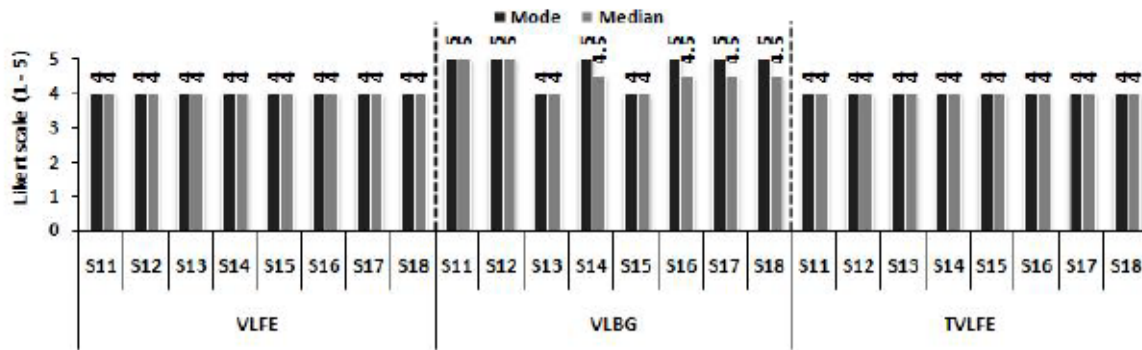
Fig. (8). A comparison of users' rating for the additional statements in the satisfaction questionnaire across the experimental platforms.
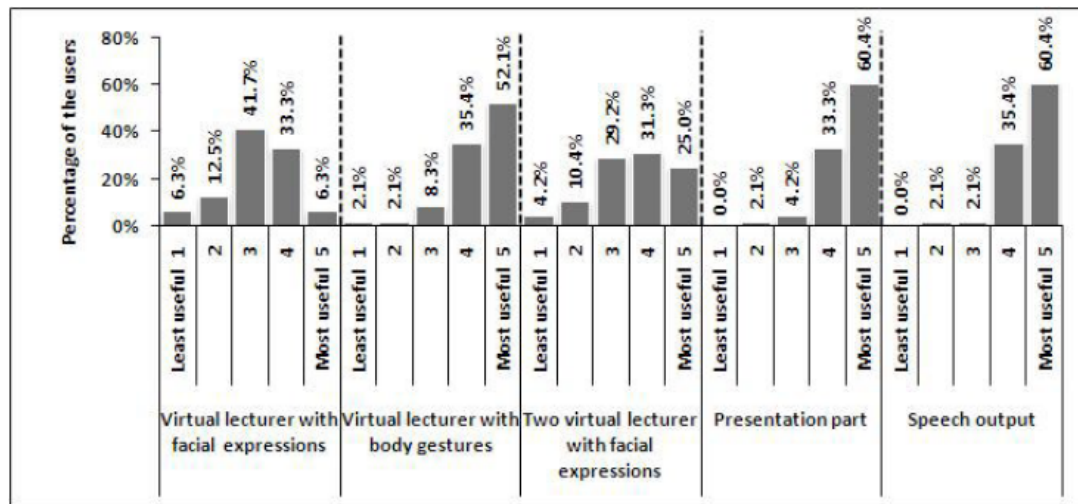

Fig. (9). Users' ratings for the usefulness of the multimodal metaphors used in the experimental platforms.

Fig. (**9**) shows how useful users found each of the multimodal features implemented in the experiment. According to users' views, incorporating tow virtual lecturers with facial expressions was found to be more impressive than the use of only one virtual lecturer with facial expressions as the observed most useful rate was VLFEP about 6% and TVLFEP 25%. In comparison, employing body gestures by the virtual lecturer found the most useful for users where slightly more than half of them consider their learning to be substantially assisted by this mode of presentation. In addition to the speech output, the textual brief notes and graphical illustrations displayed in the PowerPoint presentation part have been used in the same manner in the three platforms and seem to contribute beneficially in users' learning as both modalities attained 60% most usefulness rate. According to users' feedback, the most preferred e-learning platform was the VLBGP obtaining 79% preference rate among users. This percentage dropped dramatically to 19% for TVLFEP and 2% for VLFEP.

DISCUSSION

The experimental study reported in this paper investigated three different modes of employing avatars as virtual lecturers in the presentation of the learning material. Namely, these modes were virtual lecturer with facial expressions platform (VLFEP), virtual lecturer with full body gestures platform (VLBGP), and tow virtual lecturers with facial expressions platform (TVLFEP). The empirically obtained results have been used to compare the three ways of presentation in terms of efficiency, effectiveness and user satisfaction.

The difference among the three experimental platforms with respect to each of this usability parameter has been predicted in the main hypothesis. In answering the required evaluation 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 were differed across the three conditions. This difference was found to be significant by ANOVA calculations. As a result of multiple comparisons among the three platforms, the VLBGP was found to be the most efficient and most effective as well as the most satisfactory presentation mode. The way used in VLBGP platform to present the learning material enabled the users to be engaged in learning environment similar to the real lecturer-to-learner human interaction take place in the traditional class rooms. In addition, users were more attracted, excited and interested about the presentation (refer to (Fig. **8**) S11, S14, S16 and S18). Furthermore, presenting the learning material in the background of the virtual lecturer within the same interface component helped the users to watch both at the same time reducing their visual over-

load compared to VLFEP and TVLFEP. Hearing the auditory spoken explanations concurrently with watching the learning material could help in understanding the presented information. Therefore, users were more attentive and better concentrated to what is being presented and spoken. Also, they kept involved in cognitive processing of the delivered learning information and got better understanding of it (Fig. **8** S12 and S16). 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 the virtual lecturer either in VLFEP or TVLFEP. Additionally, the VLBGP significantly outperformed the other tow experimental conditions in terms of the correctly answered questions and user satisfaction. Moreover, the experimental results demonstrated how users rated the VLBGP to be the most useful and preferred interface.

The comparison between the VLFEP and the TVLFEP platforms revealed that the usability levels in terms of the three parameters (efficiency, effectiveness and user satisfaction) was equivalent and no significant differences have been noted at all. Even though, the use of tow facially expressive virtual lecturers performed better than utilising only one, and its usefulness was rated higher. During the experiment, it was observed that users' concentration has been spread out with the use of TVLFEP platform. It seems that the existence of tow facially expressive virtual lecturers and additional tow virtual students in different interface components (refer to Figure 3) 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 the talking head of a single virtual lecturer with facial expressions in an interface component different from that used to present the textual notes and graphics did not attract the users as much as the VLBGP and this was evidenced by the results shown in Figure 8 (S11 - S12, S14, S16 and S18) where users judged both VLFEP and TVLFEP equally in terms of learning experience.

With respect to the question type (recall and recognition), the efficiency and effectiveness of the three experimental platforms was varied. There was a difference across the three 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. As expected in H7, these differences were found to be significant for the recall questions only however 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 other type of questions. In recall questions, users may have taken more time in trying to retrieve the required information to answer, 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. In recognition tasks, users had to successfully select 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, the recall question are more difficult to be correctly answered because no options are provided and users had to rely only on their memory to retrieve the correct answer which is far to occur due to chance.

Nevertheless, the statistical calculations showed significant difference across the three experimental conditions in the recall questions results in favour of the VLBGP either in terms of answering time or correctness of answers. However, no significant difference has been revealed in the recognition questions results regarding both measures; time and correctness. These results demonstrated the obvious effect of including the virtual lecturer with body animations, as applied in the VLBGP, on users' performance in responding to the recall activities faster with higher accuracy. At the same time, this effect did not expand to users' performance in the recognition questions. Comparing the results of both VLFEP and TVLFEP, no significant difference has been achieved in both types of questions indicating equal effect of including one or tow facially expressive virtual lecturers either on users' efficiency or users' effectiveness.

Users' attitude towards each of the three experimental platforms was found significantly different confirming what has been hypothesized. In accordance with the post-hoc statistical tests applied on the SUS satisfaction questionnaire, the VLBGP interface was significantly more satisfactory to users comparable to the VLFEP and the TVLFEP interfaces. Also, the satisfaction results shown in Figure 8 offered additional support to the main hypothesis. It should be remembered that 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 in regards to different aspects of the VLBGP interface as well as to the learning experience they gained using this interface. When talking head of facially expressive virtual lecturer has been used, users expressed similar levels of satisfaction with both interfaces; VLFEP and TVLFEP.

CONCLUSION

This paper documented an experimental work conducted to investigate the usability aspects 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 a talking head of single facially expressive avatar while the second interfaces made use of a full body animated avatar. In the third interface, the talking head of tow facially expressive avatars; male and female, were included. 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 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 tow investigated e-learning interfaces. Using the talking head of facially expressive avatar as virtual lecturer was shown to be as efficient, effective and satisfactory as the use of tow talking heads of virtual lectur-

ers. Also, the multimodal audio-visual presentation of the learning material as applied in the VLBGP experimental interface, contributed particularly in memory recall activities much more than the recognition ones. However, the obtained results invoked questions such as: Does the addition of accompanying auditory technology such earcons and auditory icons to the VLBGP interface could contribute in enforcing the influence of avatars body gestures and strengthen it to comprehend both recall and recognition learning activities? Does it make deference if additional full body animated female virtual lecturer included in the interface of VLBGP? Therefore, the next experimental work will be designed to answer these tow research questions. It is believed that the results of this research program will help in producing a set of empirically derived guidelines for the design of multimodal e-learning interfaces.

REFERENCES

[1]   S. Alexander, "E-learning developments and experiences", *Educ. Train.,* vol. 43, pp. 240-248, 2001.
[2]   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)*, vol. 2, pp. 455-459, 2006.
[3]   M. Alseid and D. Rigas, "Efficiency of Multimodal Metaphors in the Presentation of learning Information", *In Proceedings of the 22nd Annual Conference of Interaction, People and Computers XXII, HCI 08*, vol. 2, pp. 107-110, 2008.
[4]   D. Rigas and M. Alseid, "Multi-modal aided Presentation of Learning Information: A Usability Comparative Study," *In Proceedings of IADIS International Conference on Interfaces and Human-Computer Interaction*, pp. 234-238, 2008.
[5]   R. Hamilton, C. Richards, and C. Sharp, "An examination of e-learning and e-books," 2001.
[6]   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.
[7]   J. J. Bjur, "Auditory Icons in an Information Space", *Department of Industrial Design, School of Design and Craft, Goteberg University, Sweden*, 1998.
[8]   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.
[9]   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.
[10]   G. Theonas, D. Hobbs, and D. Rigas, "Employing Virtual Lecturers' Facial Expressions in Virtual Educational Environments", *Intern. J. Virtual Real.*, vol. 7, pp. 31-44, 2008.
[11]   L. Brady, "Fault lines in the terrain of distance education", *Comput. Composit.*, vol. 18, pp. 347-358, 2001.
[12]   G. Shaw and N. Marlow, "The role of student learning styles, gender, attitudes and perceptions on information and communication technology assisted learning", *Comput. Educ.*, vol. 33, pp. 223-234, 1999.
[13]   T. Govindasamy, "Successful implementation of e-Learning Pedagogical considerations", *Internet Higher Educ.*, vol. 4, pp. 287-299, 2001.

[14]   H. Tabbers, L. Kester, H. Hummel, and R. Nadolski, "Interface design for digital courses", pp. In *Integrated e-learning: implications for pedagogy, technology & organization,* W. Jochems, J. van Merriënboer, & R. Koper (Eds). London: Routledge Farmer, pp. 100-111, 2005.
[15]   N. B. Sarter, "Multimodal information presentation: Design guidance and research challenges", *Int. J. Ind. Ergon.*, vol. 36, pp. 439-445, 2006.
[16]   A. Dix, G. Abowd, J. Finlay, and R. Beale, *Human-Computer Interaction,* 3rd ed. Prentice Hall, 2004.
[17]   J. Preece, Y. Rogers, H. Sharp, D. Benyon, S. Holland, and T. Carey, "Human Computer Interaction," Addison-Wesley, New York, 1994.
[18]   L. Lines and K. S. Hone, Eds. "Older adults' evaluations of speech output," *presented at Proceedings of the 5th International ACM SIGCAPH Conference on Assistive Technologies (ASSETS'02)*, Edinburgh, Scotland, 2002.
[19]   E. Janse, "Time-Compressing Natural and Synthetic Speech," *presented at Proceedings of 7th International Conference on Spoken Language Processing*, Denver, USA, 2002.
[20]   A. Ciuffreda and D. Rigas, "A usability Study of multimodal interfaces for the presentation of Internet Search Results", *Intern. J. Comput., NAUN*, vol. 2, pp. 120-125, 2008.
[21]   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.
[22]   R. Sheth, "Avatar Technology: Giving a Face to the e-Learning Interface", *eLearn. Developers' J.*, 2003.
[23]   J. Beskow, "Animation of Talking Agents," *Proceedings of AVSP*, vol. 97, pp. 149-152, 1997.
[24]   J. Beskow, "Talking heads-communication, articulation and animation," presented at Fonetik 96, Sweedish Phonetics Conference, Sweden, 1996.
[25]   N. Gazepidis and D. Rigas, "Evaluation of Facial Expressions and Body Gestures in Interactive Systems", *Intern. J. Comput., NAUN*, vol. 2, pp. 92-97, 2008.
[26]   M. Alotaibi and D. Rigas, "The Role of Avatars with Facial Expressions to Communicate Customer Knowledge", *Intern. J. Comput., NAUN*, vol. 3, pp. 1-10, 2009.
[27]   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*, St. Pölten, Austria, 2005.
[28]   A. L. Baylor, "The impact of three pedagogical agent roles," *presented at Autonomous Agents & Multi-Agent Systems Conference*, Melbourne, Australia, 2003.
[29]   J. Holmes, "Designing agents to support learning by explaining", *Comput. Educ.*, vol. 48, pp. 523-547, 2007.
[30]   R. Moreno and R. Mayer, "Interactive Multimodal Learning Environments", *Education. Psychol. Rev.*, vol. 19, pp. 309-326, 2007.
[31]   L. A. Annetta and S. Holmes, "Creating Presence and Community in a Synchronous Virtual Learning Environment Using Avatars", *Intern. J. Instruct. Technol. Dist. Learn.*, vol. 3, pp. 27-43, 2006.
[32]   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.
[33]   T. C. Lethbridge and R. Laganiere, *Object-oriented software engineering*: McGraw-Hill Education, UK, 2001.
[34]   "GainingFace, http://ccoder.com/GainingFace.". [Accessesed Jan. 2008].
[35]   A. Pease, "Body Language - How to read others' thoughts by their gestures," *Sheldon Press*, 1981.
[36]   A. P. Field, *Discovering statistics using SPSS*: Sage Publications Inc, 2005.
[37]   J. Nielsen. *Usability Engineering*: Academic Press Inc., US, 1993.
[38]   J. Brooke, "SUS: a" quick and dirty" usability scale", *Usability Evaluat. Indus.*, pp. 189-194, 1996.