Multimodal e-Assessment Interfaces: An Empirical Investigation

Dimitrios Rigas and Amirah Algahtani

***Abstract*—**Due to the availability of technology, there has been a shift from traditional assessment methods to e-assessment methods designed to support learning. With this development there is a need to address the suitability and effectiveness of the e-assessment interface. One development in the e-assessment interface has been the use of the multimodal metaphor. However, the associated effectiveness of multimodality in terms of usability and its suitability in achieving assessment aims has not been fully addressed. Thus, there is a need to determine the impact of multimodality on the effectiveness of e-assessment and to identify the benefits to the user. This paper investigates the role and effectiveness of multimodal metaphors in e-assessment by empirically evaluating the effect of multimodal metaphors in combination or on their own. Usability includes efficiency, effectiveness and user satisfaction. The empirical research described in this study consisted of two experiments with 30 users each to evaluate the effect of multimodality, avatars with whole body gestures, earcons and auditory icons. The experimentation also assessed the role that an avatar could play as a tutor in e-assessment interfaces. The results demonstrated the effectiveness and applicability of metaphors to enhance e-assessment usability. This was achieved through a more effective interaction between the user and the assessment interface.

***Keywords*—**e-assessment, multimodality, avatars, earcons, auditory icons.

I. Introduction

Learning and assessment are complementary to each other. Developments in user interfaces and the way that information is communicated continue to influence e-learning or e-assessment systems. Computer-mediated assessment, computer-aided assessment, online assessment and e-assessment are interconnected terms and often used in relation to the utilisation of information technology [1]. Assessments are generally conducted to assess students' progress and to assist on-line student learning. The design of the assessment is widely recognised as a challenge for e-learning systems. It is often an integral part of the learning software [2]. The enhancing of the quality of the learning experience is an important factor. Several pedagogical principles have been suggested to enhance the learning experience including assessment [3]. Several definitions have been introduced for e-assessment but in essence is the use of computers to elicit that a particular level in education has been achieved [4-7]. Usability is an important evaluating parameter in the development of interfaces for e-assessment. Usability examines the effectiveness, efficiency and user satisfaction of a user interface [8 and 9].

Professor Dimitrios Rigas is a University-wide professor at the University of West London, UK. He also supports the Senior Pro-Vice Chancellor (Research & Enterprise) at an associate level. Professor Rigas is also a visiting Professor at De Montfort University, Leicester, UK (e-mail: dimitrios.rigas@uwl.ac.uk).

Amirah Nasser Algahtani is conducting her PhD research studies under the supervision of Professor Dimitrios Rigas at De Montfort University.

II. Literature Review

A. Multimodality

In overall terms, the literal sense of mode is the technique through which a certain work is accomplished. The term "multimodal" refers to accomplishing a task via the use of a number of methods all combined together. Multimodal is in effect the coexistence of more than one communication metaphors [10]. In some cases, multimodality may prevent information overload to the user [11].

The term involves the use of multiple communication metaphors that are mapped to the human senses (hearing, touch, olfaction and taste) but several researchers distinguish between computing modalities and the sensory modalities of psychology. Sharon Oviatt suggest that multimodalities (e.g. speech, touch, hand gestures, eye gaze and head and body movement) are multimedia schemes of output" [12] and [13]. Generally, a multimodal interface is a human-machine interface that uses multiple channels of communication between user and the machine [14].

In this study, the term multimodal metaphor is used to indicate the use of auditory and visual metaphors to represent the information to be used in e-assessment methods. When designing multimodal user interfaces, the selection of modalities, combination and synchronization of the presentation of the modalities is considered to be important [15].

A speech modal is a channel that is used to represent particular information to users using voice [16]. Natural speech involves the use of recorded speech that is recorded, stored and played back [13]. The presentation of information using sound assists to decrease the amount of text and graphic required in the interface [16]. Also, this will utilise other senses such as hearing and sight. Non-speech sound metaphors in auditory channels are non-verbal cues that transmit information around objects in the computer interface. These can be made of digitally recorded or synthesised musical instruments, everyday sound effects, or electronically produced pure tones [17-19]. There is a growing demand for research that recommends merging non-speech sounds (earcons and auditory icons) with graphical interfaces to decrease the visual workload, which impact the users' performance [20]. According to [21] auditory icons are defined as "everyday sounds mapped to computer events by analogy with everyday sound producing events". They provide a method that sounds natural in representing data that is dimensional and also the represents the objects that are conceptual in specific computer schemes. The auditory icons allow the data to be categorised into different sets using a single sound [22]. One of the most important advantages of using these is that the sounds used in them are those which people hear in their daily lives, and link them with a specific action [23]. An example of this can be found in the virtual world where we would hear the sound of an object crashing into a wastebasket when the file is deleted, or marked for deletion. This category of auditory icons is like the sound effects, which complement the visual events with an appropriate sound in a computer scheme. Their purpose is not just simply to serve as entertainment tools but also to convey very important information regarding the events taking place in a computer scheme – this allows the user to listen to the sounds from a computer as he does from the everyday world.

Systems like EAR (Environmental Audio Reminders) play a variety of the non speech audio cues for offices and the common areas within EuroPARC in order to keep us up to date regarding the various events taking place around its building; Share Mon utilises background sounds in order to spread awareness; Sound Shark, the sonic finder, is useful when incorporating the auditory icons in an interface that is well known and used often – the simplicity of it leads people to underestimate the functions that auditory icons are capable of. For this reason, Gaver and Smith [24] demonstrated auditory icons used in a large-scale, multiprocessing, collaborative system called SharedARK, and called the resulting auditory interface SoundShark [25]. However [26] said the analysis of both source and sound are not usually significant although that [26] has introduced an ad-hoc synthesis to let users recognise sound instead of the analysis of source and sound. These systems display the extensive range of functions performed by the auditory icons. These include provision of information regarding the user's actions, the possibility of new actions and also the object's attributes that are not visible in the system. They also provide the background information regarding the modes as well as processes in a system that is more complex.

Earcons are short, non-speech, musical sounds that are used in the interaction processes between computers and users [27 and 28]. Earcons are associated with either objects or actions presented in a computer interface. As earcons require abstract associations with data, users must learn them in an initial training process [29]. Avatars are classified as naturalistic, abstract or realistic. It simulates a person as a graphical image of a user [30]. The avatar can be either the head of a man or woman, or a whole body. The idea behind the avatar is to simulate a real life person who naturally interacts with the user. For example, in e-learning, it can be used as a virtual lecturer [31 and 32]. Avatars often simulate body-gestures in order to better mimic human behavior. Body gestures are part of non-verbal messages. Non-verbal messages communicate a significant amount of information [33, 34]. Although body gestures are culture-dependent, strong messages of emotion and attitudes are communicated [35]. Body gestures in avatars are used to enhance speech and add emphasis [33, 36]. By using our hands, heads and feet, we can represent a very wide range of signs, signals and movements [35]. Avatars also help to "humanise" user interfaces. Humanisation has two objectives; to make the interfaces easier and more enjoyable to use and to make the interface more similar to humans [37]. The process of anthropomorphism offers interfaces to computer schemes via the provision of some human-like characteristics [38].

III. Aims and Objectives

The research question of this study is whether the use of multimodal metaphors (images, earcons, and avatars) has a positive effect on users' learning achievements in an e-assessment interface. The objectives were to investigate:

1) Impact of individual modalities on the effectiveness (i.e. task completed successfully), efficiency (i.e. time taken to complete tasks), user performance (i.e. score achieved by users), and user satisfaction using a post-experimental questionnaire.
2) The suitability of metaphors for specific types of e-assessment interface circumstances.
3) User performance of recall and recognition tasks of use in the presence or absence of expressive avatars with full body gestures, earcons and auditory icons.
4) User performance during the execution of simple, moderate and complex e-assessment interaction tasks.
5) The combined effect of the multimodal metaphors to e-assessment.

IV. Experimental Conditions

The hypotheses were based on the ease of answer, Efficiency and effectiveness of use, and user satisfaction. The assertion is therefore that the presence of multimodality in an e-assessment interface will provide better results that the absence of multimodality in the same interface. The e-assessment experimental platform was designed to provide a 'text only interface' version, and multimodal conditions. All condition presented to users the same information about the test but using different means to communicate this information to users. The types of questions used were true or false questions and multiple-choice. Users had to answer three difficult, three moderate and three easy questions. The non-multimodal interface (VOAP condition) is a text-only version (i.e. multimodal objects removed) but retaining the same feel, assessment, order of chapters and level of questions. The multimodal interface (VMAP condition) introduces avatars, textual description and images.

These avatar's expressions were specifically chosen based on the expressions typically used in daily life to express human feelings [116]. The plan is to use the facially expressive life-like avatar to narrate the explanation of small pieces of information for questions together with an interesting video, where the user may move the mouse cursor over the question and, following that, move the mouse cursor over the available answers from the multiple choices. The avatar will occupy the right side of the screen, so as to suggest to the user that it might assist them in selecting a correct answer and encourages the user to move the mouse cursor over the button to answer the question. In every interaction instance, the design has incorporated the feature by dividing the screen carefully, question by question, in order to avoid overlapping of questions, so the user can easily select answers on the screen. The left part shows a text of the question on a blue background, with a font size of 18 for the test. The avatar occupied the right side of the screen on a black background. When placing the mouse cursor on play in the video, the facially expressive avatar started to speak about and explain the question. The right part of the screen shows the multiple choices available and allows the user to click the correct button. After the user had finished reading and listening to the information, and the user had selected an answer they thought was correct, the user moves the mouse cursor over to the button for the next task. The modality used in the interface communicated the question to assist the user to select the right answer. The use of expressive avatars with body gestures together combined with earcons and auditory icons provides an investigation platform in e-assessment interfaces. The research assumptions are:

1) The provision of a realistic interaction with the user that resembles a face-to-face interaction.
2) Making the learning process easier and increase the user's interest, motivation, and learnability.

V. Assessment Types and Multimodality

There are six types of assessment that communicate information to users in the e-assessment interface. These are error, comment, thinking, explain questions, suggestions and mark. Earcons communicated the correct answer to the question when spoken by the avatar. The aforementioned six types of assessment were grouped in three levels in terms of their ability to help; 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. This ranking refers to the potential of each metaphor (earcons and auditory icons) to assist in communicating the correct answer. For example, the first earcon consisted of only one note to communicate low ability, the second earcon consisted of two notes to show a medium rank.

Auditory icons were also used. The sound of "glass breaking" communicated an error, "opening a bottle lid" communicated that a comment is about to start, a "honking horn" indicated that the thinking has started, "a closing window" the explanation of questions, "door opening" that a suggestion is about to start, and a "hand clapping" that a mark is about to be communicated. Earcons and auditory icons were presented during the pause intervals so that they do not interfere with the spoken messages of the avatar.

VI. Sample and Experimental Procedure

A group of users (n=30) assessed the experimental interface in order to obtain an overall viewpoint of the suitability of the metaphors used. The procedure followed during the experiment is presented below:

1) Anonymous gathering of the sample profile (e.g. educational level).
2) Recording previous knowledge in relation to e-assessment interfaces, expressive avatars, earcons and auditory icons.
3) A short demonstration video introducing the e-assessment interface.
4) Presentation of example instances of the e-assessment interface with particular emphasis upon the multimodal metaphors used. The object of this training was to ensure the user's ability to understand and interpret each of these multimodal metaphors.
5) User performed the experimental tasks and all relevant user data was recorded. Each user was asked to answer 6 questions connected to the delivered assessment type. The questions were of two types; recall and recognition.
6) Post-experimental user questionnaire that gathered their views regarding the various multimodal metaphors used.

Users also had the opportunity to provide suggestions or comments for improvement. The independent variables were:

1) Multimodal metaphors. These were earcons, auditory icons and expressive avatars with body gestures.
2) Types of assessment. These were error, comments, thinking, explain question, suggestions and mark.
3) Assessment questions. These questions included recall and recognition in order to evaluate the users' learning achievement attained from the information presented by the tested e-assessment interface.

The dependent variables were:

1) Completion level (correct answers): This is the number of successfully completed tasks. It was measured by the frequency of correct answers to the recall and recognition questions linked to the communicated assessment.
2) Involvement of users with the type of assessment: This was measured by the number of users who correctly indicated these features after being communicated by the non-speech sounds.
3) Users' recognition of earcons and auditory icons: This was measured via the number of users who successfully interpreted the auditory stimuli in the context of being

communicated in the experimental interface.

4) User Satisfaction: It was measured using questionnaires to gather the views of users.

The following provide descriptive and statistical analysis of the results obtained from the experiment in terms of achievement level, involving (in terms of correct and incorrect users' answers) in addition user satisfaction, and users' views regarding the non-speech sounds that accompanied the avatar body gestures as assist. This was the results of the experimental group consisting of 30 volunteers who took part in the study. In addition, the levels of significance in students responses was examined using the nonparametric Chi-square statistical test at α = 0.05 indicating significant difference when p-value was found less than 0.

VII. Multimodal vs non-multimodal e-Assessment

Overall, 30 users participated in this experiment. The users were randomly assigned to the control (VOAP) and experimental (VMAP) user groups (n=15). Most of the users in each group had some prior experience of e-assessment, indicating that they are likely to rely on the communicated information to answer the required task. The age range in the control group was varied, with around 60% within the 18– 24 demographic and 40% consisting of 25 - 30 years old; 60% of students who participated in this study were males (9 altogether) and 40% (6 altogether) were female; these constituted members of the control group. In the experimental group, the ages were 80% within 18 – 24, 20% within 25 – 34, and 0.0% 31 – 40 years of age. The education stage was found to be predominantly undergraduates with 67 % present in the control group, the remaining 33.3% being postgraduates. In the experimental group, 73.3% of the participants were undergraduates with 26.7% taking part as postgraduates. 13.3% of the control group reported to use computers for 1-5 hours a week, and 40% reported to use them between 6 – 10 hours a week. Furthermore, 46.7% of participants said they used computers for more than ten hours a week and 26.7% for 1-5 hours and a further 26.7% between 6 – 10 hours per week. The remaining 46.7% reported to use computers for more than ten hours a week. With respect to e-learning system experience, around 66% of the control group had more experience compared with 46.7% who had some experience with HCI. It was also found that the most popular reason for internet use was surfing with 53.3% of the control group claiming this was the main usage of the internet and a further 60% of the experimental group claiming the same. Furthermore, more than 66.7% and 46.7% had good multimodal knowledge in both groups, respectively, and at the very least, users had a limited background in multimodal knowledge in both groups.

Figure 3.3 shows the overall time spent by every user in all groups to answer all of the nine questions. Users of VOAP spent more time on the questions compared to users of the VMAP. However, the observed time differences between tasks are highly significant, with the lowest and highest recorded times for the first condition control group being 4.40 minutes (User 7) and 7.04 minutes (User 8) respectively; the mean time being 5.46 minutes. In the second condition experimental group, the maximum time recorded was slightly lower (2.49 minutes by User 4) and the minimum time (4.36 minutes by user 15). The mean time recorded was 3.22 minutes. In short, using multimodal metaphors in communicating the e-assessment material enabled the users in the experimental group to outperform their counterparts in the control group in time spent answering the required questions.

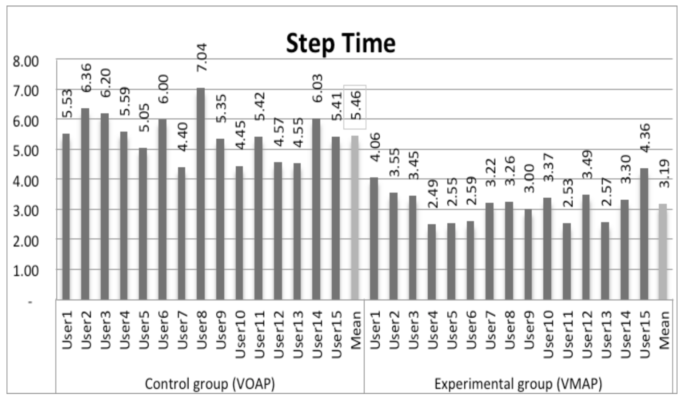

Fig. 1 Mean values of time taken user per group to complete experimental tasks and answer questions

Fig 1 shows the mean value of the time taken by users to answer the questions this was used as a measure of efficiency. This measure was considered for both VOAP and VMAP groups, for all tasks, in accordance with the question complexity (moderate/easy/complex), for all questions and all users in both control and experimental groups. It can be seen that overall, the time taken to answer questions was shorter in the VMAP group. Experimental observations showed that users in the control group regularly divided their visual attention between the symbols provided, which indicated assessment code and assessment content to understand the presented information, and in some cases a visual overload may have occurred. The users in the experimental group, however, kept their visual attention directed towards the assessment content. Fig. 2 explains the answering time grouped by the complexity of the questions which were designed to increase in difficulty and were equally divided into 3 categories: easy, moderate and difficult. All users had to answer nine questions in total. Fig 3 illustrates the mean time taken to answer all questions using the VMAP condition. This did not include reading time as well as the time taken to fill out the pre-task and post-task questionnaire.

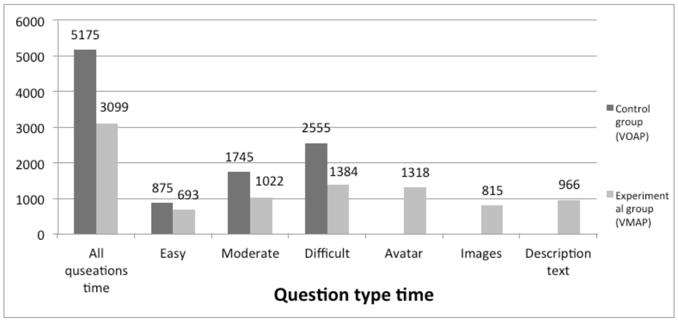

Fig 2 Mean values of time taken by users in both groups to answer all and grouped through question complexity

Overall, the total time taken recorded by users of the VOAP in the control group was 5175 seconds, averaging 5.46 minutes for each user, compared with users of the VMAP in the experimental group who took a total of 3099 seconds to answer questions, averaging 3.19 minutes for each user. It can be observed that users of the VMAP were 2076 seconds faster than those who used the VOAP. The t-test calculations illustrate that the difference between both groups in answering time was significant ($t = 32.252$, $MD = 345$, $sig < 0.5$). The experiments revealed that users in the VOAP group directed their vision towards the questions located in the text box. However, users in the VMAP group maintained their visual awareness to the images and description text, though they were listening to the avatar's messages, which helped them to enhance their focus on the delivered answer ($t = 25.246$, $MD= 206$ $sig < 0.5$).

Fig 2 illustrates the total time spent by each user in the experimental group VMAP to answer all questions. It demonstrates the result of each multimodal implementation (avatar, images, description text), which all included 3 questions. Users of the images multimodal observed a time which was slightly shorter (815 seconds) when compared with users of description text and avatar multimodal in the experimental group. The second shortest answering time was by users of the description text multimodal who took an average of 966 seconds, the difference between images and description text being 151 seconds. Finally, users of the avatar multimodal took the longest (1318 seconds) to answer the questions. In short, users of the images multimodal using VMAP were 815 seconds quicker compared to users using description text and the avatar multimodal to assist in answering the questions.

Fig. 3.5 illustrates the total time spent by each user in the experimental group VMAP to answer all questions. It demonstrates the result of each multimodal implementation (avatar, images, description text), which all included 3 questions. Users of the images multimodal observed a time, which was slightly shorter (815 seconds) when compared with users of the description text and avatar multimodal in the same experimental group. The second shortest answering time was by users of the description text multimodal who took, on average 966 seconds, the difference between images and description text being 151 seconds.

Finally, users of the avatar multimodal took the longest (1318 seconds) to answer the questions. In short, users of the images multimodal using VMAP were 815 seconds quicker than users using description text and avatar multimodal to assist in answering the questions. The percentage of correctly answered questions was used as a measure of effectiveness. This measure was considered for all the tasks in total, according to the multimodal type and question complexity (easy, moderate and difficult) as well as for each user in both control and experimental groups.

Fig. 3 presents the variation between users' performance in relation to the different multimodal, namely; avatar, images, and description text, in terms of correct answers provided by users. In consideration of this, for those who used images, 80% gave correct answers, compared with a 75% for those who used the avatar multimodal and 71% for those who used description text.

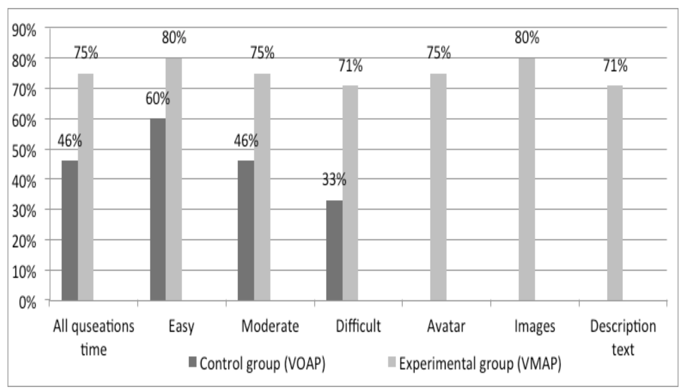

Fig 3 Results by complexity level

Correctly answered questions were used to measure the effectiveness of the metaphors. Figure 3.5 presents the correct answers, as a percentage, for questions in both the VOAP and VMAP. For the latter 75% answered correctly, more than in the VOAP at 46%. The total percentage of correct answers given by users of the images multimodal were 36% indicating that users find information, communication via images to be easier than via avatars and description text. Subsequently, users of the avatar multimodal achieved a 33% score with the lowest percentage of correct answers given by users of the description text multimodal at 32%, making it harder to complete. The mean values of correctly answered questions in the VMAP and VOAP was 7 and 5 respectively.

Fig. 3 shows that the integration of more than one communication metaphor of a different nature in the VMAP model assisted users in the experimental group in highlighting the different kinds of information, which had been delivered through each of the metaphors (description text, avatar, images). As a result, they outperformed the users of the VOAP who received the learning information by images firstly, followed by an avatar and finally by description text. Conclusively, the multimodal interaction metaphors used in the VMAP were more effective in communicating the learning material and considerably assisted the users in the experimental group to achieve a more effective rate, as opposed to the control group users.

Fig 3. illustrates the percentage of correctly answered questions for all levels of complexity, namely easy, moderate and difficult questions for both groups. The results show that the control group was outperformed by the experimental group, this was particularly noticeable for difficult questions. What is more, for easy questions in the VMAP a score of 80% was achieved, more than that about the VOAP condition. However, a larger difference between the two groups was observed for moderate questions and the largest difference was noted for difficult questions. For the VMAP condition, the users in the experimental group answered questions correctly with a score of 80%, 75% and 71% for easy, moderate and difficult questions respectively. In contrast, the users of the VOAP produced a score of 60%, 46% and 33% for the easy, moderate and difficult questions respectively. In summary, it is clear that for easy questions both groups of users achieved equivalent levels of accuracy. However, multimodal metaphors contributed significantly more to a better score with higher complexity questions.

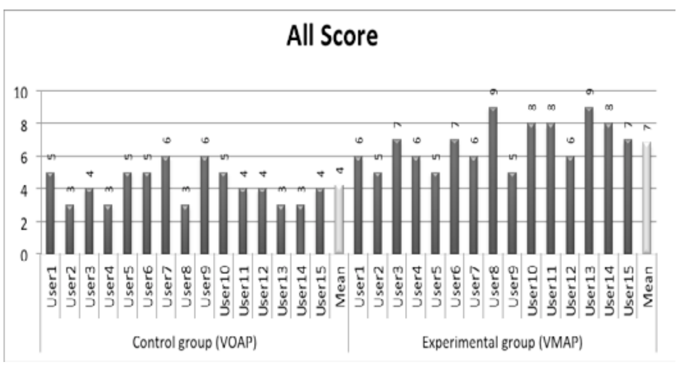

Fig 4 Mean values of user response per condition.

Figure 4 illustrates the total number of correct answers given by each user in both groups: VMAP and VOAP. It is commendable to note that 2 users (8 and 13) of the experimental group correctly answered all nine questions, with a further users (10, 11 and 14) answering all but one (question 8) correctly. However, none of the control group users were able to reach this level of performance. In fact, the highest achievement recorded was 4 correct answers given by users 7 and 9. In addition to this, the users with the lowest score in the VMAP group (2, 5 and 9) were able to correctly answer 5 questions, whereas the users with the lowest score from the VOAP group (2, 4, 8, 13 and 14) were only able to correctly answer 3 questions. The mean, as shown in Fig 5, was higher in the VMAP group (9) compared to the VOAP group (6). In summary, the use of multimodal in communicating the information enabled the users of VMAP to outperform the users of VOAP in answering the questions correctly.

Users were able to express their attitudes to pre-selected statements via the post-experimental element of the questionnaire using the five-point Likert scale, thereby enabling the measurement and recording of user satisfaction. The pre-selected statements were associated with ease of use, complexity, confidence, ease of learning, and also general satisfaction. Each statement was scored using the five-point Likert scale with options ranging from 1 (strong disagreement) to 5 (strong agreement). When calculating the overall satisfaction score, this was conducted using the SUS (System Usability Scale) system [122]. The scoring method followed here is to take the average score for each statement. This mostly results in a positive impact where users liked the VMAP condition more than VOAP condition.

Statistically, the t-test proved that the difference in users satisfaction among both groups was significant (U = 50, CV = 72, $p < 0.05$). In other words, the VMAP was more satisfactory than the VOAP. Figure 3.7 shows the frequency of the user agreement with every statement in the satisfaction questionnaire. High levels of agreement were shown via the users in control and experimental groups for difficulty using the system (S8). Nevertheless, the VMAP was less time while interacting with the system where I felt worried (S4), and high levels of agreement could be easier to learn (S7) as opposed to the VMAP.

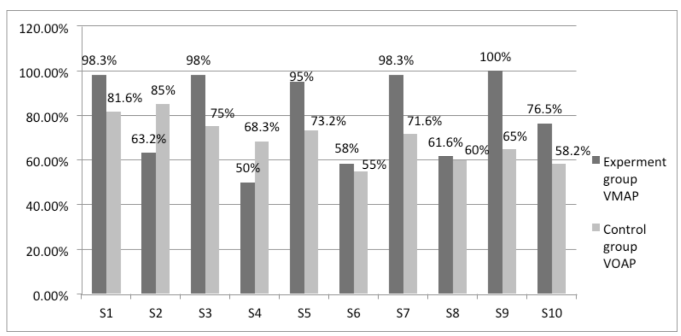

Fig 5 Frequency of users' agreement to each satisfaction statement in both VOAP and VMAP condition

In the first statement (S1), 98% and 81% of users in both groups agreed that the tested e-assessment interfaces were easy to use. The statement (S2) asked the users whether they discovered the system to be unnecessarily complex. Users of the VMAP show a slightly higher level of disagreement 63.2% than the users of the VOAP 85%. It can be noticed that 98% of users for S3 in VMAP thought was easy to use the system, but in VOAP this was 75%. In relation to (S5), the entire VMAP 95% of users found that all functions were well integrated compared with 75% for VOAP. The users found the VMAP condition was easy to understand where the mean satisfaction in the experimental groups was 79% compared to 69% for the users in the control group. The users' satisfaction is significantly enhanced in the VMAP interface in comparison with the VOAP.

VIII. AVATAR-BASED EXPERIMENTS

The test sample consisted of 30 users who took part in the experiment individually. The age profile of the sample consisted of 18 – 24 (13%), 25-30 (26%) and 31- 41 (60%). The gender ratio was 60% male and 40% female. The educational profile of the sample consisted of 14 users (43%) at a postgraduate level and 17 users (56%) at an undergraduate level. 26% of users used computers for between 1 and 5 hours per week, 20% for 6 to 10 hours and 53% for more than 10 hours and 76% of the sample had knowledge about multimodality and e-learning applications. 66% of the sample used internet for surfing and 23% for education. At a pre-experimental level, only 26% of the sample thought that e-assessment would enhance on-line e-learning applications.
The frequencies of correct answers to the assessment questions were used to assess the users' achievement. Each user answered nine questions using recall and recognition methods. The total number of questions was 180 (30 users x 6 questions per user) equally distributed over the two types.

Fig. 6 shows the percentage of successfully completed tasks (correct answers were provided by users) by users grouped according to assessment and question types. The percentage of successfully completed tasks with correct answers was 78%. This statistically significant with chi-square value at 0.200, cv = 3.84, p<0.05. For recall tasks or questions, the successful rates were higher than that for the recognition tasks or questions. The response to the 90 questions was 78.8% for recall and 87.7%. The difference between correct and incorrect answers was significant in both assessment question types; recall chi-square value at 16.8, cv=.200 p<.05 and recognition 7.4, cv=0.200, p<0.05. The percentage of the sample who correctly answered questions linked to "involved thinking" was 86.7% and for "error" 83.3%. The other results were 73.7% for "more suggestion", 70% for "explain question", 60% for "mark" and 53.3% for "comment".

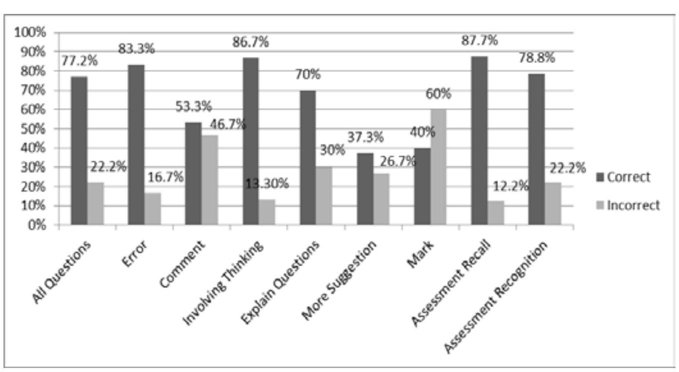

Fig. 6 Correct and incorrect percentages of answers achieved by users for all questions, assessment types and assessment question types

Table. I Chi-square values and significance levels relating to the achievement level

Variable	Chi-square value	Asymp. Sig.	Significance
All assessment question	.200a	.905	No
Assistance Type			
Error	13.333[a]	.000	Yes
Comment	.133[a]	.715	No
Involving Thinking	16.133	.000	Yes
Explain Questions	4.800a	.028	Yes
More Suggestion	6.533a	.011	Yes
Mark	1.200a	.273	No
Assessment questions			
Recall	16.800[b]	.000	Yes
Recognition	7.400[b]	.025	Yes

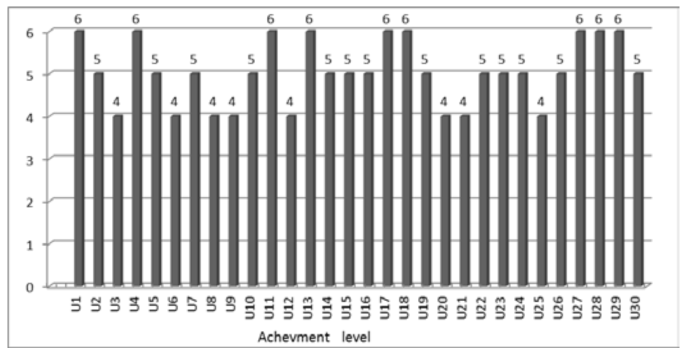

Fig. 7 The number of correct answers provided per user

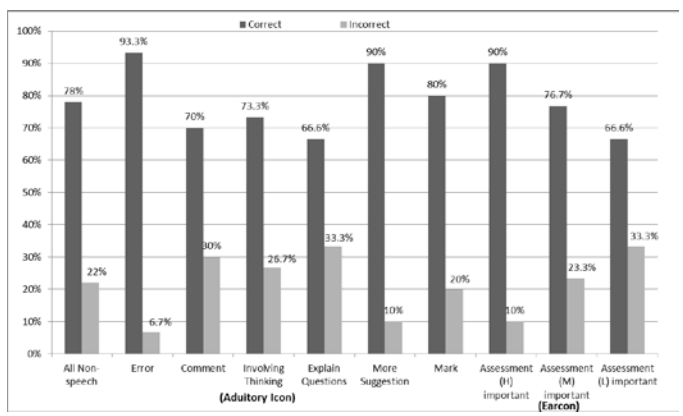

Fig. 8 Correct recognition of users of the assessment types that were communicated by non-speech sounds, earcons and auditory icons

Table I shows that the outcomes were significantly dissimilar for correct and incorrect answers for error, thinking, explain question and more suggestion but there was no significance for comment and mark. Fig. 6 shows the correct answers per user in the sample. Nine users (1,4,11,13,17,18,27,28, and 29) answered each question successfully. The multimodal metaphors used (expressive avatars with full body gestures, auditory icons and earcons) improved the delivery of the assessment content in the e-assessment interface. The auditory messages increased the volume of information communicated by the avatars but did not cause an information overload to users.

IX. User involvement

On completion of the achievement tasks, users were asked to do two more "involving" tests. Users were provided with six different assistance messages using auditory stimuli and they were requested to indicate the type of non-speech sound that was the most effective. The total number of questions was 180 (30 user x 6 questions per user). Fig. 8 shows the correct responses of users to this task related to all non-speech sounds, earcons and auditory icons.

The results ware statistically significant ((1)=15.6, cv=3.84, $p<0.05$). The majority of the users recognised correctly the assessment types communicated via auditory icons. More specifically, 93.3% of the sample (28 users) correctly recognised "error" message using an auditory icon similar to a "broken glass", 90% (27 students) accurately determined the "suggestion" message using a sound similar to a "bottle opening" and 80% (24 users) recognised the "mark" message using a "hands clapping" sound. This percentage decreased to 73.3% and until 66.6% for the remaining assessment types.

Users were requested to perform three tasks with questions that were communicated using non-speech stimuli in order to determine the high, medium or low level of the provided assistance. 90% of the sample (27 users) correctly identified the "high importance" message type compared to 76.7% (23 users) for the "medium importance" and 66.6% (20 users) for the "low importance". In a subsequent evaluation, three types of auditory stimuli were played for each of the assessment types and the importance level of the assistant type. Users had to distinguish the sound that linked each of the assessment types and its level of importance. The obtained results for the non-speech sounds, earcons and auditory icons were encouraging. In total, 84% of the tested sounds were correctly recognised by users. This outcome was highly significant ((1) = 15.6, cv = 3.84, $p<0.05$). 100% (30 users) of the sample correctly recognised the auditory icon that sounded like a "broken glass" to communicate an error and 93.3% for the sound that resembled "opening a bottle" to communicate a suggestion. However, the percentages for the other auditory icons used were 76.6% and 70%.

The earcons used to communicate high, medium and low importance of a message were correctly recognised by all users. These results suggest that the tested auditory icons and earcons were successfully interpreted and more easily remembered by users when utilised in e-assessment condition to signal the importance of particular content delivered by a body gesture. The responses of users were positive (see Fig. 9) with respect to their views and feelings about earcons and auditory icons used interactively. However, 70% of the users felt irritated when they heard the sounds through the experiment. It is noteworthy that there was some difference in user frustration. There was a small difference between agreement and disagreement of 53.3% and 46.7% respectively. For usefulness, 86.6% of users found these sounds to be helpful and 76.6% of users felt that the presentation of sound assisted them to concentrate with the presented content.

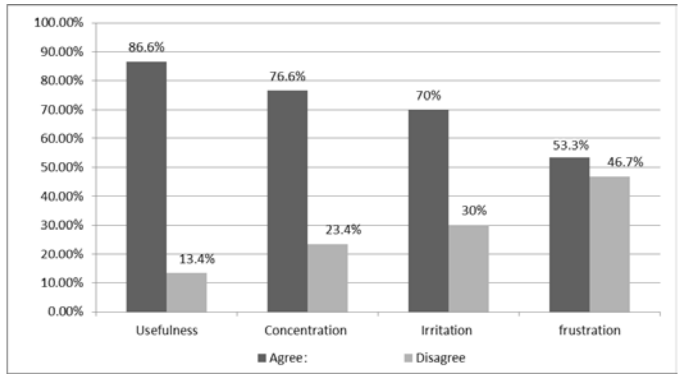

Fig. 9 Results of the user evaluation toward the non-speech sound

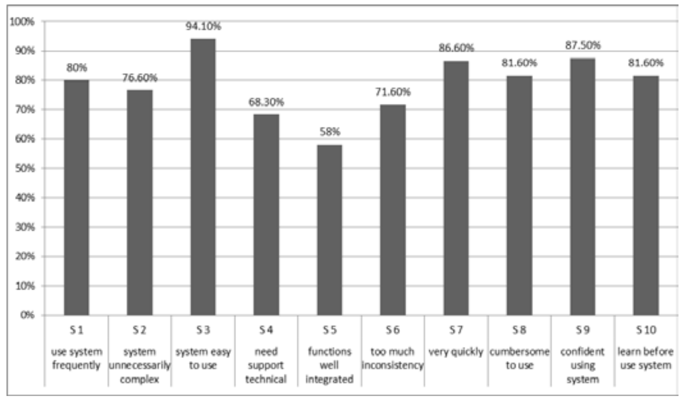

Fig. 10 User response to each satisfaction statement.

X. User satisfaction

User satisfaction was measured using a questionnaire composed of 10 statements. Users provided an answer using a 5-point Likert scale [39, 40] ranging from strong disagreement to strong agreement. These findings provided an overall viewpoint of the users' attitude towards the different aspects of the Auditory Avatar Body Gestures condition.

Fig. 5 shows the user views on their satisfaction for particular aspects of the experimental platform. The mean score for user satisfaction was 81%. The parameters expressed in statements S1, S3, S5, S7, and S9 were the most agreed by users. 85% of the users agreed that the system functions were well integrated (S5) and that most users are likely to learn quickly (S7). 78.5% of users welcomed the use of the auditory icons and expressive avatars (S9). 94.1% of users would use the e-assessment interface again and 80% thought that the interface was simple to use.

Some users disagreed with statements S2, S4, S6, S8, and S10 with rates fluctuating between 68.2% and 81.5%. 81.5% of users needed training to use the e-assessment interface (S10) and 68.2%, disagreed that using the interface requires the need for technical support (S4). Overall, users welcomed the use of the expressive avatars with spoken messages, auditory icons and earcons. On balance, the user satisfaction results were more positive than negative. This demonstrates, from the user satisfaction prospective, that there is a clear prima facie for the inclusion of multimodal metaphors in e-assessment and e-learning applications.

XI. Discussion

This experiment showed that the users had an increased level of concentration on the delivered assessment content. This increased concentration was due to the inclusion of interaction metaphors of diverse modalities in the tested condition. The textual metaphors combined in the condition with body gestures of the assistant avatar contributed to capturing the user's visual attention towards the provided information. At the same time, additional auditory explanations about this information were presented by the voice of the full body gesture avatar. Non-speech sounds did not appear to influence concentration as users were engaged with the assessment content communicated via auditory stimuli. Consequently, users were able to present the correct answer. The results of this experiment showed that user achievement levels were significantly assisted by the addition of earcons and auditory icons that aided the contribution of the body animated virtual instructor to achieve both types of the required evaluation tasks.

Auditory icons significantly assisted users to successfully complete recall and recognition questions. However, earcons were more effective in recall questions than in recognition questions. The earcons used in this experiment were less helpful compared to auditory icons. The outcomes of the experiment indicated that users were satisfied significantly with the inclusion of auditory icons and earcons in evaluating the e-assessment interface (see Fig. 10). Most users stated that these sounds assisted their involvement and did not divert their concentration. Moreover, the auditory icons were chosen because they were the closest environmental sound mapping for the communicated information. Additionally, each of these sounds indicated one meaning at a time and they were used consistently throughout the auditory body gestures avatar interface. This multimodal approach to the e-Assessment interface generated a generally improved user satisfaction and performance. Finally, the obtained results suggest that utilising non-speech sound with body gestures in the form of avatars enhances, to a large extent, the usability and user involvement within the delivery of information in e-Assessment learning interfaces.

XII. Conclusion

The experiment presented in this paper investigated the achievement level and user involvement with the use of earcons and auditory icons used as complementary auditory signals to indicate the dissimilar assessment types as presented by a virtual instructor. The experiment also investigated users' satisfaction. A total of 30 students took part in the experiment to assess the e- Assessment interface as an extension to the interface tested in the previous experiment by adding of Non-Speech sounds. The results showed that these sounds were effective in directing the users' attention to important parts of the Assessment, and contributed positively to enhance user achievement levels in different learning activities. Furthermore, these sounds were memorable, understood, and increased user satisfaction and enjoyment. Consequently, the use of these metaphors was discovered to be significantly useful to enhance the usability of an e- Assessment interface. Ultimately, this study showed the addition of auditory non-speech metaphors to an Avatar Body Gestures condition to allow the user to engage with diverse types of Assessment and questions. Three types of multimodal metaphors were presented which were included in the interface: visual-only metaphors (text which is Assessment type content), audio-visual metaphors (speaking avatar with body gestures) and auditory metaphors (earcons and auditory icons). The collection of experimental data was mostly focused on observations and questionnaires and contributed to the valuation of user's involvement and enhanced user ability performance, such as achievement level and user satisfaction. The results indicated that the users were satisfied, significantly with the inclusion of auditory icons and earcons. Mostly of students stated that these sounds were neither irritating nor frustrating, helped their involvement and did not divert their concentration.

The results of this study highlight the significance of the multimodal metaphors in enhancing learnability performance, as well as the usability of e-Assessment interfaces, in terms of achievement level and user satisfaction. The main limitation of the experiments is that more combinations of multimodal metaphors could have been included. The reason this is important is because it will extend the experimentation to more combination possibilities, thus increasing the validity of the proposed derived guidelines. Although the experiments do determine the best types of multimodal metaphors, both singular, such as images in experiment one, and in combination, such as images and avatars in experiment two and avatars (body gesture) with earcons and auditory icons in experiment three, there are obvious further combinations that be assessed. Currently there is shift in progress between traditional methods for assessing learners and e-assessment using multimodal and other emerging technologies. However, there is still a need to understand the use of technology in this area, especially its appropriateness, applicability and effectiveness in the area of assessment. This study has served to address these issues for the use of multimodality in e-assessment interfaces and will be invaluable to those who are responsible for bringing together technology with assessment, namely, academics and interface designers. It has progressed previous work in the literature [41, 42 and 43] and extended guidelines presented in [44, 45 and 46]. These results were instrumental in the development of set guidelines for the development of e-assessment interfaces using multimodal metaphors, which will be of particular use to developers and software designers. Moreover, the results contribute significantly to the research literature and offer numerous suggestions for future study to take the ideas presented here further, or to overcome present limitations.

References

[1] L. Jantschi, C. E. Stoenoiu and S. D. Bolboaca, "Linking assessment to e-learning in microbiology and toxicology for undergraduate students,"

in EUROCON, 2007. the International Conference on &# 34; Computer as a Tool&# 34; 2007, pp. 2447-2452.

[2] H. Barbosa and F. Garcia, "Importance of online assessment in the e-learning process," in Information Technology Based Higher Education and Training, 2005. ITHET 2005. 6th International Conference On, 2005, pp. F3B/1-F3B/6.

[3] C. N. Quinn, "Seven steps to better e-learning," eLearn, vol. 2006, pp. 2, 2006.

[4] S. Iram, D. Al-Jumeily and J. Lunn, "An integrated web-based e-assessment tool," in Developments in E-Systems Engineering (DeSE), 2011, 2011, pp. 271-275.

[5] G. Busuttil-Reynaud and J. Winkley, "JISC e-Assessment Glossary," 2006.

[6] S. Jordan, "E-assessment: Past, present and future," New Directions, vol. 9, pp. 87-106, 2013.

[7] R. Stowell and R. Lamshed, "E-assessment guidelines and case studies," Interim Report, Australian Flexible Learning Framework and National Quality Council, 2011.

[8] M. S. Crowther, C. C. Keller and G. L. Waddoups, "Improving the quality and effectiveness of computer-mediated instruction through usability evaluations," British Journal of Educational Technology, vol. 35, pp. 289-303, 2004.

[9] A. Koohang and J. Du Plessis, "Architecting usability properties in the e-learning instructional design process," International Journal on E-Learning, vol. 3, pp. 38-44, 2004.

[10] A. Gibbons, Multimodality, Cognition, and Experimental Literature. Routledge, 2012.

[11] C. Raymaekers, "Special issue on enactive interfaces," Interact Comput, vol. 21, pp. 1-2, 2009.

[12] Z. Obrenovic and D. Starcevic, "Modeling multimodal human-computer interaction," Computer, vol. 37, pp. 65-72, 2004.

[13] A. Jaimes and N. Sebe, "Multimodal human–computer interaction: A survey," Comput. Vision Image Understanding, vol. 108, pp. 116-134, 2007.

[14] L. Brown, S. Brewster, R. Ramloll, W. Yu and B. Riedel, "Browsing modes for exploring sonified line graphs," in IN PROCEEDINGS OF BCS-HCI 2002.

[15] N. B. Sarter, "Multimodal information presentation: Design guidance and research challenges," Int. J. Ind. Ergonomics, vol. 36, pp. 439-445, 2006.

[16] S. Brewster, Providing a Model for the use of Sound in User Interfaces. University of York, Department of Computer Science, 1992.

[17] W. W. Gaver, "Auditory icons: Using sound in computer interfaces," Hum. -Comput. Interact., vol. 2, pp. 167-177, 1986.

[18] 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.

[19] G. Kramer, "An introduction to auditory display," in Auditory Display: Sonification, Audification and Auditory Interfaces, SFI Studies in the Sciences of Complexity, Proceedings, 1994, pp. 1-77.

[20] B. N. Walker, A. Nance and J. Lindsay, "Spearcons: Speech-based earcons improve navigation performance in auditory menus," in Proceedings of the International Conference on Auditory Display, London, UK, 2006, pp. 63-68.

[21] W. W. Gaver, "Auditory interfaces," Handbook of Human-Computer Interaction, vol. 1, pp. 1003-1041, 1997.

[22] E. Brazil, M. Fernström and J. Bowers, "Exploring concurrent auditory icon recognition," in Proc. of ICAD, 2009, .

[23] P. Polotti and D. Rocchesso, Sound to Sense, Sense to Sound: A State of the Art in Sound and Music Computing. Logos Verlag Berlin GmbH, 2008.

[24] W. W. Gaver, R. B. Smith and T. O'Shea, "Effective sounds in complex systems: The ARKola simulation," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1991, pp. 85-90.

[25] M. G. Helander, T. K. Landauer and P. V. Prabhu, Handbook of Human-Computer Interaction. Elsevier, 1997.

[26] S. Conversy, "Ad-hoc synthesis of auditory icons," Proceedings of of ICAD, 1998.

[27] K. Tuuri, M. Mustonen and A. Pirhonen, "Same sound–different meanings: A novel scheme for modes of listening," Proceedings of Audio Mostly, pp. 13-18, 2007.

[28] A. Darvishi, E. Munteanu, V. Guggiana, H. Schauer, M. Motavalli and M. Rauterberg, "Designing environmental sounds based on the results of interaction between objects in the real world," in Human-Computer Interaction–INTERACT, 1995, pp. 38-42.

[29] M. A. Garcia-Ruiz and J. R. Gutierrez-Pulido, "An overview of auditory display to assist comprehension of molecular information," Interact Comput, vol. 18, pp. 853-868, 2006.

[30] M. N. A. Mazlan and L. Burd, "Does an avatar motivate?" in Frontiers in Education Conference (FIE), 2011, 2011, pp. T4J-1-T4J-6.

[31] H. Choi, "Social learning through the avatar in the virtual world: The effect of experience type and personality type on achievement motivation," in Society for Information Technology & Teacher Education International Conference, 2010, pp. 1866-1873.

[32] Y. Jang, W. Kim and S. Ryu, "An exploratory study on avatar-self similarity, mastery experience and self-efficacy in games," in Advanced Communication Technology (ICACT), 2010 the 12th International Conference On, 2010, pp. 1681-1684.

[33] T. Miller, "Formative computer-based assessment in higher education: the effectiveness of feedback in supporting student learning," Assessment & Evaluation in Higher Education, vol. 34, pp. 181-192, 2009.

[34] G. Castellano, G. Caridakis, A. Camurri, K. Karpouzis, G. Volpe and S. Kollias, "Body gesture and facial expression analysis for automatic affect recognition," Blueprint for Affective Computing: A Sourcebook, pp. 245-255, 2010.

[35] B. De Gelder, "Towards the neurobiology of emotional body language," Nature Reviews Neuroscience, vol. 7, pp. 242-249, 2006.

[36] N. Gazepidis and D. Rigas, "Evaluation of Facial Expressions and Body Gestures in Interactive Systems," NAUN International Journal of Computers, vol. 1, pp. 92-97, 2008.

[37] S. A. Brewster, P. C. Wright and A. D. Edwards, "The design and evaluation of an auditory-enhanced scrollbar," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1994, pp. 173-179.

[38] B. Duggan and M. Deegan, "Considerations in the usage of text to speech (tts) in the creation of natural sounding voice enabled web systems," in Proceedings of the 1st International Symposium on Information and Communication Technologies, 2003, pp. 433-438.

[39] I. Nix and A. Wyllie, "Exploring design features to enhance computer-based assessment: Learners' views on using a confidence-indicator tool and computer-based feedback," British Journal of Educational Technology, vol. 42, pp. 101-112, 2011.

[40] B. Price and M. Petre, "Teaching programming through paperless assignments: An empirical evaluation of instructor feedback," in ACM SIGCSE Bulletin, 1997, pp. 94-99.

[41] D. Rigas and N. Gazepidis. A Further Investigation of Facial Expressions and Body Gestures as Metaphors in E-Commerce. Proceedings of the 7th WSEAS International Conference on Applied Informatics and Communications (Editors: Minh Hung Le, Metin Demiralp, Valeri Mladenov and Zoran Bojkovic). WSEAS Press Vouliagmeni, Athens, Greece, p.p. 150-155, August 2007, ISSN: 1790-5117, ISBN: 978-960-8457-96-6

[42] M.M. Alsuraihi and D. Rigas. Efficiency on Speech Recognition for Using Interface Design Environments by Novel Designers. Proceedings of the 7th WSEAS International Conference on Applied Informatics and Communications (Editors: Minh Hung Le, Metin Demiralp, Valeri Mladenov and Zoran Bojkovic). WSEAS Press Vouliagmeni, Athens, Greece, p.p. 156-161, August 2007, ISSN: 1790-5117, ISBN: 978-960-8457-96-6

[43] D. Rigas and A. Ciuﬀ reda. An Empirical Investigation of Multimodal Interfaces for Browsing Internet Search Results. Proceedings of the 7th WSEAS International Conference on Applied Informatics and Communications (Editors: Minh Hung Le, Metin Demiralp, Valeri Mladenov and Zoran Bojkovic). WSEAS Press Vouliagmeni, Athens, Greece, p.p. 196-201, August 2007, ISSN: 1790-5117, ISBN: 978-960-8457-96-6

[44] M. Sallam and D. Rigas (2010). Multimodal Metaphors in E-Learning Note Taking. International Journal of Education and Information Technologies, 1(4), p.p. 24-32, ISSN: 2074-1316 2010.

[45] D. Rigas and M. Sallam (2010). A User Satisfaction Approach to Multimodal Interfaces for Note-Taking. International Journal of

Education and Information Technologies, 1(4), p.p. 41-48, ISSN: 2074-1316 2010.
[46] M. Sallam and D. Rigas (2010). Comparing Effectiveness and Efficiency between Multimodal and Textual Not-Taking Interfaces. International Journal of Computers, 2(4) p.p. 70-77, ISSN: 1998-4308, 2010.