Multimodal metaphors in e-learning note-taking

Mohamed Sallam, Dimitrios Rigas

Abstract—This paper introduces an empirical study conducted to investigate the use and effect of multimodal metaphors in e-learning applications. This investigation involved two different interface platforms performed by forty users. The first interface platform (textual interface), based on three input modalities, namely text, graphic, and speech, was used to deliver information about note-taking. The second platform interface (multimodal interface) was based on five input modalities, including; text, graphic, speech, avatar and earcons to deliver the same information. The aim of the experiment was to measure and compare the level of usability of textual and multimodal interfaces. The usability parameters, which are efficiency, effectiveness, and users' satisfaction, were considered in the experiment. The results obtained from this investigation have shown that the multimodal e-learning interface increased the level of usability as users took significantly less time to complete the tasks, performed successfully in a higher number of tasks, and were more satisfied than when using the textual interface. Also, the results indicate that users most preferred the avatar as choice of input modality, while earcons were the second most preferred option for representing information. These input modalities could be used to improve the attractiveness of note taking which in turn will be reflected in increasing users' motivation and interest in the presented learning material.

***Keywords*—** E-learning, Usability, User interface, Multimodal Interaction**.**

I. INTRODUCTION

In recent years there has been a marked increase in the number of e-learning websites.[1]. However, only a few of these are consistent with basic Human Computer Interaction (HCI) guidelines, in terms of multimedia metaphors (e.g. speech, avatars, and earcons). e-learning is considered one of the main facilities available to students to help them with their education [1, 2].

As more and more institutions are relying on presenting their material online, the need to present this material in a more efficient way has also increased [3, 4]. The development of information and communication technology available provides a great opportunity to extend the knowledge and skills of individuals [5, 6]. The reliance on this medium has increased almost exponentially with the popularity of the internet. In order to make good use of this resource we also require a tool to present this information in a way that can be understood and used more easily.

However, the method of transference of this information from the screen to the user has remained largely unchanged as the communication has only been visual[7]. It has been shown that overuse of textual information is not the most effective method of learning online[8]. Other human senses, if used alongside the visual sense, can greatly assist in understanding the present information [9]. Some of these include metaphors such as speech, recorded sound and graphics [10]. Two interfaces were set up; one textual interface which include text, graphic and speech and the other were multimodal interface consisting of text, graphic, speech, avatar, and earcons with the objective to measure usability and advantages of these metaphors within an e-learning tool to aid learning. The experimental study discussed in this paper was intended to investigate the use of multimodality to assist in e-learning applications in the context mentioned above.

Manuscript received February 25, 2010;

Mohamed Sallam is with the Department of Media Technology, De Montfort University The Gateway Leicester, LE1 9BH UK, BD7 1DP UK (e-mail: mohamedssallam@DMU.ac.uk).
Professor Dimitrios **Rigas** Head of Department of Media Technology. De Montfort University The Gateway Leicester, LE1 9BH UK +44 (0) 116 257 7485, email: d.rigas@DMU.ac.uk).

II. E-LEARNING

A. What is E-learning

E-learning is a collective term that describes learning with the use of internet technologies that allowing learning to take place without being constrained by time or location [11, 12]. This makes e-learning a powerful medium for learning [12]. E-learning has become a popular method of training within academic institutions and organizations [13, 14]. Rosenberg (2001) argues that "*e-learning refers to the use of the internet technologies to deliver a broad array of solutions that enhance knowledge and performance*" [8]. E-learning is rich, dynamic, and can be as effective as traditional methods of learning such as the classroom. E-learning applications include web-based learning, computer-based learning, virtual classrooms, digital collaboration audio or video tape, and CD-ROM. Advantages of E-learning include self paced learning modules, the facilitation of learning to suit different learning styles, content delivered in a more engaging fashion and empowering and encouraging students to take responsibility for their learning [8, 15]. The constant advance in the technologies which has facilitated e-learning has led to content

being made available at faster speeds and in larger volumes. This has led to e-learning becoming a significant area of research [12].

B. Why E-learning

web-based technology afford more opportunities to people who seek to learn with little income for whom the payment of traditional course fees is difficult [16]. Time is also important for those individual who cannot attend traditional classes at specific time. Thus, e-learning has offered many people the chance to educate themselves [17].

Web-based learning has offered different ways to deliver courses and students can make use of electronic resourres available on the web[18]. E-learning can be applied to either synchronous (same time but different place) or asynchronous (different time and different place) learning.

Synchronous learning

In synchronous e-learning students attend class via a computer connected to the internet and communicate with the tutor by email or in real time using live chat. This type of learning seems more suitable for those students who cannot attend the traditional class room as it simulates an atmosphere to that in the conventional classroom[19].

Asynchronous Learning

In asynchronous e-learning, the students and the tutor participate with each other, although not at the same time. Students can attend the class at any time they wish, and this approach to learning allows support and feedback to be given by the tutor at any time.[20].

E-learning is still an emerging field of research and provides many benefits that are totally different from a conventional classroom based learning environment[21]. Electronic technologies offer a wide range of benefits such as standardisation, cost-effectiveness, flexibility, and scalability. The benefits of e-learning are summarized as [22, 23]: (1) contents can be easily be kept up-to-date, because the information always comes from one central source; (2) course materials are available at any time and from any place; (3) there are links to other websites for additional explanations and help; (4) students are able to communicate with their instructors and classmates; (5) students are able to plan the learning program according to their style and pace.

III. MULTIMEDIA METAPHORS

Recent studies undertaken in human-computer interaction on the use of multimodality have shown that use of multi-modal metaphors have positive effects on the usability of interactive computer systems [24, 25]. Other studies performed in this area have concluded that multimodality applications can be used to assist users for improving learnability [26].

Rigas et al investigated the application of multimodal metaphors, including, speech along with non-speech sounds [27, 28]. The research concluded that the combination of non speech (earcons) with speech was a successful and effective method for conveying information to the user [29]. In software applications, speech and sounds, after visual output, are the most common methods for communicating a response to the user. Auditory stimuli, comprises of two groups. One of these is auditory icons [30, 31], which refers to environmental sounds, and the other group is earcons [32, 33], which refers to specialized musical sounds. Environmental sounds are heard in everyday life and are generated in response to the user's input actions on the computer. While some of these sounds are quickly able to grab the user's attention, other sounds cannot. These sounds can have the same effect in computer interfaces. Earcons sounds are produced using musical instruments and the qualities of an earcon include rhythm and pitch. The pitch and rhythm can be varied according to the type of data that is being communicated. Moreover, it has also been shown that a multimodal learning application allows for more engaging learning experiences.

A. Anticipated Benefits

According to the literature studied, use of multimodality can influence a student's learning [25]. As e-learning applications are widely used, the expected benefits of using multimedia in e-learning systems, in particular within the area of note-taking, will result in improving student performance by reducing the time required to complete the tasks with fewer errors and to enhance student understanding and satisfaction [34]. Also, it will provide additional usability guidelines for development of multimodal metaphors in e-learning applications.

IV. EXPERIMENT PLATFORM

The aim of this experiment was to investigate the effect of using multimodal metaphors in an e-learning system. Two different versions of the experimental e-learning tool were developed for this empirical study. The first platform used text, graphics and speech. The second platform used five input modalities for the interface to improve efficiency of e-learning, and these were text, graphic, speech, avatar and non-speech sounds (earcons). Each multimodal application was tested with five tasks. In the text based interface, participants were required to make notes about a specific word. For example, the participants were required to read and select a word from a passage of text and then to write some notes relating to the selected word. The same task was then replaced with a graphic for adding notes. For example, participants selected a word and then right clicked the mouse to display a menu of options from which a graphic and/or an image could be chosen and inserted. In the third task, notes were added by recorded speech. Three earcons were employed in the multimodal interface and created using musical tones. Each earcon was utilized in the fourth task to represent the number of notes. For example when the user added two notes to one word and three note to another, the earcons would make two short sounds for first word and three short sounds for the second word. In the last task, a human-like avatar was included in the multimodal interface to represent the recorded speech.

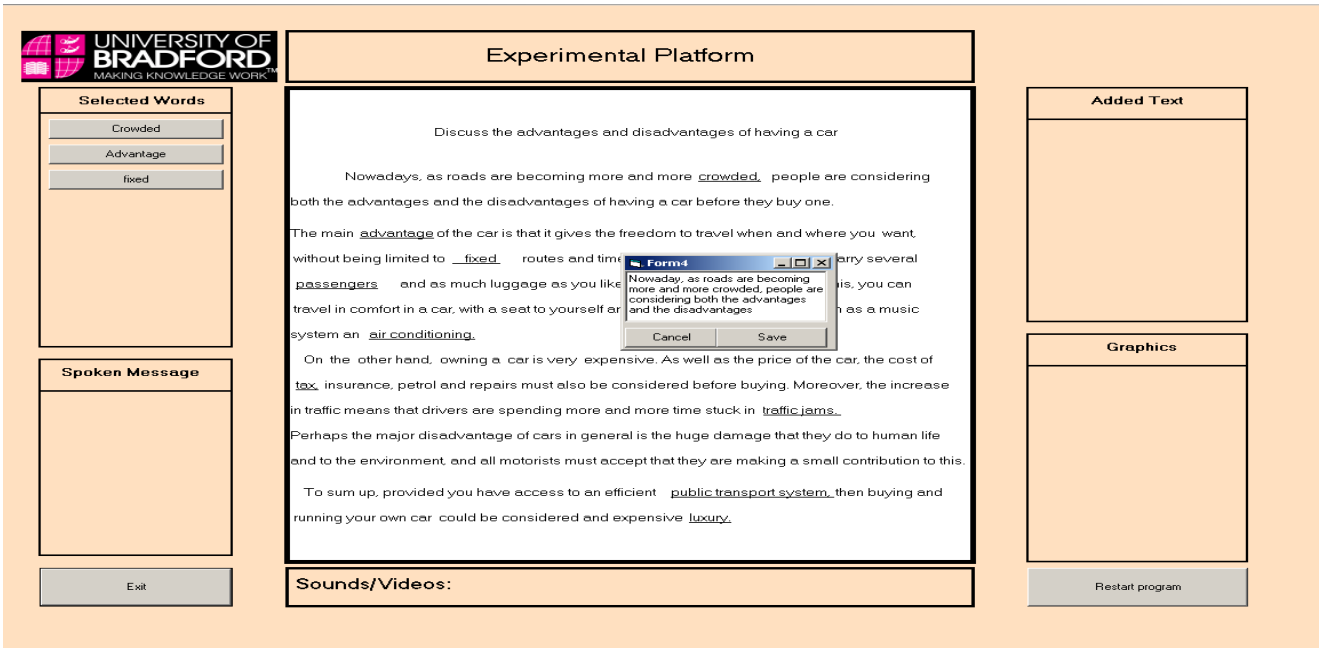

Figure 1. Shows the textual Interface

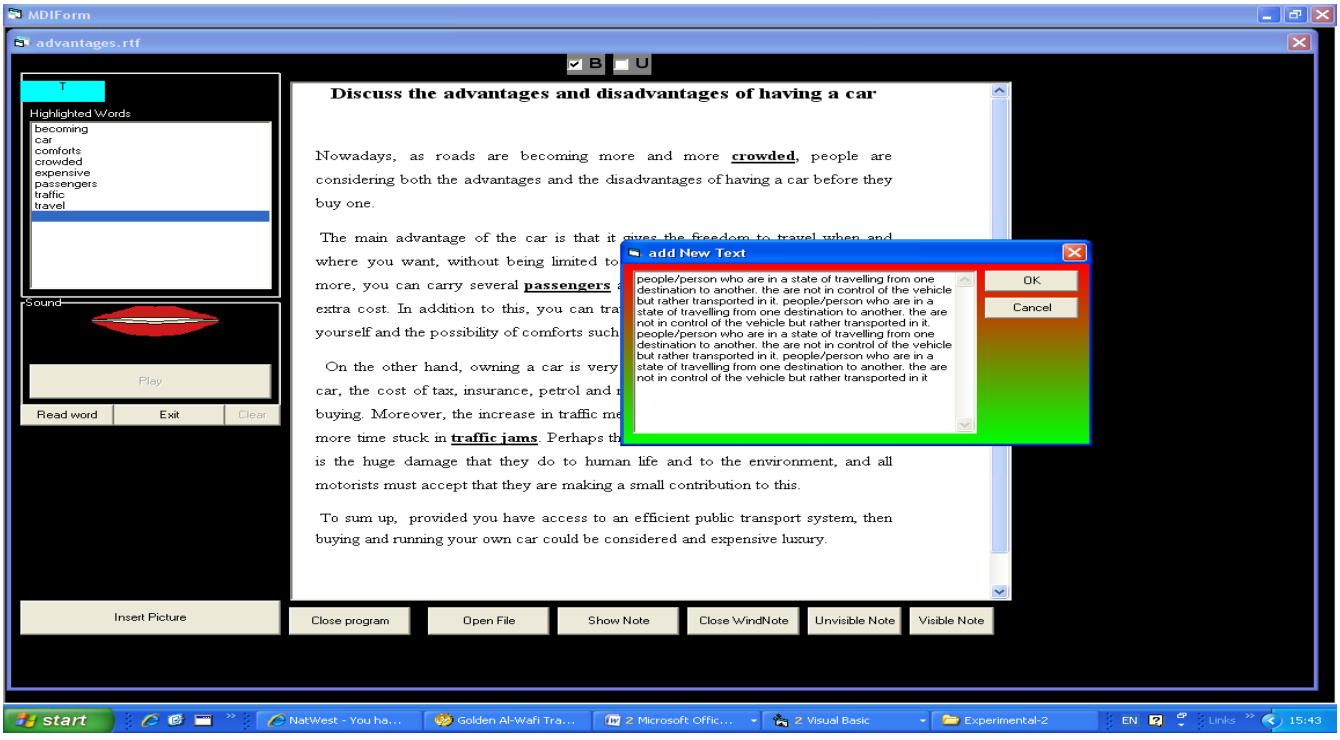

Figure 2. Shows the Multimodal Interface

A. Participants

Forty participants, consisting of under-graduates and post-graduates were selected to investigate the effect of including multimodal metaphors usability of e-learning interfaces. A post-experimental questionnaire at the end of the experiment was answered by all participants. Participants were 15% of them had a bachelor's degree, about 18% had doctor's degree and the remaining percentage had master degree 65%.

The participants were grouped into three categories on the basis of their age. Figure 3 shows that the majority aged are between 25 and 34 years old (48%) followed by those between 35 and 44 (35%), and the remaining percentage were over 35 years of age.

The participants comprised of 85% males and 15% females. The reason for a low number of female participants was due to the lack of females who satisfied the criteria of English being a second language, and also having basic competency with

computers. The participants were also required to have not previously used the experimental platform. Figure 3 shows that the number of participants who have a limited knowledge of human computer interaction was 23%, the number with good knowledge was 15%, and about 30% had no knowledge. In order for the experiment to be successful, all participants had to fulfill a certain set of criteria. The requirements were; (a) be computer literate (i.e. use computers for more than 10 hours a week); (b) had not used the experimental platform before; and (c) spoke English as a second language. Approximately 45% of users studied computing, and 15% studied telecommunications. The rest of the users studied electronics, and engineering. An analysis of the participants showed that 70% used a computer for more than 10 hours per week. The number of participants who used a computer for 1-5 and 6-10 hours per week was about the same. Only 3% of participants had never used a computer. The number of participants who used the internet for less than 10 hours per week was very small at only 15%. The number increases to 70% for those who use the internet on average more than 10 hours a week.

B. Methodology

Three criteria were chosen for measuring the level of usability of the two interfaces: effectiveness, efficiency and users satisfaction. The relationship between the communication metaphors, used in the applied interface version, and each of usability parameters was required to be evaluated and discussed. Efficiency was measured by the time taken by users to complete the required tasks. The effectiveness was measured by the number of tasks performed successfully and the number of errors made by users when using both interfaces. Satisfaction was evaluated by the users' responses to the post-experimental questionnaire, which was scored using the Likert Scale from 1- 5. Users were required to indicate their agreement to fourteen statements regarding each interface. These statements were about the ease of use, ease of learning, and usefulness of each metaphor. Therefore, the main hypothesis stated that the multimodal e-learning interface would be more efficient, effective, and satisfactory than the e-learning interface that used text and graphic input modalities.

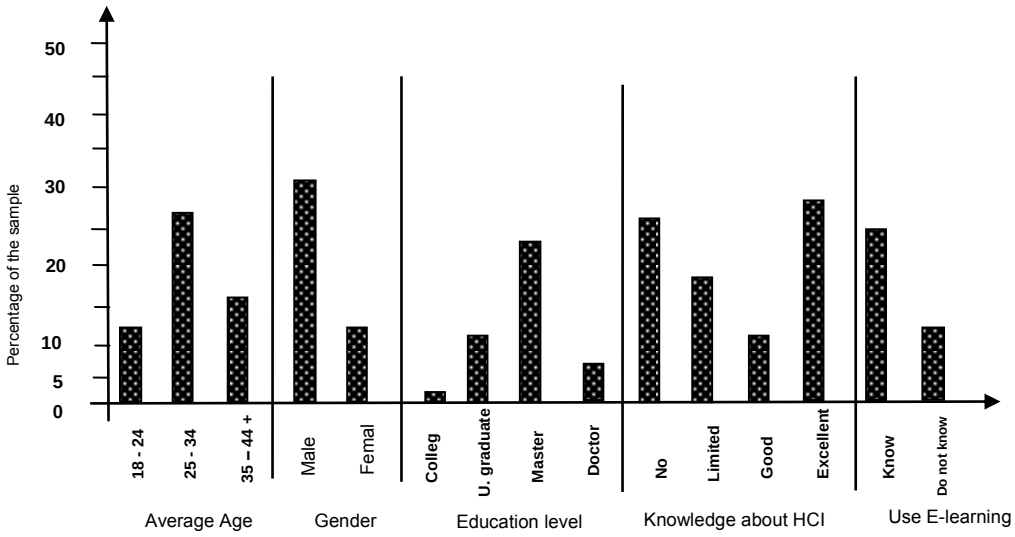

Figure 3. The average number of Age, Gender, Education, Knowledge and use e-learning

C. Tasks

The users who participated in the experiment were asked to complete five tasks in each interface. The tasks were designed to test all the 5 different modalities listed above for the multimodal interface. For the textual interface, the steps were exactly the same for each task. Participants were given a set of pre-selected words and some notes to add as comments to the words. These tasks gradually increased in terms of complexity from easy, to moderate, and to difficult. Each task comprised a set of requirements which asked the user to place the mouse cursor over a selected word. In the textual interface the user was required to make notes about a specific word. For example, the user was asked to read and select a word from a passage of text and then to write some notes relating to the selected word. The same task was then replaced with a graphic, speech, avatar, and earcons, for adding notes. Each user had to perform these steps for 3 words in each interface. The number of requirements in each task was proportional to the level of task complexity. After the completion of the requirements, the time taken to perform the tasks and the errors incurred while performing the tasks was recorded.

D. Procedure

The participants were briefed on the procedure for the experiment and given approximately ten minutes to read and understand the text they were expected to use in the experiment. Further, participants were given a quick demonstration of the procedure to ensure that all participants fully understood what was required from them. The time spent on the demonstration was not included in the timings for the actual experiment. Each user had to perform each of the specified tasks individually. The experiment time was recorded for each individual task and also for the overall experiment. The efficiency and frequency of errors were also recorded for all tasks. Efficiency was measured by timing how long a user took to complete each task. On completion of the experiment, the participants were asked to answer the satisfaction questionnaire. This questionnaire was designed to ask the participants to identify multimedia metaphors, as part of an e-learning system, by ranking them in the order of their preference.

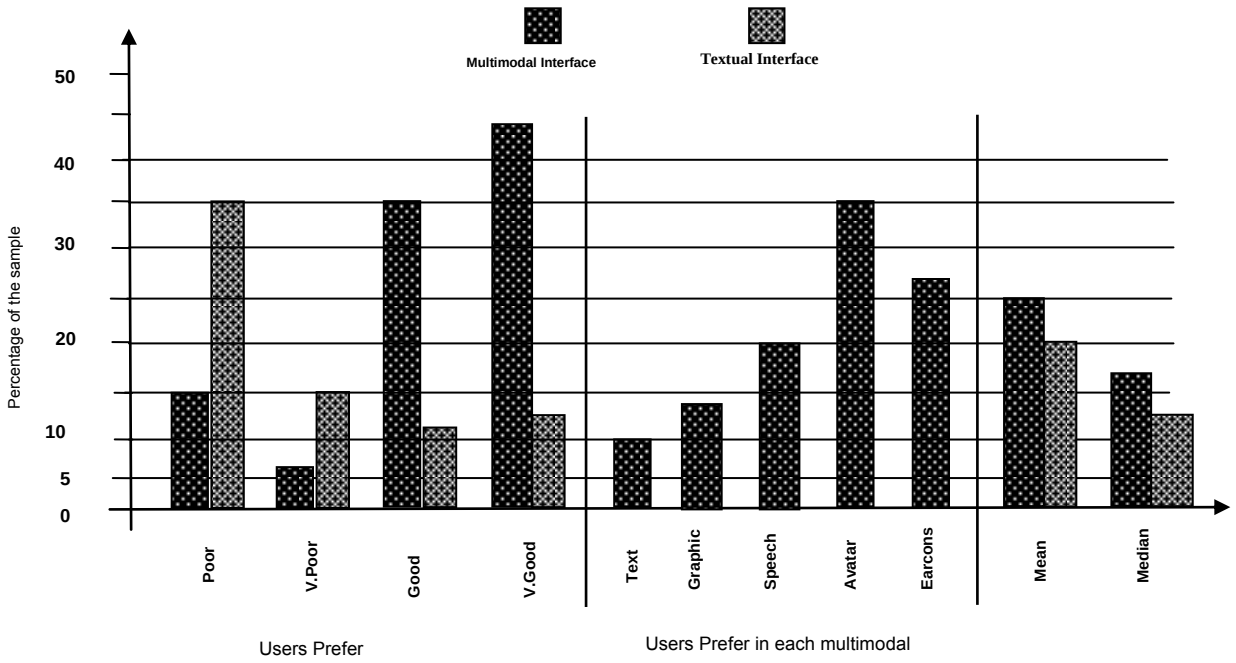

Figure 4. Shows the users Preference

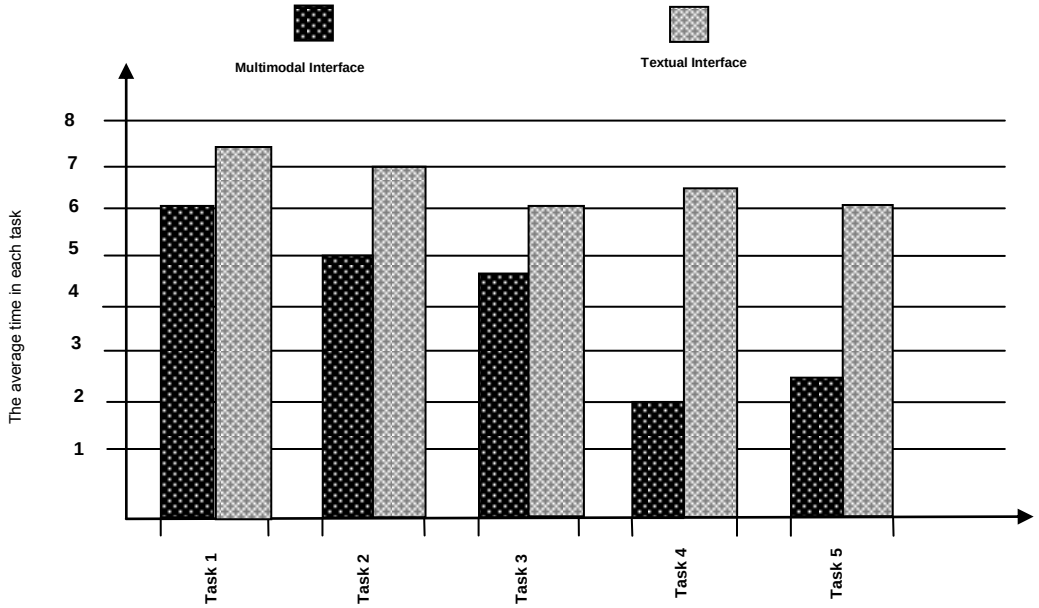

Figure 5. shows the mean time in each task

V. Results and Discussion

The experiment was conducted with 40 participants. The usability was measured by efficiency, effectiveness, and user satisfaction. The results of the experiment show that users prefer to use modalities such as avatar, earcons, text, graphic, and recording speech when they want to add comments. The users of the multimodal interface managed to complete tasks more successfully than the users of textual interface (approximately 91% or 182 out of 200 tasks). The overall mean completion time for all tasks was significantly lower in the multimodal interface in comparison with the textual interface ($t = 2.76$, $cv = 1.68$, $p<0.05$). Also, as shown by figure 4, 61% and 27% of the users in the multimodal interface preferred to use the application using the earcons, avatar, and recorded speech input modalities. The remaining 12% of the users preferred to use the application using text and graphic. However, users suggested that there should be some options in the textual interface for adding multimedia metaphors to help them add comments. On the other hand, approximately 74% of participants described the multimodal interface as very good. Only a small percentage expressed a negative view of such use. The multimodal interface was ranked positively by almost every user. Conversely, about 56% of participants described the textual interface as poor and only 9% described it as good, and very good.

There was a noticeable difference in the successful completion of tasks that involved complex tasks between the multimodal interface and textual interface. In those tasks, it appears that for most of the tasks the users performed better when using the multimodal interface compared to the textual interface. However, the results show that users performed better in the textual interface when performing the simple tasks.

A. *Task Completion time*

The efficiency was measured as the total time taken to complete the tasks and the total time the user spent looking for information. The performance of each user was observed, recorded, and noted in an evaluation form. Also, the result shows that the mean completion time for all tasks in the multimodal interface was lower than in the textual interface.

The overall mean completion time for all tasks was significantly lower in the multimodal interface compared to the textual interface ($t = 2.76$, $cv = 1.68$, $p<0.05$). The main reason for this was that the multimodal interface included more modalities such as avatar, earcons, speech, and graphic. The percentages of task completion time for the multimodal and textual interfaces were participants used text. The results show that users managed to perform slightly better when using the multimodal interface. However, there was a noticeable difference when the participants used input multimodalities such as avatar, earcons, speech and graphic.

This difference was found to be more significant between the task where participants were using graphic ($t = 2.08$, $cv =$

1.68, p<0.05) and the task using speech (t = 2.16, cv = 1.68, p<0.05). During the experiments, it was observed that the time taken to perform the tasks in the multimodal interface was lower. The number of mouse clicks performed in the two experimental interfaces was recorded. The results of the experiment show that users of the textual interface recorded more mouse clicks than users of the multimodal interface. The number of mouse clicks recorded for the text, graphic and speech tasks was 4, 5 and 8 clicks respectively, while the total number of mouse clicks performed for the completion of task in the multimodal interface was 8. High statistical differences were recorded when comparing the time spent performed between the multimodal interface and the textual interface.

The difference in completion time between the interfaces increased as the complexity of the task increased. This difference was found to be significantly in moderate (t= 2.02, cv= 1.68, p<0.05) and difficult (t= 2.89, cv= 1.68, p<0.05) tasks. These results show that multimedia helped participants to perform different tasks more successfully and show that 92% of participants thought that multimodal metaphors were important for e-learning.

B. Successful Completion of Tasks

The performance of each user was checked and the total number of completed tasks for each user was obtained and then used for statistical analysis.

The total number of tasks performed in both interfaces was 200 (40 multiplied by 5 tasks in each user). The number of tasks completed in the multimodal interface was 182 out of 200 tasks (91%) while the completed number of tasks in the textual interface was 131 out of 200 tasks (65%). This shows that users managed to complete more tasks successfully when using the multimodal interface than when using the textual interface. Figure 6 shows that the difference in the number of completed tasks between the two interfaces increased as the multimodal interface used more multimodal metaphors.

Figure 5 shows the average time spent on each task. The results show that the time spent in the first task (add text) was slightly higher in the multimodal interface. However, Figure 6 shows that in the last two tasks (add avatar and add earcons) there was a noticeable difference between the multimodal interface and the textual interface.

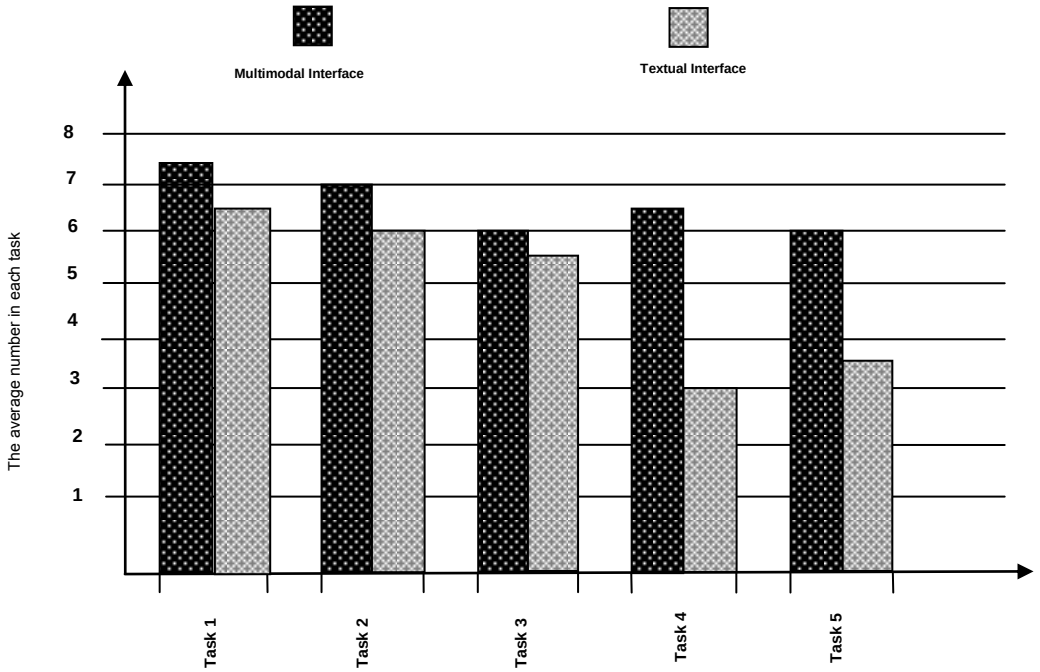

Figure 6. shows the average number of completed tasks

C. *Satisfaction*

Users were asked to select their preferred interface and provide an explanation for their choice. These ratings were used to analyse the level of user satisfaction of the two interfaces in regard to different aspects (ease of use, confusion, nervousness and overall satisfaction). This questionnaire used the five points Likert scale. Users' satisfaction of the interfaces was evaluated by obtaining users' views of each individual metaphor and interface used.

These five points were used for each statement in the questionnaire ranging from 1- the value for strongly disagree, to 5- the value of strongly agree. T-test was performed on the total number of scores to test the difference in the users' satisfaction. The results in figure 7 show that participants were significantly more satisfied when using the multimodal interface than when using textual interface ($t = 2.16$, $cv = 1.64$, $p<0.05$).

VI. CONCLUSION AND FUTURE WORK

This paper reported an empirical study that investigated the use of multimodal interaction metaphors for delivering information using two interfaces in an e-learning system. The aim of the study was to compare the usability between two different e-learning interfaces. The first interface incorporated a combination of multimodal metaphors such as recorded speech, earcons, and avatar, while the second interface was based on text, and graphic.

The experiment, consisting of 40 participants, was conducted to measure the usability of multiple input modalities in e-learning applications by making a comparative analysis of two interfaces in terms of effectiveness, efficiency, and user's satisfaction. The results show that the use of both visual and auditory metaphors, as employed in the experimental interface, significantly improved the level of usability in comparison to the textual interface.

Overall, it can be summarised that using multimodal metaphors such as recorded speech, earcons, and avatar was more efficient than when using only text, when providing information for learning.

Future research on usability will involve comparing learnability responses, in addition to other usability parameters, including effectiveness, efficiency, and satisfaction. Furthermore, adding different colours to the interface input modalities can be tested to determine the effects they have on learning. The use of speech modality has been immense. Therefore, other ways in which this metaphor can be incorporated, for example, text to speech, can be explored.

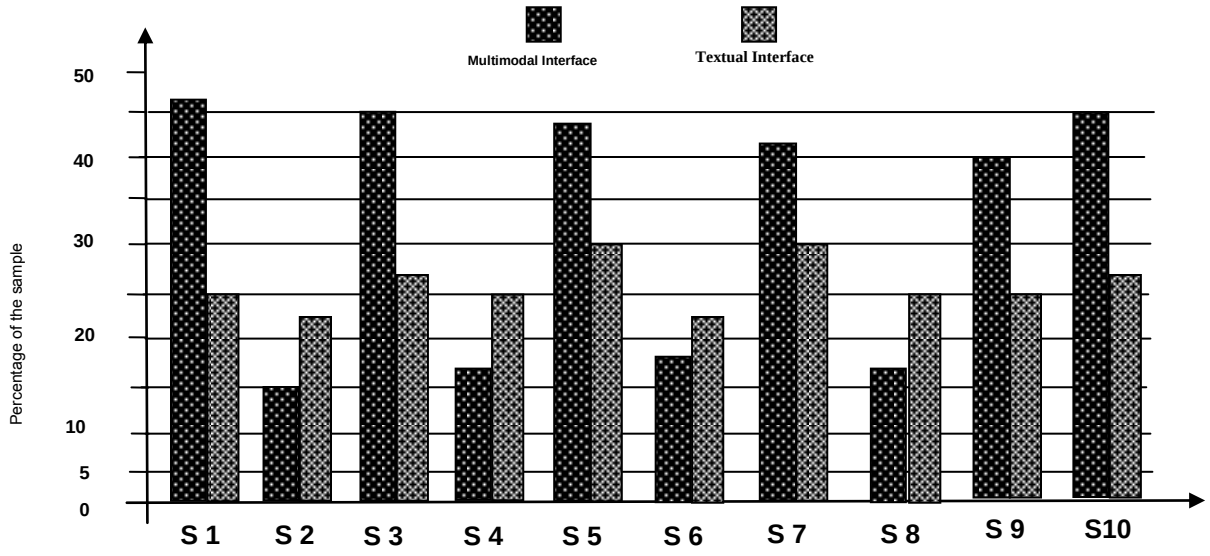

Figure 7. Show the mean satisfaction in each Interface

References

[1] a. S. Molnar, "computers in education: a brief history," the journal (technological horizons in education), vol. 24, 1997.
[2] s.alexander, "e-learning developments and experiences," 2001.
[3] s. Naidu, "designing and evaluating instruction for e-learning," designing instruction for technology-enhanced learning, 2002.
[4] b. Dahlbom and l. Mathiassen, "the future of our profession," communications of the acm, vol. 40, pp. 80-89, 1997.
[5] l. Kvasny and d. Truex, "defining away the digital divide: a content analysis of institutional influences on popular representations of technology," realigning research and practice in information systems development. Social and organizational perspective," 2001.
[6] m. Paiting, "e-learning: is it really the best thing since sliced bread," 2002.
[7] v. Jones and j. H. Jo, "ubiquitous learning environment: an adaptive teaching system using ubiquitous technology," beyond the comfort zone: proceedings of the 21st ascilite conference, pp. 468-474, 2004.
[8] j. M. Rosenberg, "e-learning: strategies for delivering knowledge in the digital age.," 2001.
[9] s. A. Brewster, "using non-speech sound to overcome information overload. Displays," 1997.
[10] d. A. D. H. Rigas, "the role of multimedia in interfaces for on-line learning.," 2003.
[11] a. W. Bates and t. Bates, technology, e-learning and distance education: routledge, 2005.
[12] p. Berman, "e-learning concepts and techniques," 2006.
[13] j. Mcgovern and k. Gray, "directions for organisation and management of university learning: implications from a qualitative survey of student e-learning," proceedings of ascilite 2005, 2005.
[14] m. Paiting, ""e-learning: is it really the best thing since sliced bread,"", 2002.
[15] r. Hamilton, c. Richards, et al., "an examination of e-learning and e-books," 2001.
[16] h. R. Pfister and m. Mühlpfordt, "supporting discourse in a synchronous learning environment: the learning protocol approach," proceedings of the conference on computer supported collaborative learning (cscl), pp. 581-589, 2002.
[17] g. Singh, j. O'donoghue, and h. Worton, "a study into the effects of elearning on higher education," journal of university teaching and learning practice, vol. 2, pp. 13-24, 2005.
[18] a. P. Rovai, "sense of community, perceived cognitive learning, and persistence in asynchronous learning networks," the internet and higher education, vol. 5, pp. 319-332, 2002.
[19] c. Zheng and g. D. Bennett, "applied contaminant transport modeling," industrial and commercial training, vol. 34, pp. 256-262, 2002.
[20] k. O'neill, g. Singh, and j. O'donoghue, "implementing elearning programmes for higher education: a review of the literature," journal of information technology education, vol. 3, pp. 313–323, 2004.
[21] f. A. L. A. Mikic, "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,," 2006.
[22] c. J. Bonk and r. A. Wisher, "applying collaborative and e-learning tools to military distance learning: a research framework," us army research institute for the behavioral and social sciences, alexandria, va., technical report, 2000.
[23] d. Murray, "e-learning for the workplace: creating canada's lifelong learners," conference board du canada, 2001.
[24] g. Shaw and n. Marlow, "the role of student learning styles, gender, attitudes and perceptions on information and communication technology assisted learning,," 1999.
[25] n. B. Sarter, "multimodal information presentation: design guidance and research challenges,," 2006.
[26] d. H. G. Theonas, d. Rigas, "the effect of facial expressions on students in virtual educational environments," 2008.
[27] d. Rigas, d. Memery, et al.,, "experiments in using structured musical sound, synthesised speech and environmental stimuli to communicate information: is there a case for integration and synergy?," proceedings of international symposium on intelligent multimedia: video and speech processing, kowloon shangri-la, hong kong, pp. 465-468, 2001.
[28] d. I. A. D. M. Rigas, "utilising audio-visual stimuli in interactive information systems: a two domain investigation on auditory metaphors," in international conference on information technology: coding and computing,, pp. 190-195, 2002.
[29] m. L. A. S. A. Vargas, "combining speech and earcons to assist menu navigation," 2003.
[30] m. Cohen, "throwing, pitching and catching sound: audio windowing models and modes," nternational journal of man-machine studies, vol. 39, p. 269, 1993.
[31] w. W. Gaver, r. B. Smith, et al., "effective sounds in complex systems: the arkola simulation," in proceedings of the sigchi conference on human factors in computing systems: reaching through technology new orleans, louisiana, usa, pp. 85-90, 1991.
[32] m. M. Blattner, d. A. Sumikawa, et al.,, "earcons and icons:their structure and common design principle," human computer
Vol. 4, pp. 11-44, 1989.
[33] d. I. A. J. L. A. Rigas, "the rising pitch metaphor: an empirical study," international journal of human-computer studies, vol. 62, pp. 1-20, 2005.
[34] r. Wagner, j. Werner, and r. Schramm, "an evaluation of student satisfaction with distance learning courses," 2002.