Comparing effectiveness and efficiency between multimodal and textual note-taking interfaces

Mohamed Sallam, Dimitrios Rigas

***Abstract*—** this paper describes an experimental study conducted to investigate the use of multimodal metaphors in the interface of e-learning applications. This investigation involved two different interface versions of the experimental e-learning tool. In the first interface platform (textual interface), three input modalities, were used to deliver information about note-taking: text, graphic, and image. The second version of the interface application (multimodal interface) offered a combination of multimodal metaphors such as recorded speech, video, and avatar with simple facial expressions to communicate 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 user' 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. 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 learning material presented.

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

I. INTRODUCTION

Every day, technologies are developing in different fields of our life and computers play an important part in our lives. In the education system, for instance, the use of the computer is becoming very significant. Software applications are extremely crowded with functions, menus, icons and toolbars. E-learning is one of the main facilities available to students in order to help them with their studies. As academic institutions are now presenting their material online, the need to present this material in a more efficient way has also increased [1, 2]. There has been a marked increase in the number of e-learning websites in recent years [3].

Moreover, the e-learning application is also crowded with visual user interface and content. This makes searching for information in the web more difficult. Some of these websites are consistent with the principles of Human Computer Interaction, in terms of multimedia such as speech, avatar, and video [4]. Rigas et al, suggest that the use of multimodal metaphors in application learning interfaces can be more useful to communicate the information that 'needs' to be communicated to the user [5, 6, 7]. Several other studies have been carried out to test the use of multimodal metaphors in visual user interface and to evaluate and examine the affect of these metaphors on the usability of computer applications [8]. Some of these studies suggest that the use of multimodal metaphors such as speech, sounds, non speech sound, and avatar could improve the usability of computer interfaces in many ways, including in e-learning application. Nevertheless, more investigations sill need to be conducted in this field [9].

Two different interfaces platforms tested by forty users were designed to investigate the use and effect of multimodal metaphors in an e-learning application. The first interface platform (textual interface), based on three input modalities, namely, text, graphic, and image was used to deliver information about note-taking. The second platform interface (multimodal interface) was also based on three input modalities: speech, video, and avatar.

The aim of this paper is to discuss the results of the experiment regarding the usability parameters.

The usability parameters considered in the study are efficiency, effectiveness, and users' satisfaction. The following sections provide more details about the related work in e-learning and multimodality, experimental platform and design, and the discussion of the obtained results.

II. E-LEARNING

E-learning can be defined as the process of delivering, supporting and administering of the learning opportunity using electronic means. E-learning has become a popular method of training within academic institutions and organizations [10, 11]. Rosen-berg (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" [12]. Also, e-learning affords advantages such as allowing students to plan their learning program according to their strengths and weakness [12]. Students are able to access the online learning facilities at any time and from any location, thus benefiting students who cannot attend the traditional classroom method of teaching. [13]. learning accommodates students where

Manuscript received March 10, 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).

course fees for the traditional teaching program can prove to be expensive for them [14]. E-learning has offered new opportunities to learn. Studying via internet technologies can complement the traditional class method of learning.

III. MULTIMODAL 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 [15, 16]. Other studies performed in this area have concluded that multimodality applications can be used to assist users for improving learnability [17, 18]. For example, users' visual channel could be overloaded [19, 20] and important information that needs to be communicated could be missed [20].

In our interaction with the word around us, we use many senses to get information from the surrounding environment. Users will feel that they interact with computer more naturally, when these senses are used in human computer interaction [17]. Several other studies have been carried out to test the use of these senses. They show that the amount of information delivered by one sense will be reduced [21].

Using more than one communication channel to receive different types of information will enhance usability of computer systems interfaces [22].

In software applications, speech and sounds, after visual output, are the most common methods for communicating a response to the user [23, 24]. It was found that earcons with short musical sounds are a more efficient and effective metaphor for interaction with users in the interface[28]. Moreover, using recorded speech and earcons in the interface of multimedia e-learning application assisted users to successfully complete more complex learning tasks [25].

Another visual and auditory interaction involved in this experiment was the avatar. This is a computer based character that could be used to play the role of human being and has the ability to express feelings, emotions, and other linguistic information via various facial expressions [26]. It could be utilized in e-learning environments to enhance the usability [27]. In this experiment, video, speech, and avatar are used for enhancing the efficiency and effectiveness of the interface, and users' satisfaction. This study investigates the comparison of the usability experience using the above mentioned modalities by using visual-only interaction metaphors: text, graphic, and image.

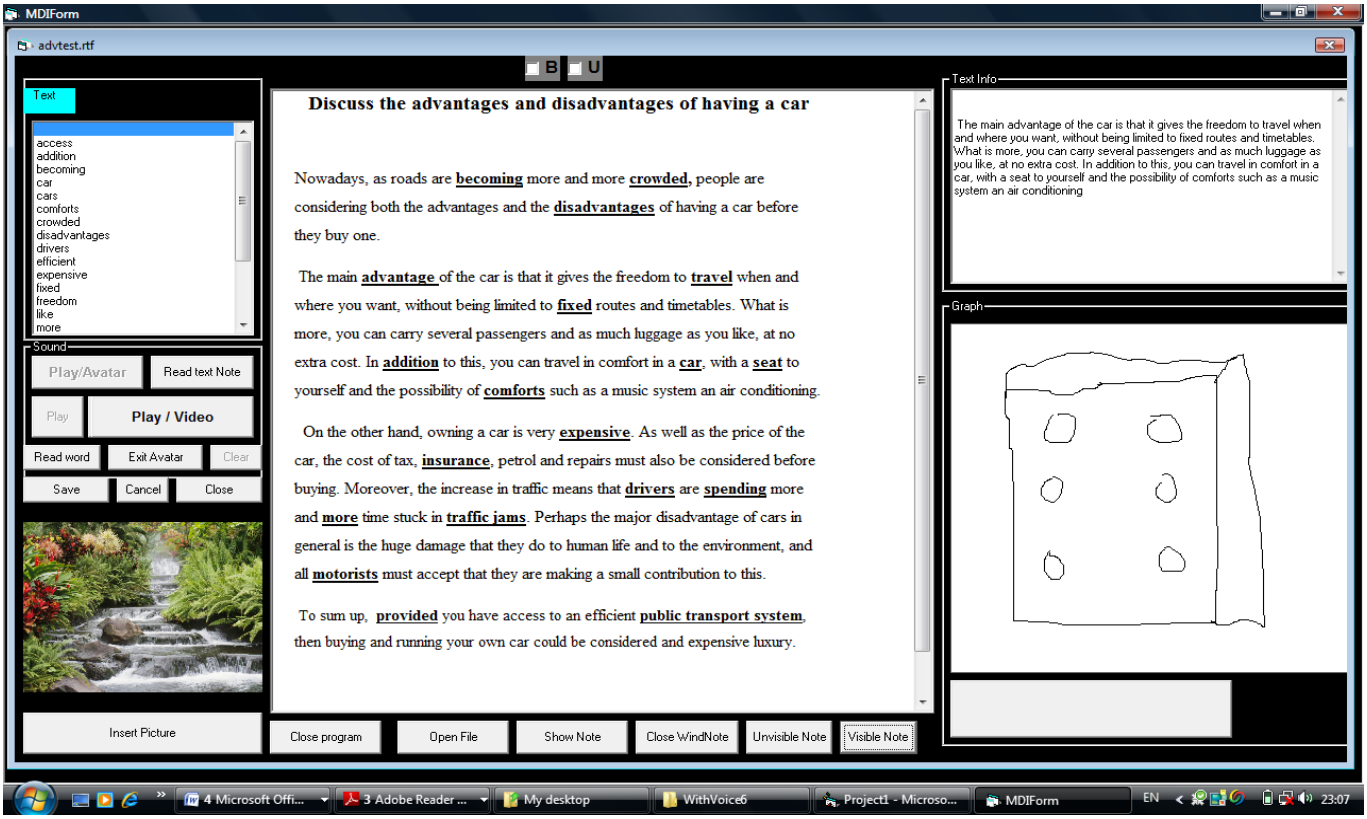

Fig. 1 The textual interface

enhance usability of an e-learning system?

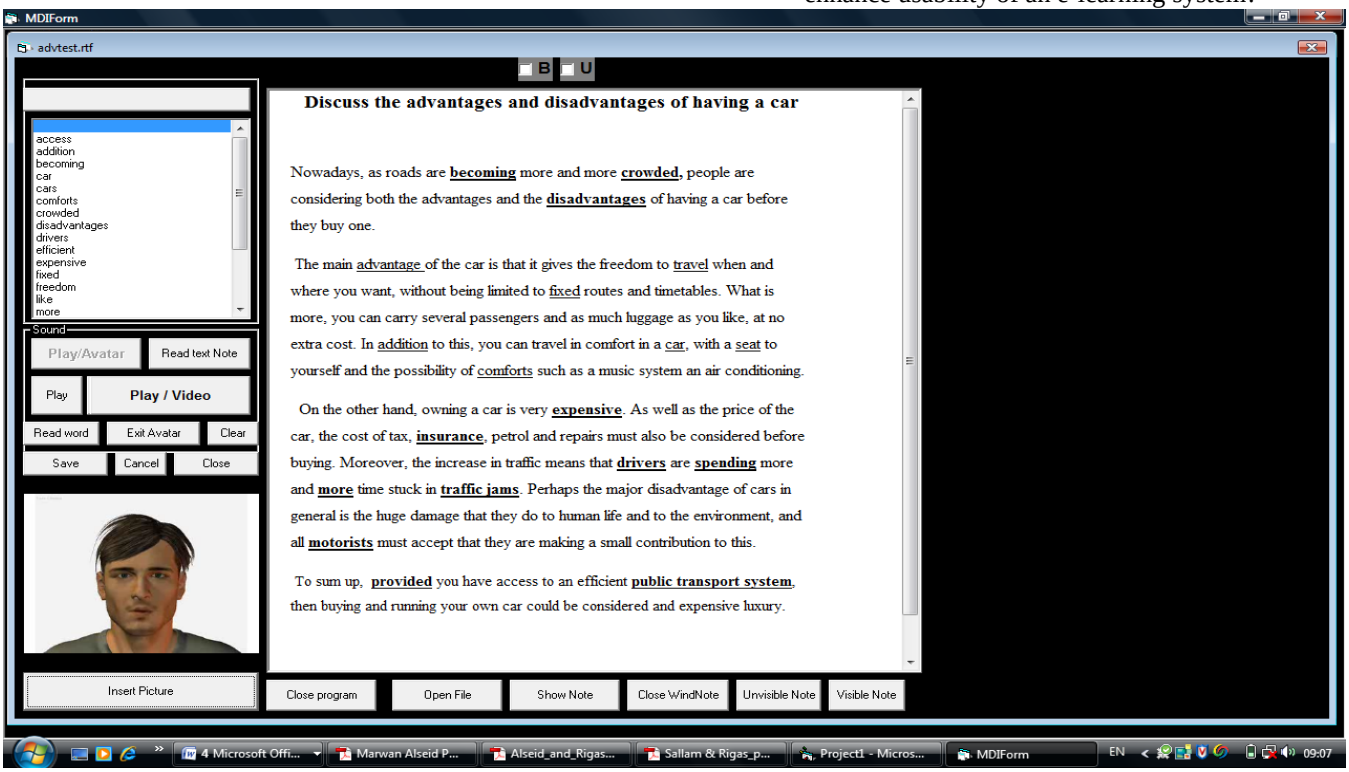

Fig. 2 The multimodal interface

IV. EXPERIMENT

Two different interfaces of the experimental e-learning tool were designed to be used for this study. The first interface platform (textual interface) based on three input modalities, namely, text, graphic, and image was used to deliver information about note-taking. The second platform interface (multimodal interface) was also based on three input modalities including; speech, video, and avatar to deliver the same information. In each interface, participants were required to make notes about specific words by selecting a word and then right clicking the mouse to display a menu of options. In the multimodal interface this included add speech, add video, and add avatar. In the textual interface this included add text, add graphic, and add image. For example, in the multimodal interface the participant was required to read and select a word from a passage of text and then make some notes, by speech, relating to the selected word. The same task was then replaced with a recorded video for adding notes. In the last task, a human-like avatar was included in the multimodal interface to represent the recorded speech.

A. *Research question*

1. Does the use of avatars with human like facial expressions and natural speech have a positive effect within an e-learning interface?
2. Would the use of recorded video for making notes enhance the usability of visual interfaces more than using graphic only?
3. Would using speech for input and output significantly

B. *Objectives*

According to the literature studied, use of multimodality can influence students learning. As e-learning applications are widely used the expected benefits of using multimedia in e-learning systems, in particular within the area of text annotation are immense. These can be achieved by building improved user interfaces, enhancing usability, and improving interaction. Measuring the efficiency of each task for both interfaces by the time users spent to complete the required tasks in required time. Measuring the effectiveness of each task for both interfaces by calculate the percentage of tasks successfully completed by users. Measuring the user satisfaction was in regarding to the different aspects (ease of use, confusion, nervousness and overall satisfaction).

C. *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 users took to complete the required tasks. The effectiveness was measured through the number of successfully performed tasks and the number of error made by users about the interfaces and the metaphors. Satisfaction was evaluated by the users' responses to the post-experimental questionnaire. This questionnaire was scored 1- 5 on the Likert Scale with fourteen statements in

each interface, which fitted all experimental conditions, and the users were required to specify their agreement to these statements. These statements were mainly about the ease of use, ease of learning and usefulness of each metaphor. Therefore, the main hypothesis stated that the multimodal e-learning interface which used recorded speech, video, and avatar would be more efficient, effective and satisfactory than the e-learning interface that used text, image, and graphic input modalities.

D. Questionnaire

The participants were asked to read the instructions in the questionnaire and type their name in the space provided. The above action was to enter their gender and area of study. The participants were also asked to go through five tasks. The experiments measured effectiveness, interactivity and user satisfaction. The questionnaire was the same for all two interfaces except that the post –question varied.

This was done because there was a difference between the multimodality which was added in each interface.

The first platform interface (textual interface) was based on three input modalities; text, graphics, and images. The second platform interface (experimental interface) was based on three input modalities, namely recorded speech, video, and avatar. On completion of the tasks, the participants were required to write down their views about the experiment.

The tasks were designed with the objective of measuring ease of information access and how clearly the descriptions matched the main form. The participant was asked to make some notes on both interfaces. Interface one by either by adding text, graphic, and image. In interface two by adding recorded speech, video, and avatar.

The questionnaire attached to the prototype contained detailed instructions and the six tasks the participants were asked to perform. The participant was asked to select a word after reading the text and then write some notes which related to the word selected. For example, in the first interface platform (multimodal interface) The first task enabled recording of speech, for example, click on the word 'crowded' and then right-click on the mouse to display a menu of options and choose add speech.

The user recorded his/her notes verbally which related to the word selected.

In the last task, a human-like avatar was included in the multimodal interface to represent the recorded speech. After performing each task the participants were asked to rate the user satisfaction, perception of information, and description. Using the scale below to indicate how much he/she agreed or disagreed with the statements by circling the number that most closely describes their view.

E. Participants

Forty participants, consisting of under-graduates and post-graduates were selected to investigate the effect of including multimodal metaphors usability in e-learning interfaces. A post-experimental questionnaire was answered by all participants. Of the participants, 15% had a bachelor's degree, 30% a PhD, and 55% a master degree.

The participants were grouped into three categories on the basis of their age; 43% were aged between 25-34 years, 38% were aged between 35-44 years, and 19% were aged over 35.

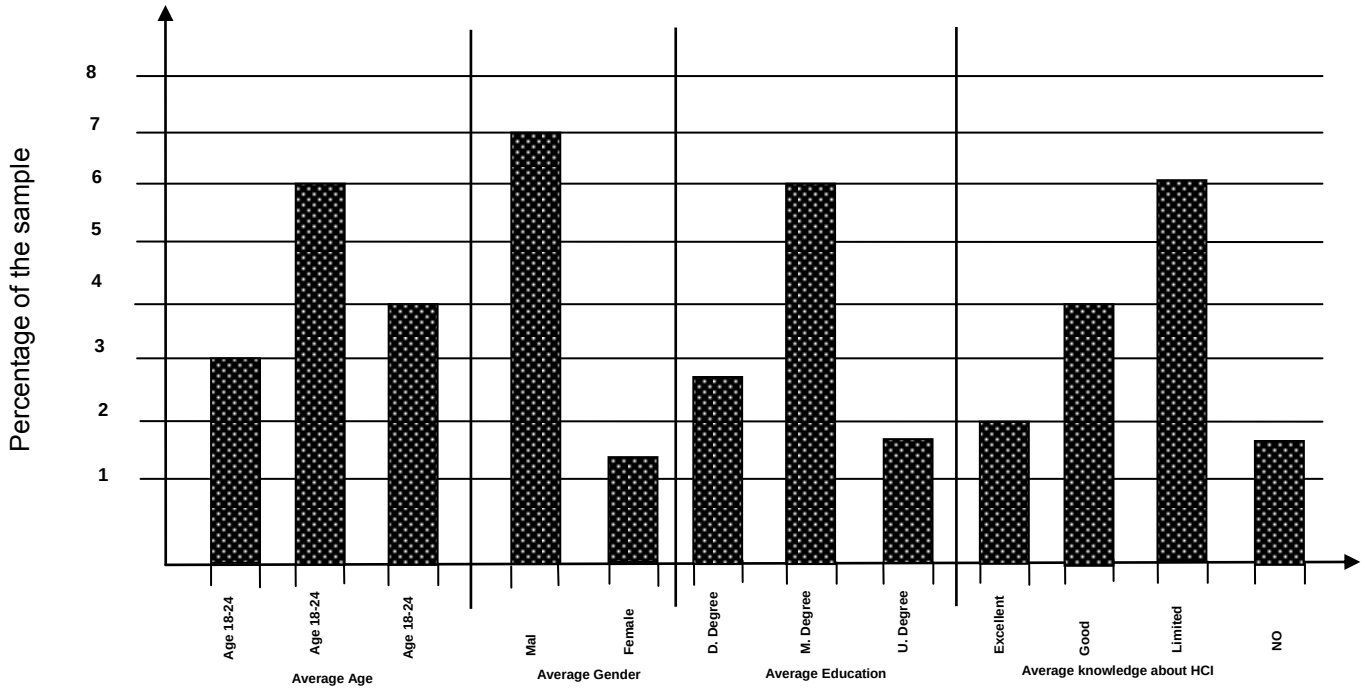

Fig. 3 The mean gender, age for both interfaces

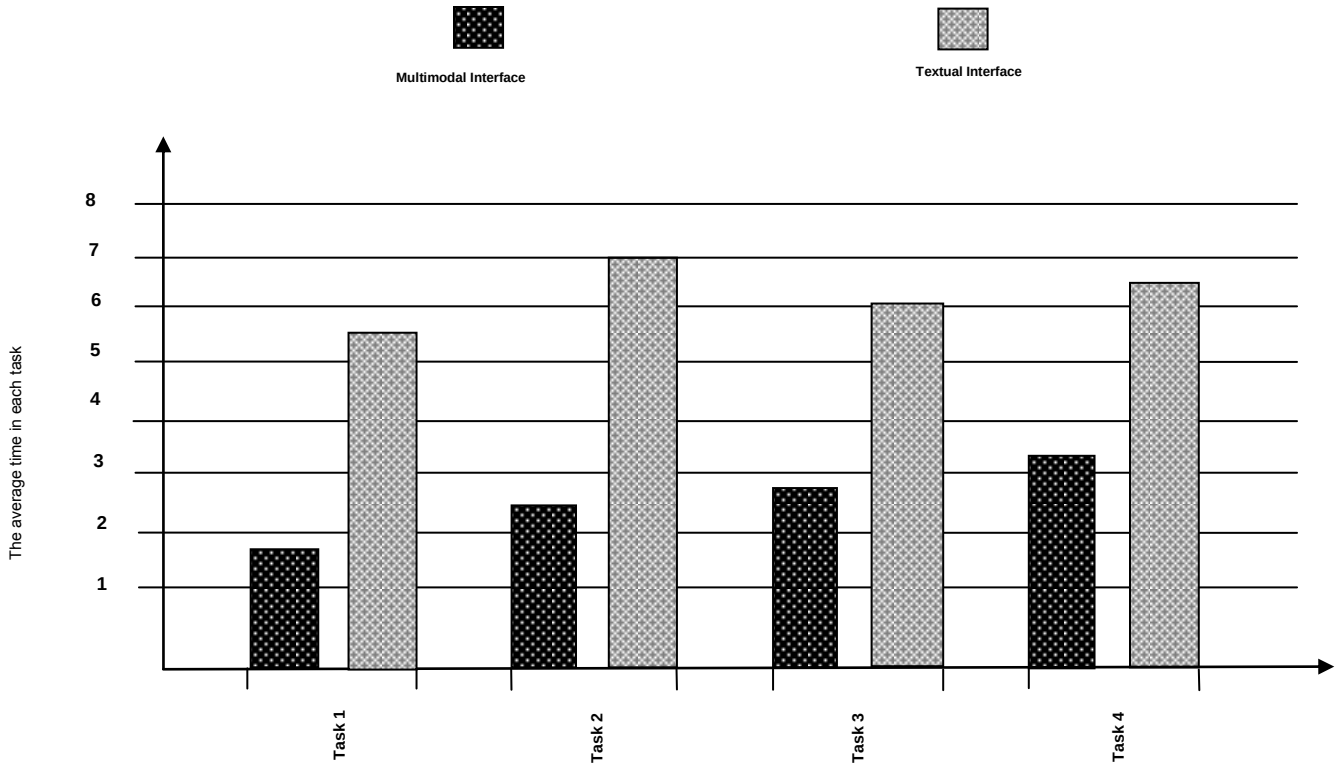

Fig. 4 The mean time of users in each task

The participants comprised of 78% males and 23% females. The reason for a low number of female participants was due to the scarcity of females satisfying the criteria of having English as a second language and some basic computer competency. The participants also had a scientific background and were using the experimental platform for the first time.

Figure 3 shows that the number of participants who had limited knowledge about human computer interaction in the experiment was 35%, the percentage with good knowledge was 35% and about 25% 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) computer literacy (i.e. used computers for more than 10 hours a week), b) had not used the experimental platform before. c) spoke English as a second language. Approximately 43% of users' area of study was computing and 13% was telecommunication. The rest of the samples were based in electronics, engineering, networking and communication.

The analysis of the respondents found that 84% of the total used a computer for more than 10 hours.

The numbers who had used a computer for between 1-5 and 6-10 hours were nearly equal however, only 2% of the total selected said that they never used a computer. The average number of participants who used the internet for less than 10 hours a week was 28%. The number increased to 73% for those who used the internet for more than 10 hours a week.

F. Program Language

The experimental platform was developed using the visual basic programming language from Microsoft Visual Basic 6.0 because it was recommended to be useful software. This work was supported by the fife Frameworks Program.

First frame box was the main interface which presented the output of textual interface text, graphic, and image; the others frames used as input of text, graphic and image.

In the multimodal interface it was four frameworks program. The first frame was the main interface that presented the output speech, video, and avatar.

G. Tasks

The participant was asked to complete four tasks in each interface. The tasks were designed with the objective of testing all the three different modalities, as mentioned listed above for the multimodal interface. For the textual interface the steps were exactly the same for each task. They were given a set of pre-selected words and some notes to add as comments for them. These tasks were gradually increasing in terms of complexity; thus, were equally divided into easy, moderate, and difficult. Each task comprised a set of requirements which asked the user to place the mouse cursor over a selected word.

In the first platform interface (textual interface) the user was required to make notes about a specific word by text, graphic, and image. In the second platform interface (multimodal

interface) required to make notes by speech, video, and avatar. Each user had to do these steps for 3 words in each interface. The number of requirements in each task was proportional with the level of task complexity.

After completion of the requirements, the time taken to perform the tasks and errors while performing tasks was recorded.

H. Procedure

The participants were briefed on the procedure and then given approximately ten minutes to read and understand the text that they were expected to use in the experiment.

Further, a quick demonstration of the procedure was given to ensure all participants fully understood what was required from them. This time spent was not included in the timings for the actual experiment. Each of the users had to individually accomplish each of the specified tasks.

The experiment time was recorded for each individual task and also for the overall experiment.

V. RESULTS AND DISCUSSION

A. Task Completion Time

The overall mean completion time for all tasks was significantly lower in the multimodal interface compared to the textual interface ($t = 2.92$, $cv = 1.68$, $p<0.05$). The main reason for this was that the multimodal interface included more modalities such as speech, video, and avatar.

The percentages of task completion time participants using speech in multimodal interface and using text in textual interface ($t = 2.21$, $cv = 1.68$, $p<0.05$).

The results show that users managed to perform slightly better when using recording speech in the multimodal interface and text in the textual interface. However, there was a noticeable difference when the users used multimodalities such as avatar and video. This difference was found to be more significant between the task where users were using recording video ($t = 2.76$, $cv = 1.68$, $p<0.05$) and the task using avatar ($t = 2.62$, $cv = 1.68$, $p<0.05$). During the experiments, it was observed that the time taken to perform all 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 image tasks was 4, 5 and 5 clicks respectively in textual interface, while the total number of mouse clicks performed for the completion of tasks in the multimodal interface was 7.

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 ($t= 1.81$, $cv= 1.68$, $p<0.05$) and difficult ($t= 2.39$, $cv= 1.68$, $p<0.05$) tasks. These results show that multimedia helped users to perform different tasks more successfully and show that 96% of users thought that multimodal metaphors were important for e-learning.

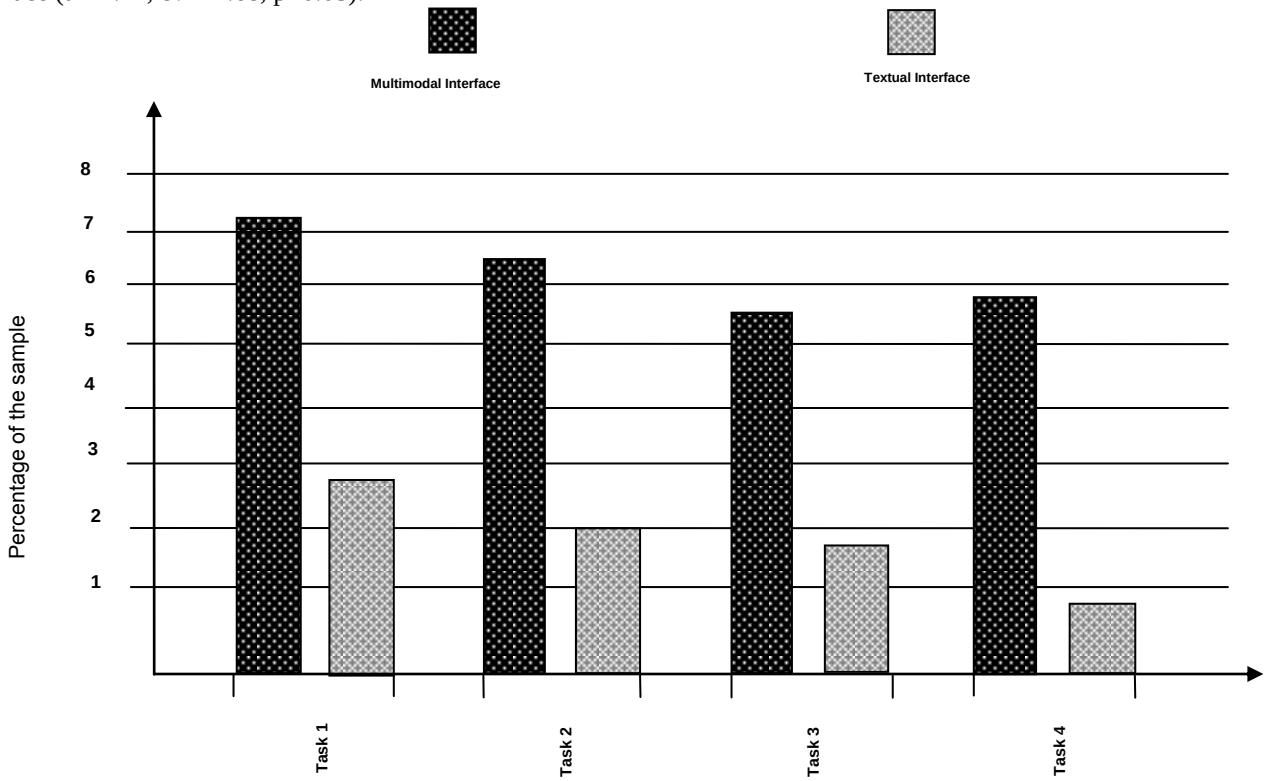

Fig. 5 The mean number of completed task

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 by all users in both interfaces was 160 (40 multiplied by 4 tasks in each interface). The number of tasks completed in the multimodal interface was 152 out of 160 tasks (95%) 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 5 shows that the difference in the number of completed tasks between the two interfaces increased as the multimodal interface used more multimodal metaphors such as speech, video, and avatar. Also the result shows that the mean value of time spent by users on each task. The results show that the time spent in the first task (add speech) was slightly higher in the multimodal interface. However, in the last two tasks (with avatar and video) there was a noticeable difference between the multimodal interface and the textual interface.

VI. CONCLUSION

An empirical study involved two different interface platforms performed by forty users. The first interface platform (textual interface) based on three input modalities, namely, text, graphic, and image was used to deliver information about note-taking. The second platform interface (multimodal interface) was based on three input modalities as well, including; speech, video, and avatar to deliver the same information. The results in this paper have shown that incorporating speech, video, and avatar can improve the efficiency of e-learning applications. The overall time taken to complete the required tasks was significantly less when these multimodal interaction metaphors were utilized to communicate information about electronic notes.

References

[1] s. Naidu, "designing and evaluating instruction for e-learning," designing instruction for technology-enhanced learning, 2002.

[2] b. Dahlbom and l. Mathiassen, "the future of our profession," communications of the acm, vol. 40, pp. 80-89, 1997.

[3] a. S. Molnar, "computers in education: a brief history," the journal (technological horizons in education), vol. 24, 1997.

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

[5] D. I. Rigas and D. Memery, "Utilising audio-visual stimuli in interactive information systems: a two domain investigation on auditory metaphors," Proceedings of the International Conference on Information Technology: Coding and Computing.

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

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

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

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

[10] 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.

[11] M. Paiting, ""E-learning: is it really the best thing since sliced bread,"", 2002.

[12] J. M. Rosenberg, "E-Learning: Strategies for Delivering Knowledge in the Digital Age," 2001

[13] Singh, G., J. O'Donoghue, and H. Worton, *A study into the effects of elearning on higher education.* Journal of University teaching and learning practice, 2005. **2**(1): p. 13-24.

[14] R. Hamilton, C. Richards, et al., "An examination of e-learning and e-books," 2001.

[15] G. shaw and N. Marlow, "the role of student learning styles, gender, attitudes and perceptions on information and communication technology assisted learning,," 1999.

[16] N. B. Sarter, "Multimodal information presentation: Design guidance and research challenges. International Journal of Industrial Ergonomics," 2006.

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

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

[19] R. Low and J. Sweller, "The Modality Principle in Multimedia Learning," The Cambridge handbook of multimedia learning, pp. 147–158, 2005.

[20] R. E. Mayer, "principles for managing essential processing multimedia learning: segmenting, pretraining, and modality principles," Cambridge handbook of mulrimedial learning, pp. 169-182, 2005.

[21] Brown, M. L., S. L. Newsome, et al. (1989). "An experiment into the use of auditory cues to reduce visual workload." Proceedings of the SIGCHI conference on Human factors in computing systems: Wings for the mind: 339-346.

[22] Sarter, N. B. (2006). "Multimodal information presentation: Design guidance and research challenges." International Journal of Industrial Ergonomics 36(5): 439-445.

[23] D. I. Rigas and D. Memery, "Utilising audio-visual stimuli in interactive information systems: a two domain investigation on auditory metaphors," Proceedings of the International Conference on Information Technology: Coding and Computing.

[24] 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

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

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

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

[28] 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.