An Initial Study into a Multimodal Note-Taking System

MOHAMED SALLAM, DIMITRIOS RIGAS
mssallam@bradford.ac.uk, d.rigas@dmu.ac.uk
Faculty of Technology, De Montfort University
UNITED KINGDOM

Abstract: This paper introduces an empirical study to investigate the effect of including multimodal metaphors such as text, graphic, avatar, earcon and speech in an e-learning application. The aim of the experiment was to measure and compare the level of usability of textual and multimodal interfaces. The usability parameters considered in the study are efficiency, effectiveness, and users satisfaction. The first platform interface (textual interface) was based on three input modalities, namely text, graphic and speech. The second platform interface (experimental interface) was based on five input modalities, including; text, graphic, speech, avatar and earcons. The results obtained from this investigation have shown that the experimental multimodal e-learning interface increased the level of usability as users took less time to complete the experimental tasks, performed successfully in a higher number of tasks, and were more satisfied than the users of the textual interface. That could be used to improve the attractiveness of note taking which in turn will be reflected in increasing users’ motivation and interest about the presented learning material.

Key-Words: - E-learning, Usability, User interface, Multimodal Interaction.

1. Introduction

In recent years there has been a marked increase in the number of e-learning websites [1, 2]. 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) [3, 4].

Currently, user interfaces for e-learning applications typically use text, graphics and the occasional or limited use of multimodal metaphors including speech, earcons or avatars, to deliver information. Many of these different types of modalities were found to be particularly useful in other applications [5, 6].

2. 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 [7, 8]. This is the reason that makes e-learning a powerful medium for learning [8]. E-learning has become a popular method of training within academic institutions and organizations [9, 10]. 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” [11]. 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 empowers and encourages students to take responsibility for their learning[11, 12]. The constant advances 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[7].

3. Multimedia Metaphors

Recent studies undertaken in human-computer interaction on the use of multimodal metaphors have shown that use of multi-modal modality have shown positive effects on the usability of interactive computer
systems [13-15]. Other studies performed in this area have concluded that multimodality applications can be used to assist users for improving learnability [16].

Rigas et al investigated the application of multimodal metaphors, including, speech along with non-speech sounds [17-22]. The research concluded that the combination of non speech (earcons) with speech was a successful and effective method for conveying information to the user [23]. 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 [24-28], which refers to environmental sounds, and the other group is earcons [29-33], 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 get 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. The qualities of an earcon include rhythm and pitch. The pitch and rhythm can be varied according 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.

4. Experiment

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 5 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 participant was 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, by selecting a word and then right clicking the mouse to display a menu of options from which a graphic could be chosen and/or insert an image of the users own choice, related to their selected word. In the third task, notes were added by speech (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 each note. 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.

4.1 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- 6 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 would be more efficient, effective and satisfactory than the e-learning interface that used text and graphic input modalities.

4.2 Participants

Forty participants, consisting of under-graduates and post-graduates, all of scientific backgrounds, were selected to investigate the effect of including multimodal metaphors on the usability of e-learning interfaces. All participants used the experimental platform for the first time. The majority of them were regarded as experts as they used a computer for a period of ten hours or more a week.
4.3 Procedure
The users were briefed on the procedure for the experiment and were 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 provided, to ensure that all the participants fully understood the requirements of the task. This time spent was not included in the timings for the actual experiment. Each user had to accomplish the specified tasks individually. The 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. Following the experiment, the participants were asked to answer the satisfaction questionnaire.

4.4 Hypothesis
The main hypotheses for this experiment is that the multimodal interface which include avatar, earcons, speech and graphic users will spend lower time to complete the required tasks and with less errors in comparison with textual interface.

5. RESULTS AND DISCUSSION
Figure 1 show the mean completion time for all tasks in the multimodal interface was lower than the textual interface. The difference was found to be significant \( t = 2.14, \text{ cv } = 1.65, \ p < 0.05 \). This is attributed to the inclusion of earcons and avatar in the multimodal interface. The main reason of this result was that the multimodal interface involved more modality such as avatar, earcons, speech and graphic. The percentages of task completion time were slightly better with multimodal interface when users used add text in both interfaces. However, there was a noticeable difference when users implemented more modality such as such as avatar, earcons, speech and graphic.

The total number of completed tasks for each user in each interface was obtained and then used for statistical analysis. The total number of tasks performed in both interfaces was 400 tasks; moreover the result shows that the users managed to finish more tasks when using the multimodal interface, which included avatar and earcons rather than textual interface.

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

These points were used for each statement in the questionnaire ranging from 1-strongly disagree, to 5-strongly agree. T-test was performed on the total number of scores to test the difference in users satisfaction. The result showed that the multimodal interface was significantly more satisfactory than the textual interface \( t = 1.87, \text{ cv } = 1.65, \ p < 0.05 \). The results also show that ease of use was greater with the multimodal interface. The results of the post experimental questionnaires show that almost all participants involved in the experiment preferred the application with multimodal metaphors in the following order; avatar, earcons, speech, graphic, text.
6. Conclusion and Future Work
This paper has shown that incorporating avatar, earcons and speech 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 utilised to communicate information about electronic notes. Future work will explore the use of, different combinations of these modalities in the interface of e-learning applications, the effect of adding different colours, and using text to speech in an e-learning application.

REFERENCES:


