# Tracing Users' Behaviors in a Multimodal Instructional Material: An Eye-Tracking Study

Esra Yecan, Evren Sumuer, Bahar Baran, and Kursat Cagiltay

Computer Education and Instructional Technology,
Middle East Technical University, 06531 Ankara Turkey
{yecan,sumuer,boztekin,kursat}@metu.edu.tr

**Abstract.** This study aims to explore user behaviors in instructional environments combining multimodal presentation of information. Cognitive load theory and dual coding theory were taken as the theoretical perspectives for the analyses. For this purpose, user behaviors were analyzed by recording participants' eye movements while they were using an instructional material with synchronized video and PowerPoint slides. 15 participants' eye fixation counts and durations for specific parts of the material were collected. Findings of the study revealed that the participants used the slide and video presentations in a complementary way.

**Keywords:** Producer, PowerPoint, video, eye tracking, cognitive load, dual coding, multiple channels.

## 1 Introduction

Instructional materials containing various media - like text, image, audio, video etc- allow learners receive information through multiple channels. The use of multiple representations has been regarded as powerful way of facilitating understanding for many years [1]. According to Mayer [2], learners learn better in well-designed multimedia environments which integrate different media types than in traditional learning environments. Multiple representations can make presentation of an application domain more complete than a single traditional source of information because they can complement each other [3]. On the other hand, multimodal approach requires learners to additional cognitive demand since they have to give attention simultaneously to different representations [4]. Therefore, these cognitive demands cause learners to experience cognitive overload, result is less learning.

In the use of multimodal representation, it is important to take cognitive processing theories into account. Two important cognitive processing theories are important in the scope of this study; namely cognitive load theory and dual coding theory [4], [5]. Cognitive load theory predicts that learner can hold few elements in the working memory because of limited amount of information that can be processed in visual and verbal channels at one time. Sweller and Chandler [6] presented two concerns for cognitive load theory; split attention and redundancy effect. Split attention effect occurs when learners are required to split their attention and combine multiple representations of information mentally [6]. Redundancy effect occurs when

J. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2007, LNCS 4552, pp. 755–762, 2007.

additional information are presented with two or more different media since organizing redundant information with essential information increases cognitive load [7]. On the other hand, dual coding theory suggests that the capacity of working memory is stretched by using both visual and verbal storage systems simultaneously [8]. When both visual and verbal elements are processed at the same time, the available amount of working memory is maximized, thereby promoting learning.

In the design of multimodal instructional materials, these issues are needed to be considered. Today, many universities and corporations provide multimodal instructional materials which may enhance learning by supporting the use of presentation slide sequences integrated with video lectures [9], [10]. Learners gain information from such materials through two sensory memory; eyes and ears. This information is processed in both visual and verbal processing areas simultaneously in the working memory.

In this study, user behaviors were explored by eye-tracking method while using an instructional material with synchronized video and PowerPoint slides. The aim is to explore learners' behavior patterns while using the instructional material. For this purpose, eye fixation counts and durations for specific parts of the material were collected. Based on cognitive load theory and dual coding theory, results are expected to be helpful in providing evidence related to the design of environments combining multimodal presentation of information.

## 2 Methodology

### 2.1 Participants

The participants were 15 first year undergraduate students from a major university in the central region of Turkey. The participants were students in the department of Computer Education and Instructional Technology. There were 6 females and 9 males, ranging from 18 to 21 years old. All students voluntarily participated to this study.

### 2.2 Material

An instructional material was developed in Microsoft Producer, a Microsoft PowerPoint add-in that makes it easy to produce engaging rich media presentation by capturing and synchronizing audio, video, slides and images [5]. The selected presentation topic was "Introduction to Instructional Technology (IT)". The content covered the definition and goals of IT, and a very brief summary of three main learning theories. The material consisted of three parts which were a video of the presenter, PowerPoint slides explaining the content in text format, and a navigation menu part presenting the links to the content (Fig. 1). PowerPoint slides were synchronized with video on computer as the lecture. Therefore, the material requires students to use visual and auditory sensory channels in parallel. Total length of the material is 8 minutes and 33 seconds consisting of 8 slides.

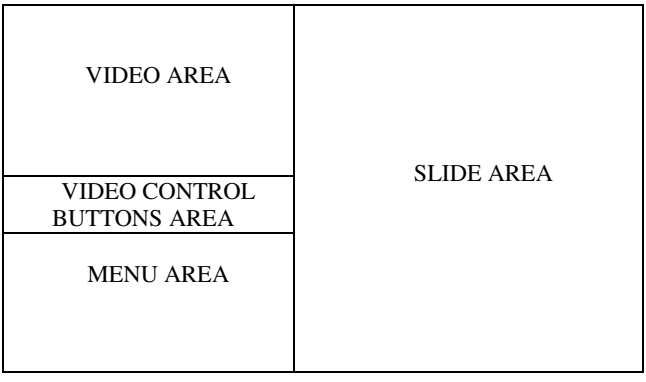


**Fig. 1.** Areas of interest (AOI)

### 2.3 Data Collection

The sessions were conducted in the Human-Computer Interaction laboratory. Data were collected through an eye-tracker device (The Tobii 1750 Eye Tracker, Tobii Technology). The device has an eye tracker that discretely integrated into a monitor without any visible "tracking devices" so this non-intrusiveness enables user to behave in a natural manner. It can collect the records of eye gaze location at 50 Hz. Data on fixation places and durations of the users were generated by the help of eye-tracking data analysis software. Eye tracking provides both qualitative and quantitative data.

### 2.4 Data Analysis

Before the analysis, areas of interests (AOI) on the screen were determined. The video area, video control buttons area, PowerPoint slides area, and menu area were determined as the main areas of the material, and each were defined as an area of interest (Fig. 1). This enables to analyze participants' fixation counts and durations on each AOI.

Descriptive and inferential statistics were applied to analyze the data of fixation counts and durations on the AOIs. In order to understand participants' behavior patterns in detail while using the instructional material, slide-based analyses were conducted. In addition, qualitative analyses of hotspot and gaze replay data were used.

## 3 Results

### 3.1 Total Eye-Fixation Durations and Fixation Counts for the Material

All participants' total fixation durations and fixation counts on defined areas of interests were examined, namely slide area, video area, video control buttons area and

menu area. The total fixation duration of AOI frames revealed that the participants focused especially on presentation slide and video screen rather than video control buttons area and menu area. There was a small difference between total fixation durations of the video and slide screens for all participants (2,511,644ms for video screen, 2,424,797ms for slide screen, 318,497ms for video toolbar and 227,146ms for menu). This result may show us that the participants tried to follow these two screens together.

When the fixation duration means were taken into consideration, the highest fixation duration mean was for video screen. Furthermore, the highest fixation count comes out in slide screen (Table 1). Contrary to total fixation duration, means of fixation duration of the AOI and the counts of fixations according to the AOI has different results for video and slide screens. The mean of the fixation duration of the video screen ($\overline{X}$=618.4) is significantly higher than the slide screen ($\overline{X}$=238.8). However, the slide screen had more fixation count than video screen. These results show that the participants stared at the video while they were watching the video although they more fixated while they read the text on the slide screen.

**Table 1.** Eye-fixation durations and counts for fifteen participants based on AOIs

| AOIs | Fixation count | Fixation Duration Means | Std. Deviation | Minimum fixation duration | Maximum fixation duration |
|---|---|---|---|---|---|
| Out of topic | 168 | 247.1 | 178.0 | 100 | 937 |
| video | 4061 | 618.4 | 871.4 | 100 | 15730 |
| video toolbar | 846 | 376.4 | 355.0 | 100 | 3708 |
| menu | 884 | 256.9 | 189.1 | 100 | 2532 |
| slide | 10153 | 238.8 | 184.0 | 100 | 6200 |
| Total | 16112 | 342.8 | 498.1 | 100 | 15730 |

* Durations are given as ‘seconds’.

### 3.2 Eye-Fixation Durations Based on Slide Presentation Styles

To analyze participants’ eye fixations on different slide presentation types, hotspot data were collected from a sentence-based slide, a bulleted slide, and a table-based slide. Analysis results of hotspot data indicated that participants’ eye-fixations were similar for these three types of slides, except less fixation counts on the slide part in sentence-based presentation (Fig. 2, 3, and 4). Therefore, we may claim that sentences are not likely to be read by participants as compared to bulleted and table-based presentation styles.

Data on fixation counts for each slide were used for further analysis (Table 2). The quantitative data indicated that there is a similarity in regards to fixation counts for bulleted and table-based presentation styles including slides 2,3,5,6, and 7 (Fig. 5). On the other hand, there is a similarity between slides 1, 4, and 8. The difference between these two groups of slides is the extra information given on video for slides

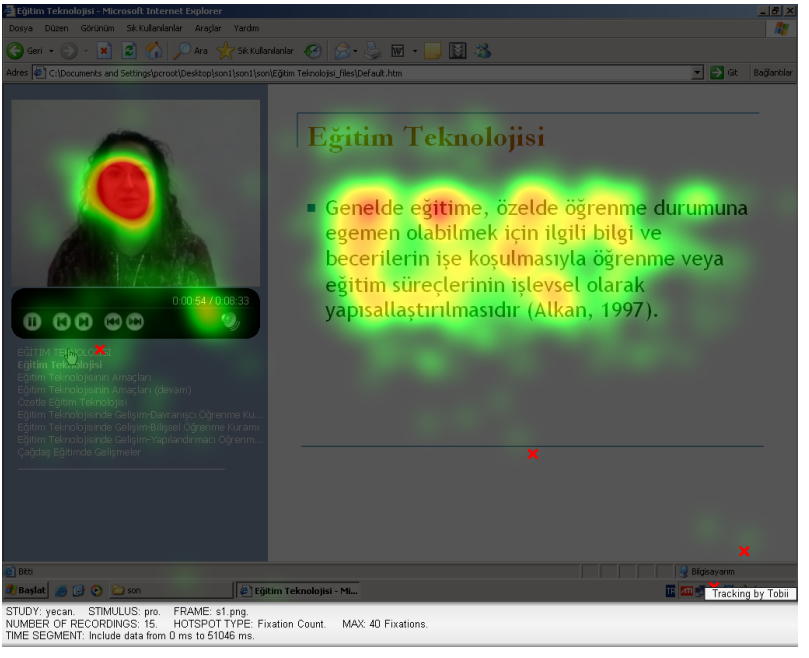


**Fig. 2.** Hotspot data for a sentence-based slide presentation (Slide 1)

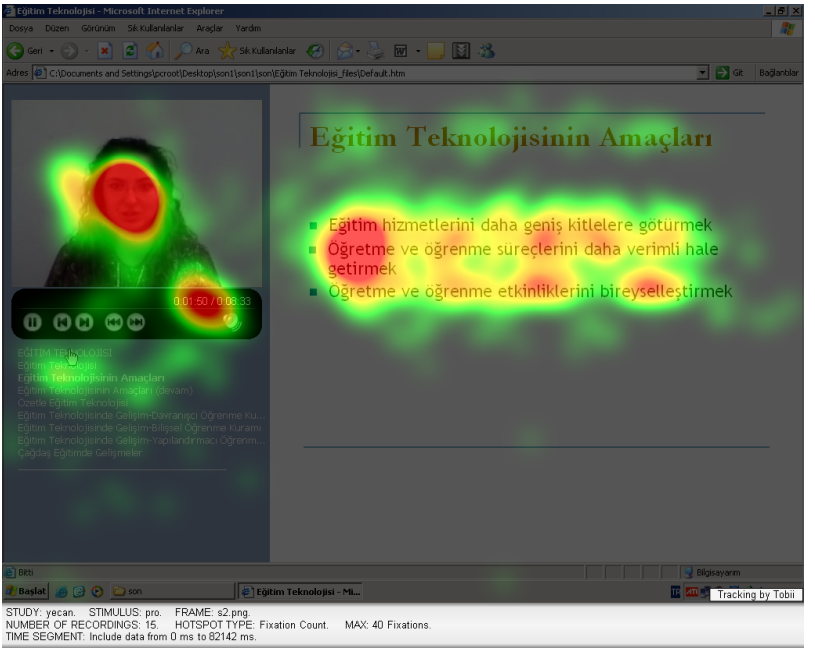


**Fig. 3.** Hotspot data for a bulleted slide presentation (Slide 2)

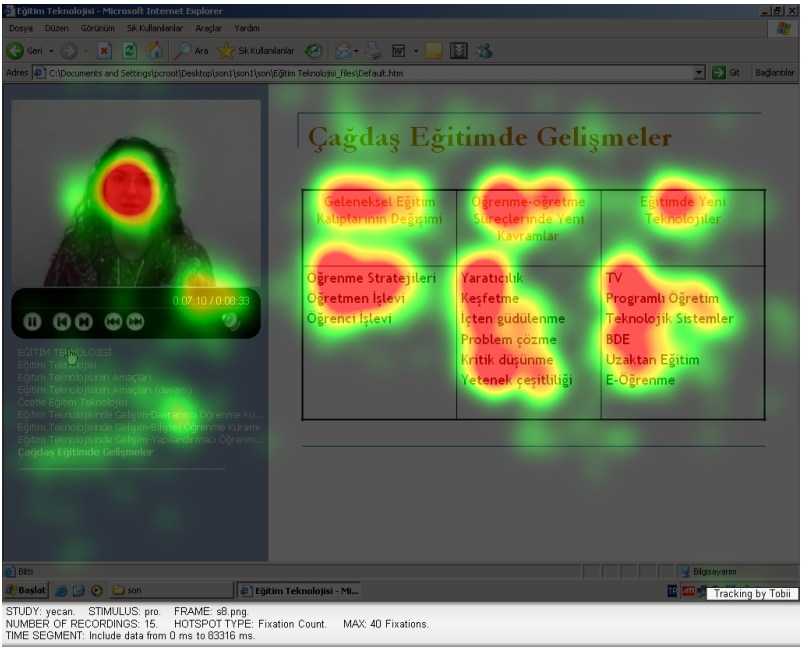


**Fig. 4.** Hotspot data for a table-based slide presentation with large amount of text (Slide 8)

**Table 2.** Fixation counts based on slides and AOIs.

| | Slide 1 | Slide 2 | Slide 3 | Slide 4 | Slide 5 | Slide 6 | Slide 7 | Slide 8 |
|---|---|---|---|---|---|---|---|---|
| Video | 370 | 730 | 714 | 76 | 841 | 491 | 334 | 314 |
| Video toolbar | 61 | 137 | 147 | 9 | 148 | 71 | 59 | 74 |
| Menu | 37 | 97 | 132 | 34 | 191 | 100 | 92 | 117 |
| Slides | 1184 | 1387 | 1373 | 419 | 1722 | 902 | 656 | 2356 |

* Slide1: Sentence based. Slides 2&3: Bulleted. Slide 4: Sentence-based. Slides 5,6 & 7: Table-based. Slide 8: Table-based with large amount of text.

2,3,5,6, and 7. While the content of the slide presentation and video are almost the same for slides 1, 4, and 8, for the rest there are some additional information and examples given by the instructor in the video, which are not presented in text format.

For further analysis, users' gaze replay records were analyzed by qualitative methods. It was observed that most of the participants first looked at the slide presentation after each slide transition. Then, they looked at the video after taking a look at the text-based material. Almost all of the participants showed continuous transitions between the slide presentation and video parts. Moreover, it was observed that participants continued to look over the PowerPoint presentation part even while the information given on the video was not available on slides. So, it could be suggested that the information given through video is searched by users in slides.

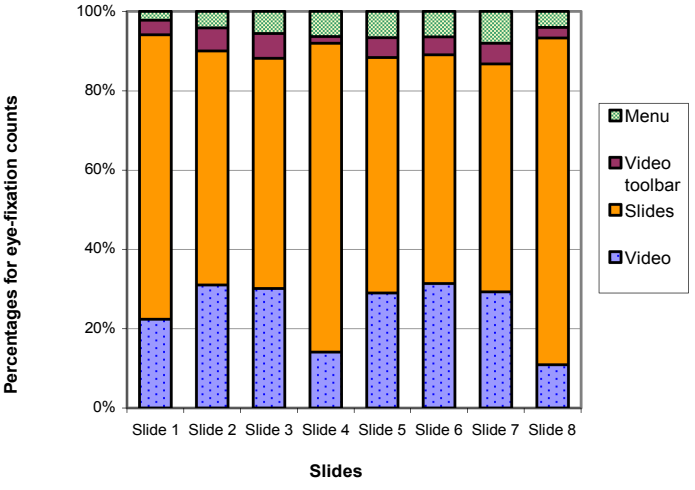


**Fig. 5.** Percentages of eye-fixation counts for AOIs of each slide

## 4 Conclusion

Multiple channel presentation of information integrating different media types may facilitate learning [1], [2]. The important point is to provide well-designed multimodal instructional environment. In this study, the researchers aimed to examine an instructional material with synchronized video and PowerPoint slides.

Findings of the study revealed that users use the slide and video presentations in a complementary way. Overall analysis of the data indicated that eye-fixation duration means are higher for video screen compared to slide presentation screen. On the other hand, fixation counts are higher for slide screen compared to video. This result indicates that participants stared at the video, while they were focusing on different places in text-based presentation.

Further analyses were conducted to explore the user behaviors based on each slide. Findings showed that participants firstly preferred to read the text at the beginning of each slide. Eye-fixation counts for video become higher if there is some additional information on video. Otherwise, the text-based material gains importance. It could be suggested that designers should put visual/verbal information as much as possible.

Moreover, the qualitative analysis of data revealed that the explanations on the video are searched by users in written material. So, there would not be missing information either in slide presentation or video. The text-based material might include all of the information or just some clues related to the information presented on the video. Dual coding theory also suggests that the available amount of working memory is maximized, when both visual and verbal elements are processed at the same time [8]. In order to propose principles for effective design for synchronized PowerPoint and video materials, different video and slide combinations should be examined in further studies.

**Note:** This study was supported by TUBITAK under grant SOBAG 104K098 and METU Human Computer Interaction research group (http://hci.metu.edu.tr).

## References

1. Ainsworth, S., Van Labeke, N.: Using a Multi-Representational Design Framework to Develop and Evaluate a Dynamic Simulation Environment. In: Paper presented at the International Workshop on Dynamic Visualizations and Learning, Tubingen, Germany (2002)
2. Mayer, R.: The promise of multimedia learning: using the same instructional design methods across different media. Learning and Instruction 13, 125–139 (2003)
3. Bodemer, D., Ploetzner, R.: Encouraging the Active Integration of Information during Learning with Multiple and Interactive Representations. In: Paper presented at the International Workshop on Dynamic Visualizations and Learning, Tubingen, Germany (2002)
4. Sankey, M., Smith, A.: Multimodal Design Considerations for Developing Hybrid Course Materials: An Issue of Literacy. In: Paper presented at the Third Pan-Commonwealth Forum on Open Learning, Dunedin, New Zealand (2004)
5. Mark, G., Hans, V.D.M., Ton, D.J., Jules, P.: Multimodal versus unimodal instruction in a complex learning context. The Journal of Experimental Education 70(3), 215–239 (2002)
6. Sweller, J., Chandler, P.: Why Some Metarial is Difficult to Learn. Cognition and Instruction 12(3), 185–233 (1994)
7. Sweller, J.: The redundancy principle. In: Mayer, R. (ed.) Cambridge Handbook of Multimedia Learning, pp. 159–168. Cambridge University Press, New York (2005)
8. Clark, R.:Six Principles of Effective e-Learning: What Works and Why. The eLearning Developers Journal, 1–8 (2002)
9. Hartle, M., Bär, H., Trompler, C., Rößling, G.: Perspectives for Lecture Videos. In: Cunha, J.C., Medeiros, P.D. (eds.) Euro-Par 2005. LNCS, vol. 3648, pp. 901–908. Springer, Heidelberg (2005)
10. Rui, Y., Gupta, A., Grudin, J.: Videography for Telepresentations. In: Proceedings of the SIGCHI conference on Human factors in computing systems, Florida, USA (2003)
11. Microsoft Producer for Microsoft Office PowerPoint. Retrieved, from (December 2007) http://www.microsoft.com/windows/windowsmedia/technologies/producer.mspx