Observation support system for recording, reviewing and sharing observed design problems

Hideyoshi Yanagisawa *, Yasuhide Kozuka, Masayuki Matsunaga and Tamotsu Murakami

** The University of Tokyo, Department of Engineering Synthesis*
Japan, hide@design.t.u-tokyo.ac.jp

Abstract: In this paper we propose a system that supports the observer to keep a record, review and share findings from user observation. Some advantages of such technical support for user observation are that it can reduce the cognitive load caused by the recording process so that the observer can focus on finding problems, and it facilitates the reviewing and sharing of the findings by employing easily understood representation of the relevant phenomena. The system automatically links the observer's written memos on digital notepad of findings during observation with a short scene video related to the phenomenon using the recorded time. The observer's gaze movement is recorded and shown on the scene video to help him/her to remember and share what he/she found. We discuss the feasibility of the system based on the results of an observation experiment using the proposed system.

Key words: User observation, scene video, gaze movement, review and share findings.

1. Introduction

In the design of matured products, it is just as important to identify the design problems as it is to find their solutions. Users often have difficulty describing design problems and, in many cases, are not even consciously aware of them. Asking the user what those design problems are may be insufficient for acquiring such design problems. Recently, user observation has been recognized as an effective approach to finding such potential problems. Observing people using a product allows us to understand how they interact with a product and the environment. This can provide clues to finding unexpected design problems.

There are several ways of classifying observations, according to the purpose and conditions. One is done in the real field of the user and the other is conducted in the laboratory, where a simulated environment is set up. In laboratory observations, the observer often employs multiple video cameras set up at different angles. The observer observes the user's activity through the video monitors. Another categorization of observation is based on the degree of freedom allowed in the user's activity. In one, the user's natural activity is observed and in the other, the user carrying out specific tasks is observed.

During observation, the observers record their findings using a combination of media, such as memos, video camera and still camera. The obtained materials are replayed to view the findings and to share them with others. Those conventional tools, however, have some potential problems. The obtained materials are not linked together, so that it is often a laborious process to collate and summarize them after observation. For example videos are helpful for recording findings that are difficult to describe in a memo, such as the user's posture and motion. It is hard to search a certain scene from recorded video which is related to a written memo because the memo and video are not linked together. Furthermore, if the observer has to pay attention to maintaining the quality of the recording, this will interfere with his/her concentration on finding relevant phenomena.

To address these problems, we developed a recording device for observation. The system automatically inserts a time index whenever the observer starts to write a memo about some finding, so that it is possible later to replay short scene videos that start a few seconds before the recorded time indexes. The related memo is displayed along with the scene video. This allows the observer to quickly replay a scene video together with the written memo and carefully describe what they found. In the system, any form of writing is allowed such as scribbles and symbols. To help the observer to remember his/her findings for which the written memo is illegible, the system displays the observer's eye-movements on the scene video.

The findings of a single observer are relatively limited. It is possible for multiple observers with different backgrounds to observe and share their findings and to find problems from different viewpoints. The automatically recorded time indexes can be used as an indicator for sharing related findings. For example, we can identify scenes where one observer finds nothing but another finds something. Multiple observers can exchange such scenes to find problems from different points of view. We believe that videos with gaze plots are helpful for understanding the findings of other observers and for recalling the original viewpoint of the observer when during replay it is unclear.

We conducted an experiment to confirm the feasibility of the proposed system under the restricted condition where the observers replay and share findings without any explanatory descriptions. We also discuss conditions where gaze-plots are effective.

2. Issues in user observation

2.1 User observation in human centered design

Although there are several design methodologies that take the user into consideration, most of them are similar to the human centered design (HCD) methodology. HCD consists of a cycle of four processes: understanding and the specifying the context of use, specifying the user and organizational requirements, producing a design solution and evaluating the solution towards the requirements[1].

User observation is widely used in such a design cycle. For example, IDEO, America's leading design firm, points out the importance of user observation in the early stage of design. In their design methodology, the designers observe real people in a real-life situation to understand unexpected phenomena and find out latent design needs not addressed by current products and services before brainstorming and designing prototypes[2]. Alternatively, user observation is used as a method to evaluate the usability of current products or prototypes of design solutions under certain conditions.

The style of observation varies depending on the task, target product and condition. We roughly divide them based on two classification schemes. One is based on whether the observer participates in the context of use or not. In field-work observations such as in ethnographic approaches, observers go into the real field of the user. Multiple methods have been proposed depending on the purpose and conditions of the observation, such as shadowing and in-situ acting self-recording[3]. The alternative approach is for the observer not to participate in the situation, allowing the user to act naturally without an awareness of the observer. In one example of this approach, multiple cameras are hung from the ceiling in a room to observe how users interact with a product, such as a copying machine, vending machine, etc.. We will call the above two approaches the "participating approach" and the "nonparticipating approach" respectively.

Another classification is based on the settings of the observed user's, i.e., whether he/she is acting in a naturalistic setting or being observed under an experiment. Naturalistic observations of the user's activities tend to be carried out in the early stages to understand the context and find the rough assumptions of the design problem. Meanwhile, to evaluate a design prototype from user observation, it is often necessary to set up experimental conditions and ask multiple participants to undertake specific tasks in order to confirm the set experimental assumptions.

2.2 Potential of technical support in user observation

Although several observational methods exist as mentioned above, the actions of observers have some things in common. During observation, the observer searches for phenomena related to his/her purpose, such as identifying unexpected design problems or evaluating some design issue in question. The observer keeps a record of the found phenomena to later remember his/her findings. After observation, the observer reviews his/her findings and shares them with other observers to discuss different points of view.

During observation the observer seeks relevant phenomena while keeping a record at the same time. Due to the limitation of human cognitive abilities such as attention and memory, there is a tradeoff relation between observing and record keeping in terms of their quality. In an ideal situation, the observer pays close attention to phenomena in search for problems while controlling the cognitive load caused by record keeping. At the same time, the record must contain information relevant to the finding so that it can be reviewed and shared with others. The record must be such that its reviewing can take place quickly and be easily understood.

Different media are used to keep records such as a notepad, still camera and video camera. Each medium has its advantages and disadvantages. For example, video is useful for recording user movements that are difficult to describe by writing. The observer usually uses a combination of media in order to complement their respective disadvantages. The problem is that such media are not connected with each other. For example, a written memo description is not related to the video scene, so that it takes time to find the relevant scene in recorded video.

The observer may pass over a phenomenon that was noticed by another observer. Such a phenomenon may contain multiple problems when observed from different viewpoints, such as efficiency, safety, health, emotions and physical load. Thus, it is beneficial to share found phenomena and discuss them from different viewpoints. To encourage diverse viewpoints, the observed phenomena should be reviewed in an easily-understood, accurate format without bias. We believe that video is an appropriate medium for representation of phenomena. However, video is tricky to use because it takes time to edit if the recording duration is long.

From the above discussion, we can summarize the advantageous possibilities of technical support in observational activities as follows.

- Minimize the cognitive load for recording findings to allow the observer to concentrate on finding phenomena
- Facilitate the reviewing and sharing of the findings
- Record and present the found phenomena in a manner that is readily understandable to oneself as well as other observers

4. Observation support system

4.1 System overview

We developed a system that supports observers in recording, replaying and sharing their findings obtained through observation. The system uses a digital notepad, digital video camera, eye-tracker and PC as hardware.

During user observation, the observer records the target users and their environment by using the video camera. Any type of camera is acceptable, such as fixed cameras in a room and handy-type cameras for ethnographic field work. For field work, we propose a head-mounted camera attached to glasses, as shown in figure 1. It allows the observer to concentrate on observation. The observer writes memos on a digital notepad, which is a pressure-sensitive device that captures drawings on paper. We can use an ordinal paper and pen and the device can be used as an underlay. The coordinate data of the drawing are transmitted to a computer.

We assume that the observer notices some phenomenon a few seconds before writing the memo. Based on this assumption, the developed software system automatically records the clock time when the observer starts to write or draw something on paper on the digital notepad. To replay the findings after observation, the system shows short videos that start a few seconds before the recorded clock times. The system allows us to view the words and figures drawn at the clock time, as shown in figure 2. Any kind of drawing is allowed. For example marks and symbols are acceptable if the observer has difficulty describing the finding with words.

The observer's gaze movements are recorded using an eye-tracking system to support later replaying of what he/she had found. The system shows the gaze movements on the recorded video. We assume that the gaze plot can help the observer and others to see how his/her finding was obtained during the scene and to recall it, especially in cases when the written memo is not sufficiently descriptive or legible.

There may be situations in which multiple observers observe the same target but from different points of view. Findings can differ from person to person. A problem found by one observer may go unnoticed by another observer.

Sharing findings among observers is beneficial because not only do one's findings complement the others' but also common problems can be seen from different viewpoints. The system displays a time chart that indicates when problems are found and by whom, as shown in figure 3. Dashed lines in the chart represent the clock times when an observer starts to draw something on the notepad.

Using this chart, we can readily recognize when someone missed a problem found by another. It can be used for setting up strategies for sharing findings among the observers; for instance, it can be used to find scenes that were noticed by few observers except for most observers. Different observers may have different viewpoints. Thus, multiple observers may notice problems in the same scene, but those problems may not be the same ones.

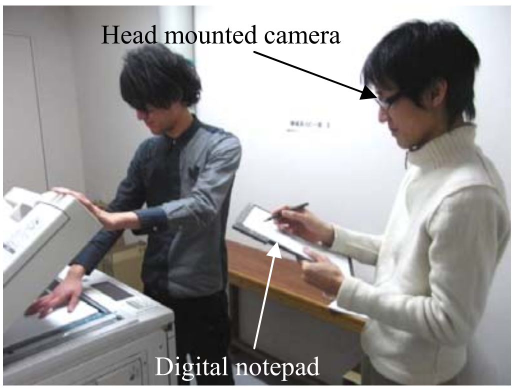

Figure 1 Hardware system used in field work observation

Figure 2 Replaying scene video (left) and related memo (right). Video is recorded using four cameras set from different angles.

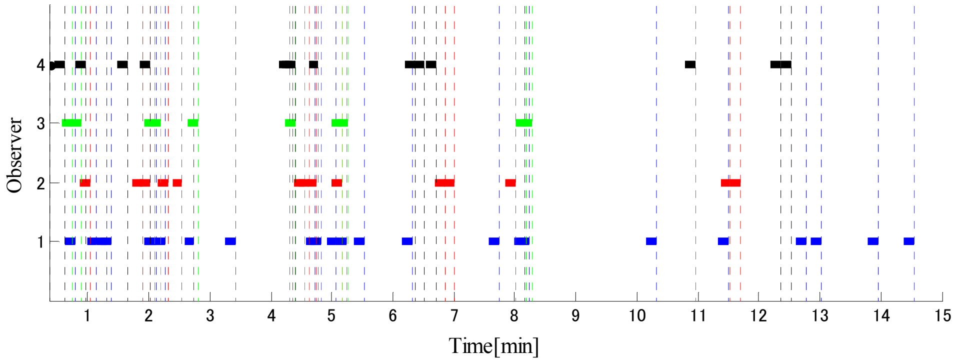

Figure 3 Example of time chart of findings for four observers

4.2 Gaze representation

The human eye has an effectively useful visual field extending about 30 degrees, while the area of high acuity covers only 4 to 5 degrees of view, which is called the foveal zone. The area of particularly high acuity within the foveal zone covers approximately 2 degrees of view. Beyond the foveal zone, the visual acuity drops off sharply. The foveal zone vision is required to perceive details of objects[4]. The observer must move his/her gaze point over a certain area of a potentially problematic phenomenon and perceive its details before he/she can recognize the existence of a problem. We believe that movements of the gaze point allow us to see what the observer perceived in the phenomenon that gave him/her a clue to finding the problem. Gaze movements consist of fixations and saccades. The fixations are the moments when the eye is relatively still and focused on a certain target. The saccades are the rapid eyeball movements between the fixations. A fixation lasts for 200-300 ms on average, while saccades usually last less than 100ms[5]. Our brain receives information from the eye only during the fixations.

We use a vision-based eye tracking system to catch eye-movements of the observer. The vision-based eye-tracker remotely tracks the gaze position using a combination of infrared light and camera images. The gaze positions are estimated from the relative positions of the pupil centre and the corneal reflection produced by infrared light shone to the eye [6]. We superimpose the fixation points and saccade lines on a scene video, as shown in figure 4. Fixation points are represented as circles. Saccades are represented as lines connecting the fixation circles. We define 300ms as the fixation time. If the fixation is longer than 1 sec, the circle is displayed in red to denote a target that was carefully watched by the observer. To show the process of gaze movements, the system retains the previous fixation points and saccade lines for certain duration. When the Euclidian distance of the gaze point from the last fixation point exceeds a given distance, the system deletes that fixation point and associated gaze lines.

Figure 4 Examples of gaze representation of observer

5. Experiment

5.1 Experiment 1: Replaying found problems without explanation

The proposed system allows any form of notation for keeping records of findings, such as words, sketches and marks. A detailed description of a finding allows one to recall the event accurately. Yet, concentrating on documentation may distract the observer's attention from carefully observing the constantly changing phenomenon. Furthermore, there are observations that are difficult to describe by written means such as the user's posture or action. The observer can concentrate on finding problems only if the proposed system enables him/her to remember what he/she found without resorting to descriptive documentation.

One of the simplest form of documentation is a mark indicating the time index of a finding. We carried out an experiment to verify how much the system enables observers to recall what they found, using only short videos with gaze plots that start a few seconds before the marks are drawn. We examined the effectiveness of gaze plots for recalling the findings as compared to when only videos are shown.

For this experiment, we employed a home kitchen environment. We set up a kitchen environment in a room as shown in figure 5. We asked the user to cook a meal (a hamburger). We took videos using four cameras hung from the ceiling at different angles while the user cooked. The signals of the four cameras were recorded as a split screen. The total length of the video is approximately 30 min.

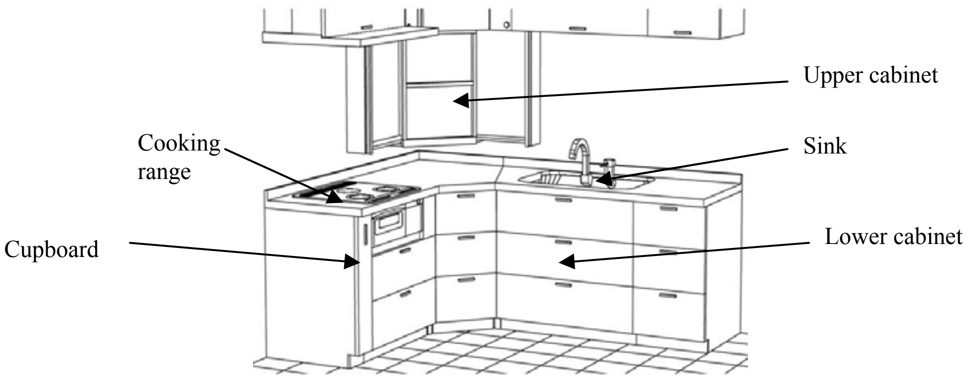

Figure 5 Layout of kitchen environment used in observation experiment

This experiment consists of two phases. In the first phase, four participants as observers independently observe a user's activity recorded on video to find design problems of the kitchen environment. Before the experiment, we give them four considerations for finding problems: efficiency, safety, health and physical load. We ask them to find design problems related to these aspects as much as possible. The observer presses a mouse button instead of checking a mark when he/she finds some clue to a problem while viewing the video. During the observation, we record the gaze movements using the eye tracker system (Tobii X120).

One day after the observation experiment, we show each participant short video segments that start 10 seconds before he/she pressed the mouse button. In the following discussion we will call such a video segment a "scene video." We ask the participants to describe what they found in the scene video. We ask three questions: the phenomenon, the problem and a possible solution. The phenomenon is the clue that leads to finding a problem, such as the user's posture or action. The participants are allowed to replay the scene videos. If the participants do not recall what they had originally found in the scene, we show their gaze movements on the video and then ask the same questions.

5.2 Experiment 2: Mutually sharing of found phenomena

Although the four participants observed the same video in section 5.1, their findings differ from person to person. We assume that it is beneficial to share those observed phenomena related to problems which were passed over by some observers because different observers may have different viewpoints. To share such findings, they must be represented in a manner that is understandable to those who passed over it. We conducted an experiment to evaluate the appropriateness of our system in terms of understandability of representation.

Using the time chart obtained from the result of experiment 1 using the proposed system, we select for each participant findings that he/she missed but were found by the other participants. We firstly show only the scene videos and ask them if he/she notices any phenomenon related to a problem. If the answer does not match the original ones given by the other observers or he/she does not respond, we show the gaze plots on the short video. Participants are allowed to repeat the video. We record the gaze movements of the participants while they view the scene videos to analyze the cause of effectiveness of gaze plot representation.

6. Result and discussion

6.1 Replaying found problems

In experiment 1, we asked the four participants to recall their found phenomenon, problem and solution using scene videos with and without gaze plots but without any descriptive memos. Our assumption in this experiment is that showing the scene video with or without gaze plots is an appropriate way to recognize findings after observation.

Figure 6 shows the number of scene videos that represent the findings for each participant. Each bar consists of the numbers of videos for three cases . In the first case the participant remembered what he/she had found by seeing only the scene video. The second case is when the participant remembered his/her findings after viewing the scene video with gaze plots. The third case is when the participant did not remember what he/she had found even when he/she viewed the scene video with gaze plots.

Participants 2 and 3 recalled every finding either from the scene video alone or with gaze plots . The other participants recalled more than 82% of their findings. As for contribution of the gaze plots, the use of gaze plots was required to recall findings in 26% to 52% of the cases.

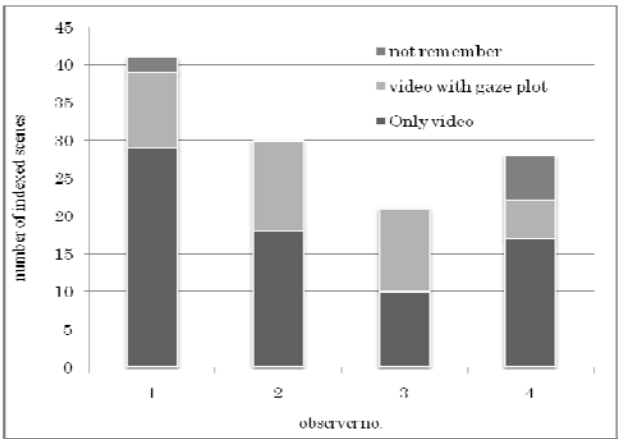

Figure 6 Number of indexed scenes for each observer.
Each bar consists of three cases of remembering.

Findings that were recalled by the participant from only scene videos tend to be associated with phenomena related to the user's action. For example one participant found that the hand-towel often becomes caught in the cabinet door when the user opens and closes the cabinet. (In a typical Japanese kitchen, people hang a towel on the lower cabinet door. The hand towel is frequently used during cooking to wipe hands.) The participant found the problem by observing a scene where the user moves the towel out of the way to take some dishes out when serving the food. His gaze movements showed that he observed the series of the user's actions such as holding the towel, opening the cabinet, removing dishes and plates out of the cabinet and closing the cabinet door.

The scene videos each last for 10 seconds. Within such a short time period, one tends to focus on phenomena that contain more information than others to understand what is occurring in the scene. In this experiment, the user's movements fall into this category. Thus, we found that the participants were able to easily recall findings related to the user's actions.

On the other hand, there are scenes in which the participant did not remember what he/she had found until he/she had viewed the accompanying gaze plots. We observed that these findings tend to be static phenomena such as the environmental setting. For example, a participant found that the upper cabinet is often left open in the scene. The user opened the cabinet door and often forgot to close it. The participant pointed out that a problem existed in the structure of the cabinet door. In the replay, he could not remember what he had originally found in the scene by seeing only the scene video, as he focused on the user's hands as she cut vegetables, i.e. the user's movement. The participant remembered his original finding only after seeing his gaze plots clustered around the cabinet.

There are very few scenes where the participants could not remember what they had found even after viewing the gaze plots on the scene video. One cause of such a failure to recall is when the finding had been noticed more than 10 seconds before the marked time. In other words, the participant did not click within 10 seconds after he/she had recognized the phenomenon. Although we set 10 seconds for the scene video as the experimental condition, the duration can easily be changed in practice. Another case is when the participant clicked because he/she felt that something was wrong but was unable to clearly identify the problem.

Figure 7 Example scene in which observer remembers finding with only scene video. (Hand-towel is often caught in cabinet door when user opens and closes the cabinet.)

Figure 8 Example scene in which observer remembers finding with scene video and gaze plot. (Upper cabinet is often left open.)

6.2 Sharing found phenomena

In the final experiment, the four participants each tried to recognize problem phenomenon he/she had missed but had been found by other participants. In the results, all participants recognized all shown phenomena by seeing the scene videos and gaze plots. Figure 9 shows the number of scene videos that each participant viewed in experiment 2. The number differs for each participant because the number of missed findings is different. Each bar consists of two cases. One is when the participant recognized a problem phenomenon by seeing only the scene video, and the other is when he/she needed the gaze plots to recognize the phenomenon. In total, the participants recognized about 90% of the problem phenomena by only seeing the scene videos.

The participants tend to recognize a problem phenomenon related to the user's movement by seeing only scene video. Figure 10 shows a scene in which a participant focused on the user who is putting away a seasoning container into the cupboard. The user often uses seasonings while cooking, so she has to move to the lower cabinet many times to get seasonings. Moreover, the user has to bend over to reach into the cupboard since its position is low. Frequent bending over may put a strain on the back, especially for elderly users. The red dot in the figure 10 represents a gaze point of the participant. We can see that the participant gaze follows the user's movement.

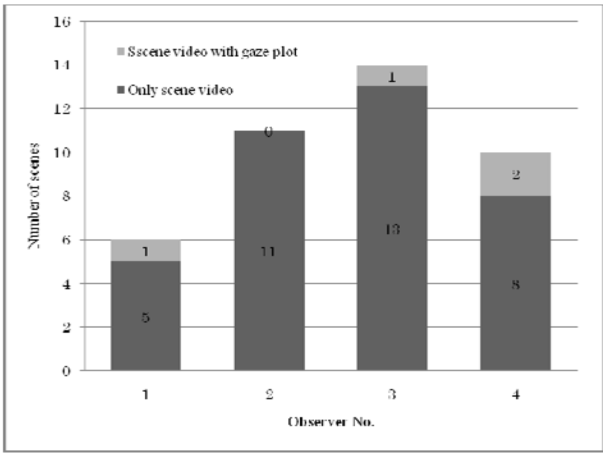

Figure 9 Number of shared scene videos for each observer. Each bar consists of cases of finding phenomena.

Figure 10 Participant focused on user who is putting away seasoning container into the cupboard.

On the other hand, there were some cases in which the participants required gaze plots to recognize the problem phenomenon. Figure 11 shows a snapshot of such scene. In this scene, the user tried to flip the hamburger in the pan. The observer's gaze plots show that he observed the area around the user's hands. However, the problem found by the original participant was that the user's head seems close to hitting the range hood. The participant recognized the problem after seeing the gaze plots of the original finder, which is localized at the area around the user's head, as shown in figure 12. As this example shows, the participants tend to focus on the user's movements and require gaze plots to recognize static phenomena related to environmental issues.

When viewing such short-duration videos, we observed two types of strategy in the gaze movements. One is to fix the gaze on a relatively few points for a certain duration. The other is that the gaze frequently moves about within the short duration, searching for problems in a larger area. We observed some cases in which the participant who took the latter strategy did not become aware of a problem phenomenon even when his/her gaze was fixed on the relevant area for a short time span. Some problems are recognized only after some cognitive process has taken place in the observer's mind, in which case the gaze must remain in the problem area for a certain duration. We believe that showing the gaze plots of the original finder is effective in encouraging others to pay close attention to the relevant areas and to recognize the phenomenon related to a problem.

Figure 11 Participant focuses on the area around the user's hands.

Figure 12 Participant recognized original finding after viewing gaze plots of the original finder.

7. Conclusions

We proposed an observation support system that aims to reduce the cognitive load caused by the recording/documentation process and increase both the efficiency and quality of record representation for reviewing and sharing findings. The system automatically extracts scene videos starting a few seconds before the observer writes on the digital notepad, and later displays the videos along with the observer's gaze plots and related memo. To evaluate the effectiveness of the system, we carried out experiments under the restricted condition whereby the observer tries to recall his/her findings and to find relevant phenomena found by others without any written description. The results show that extracted scene videos with gaze plots are effective in recalling and sharing such findings. Specifically, we found that gaze plots help to recall and share findings that are related to static environmental phenomena.

Acknowledgement

This work is supported by the Japan Society for the Promotion of Science (No. 20686012) and the Design Innovation Lab., University of Tokyo. We would like to thank TOTO Ltd. Universal Design Research Center, a partner of the Design Innovation Lab., for providing the user video used in the experiment.

References

[1] Timo Jokela, T., Iivari, N., Matero, J., Karukka, M. (2003) The standard of user-centered design and the standard definition of usability: analyzing ISO 13407 against ISO 9241-11, *ACM Int. Conference Proc.; Vol. 46 Proc. the Latin American conference on Human-computer interaction*, pp. 53-60.

[2] Kelley, T., Littman, J., Peters, T. (2001) *The Art of Innovation: Lessons in Creativity from IDEO*, Bantam Dell Pub Group, pp. 28-52.

[3] Ylirisku, S., Buur, J. (2007) *Designing with video: focusing the user-centered design process*, Springer.

[4] Duchowski, A. T. (2003) *Eye Tracking Methodology: Theory and Practice*, Springer. pp.31-38.

[5] Saslow, M G (1967) Effects of components of displacement step stimuli upon latency for saccadic eye movements, *J Opt Soc. Am* 57, pp. 1024-1029.

[6] Carlos H. Morimoto , Marcio R. M. Mimica (2005) Eye gaze tracking techniques for interactive applications, *Computer Vision and Image Understanding*, vol.98 no.1, pp. 4-24.