

A PERFORMANCE COMPARISON OF EYE TRACKING AND MOUSE INTERFACES IN A TARGET IMAGE IDENTIFICATION TASK

O. Oyekoya, F.W.M. Stentiford

Content Understanding Group, University College London,
Adastral Park, Ipswich UK IP5 3RE

Keywords: eye tracking, target identification, image search interface, scene perception.

Abstract

Eye-tracking technology offers an intimate and immediate way of communicating with human thought processes, interpreting users' behaviours and guiding a computer search through large image databases. In experiments involving image identification tasks, a target image was presented along with 24 other distractor images, while task completion times were measured for two modes of interface control and two experimental conditions. Results show faster target identification for the eye interface than the mouse for identifying a target image on a display. The effects of the order of input modes, positions of target images and the level of skills transfer produced information that is relevant to future eye tracking studies. The results of the performance analysis should also serve as a basis for the refinement and redesign of new interfaces for image retrieval.

1 Introduction

The eye is driven by attention. It can obtain relevant information quickly and in ways that do not interfere with the task itself. Eye-tracking technology offers an intimate and immediate way of communicating with human thought processes, interpreting users' behaviours and guiding a computer search through large image databases.

Duchowski's review [2] indicates that research in the applications of eye tracking is increasing and presents an account of current applications. He divides them into diagnostic and interactive applications based on real-time analysis for the latter as opposed to offline analysis for the former. The increase in interactive usage has been fuelled by the advancement of eye tracking and computer hardware in the last few decades. In its diagnostic capabilities, eye-tracking provides a comprehensive approach to studying interaction processes such as the placement of menus within web sites to influence design guidelines [7]. Eye tracking work has also concentrated upon replacing and extending existing computer interface mechanisms rather than creating a new natural form of interaction. The tracking of eye movements has been employed as a pointer and a replacement

for a mouse [3], to vary the screen scrolling speed [9] and to assist disabled users [1]. Schnell and Wu [15] apply eye tracking as an alternative method for the activation of controls and functions in aircraft. Dasher [18] uses a method for text entry that relies purely on gaze direction. The imprecise nature of saccades and fixation points has prevented these approaches from yielding benefits over conventional human interfaces. Fixations and saccades are used to analyze eye movements, but it is evident that the statistical approaches to interpretation (such as clustering, summation and differentiation) are insufficient for identifying interests due to the differences in humans' perception of image content. More robust methods of interpreting the data are needed. There has been some recent work on document retrieval in which eye tracking data has been used to refine the accuracy of relevance predictions [13].

Human eye behaviour is defined by the circumstances in which they arise. The eye is attracted to regions of the scene that convey the most important information for scene interpretation. Initially these regions are pre-attentive in that no recognition takes place, but moments later in the gaze the fixation points depend more upon our own personal interests and experience. Understanding of innate eye behaviour (i.e. fixations and saccades) is essential in drawing inferences from this gaze behaviour. Salvucci [14] presented five algorithms for identifying fixations and saccades from eye tracking protocols. The algorithms are classified with respect to spatial and temporal characteristics. The ability to track human gaze during image viewing presents an interesting area of research, where humans can retrieve a target image from an image collection. Indeed, lack of high-quality interfaces for query formulation has been a barrier to effective image retrieval [17]. Eye tracking presents an adaptive approach that can capture the user's current needs and tailor the output to fit the needs of the user through a high-quality interface.

Yamato et al [20] conducted an experiment to evaluate two adjustment techniques, in which computer users use both their eye and hand in carrying out operations in GUI environments. In the first technique, the cursor moves to the closest GUI button when the user pushes a mouse button. The second adjustment involves gross movement of cursor by the eye and the user makes final adjustments and moves the mouse onto the GUI button. The second adjustment performed better. The

input device is switched from the eye tracking device when the user moves the mouse in the manual adjustment, so the user has to be careful not to move the mouse until required. Ware and Mikaelian [19] evaluated the eye tracker as a device for computer input by investigating three types of selection methods (button press, fixation dwell time and screen select button) and the effect of target size. Their results showed that an eye tracker can be used as a fast selection device providing the target size is not too small. Eye gaze has also been shown to be faster than the mouse for the operation of a menu based interface [10]. Sibert and Jacob [16] performed two experiments involving circles and letters respectively. The former required little thought, while the latter required comprehension and search effort from participants. Eye gaze interaction was found to be faster than the mouse in both experiments.

Humans perceive visual scenes differently. We are presented with visual information when we open our eyes and carry out non-stop interpretation without difficulty. Research in the extraction of information from visual scenes (high-level scene perception) has been explored by Yarbus [21], Mackworth & Morandi [6] and Hendersen & Hollingworth [4]. Mackworth and Morandi [6] found that fixation density was related to the measure of informativeness for different regions of a picture and that few fixations were made to regions rated as uninformative. The picture was segmented and a separate group of observers were asked to grade the rate of informativeness. Scoring the informativeness of a region provides a good insight into how humans perceive a scene or image. Henderson and Hollingworth [4] described semantic informativeness as the meaning of an image region and visual informativeness as the structural information. Fixation positions were more influenced by the former compared to the latter. The determination of informativeness and corresponding eye movements are influenced by task demands [21].

Previous work [11,12] used a visual attention model to score the level of informativeness in images and found that a substantial part of the gaze of the participants during the first two seconds of exposure is directed at informative areas as estimated by the model. Subjects were presented with clear region-of-interest images and results showed that these attracted eye gaze on presentation of the images studied. This led credence to the belief that the gaze information obtained from users when presented with a set of images could be useful in driving an image retrieval interface.

In this paper, experiments are conducted to compare the speed of the eye and the mouse as an input mode to control an interface. It is expected that users in a visual search will look at any object that is similar to the target so that it can be recognised and a decision made to end the search. This natural behaviour involves the two stages of inspection and target selection respectively. Using a mouse requires inspection of the images, a mouse move and a click on the target while the eye involves inspection and fixations of the eye on the target. In this task-oriented experiment,

participants are asked to find a target image in a series of displays, with the aim of studying the response times of searching and selecting the target image using the computer mouse or the eye under varying conditions.

2 Method

2.1 Equipment and Data

An Eyegaze System [5] was used in the experiments to generate raw gaze point location data at the camera field rate of 50 Hz (units of 20ms). A clamp with chin rest provided support for chin and forehead in order to minimize the effects of head movements, although the eye tracker does accommodate head movement of up to 1.5 inches (3.8cm). Calibration is needed to measure the properties of each subject's eye before the start of the experiments. The processing of information from the eye tracker is done on a 128MB Intel Pentium III system with a video frame grabber board.

25 images were selected from the Corel image library. The initial screen (including the target image) is shown in Figure 1. These images were displayed on a 15" LCD Flat Panel Monitor at a resolution of 1024x768 pixels.

2.2 Experiment

A total of 12 participants took part in this experiment. Participants included a mix of students and university staff. All participants had normal or corrected-to-normal vision and provided no evidence of colour blindness.

Participants were asked to locate a target image from a series of 50 grid displays of 25 stimuli (24 distractors and 1 target image shown in Figure 1). On locating the target image, the participants select the target by clicking with the mouse or fixating on it for longer than 40ms with the eye. The grid is then re-displayed with the positions of the images (including the target image) re-shuffled. Participants were randomly divided into two groups (Table 1), the first group used the eye tracking interface first then the mouse, and the second group used the interfaces in the reverse order. This enabled any variance arising from the ordering of the input modes to be identified. Different sequences of the 50 target positions were also employed to identify any confounding effects arising from the ordering of the individual search tasks. All participants experienced the same sequence of target positions (Figure 2a) as well as different sequences (Figure 2b,c) while using the two input modes. Figure 3 describes the sequence of display for the images.

A typical participant in the mouse first group performed four runs: mouse (target position 1), eye (target position 1), mouse (target position 2) and eye (target position 3). There was a 1 minute rest in between runs.

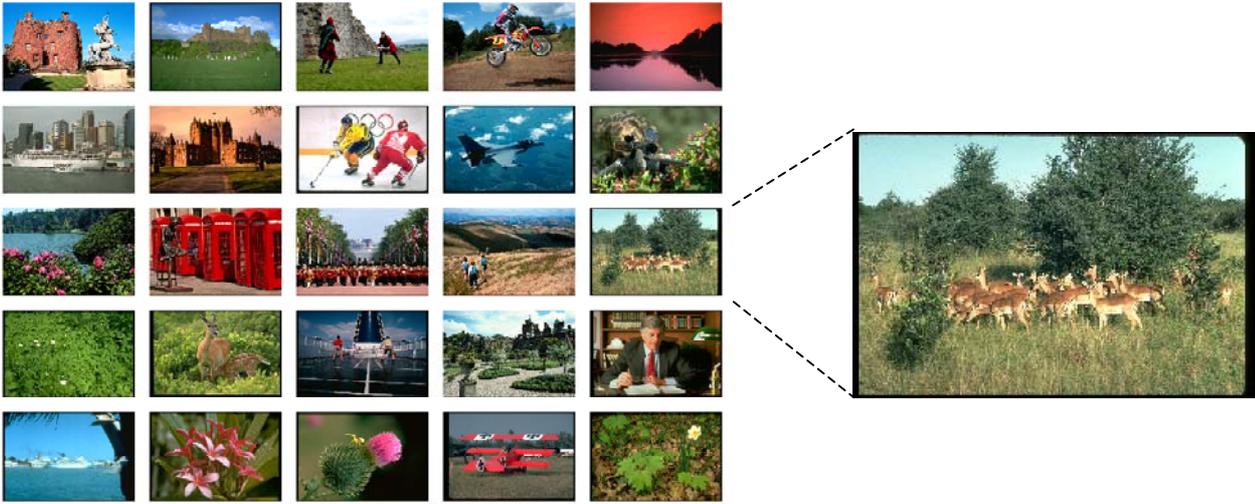


Figure 1: 25 images arranged in a 5x5 grid used in runs (target image expanded on the right)

6	12	9	24	14
30	46	35	48	26
15	2	22	5	11
39	27	32	38	42
21	8	18	16	1
37	41	29	44	34
19	23	25	20	7
45	43	47	50	28
4	17	10	3	13
31	49	33	36	40

Figure 2(a): Target Positions 1

25	8	1	12	7
41	39	27	30	23
4	18	14	17	3
43	31	35	42	38
6	20	22	5	10
37	49	33	46	28
24	11	26	19	15
45	47	44	40	32
16	2	9	13	21
34	29	50	48	36

Figure 2(b): Target Positions 2

21	5	12	8	4
47	33	26	40	23
10	15	2	20	14
35	38	31	44	30
17	19	28	11	1
43	41	49	36	27
6	24	22	7	25
32	45	34	39	42
13	3	9	18	16
29	37	48	50	46

Figure 2(c): Target Positions 3

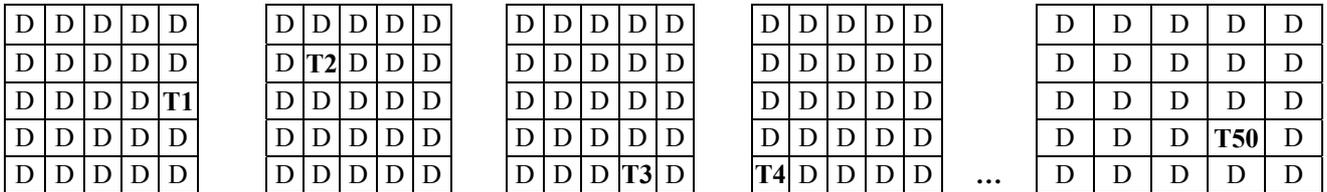


Figure 3: Sequence of displays for the target positions 1 in Figure 2a (T1=target 1; D=distractors)

3 Results

Order	Target Positions	Input Mode	Response Time	
			Mean	Standard Deviation
Mouse First (6 participants)	Same-sequence	Mouse	2.33	0.51
		Eye	1.79	0.35
	Different-sequence	Mouse	2.43	0.38
		Eye	1.96	0.42
Eye First (6 participants)	Same-sequence	Mouse	2.35	0.82
		Eye	2.29	0.74
	Different-sequence	Mouse	2.59	1.44
		Eye	2.27	0.73

Table 1: Mean response times for target image identification task

The length of time it took to find the target image from the grid display were recorded and 50 response times were

obtained for each participant's run. The mean response times are calculated and presented in Table 1. The loading

of 25 images in the 5 x 5 grid display took an average of 110ms on a Pentium IV 2.4GHz PC with 512MB of RAM. Gaze data collection and measurement of response times were suspended while the system loaded the next display.

The 48 means were entered into a mixed design ANOVA with three factors (order of input, input mode, and target positions).

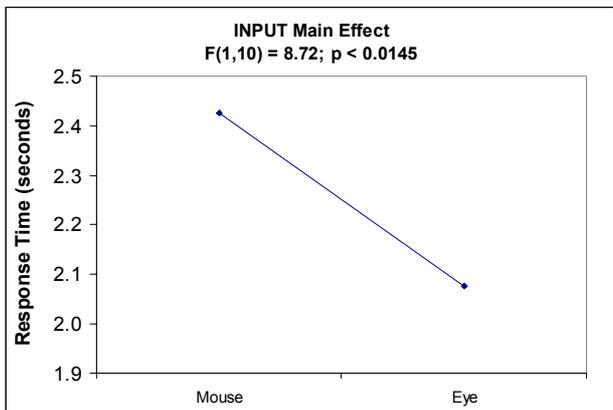


Figure 4: Mean response time by input

There was a significant main effect of input, $F(1,10)=8.72$, $p=0.015$ with faster response times when the eye was used as an input (2.08s) than when the mouse was used (2.43s) as shown in Figure 4. The main effect of the order was not significant with $F(1,10)=0.43$, $p=0.53$. The main effect of target positions was not significant, $F(1,10)=0.58$, $p=0.47$. All two-factor and three-factor interactions were not significant.

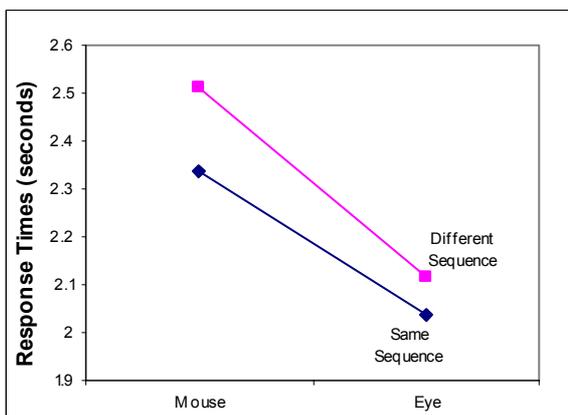


Figure 5: Mean response time by input and target position sequence

Further analysis of the first-order and second-order simple main effects was conducted individually on all levels of the three factors. The input modes influenced response time of subjects when they were presented with the same-sequence target positions, $F(1,10)=14.22$, $p=0.004$, with faster eye response times ($M=2.04s$, $SD=0.61$) than the mouse ($M=2.34s$, $SD=0.65$) as shown in Figure 5. In other words the response times were shorter by a greater amount

than the mouse when the participants experienced the same-sequence target positions as opposed to the response times for different-sequences when the difference was less ($p=0.075$). There was significant variability between the response times (Table 1).

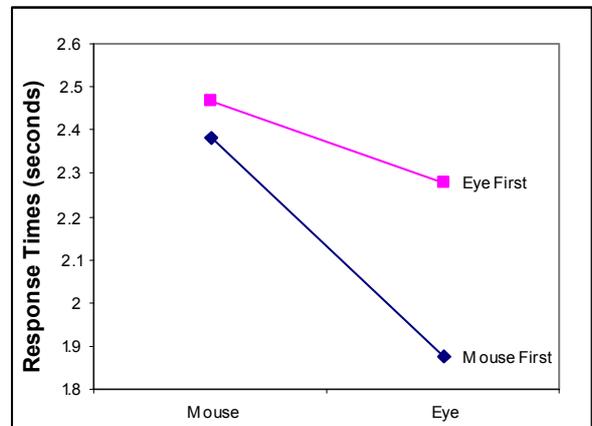


Figure 6: Mean response time by input and Mouse/Eye order

The input modes influenced the response times of subjects in the Mouse First group, $F(1,10)=9.09$, $p=0.013$, with faster eye response times ($M=1.878s$, $SD=0.381$) than the mouse ($M=2.38s$, $SD=0.43$). The response time was faster with the eye interface than the mouse when the participants used the mouse interface first and no significant difference between the eye and mouse interfaces when the eye was used first, $p=0.27$ (Figure 6)

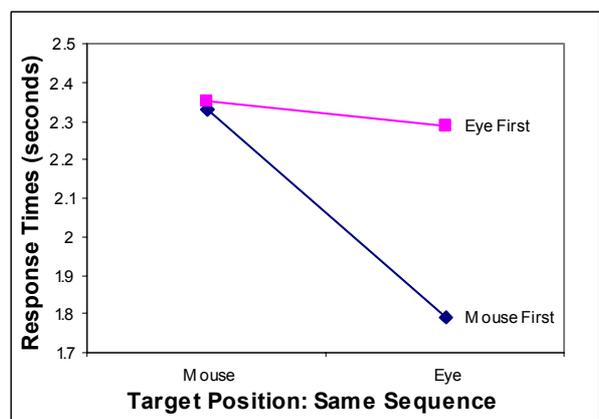


Figure 7: Mean response time by input, order and same-sequence target position

The input levels influenced the response times of subjects in the Mouse First group when they were presented with the same sequence target positions, $F(1,10)=22.81$, $p=0.001$, with faster eye response times ($M=1.79s$, $SD=0.345$) than the mouse ($M=2.33s$, $SD=0.51$). The response time was faster with the eye interface when the mouse was used first and participants experienced the same-sequence target positions (Figure 7).

There were no other significant simple main effects. A fourth factor of display was included in the mixed design ANOVA to investigate the effect of the grid display (Figure 1) changes. There was a significant main effect of display, $F(49,490)=2.39$, $p<0.0001$. This indicated that the display affected the response times, as each display is affected by previous displays. It can also be argued that the effect of change blindness [4] might be present in this study during display changes. Thus, the average response time is preferred in the three-factor design.

4 Discussion

The 25 stimuli presented to each participant and the predetermined choice of image target produced a difficult task and the runs required a high cognitive load. The participant has to search for the target and then make a selection. When using the mouse the participant has to first locate the cursor and then move the mouse to the item to be selected. This can result in slower mouse responses. An eye tracking interface requires a fixation for a fixed period to make a selection. Our results indicate slower mouse responses and is supported by the main effect of input ($p=0.015$) and is consistent with Ware and Mikaelian's conclusions [19].

It could be suggested that some of the skills gained through the use of the mouse in this task are passed on and remembered during the subsequent eye gaze task thereby obtaining a large difference in response times. On the other hand any new skills acquired during the use of the eye do not make much difference to subsequent mouse performance. This may help to confirm that simple knowledge of target positions from previous tasks is not a major confounding factor in the results.

Although, there was a simple effect of the input on the same-sequence target positions, the differences in the mean response times were similar as shown in Figure 5. There was a significant variability around the means observed on closer scrutiny of the data. The different-sequence target positions were not affected by the input and will be the choice for future experiments. The significance of both conditions together (Mouse First and same-sequence target positions) was also tested. Given that there was a simple effect individually on the Mouse First group and same-sequence target positions, the test of significance on both conditions was not surprisingly high ($p=0.001$).

During experiments it was possible that users did not always look directly at the targets. This would have meant that gaze direction did not necessarily indicate interest, only a general direction and could have confounded some results. However, the effects of covert attention were minimized by giving users clear and focussed instructions.

Eye tracking technology promises to be a fast and effective interface to computers when retrieving and categorising

images. An eyetracking interface will enable users to browse large image databases by traversing a network of pre-computed image similarities. The computer will be able to display related images in response to eye movements with a minimum of thought and manual involvement on the part of the user. The scan paths naturally provide accumulating evidence of the users' intentions and wishes and the system should converge on target images after relatively few steps.

5 Conclusion

The goal of this study was to explore the speeds of visual processing involved in an image target identification task. A significant main effect of input was found with the eye interface having faster response times than the mouse interfaces. There were significant simple main effects of the eye on the Mouse First group and the same-sequence target positions. This result confirmed that skill transfer was taking place when the mouse was used first but not when the eye was used first. This could indicate that the experience gained during visual tasks carried out using a mouse will benefit users if they are subsequently transferred to an eyetracking system.

Future work is aimed at devising new interfaces for content based image retrieval that are easier and more natural to use and which converge to the targets more rapidly through the use of behavioural information extracted in real time from eye gaze data.

Acknowledgements

The authors acknowledge the support of BT Research and Venturing, SIRA and the Engineering and Physical Sciences Research Council in this work. The work has been conducted within the framework of the European Commission funded Network of Excellence "Multimedia Understanding through Semantics, Computation and Learning" (MUSCLE) [8].

References

- [1] Corno F., Farinetti L. and Signorile I. (2002), "A cost effective solution for eye-gaze assistive technology," IEEE Int. Conf. on Multimedia and Expo, August 26-29, Lausanne.
- [2] Duchowski, A. T. (2002), "A Breadth-First Survey of Eye Tracking Applications", Behaviour Research Methods, Instruments, & Computers (BRMIC), 34(4), pp.455-470.
- [3] Hansen J.P., Anderson A.W., and P. Roed (1995), "Eye gaze control of multimedia systems", Symbiosis of Human and Artifact (Y. Anzai, K. Ogawa, and H. Mori (eds), Vol 20A, Elsevier Science, pp 37-42.
- [4] Henderson John M. and Hollingworth Andrew (1999), "High-Level Scene Perception", Annual Reviews Psychology 50:243-71.
- [5] LC Technologies Inc. - <http://www.eyegaze.com/>

- [6] Mackworth, N., And Morandi, A. (1967), "The gaze selects informative details within pictures", *Perception and Psychophysics* 2, 547–552.
- [7] McCarthy, J, Sasse, M.A. & Riegelsberger, J. (2003), "Could I have the menu please? An eye tracking study of design conventions", *Proceedings of HCI2003*, 8-12 Sep 2003, Bath, UK.
- [8] Multimedia Understanding through Semantics, Computation and Learning, Network of Excellence. EC 6th Framework Programme. FP6-507752. <http://www.muscle-noe.org/>
- [9] Numajiri T., Nakamura A., and Kuno Y. (2002), "Speed browser controlled by eye movements", *IEEE Int Conf. on Multimedia and Expo*, August 26-29, Lausanne, 2002.
- [10] Ohno Takehiko (1998), "Features of eye gaze interface for selection tasks", *APCHI'98*, Japan, 176-181
- [11] Oyekoya O. K., Stentiford F. W. M. (2004), "Exploring Human Eye Behaviour Using a Model of Visual Attention", *International Conference on Pattern Recognition*, Cambridge UK, August.
- [12] Oyekoya O K, Stentiford F W M (2004), "Eye Tracking as a New Interface for Image Retrieval", *BT Technology Journal*, Vol 22 No 3, July.
- [13] Puolamäki K., Salojärvi J., Savia E., Simola J., Kaski S. (2005), "Combining Eye Movements and Collaborative Filtering for Proactive Information Retrieval", In *Proceedings of the 28th ACM Conference on Research and Development in Information Retrieval (SIGIR) 2005*.
- [14] Salvucci, D. D., and Goldberg, J. H. (2000), "Identifying fixations and saccades in eye-tracking protocols", In *Proceedings of the Eye Tracking Research and Applications Symposium* (pp. 71-78). New York: ACM Press.
- [15] Schnell, T., Wu, T., (2000), "Applying Eye Tracking As Alternative Approach for Activation of Controls and Functions in Aircraft", *Proceedings of the 5th International Conference On Human Interaction with Complex Systems (HICS)*, April, 30 - May, 2, 2000, Urbana, Illinois, USA, pp 113.
- [16] Sibert, Linda E. and Jacob, Robert J.K. (2000), "Evaluation of eye gaze interaction", In *CHI 2000 Conference on Human Factors in Computing Systems*, April 2000.
- [17] Venters, C.C., J.P. Eakins and R.J. Hartley (1997), "The user interface and content based image retrieval systems", *Proc. of the 19th BCS-IRSG Research Colloquium*, Aberdeen, April.
- [18] Ward D.J. and MacKay D.J.C. (2002), "Fast hands-free writing by gaze direction", *Nature* 418 pp 838, Aug. 22.
- [19] Ware, C. and Mikaelian, H. (1987). "An evaluation of an Eye Tracker as a Device for Computer Input", *Proceeding of SIGCHI+GI '87*, Human Factors in Computing Systems. Also published as a special issue of the *SIGCHI Bulletin*. 183-88. April.
- [20] Yamato, M., Monden, A., Matsumoto, K., Inoue, K., Torii, K. (2000) "Button selection for general GUIs using eye and hand together", *Proc. 5th International Working Conference on Advanced Visual Interfaces (AVI2000)*, ACM Press, pp.270–273.
- [21] Yarbus, A. (1967), "Eye Movements and Vision", Plenum Press, New York.