# Find The Difference! Eye Tracking Study on Information Seeking Behavior Using an Online Game

E. Józsa[1] & B. P. Hámornik[2]

[1,2]Department of Ergonomics and Psychology, Budapest University of Technology and Economics,

Magyar Tudósok körútja 2. Q bldg, Budapest, 1117-HUNGARY

[1,2] {jozsae,hamornik}@erg.bme.hu

[1,2]*www.erg.bme.hu*

## ABSTRACT

With the advent of digital era web applications have become inevitable part of our lives. We are using the web to manage even the financially or ethically sensitive issues. For this reason exploration of information seeking behavior is an exciting area of research. Current study provides insight on information seeking behavior using a classic 'Find the Difference' game.

50 university students between the age of 19 and 26 participated in the study. Eye movement data were recorded with a Tobii T120 device. Participants carried out 4 continuous tasks. Each task included two pictures side by side with 7 hidden differences. After finishing the tasks, participants were asked to repeat the game with the same picture set. This data collection methodology allows the evaluation of learning curves. Additionally, participants were asked about their hand preference.

For the purpose of analysis the following metrics were applied: task times (including saccades), fixation count and fixation duration (without saccades). The right- and left-hand side on each picture was selected as AOI (Area of Interest) to detect side preference in connection with hand preference. Results suggest a significant difference between male and female participants regarding aggregated task times (male 58.37s respectively female 68.37s), deviation in the number of fixations and fixation duration (apparently female have less but longer fixations) and also in the distribution of fixations between AOIs.

Using eyetracking data current paper highlights the similarities and differences in information acquisition strategies respectively reveals gender and education (Arts vs. Sciences) dependent characteristics of interaction.

**Keywords: eyetracking, online games, gender differences**

## 1. INTRODUCTION

In recent years eye tracking technology gained considerable attention from researchers of various fields such as marketing, communication, consumer behaviour, user interface design or software ergonomics. Due to rapid technological advancements this promising technique has become a popular method among usability researchers, psychologists and neuroscientists. Current study establishes connection between eye tracking, psychological gender studies and psychology of information search.

Concerning the application of eye tracking in software ergonomics and usability Goldberg employs the technology to evaluate specific design features of a prototype portal application (Goldberg, 2002) and computer interface design (Goldberg, 1999). In their study Pan, Hembrooke, Gay, Granka, Feusner and Newman (2004) explores the determinants of ocular behaviour on a single web page considering individual and gender differences. The work of Meyers-Levy (1989) suggests that the process of message claims differs among genders. Apparently, relative to males, females often engage in specific message contents more in detail. Considering the picture viewing behaviour of human observers, Engelke, Zepernick and Maeder (2010) explored the impact of structural distortions and image content. Their results suggest that distortion seems to play a minor role in viewing behaviour compared to the image content.

Regarding online games the application of eye tracking is limited to use it as an input device (Isokoski & Martin, 2006, Gowases, 2007), the e-version of classic 'Find the difference!' game has remained out of research focus yet. This type of game provides an insight on different information seeking strategies and can be considered as a representation of a search, comparison, and decision situation. This comparison and decision between visually presented objects or products can be related to consumer behaviour. Contrary to the normative decision theory, people do not search objects for all possible information to make their rational decision, but they rely on preconceptions, feelings, or on impressions. Besides these factors familiarity and previous experiences (and learning as a consequence) can also play a major role in decision-making processes. Using the terminology of Thaler and Sunstein (2008) people can be characterized as 'Human' rather than the economically rational and profit maximising 'Econ'. In the online game of 'Find the difference!' this decision-making is modeled, and the behaviour of the 'Human' can be examined by eye tracking. Participants' eye movements provide information about the strategies that 'Human' use to search information, compare objects and make decisions.

Our study investigates the characteristics of picture viewing behaviour and the role of feedback in connection with task difficulty and learning effect during an online game. The study concerns whether the variations in ocular measurements related to gender or educational background differences or caused by other independent variables.

### *1.1 Related research on eye tracking measurement*

Objective studies of human eye movements date back to the turn of the twentieth century, although methods involving direct observation go back to the 18th century. Recently, rapid technological advancements in terms of increased processor speed or advanced digital video processing have both lowered the cost and dramatically increased the efficiency of eye tracking equipment (Land, 2006). These systems usually based on pupil position, made visible by illuminating the eye with infrared light to produce a 'white' pupil (also called bright pupil), which can be tracked. It gives a pictorial display of foveal gaze direction, or 'point of regard' (Duchowski, 2003).

In the extensive literature of eye related measurements several definitions have been widely accepted such as fixations and saccades. Fixation is generally defined as a 200-300 millisecond of relatively motionless dwell, while saccades are the rapid and ballistic movements of the eyes that abruptly change the point of fixation (Rayner, 1998). Saccadic movements occur so quickly that they occupy only 10% of the total time spent in eye movements whereas fixation accounts for 90% of the time. While fixations have been linked to intensive cognitive processing throughout a saccade visual perception is suppressed (Guyton, 1977). The definition scanpath, proposed by Noton and Stark (1971) is also widely accepted and applied as sequence of fixations and saccades. They performed their eye movement analysis over images and defined the observed pattern as a scanpath. Their research suggests that even without instructions participants tend to fixate on informative regions of the stimulus (e.g. the corners of a square). Fitts, Jones and Milton (1950) showed that the frequency of fixations within an Area of Interest (AOI) indicates the degree of importance while fixation duration is a measure of complexity and difficulty of visual display. Loftus and Mackworth (1978) and later Rayner (1998) have found that human eye are attracted to the most informative areas of a stimulus, which can be measured with the time spent within an AOI (Pan, 2004).

## 2. METHODOLOGY

### *2.1 Participants and equipment*

50 students were recruited from the Budapest University of Technology and Economics (BUTE) through university mailing lists. Students participated voluntarily in the study. Participants were screened for normal (or corrected to normal) visual acuity. Considering the limitations of the eye tracker, the data of participants with contact lenses were not further analyzed. Due to various issues (low data quality, contact lenses) recordings from 7 users were later discarded from the analysis. Eye movement recordings of 29 women and 14 men between the age of 19 and 26 have been evaluated as demonstrated in Table 1.

**Table 1.** *Participant groups.*

| Participant Groups | Education | Total |
|---|---|---|

| | | Arts | Sciences | |
|---|---|---|---|---|
| Gender | Female | 13 | 16 | 29 |
| | Male | 3 | 11 | 14 |
| Total | | 16 | 27 | 43 |

For the purpose of eye movement recording a Tobii T120 device was applied. In the T120 model the eye tracker is integrated in a 17-inch TFT monitor and allows the freedom of head movements within a 30×22×30 cm virtual box. The data sampling occurs at 120 Hz. For the remote control of the eye tracker and running the Tobii Studio data analysis software a desktop computer was applied. We also used Tobii Studio software to export eye movement data in txt format.

### *2.2 Procedure and stimulus description*

The study took place in the usability laboratory of the Faculty of Ergonomics and Psychology at BUTE. Before each data recording the session coordinator prepared the stimuli. Upon arrival participants were greeted by the session coordinator and were asked to comfortably take a seat in front of the tracker device. Then the position of the eye tracker device was adjusted to the height of the participant. After, Tobii's built-in 5-point calibration method has been applied. Once the calibration succeeded participants were asked to start the game in full screen mode.

A free online game-Museum of Thieves-formed the basis of current study. The game can be accessed through Google Chrome's web store. After calibration participants were asked to start the game on level 1 and find the seven hidden differences on four subsequent levels (one mirrored picture on each level, see Figure 1.) until they reach the end of level 4. On each level participants had to reveal all the differences in order to finish the task and reach the next level. This first part of data recording referred as phase A or presentation A. Upon participants reached the end of level 4 the session coordinator gave the instruction to quit and start the second presentation (referred as presentation B or phase B), which was prepared on the computer. Due to the game's construction differences can be eliminated on both side of the picture.

In his study, Hercegfi (2011) demonstrated cognitive-style based differences in web searching tasks. He used a customized version of MBTI psychological test and found that thinking-type users understood the logic of the content and the user interface easily while the apparent intactness of the page layout caused more mental effort for users with the feeling-type cognitive style. In order to avoid bias, caused by different cognitive styles (e.g. the deviation in number of objects could be detected easier for one type than for the other), hidden differences changed randomly across participants and presentations. Apart from the hidden differences the same pictures were displayed for all the participants in both presentations. The following main types of differences appeared throughout the game: difference in size of the objects (e.g. length of rug), difference in number of objects (2 vs. 3 chess pieces), existence of objects and content of objects (e.g. painting with the same frame but different content). The random selection of differences has been made automatically from a pool of 16 choices in case of level 1; 14 choices in case of level 2; 15 choices in case of level 3 and 16 choices in case of level 4.

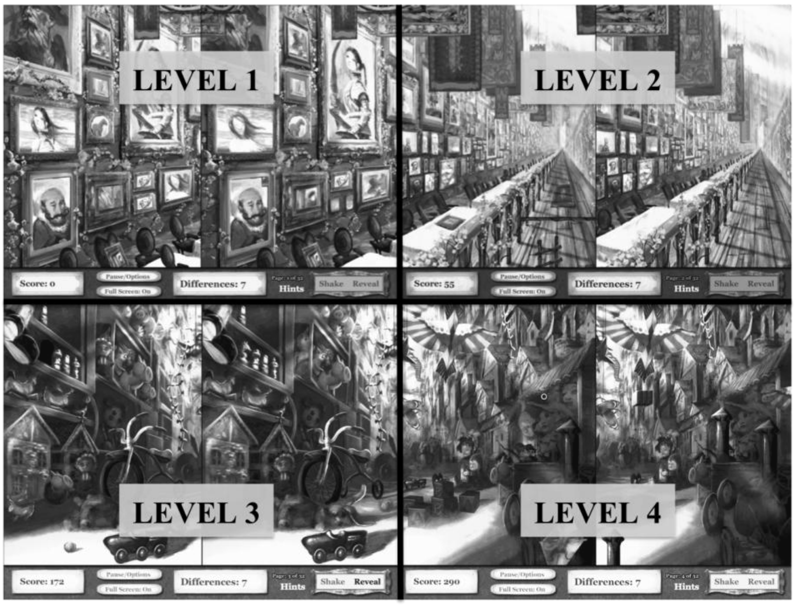


**Figure 1.** *Level 1, 2 ,3 and 4 also referred as Picture 1, 2, 3 and 4.*

### *2.3 Measures*

During the data analysis gender and education background served as independent variables. Participants were grouped into 'Arts' and 'Sciences' sections on the ground of their university major. The 'Arts' section harbored students with psychology and communication major while 'Sciences' section comprises industrial design, mechanical engineering and electrical engineering majors.

Based on the literature current study applied the following dependent variables as determinants of ocular behavior: total visit duration, which is a measure of duration of all visits within an Area of Interest (referred as task time in the Results section). The results of Nakayama (2002) suggest that this measure is negatively related to task difficulty. Fixation duration, which measures the duration of each individual fixation within an AOI, and according to Pelz, Canosa and Babcock (2000), it is influenced by the complexity of the task. Fixation count measures the number of times participants fixate on an AOI. We also computed mean fixation duration, which is frequently used as an indicator of information complexity and task difficulty (Rayner, 1998). We applied these measurements on Areas of Interest 1, 2, 5 and 6, demonstrated on Figure 2.

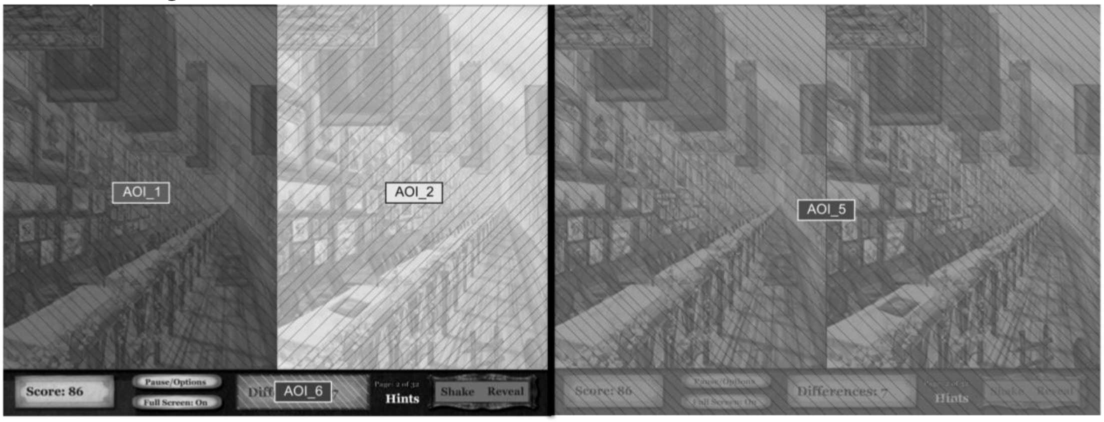


**Figure 2.** *Analyzed Areas of Interest.*

# 3. RESULTS

### *3.1 Total visit duration: The task times*

Total visit duration was taken on the whole picture stimuli (AOI5 see Figure 2.), thus it is representing an aggregated task time passed during each find the difference game level.

**Table 2.** *Mean total visit duration values and standard deviations of phase A and B pictures.*

| Total Visit Duration | Phase A | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Female** | **SD** | **N** | **Male** | **SD** | **N** | **Total A** | **Total A SD** | **Total A N** |
| **Picture 1** | 70,95 | 7,11 | 29 | 77,29 | 3,86 | 14 | 74,12 | 5,49 | 43 |
| **Picture 2** | 76,51 | 8,08 | 29 | 58,98 | 12,40 | 14 | 67,74 | 10,24 | 43 |
| **Picture 3** | 77,92 | 9,60 | 29 | 60,52 | 14,10 | 14 | 69,22 | 11,85 | 43 |
| **Picture 4** | 50,55 | 3,86 | 29 | 43,89 | 16,74 | 14 | 47,22 | 10,30 | 43 |
| **Total** | 68,98 | 7,16 | 29 | 60,17 | 11,78 | 14 | 64,58 | 9,47 | 43 |
| **Total Visit Duration** | **Phase B** | | | | | | | | |
| | **Female** | **SD** | **N** | **Male** | **SD** | **N** | **Total B** | **Total B SD** | **Total B N** |
| **Picture 1** | 52,58 | 4,88 | 29 | 82,52 | 8,51 | 14 | 61,92 | 29,29 | 43 |
| **Picture 2** | 53,07 | 5,16 | 29 | 37,83 | 9,01 | 14 | 45,83 | 26,06 | 43 |
| **Picture 3** | 43,87 | 4,21 | 29 | 64,97 | 7,34 | 14 | 49,70 | 25,08 | 43 |
| **Picture 4** | 43,35 | 3,57 | 29 | 46,54 | 6,23 | 14 | 41,93 | 19,72 | 43 |
| **Total** | 48,22 | 4,46 | 29 | 57,97 | 7,77 | 14 | 49,84 | 25,04 | 43 |
| **Grand Total** | 58,60 | 5,81 | 29 | 59,07 | 9,77 | 14 | 57,21 | 17,25 | 43 |

We used 2x2x4 repeated measures ANOVA to analyze the differences and changes between gender and education groups across the two presentations. Preliminary results suggested that there is significant difference in task times across genders. However, detailed statistical analysis revealed that this phenomenon is caused by individual (not gender dependent) deviations.

Significant within subject differences have been found in total visit duration across stimuli (pictures of tasks, for all 4 tasks) in both presentation rounds. In the first (A) presentation of the task, pictures caused different total visit duration times across all participant groups: a significant within subject main effect of pictures appeared (Greenhouse-Geyser corrected $F(2,171)=3,014$, $p‹0,05$). In the pairwise comparisons we found significant differences between picture 4-1, 4-2, 4-3 (LSD test, $p‹0,05$ for each). As demonstrated in Table 2., level 4 indicated the lowest total visit duration. Thus it can be interpreted as the easiest task, which can be solved fastest among the four levels in the first presentation. There was no significant between subject effect in this round, the difficulty was perceived similarly across gender and education groups.

In the second (B) presentation round of the tasks, stimuli caused different total visit duration times across all participant groups: a significant within subject main effect of pictures appeared (Greenhouse-Geyser corrected $F(2,512)=6,382$, $p‹0,05$). Also a significant picture stimuli and gender interaction raised in the results of the ANOVA: Greenhouse-Geyser corrected $F(2,512)=5,728$, $p‹0,05$. There were differences among total visit duration between the pictures, and also gender specific differences appeared. In the pairwise comparisons of total visit duration, the following significant differences were found between pictures: 1-2 and 1-4 (LSD test, $p‹0,05$ for each) displayed in Table 2. Level 1 indicated longer total visit durations than level 2 and 4. It can be interpreted as level 1 in the second round was the most difficult across stimuli. Taking the two presentations together level 1 appeared the most difficult stimulus, caused the longest total visit duration times, while level 4 appeared to be the easiest one (with the lowest total visit duration times). It is important to note that fatigue cannot be accounted for these results. The structure of the game did not allow participants to speed up the task solving in the second round with clicking on the new level. They had to find all the differences to reach the next level.

The gender interaction of this main effect is presented in Table 2. and on Figure 3: differences in the total visit durations were higher in the male group of participants, level 1 indicated the most time to solve the task (in pairwise comparisons: $F(1)=5{,}55$, $p<0{,}05$). One possible explanation of this phenomenon can be that female participants, when they found out that they have to repeat the task in presentation B, were more approachable while male participants became unobservant and accomplished level 1 slower for the second time. Participants had no previous knowledge about the structure of the experiment. They were asked to repeat the task after they finished the four levels in presentation A.

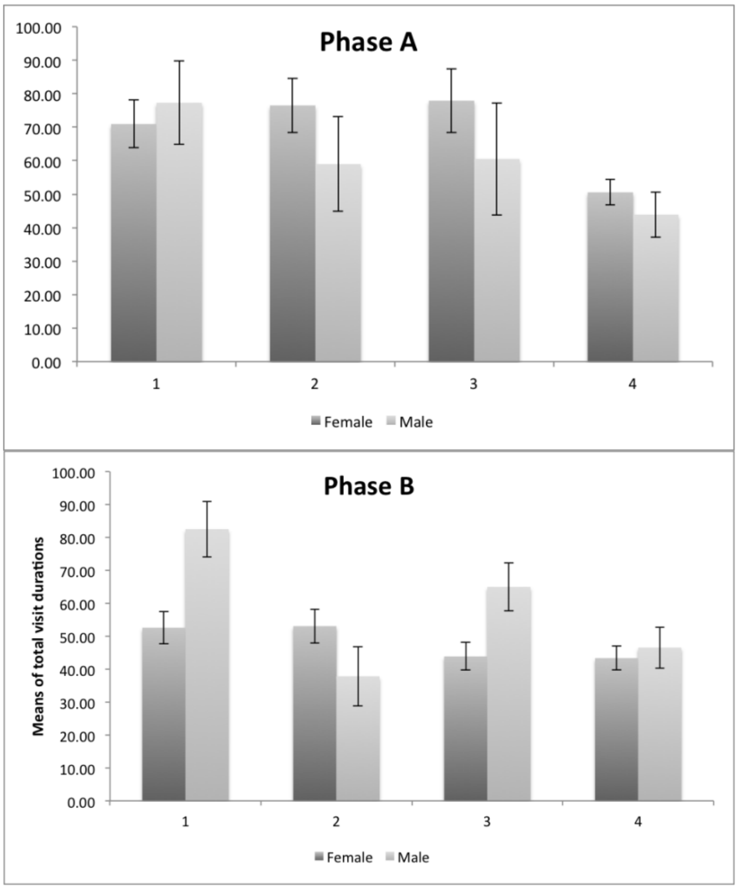


**Figure 3.** *Bar chart of mean total visit duration values compared across gender with standard deviations of phase A and B pictures.*

Comparing the pictures across presentations, searching for the effect of repeated presentation, and a possible learning curve, the same ANOVA revealed significant differences between the total visit durations of level 2 across first and second presentations ($F(1)=8{,}296$, $p<0{,}05$). However, no other significant differences appeared between the two presentations, it had no effect on total visit duration (aggregated task time). In case of picture 2 a decrease of total visit duration appeared from 67,745s (SD=8,127s) to 45,450s (SD=5,192s), participants solved the task faster for the second time, but this learning effect is only limited to picture 2. The cause could be related to the features of this picture: it contained fewer types of differences (length of carpets) and 7 differences out of the total 14 located in the upper section of the picture.

### *3.2 Fixation count and fixation duration*

As demonstrated on Figure 2., besides AOI5 we evaluated the left (AOI1) and right (AOI2) side of each picture. Fixation count and fixation duration data were analyzed by 2x2x4 repeated measure ANOVAs.

For the whole stimuli area (AOI5) there was no significant main effect of the repetition of the task (within subject main effect) neither between subject main effects of gender or education in mean fixation duration. We took the mean fixation duration as a measure of complexity (Rayner, 1998) and there were no differences found: even the repetition of the same pictures cannot make the task easier if the differences moved to other locations. Comparing the two sides of the stimuli picture (AOI1 and AOI2) using fixation count we found no side preferences among participants in general, as well as in education or gender groups. During demographic data sampling, we recorded participants' hand preference and registered all users as right-handed. The repetition phase had no effect on (lack of) side preferences, thus it is stable in time and across groups.

We propose that if there is no instruction about how to find the difference (on which side to find it), there is no coherent preference of sides in terms of fixation count and duration. Due to the construction of the game differences could be eliminated on both sides.

*3.3 The effect of feedback*

The task picture screen contained a feedback panel that showed the participant the number of differences left to finish the task. The fixation measurements (count and duration) on this field (AOI6) inform us about the usage of feedback during the tasks.

A 2x2x4 repeated measures ANOVAs were used to analyze the fixation count and fixation duration. The analysis revealed that gender and education had no effect on these metrics. In order to make our analysis more accurate we applied one-way ANOVA to test the effect of presentation rounds. The fixation data on the feedback area was relatively sparse for several participants (they did not fixated on it at all), thus the group size varied across analyses due to these missing values. There was an apparent decrease in fixation count between the first and second (A-B) presentations, but it only turned to be significant for picture 1 (on 46 participants, Robust Welch test, $D(1)= 7{,}94$, $p<0{,}05$, see Figure 4). There was also a marginally significant decrease in fixation count for picture 2 (on 42 participants, Robust Welch test, $D(1)=3{,}98$, $p=0{,}059$). This can be interpreted as a tendency towards learning of completion the task with less usage of feedback for the second time. Participants learned the possible amount of differences on a stimulus.

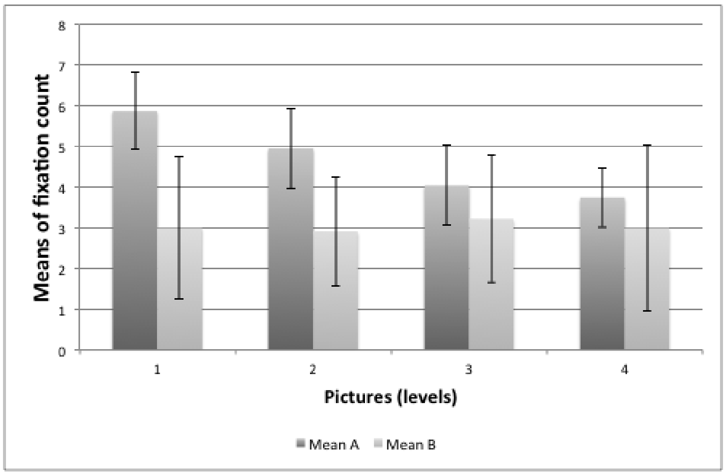


**Figure 4.** *Bar chart of mean fixation count values compared across the two presentation rounds with standard deviations.*

In case of the fixation duration we applied the same one-way ANOVA as described in the case of fixation count. A significant decrease appeared between the first and second (A-B) task phases for picture 1 and picture 2 (for picture 1: 46 participants, Robust Welch test, $D(1)=7{,}03$, $p<0{,}05$, for picture 2: 42 participants, Robust Welch test, $D(1)=4{,}49$, $p<0{,}05$, see also Figure 5). It is a congruent tendency with the fixation count.

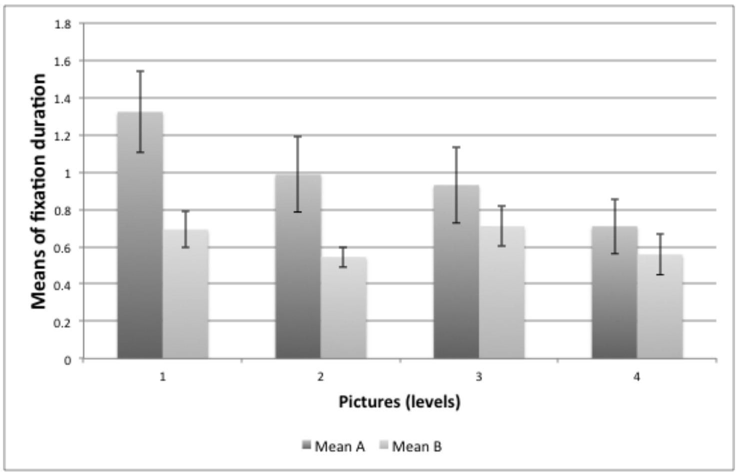


**Figure 5.** *Bar chart of mean fixation duration values compared across the two presentation phases with standard deviations.*

These results show a decrease in the usage of feedback, irrespectively to gender and education groups on every measure. Results suggest that the easiest task, in this case level 4, requires least feedback in terms of fixation duration and fixation count. However, this result has to be clarified on bigger sample, which compensates missing values for feedback area.

## 4. CONCLUSIONS

### *4.1 Discussion*

Summarizing our results, we can assume from total visit durations, as the measure of task time, that there is no general learning effect from the first to the second presentation of the task. The relocation of differences on even familiar pictures is able to suppress the learning effect (if it exists). In our study we defined familiarity as the perceived knowledge on the major dimensions of the stimuli (Park, 1981). This familiarity seems to have a weak effect on task solving times, if the small differences are relocated. Taking into consideration the spatial nature of mental representations contrary to the abstract character of them, these results are parallel with the classical evidences of pictorial-spatial mental representations (Shepard, 1978; Kosslyn, 1978). If participants could have applied their spatial (pictorial) representations of the familiar picture, they would have completed the task faster (with lower total visit duration) by searching for the differences in the known places. However, the differences were relocated, new differences appeared on a familiar picture, and the mental representation has not been useful: the previously known and represented places contained no differences, thus the participant had to search though the whole stimulus again.

Results connected to measurements on the feedback panel confirm that less difficult task reduces the need of feedback (e.g. easier task requires less feedback).

The fragmented gender and education related results indicate that deviations in fixation count and fixation duration cannot be directly explained by these independent variables. Possible future extension of current study requires the application of new grouping variables such as cognitive style (Embedded Figure Test) and the detailed analysis of scanpaths.

**Acknowledgements:** The authors would like to thank to Károly Hercegfi, PhD his remarkable proposals on research methodology. We also would like to thank Szilvia Kucsora's (MSc student) contribution on the illustration and Máté Köles's (MA student) comments on the results. The authors are also grateful for the efforts of all volunteer participants.

## 5. REFERENCES

A Duchowski (2003), *Eye Tracking Methodology: Theory and Practice*, Springer, Berlin.

U Engelke, H-J Zepernick, A Maeder (2010), Visual Fixation Patterns in Subjective Quality Assessment:

The Relative Impact of Image Content and Structural Distortions, *International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2010),* IEEE.

P M Fitts, R E Jones, J L (1950), Eye Movement of Aircraft Pilots during Instrument-Landing Approaches, *Aeronautical Engineering Review, 9,* pp. 24-29.

J H Goldberg, P K Xerxes (1999), Computer interface evaluation using eye movements: methods and constructs, *International Journal of Industrial Ergonomics, 24*, pp. 631-645.

J H Goldberg, M J Stimson, M Lewenstein, N Scott, A M Wichansky (2002), Eye Tracking in Web Search Tasks: Design Implications. *Proceedings of the 2002 symposium on Eye tracking research & applications,* pp. 51-58, ACM.

T Gowases (2007), *Gaze vs. Mouse: An evaluation of user experience and planning in problem solving games*, University of Joensuu, Department of Computer Science.

A Guyton (1977), *Basic human physiology: Normal function and mechanisms of disease,* Saunders, Philadelphia.

K Hercegfi (2011), Event Related Assessment of Hypermedia Based E-Learning Materials with an HRV-Based Method That Considers Individual Differences in Users, *International Journal of Occupational Safety and Ergonomics,* **2**, *17, pp.* 119-127.

P Isokoski, B Martin (2006), Eye Tracker Input in First Person Shooter Games, *Communication by Gaze Interaction,* Turin.

A M Kosslyn, T M Ball, B J Reiser (1978), Visual images preserve metric spatial information: Evidence from studies of image scanning, *Journal of Experimental Psychology: Human Perception and Performance,* ***1,*** *4*, pp. 47-60.

F M Land (2006), Eye movements and the control of actions in everyday life, *Progress in Retinal and Eye Research, 25*, pp. 296-324.

G R Loftus, N H Mackworth (1978), Cognitive determinants of fixation location during picture viewing, *Journal of Experimental Psychology: Human Perception and Performance,* ***4****, 4* , pp. 565-572.

J Meyers-Levy (1989), Gender Differences in Information Processing: A Selectivity Interpretation, In *Cognitive and Affective Responses to Advertising* (P Cafferata, A Tybout), Lexington, MA.

M Nakayama, K Takahashi, Y Shimizu (2002), The Act of Task Difficulty and Eye-movement Frequency for the 'Oculo-motor indices', *Eye Tracking Research & Applications Symposium,* pp. 43-51, ACM.

D Noton, L Stark, L (1971), Scanpaths in Saccadic Eye Movements While Viewing and Recognizing Patterns, *Vision Research, 11*, pp. 929-942.

B Pan, H A Hembrooke, G K Gay, L A Granka, M K Feusner, J K Newman (2004), The Determinants of Web Page Viewing Behavior: An Eye-Tracking Study, *Proceedings of the 2004 symposium on Eye tracking research & applications,* pp. 147-154, ACM.

W Park, V L Parker (1981), Familiarity and Its Impact on Consumer Decision Biases and Heuristics, *Journal of Consumer Research,* **2,** *8,* pp. 223-231.

J B Pelz, R Canosa, J Babcock, J, (2000), Extended Tasks Elicit Complex Eye Movement Patterns, *Eye Tracking Research & Applications Symposium,* pp. 37-43, ACM.

K Rayner (1998), Eye Movements and Information Processing: 20 Years of Research, *Psychological Bulletin,* **3,** *124*, pp. 372-422.

R N Shepard (1978), The mental image, *American Psychologist,* **2**, *33*, pp. 125-137.

R M Steinman (2003), Gaze control under natural conditions. In *The Visual Neurosciences* (L M Chalupa, J S Werner), MIT Press, MA, pp. 1339-1356.

R H Thaler, C R Sunstein (2008), *Improving decisions about health, wealth and happiness,* Yale University Press, New Haven, CT.

Tobii product description: http://www.tobii.com/en/eye-tracking-research/global/products/hardware/tobii t60t120-eye-tracker/