HOW TASK DIFFICULTY INFLUENCES EYE MOVEMENTS WHEN EXPLORING NATURAL SCENE IMAGES

Hélène Devillez, Anne Guérin-Dugué and Nathalie Guyader

GIPSA-lab, Grenoble Images Parole Signal Automatique
Grenoble-Alpes University, Grenoble, France

ABSTRACT

Since Yarbus [1], it is well known that task greatly influences the visual exploration of a scene. In this paper, we quantify the influence of three different cognitive tasks on the visual exploration of various natural scene images. Eye movements made to solve the three tasks were compared to the ones recorded during a free-exploration condition. Eye movement parameters and eye fixation positions were compared during the time course of the exploration according to fixation/saccade rank. Tasks were chosen because they implied different visual processing and might be classified according to their difficulty. We found that task difficulty impacts eye-movement characteristics at the very beginning of exploration. The variability between eye fixations of observers is also impacted. Finally, eye-movement characteristics might reveal the tasks solving.

***Index Terms*—** Eye movement, Visual Attention, Tasks, Saliency

1. INTRODUCTION

To perceive our visual environment we sequentially fixate particular areas. These fixated areas are seen with a high visual acuity allowing for a detailed analysis during fixations. Fixations are characterized by their spatial location (where they take place) and their duration (how long they last) in the visual stimulus. Many studies have been working on finding the factors that influence *fixation positions*. Regions that attract observers' gaze are called salient and were found to have higher spatial contrast or higher spatial frequencies than random locations [2, 3, 4]. These statistical properties are in agreement with the Treisman's Feature Theory [5]. Based on this theory and the physiological properties of visual cortical cells, several saliency models have been proposed [6, 7, 8]. These models, based on visual feature analysis, are efficient to predict fixation positions during the free exploration of static and dynamic scenes, at least at the beginning of the exploration. Since Yarbus [1], it is now well known eye fixations are guided not only by the visual features but also by the task [9, 10, 11]. However, very few visual attention models incorporate the task influence to predict eye fixations; Torralba and colleagues proposed a "Contextual Guidance Model" that integrates image saliency and scene priors [6] to predict eye fixations of observers searching for a target. On parallel, eye tracking experiments have been proposed to understand which type of factors influences fixation durations (for reviews see [11, 12, 13]), while this aspect has been almost ignored in visual attention modelling. Longer fixations are observed when the luminance contrast of the fixated region is reduced or contained noise [11]. Like for fixation positions, not only visual properties of the signal but also task demands influence fixation durations. Castelhano and Rayner [13] reported that different tasks greatly modify fixation durations with shorter fixations for visual search task compared to free scene viewing. More recently, Mills and colleagues [14] used four tasks (memorization, pleasantness, visual search of a target letter added in the scene and free-viewing) to show that task affects fixation duration over the time course of viewing but not saccade amplitude. Finally, it have been shown that the level of difficulty of a visual search task induce different visual strategy [15, 16].

The aim of the present paper is to also quantify the influence of different cognitive tasks on eye movement parameters but also on fixation positions. We also want to go further analyzing more specifically how the difficulty of the task modifies the eye movements. We compared three cognitive tasks to a free exploration condition. The tasks differed in their goal as well as in their difficulty. Eye-movement parameters (fixation duration, saccade duration and saccade amplitude) were analyzed across the four conditions and along the exploration by fixation/saccade ranks, allowing inference on the visual attention deployment. Moreover, the variation of fixation positions across participants along the exploration gave indications on how the spatial exploration was coherent or not.

2. EYE TRACKING EXPERIMENT

In this study, observers were eye-tracked when exploring natural scene images without any particular task i.e. free exploration (FE), or to solve a specific task: (1) to report the scene category (indoor or outdoor categorization: CAT), (2) to report if a target object was on the right or

This work was supported by a grant from the ANR Gaze-EEG "Joint synchronous EEG signal and eye tracking processing for spatio-temporal analysis and modeling of neural activities".

978-0-9928626-3-3/15/$31.00

left side of another target object (spatial organization: SO) and (3) to report if a target object was or not present in the scene (visual search: VS task). In our experimental setup, scenes were randomly attributed to one task with the constraint to have: for the CAT task, 30 indoor and 30 outdoor scenes, for the SO task, 30 scenes with the answer "right" and 30 scenes with the answer "left" and for the VS task, 30 scenes with the object present and 30 scenes without the object.

Those tasks were chosen because they imply different levels of visual processing [15]: the category of a scene is extracted very quickly, even without eye movement, whereas to find an object takes more time and needs an active visual exploration. In our experimental setup, observers were asked to explore the scene during 4 sec. This allows us to measure the task influence over all the time course of viewing.

CAT task was classified as the easiest task due to the categories that we chose (outdoor and indoor) and the fact that it might be solved at a glance. The SO task was more difficult than CAT task because it needs eye movements to identify and find the two target objects of the question. However, the two target objects have large size (61.03±110.17°) which makes the task very feasible. Finally, the most difficult task was VS task for which the target object was smaller (6.82±12.46°). Furthermore, for VS task, the answer was not certain because in half of the trials the answer was "no", i.e., the target object was not present. According to these considerations we used three cognitive tasks with different "levels of difficulty" (with the difficulty defined by the time needed to find the answer and the size of the target objects). Each participant viewed scenes under the three cognitive tasks and also in a free exploration condition (in total, four conditions), which allows us to analyze the task effect for the same observers.

2.1. Participants

Thirty-nine healthy adults participated in the experiment (22 female; age range: 20 – 36; M=24.69; SD=3.49). They had a normal or corrected-to-normal vision and gave their written and informed consents.

2.2. Stimuli

240 color images representing 120 indoor and 120 outdoor scenes were used (Fig. 1a). Images with various contents were selected from several sources: the Oxford Buildings Dataset[1], different websites (not copyrighted) and an in-house database. Images have a resolution of 768 × 1024 pixels, subtended 30 × 40 degrees of visual angle.

2.3. Protocol

The experiment was run using SoftEye software [16]. Each participant viewed 4 blocks of 60 scenes in the four different conditions in a random order: FE, CAT, SO and VS (Fig. 1a). Each scene was seen only once by each participants.

A trial is a succession of 3 to 5 displays, depending on the condition (Fig. 1b). Trials started with a question. There was no question for FE and the question was not displayed during CAT because the question was always the same. The question was displayed until the observer pressed a button or up to 5 sec. Then, a white central fixation cross was displayed for 800 to 1200 msec. If the gaze was stabilized on the cross, the scene was displayed for 4 sec. Finally, a screen showed again the question and the possible answers (for FE this screen was not presented). Observers had to press the right or left button of the mouse to choose the appropriate answer. If they did not know the answer, they were asked to not answer and the display disappeared after 5 sec. A grey screen appeared for 1 sec between two successive trials. For eye tracking purpose, a 9-point calibration routine was carried out at the beginning of each condition and was repeated every 20 trials or when the drift correction, performed every 10 trials, reported an error above 0.5°.

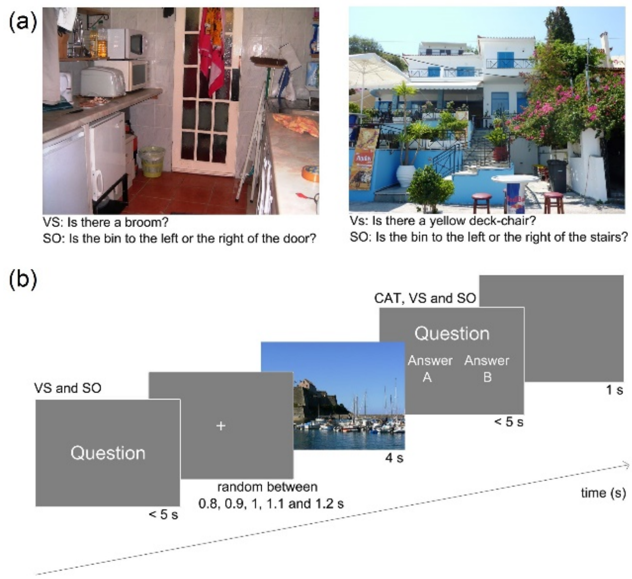

Fig. 1. (a) One indoor scene and one outdoor scene with corresponding questions for VS and SO tasks (translated from French to English). (b) The succession of displays during one trial.

Stimuli were displayed onto a 20-inch ViewSonic CRT monitor located 57 cm in front of the participants, with a resolution of 768 × 1024 pixels and a 75 Hz refresh a rate. Eye movements were recorded using the Eyelink 1000 (SR Research), in a binocular pupil-corneal-reflection mode, with a 1000 Hz sampling rate. Head was stabilized using a chin rest. Only the dominant eye of each participant was analyzed.

3. ANALYSIS AND RESULTS

First, to validate the chosen tasks, we computed the correct answer rate for each task to assess the task difficulty. The task effect was analyzed, globally and along the scene exploration, using various ocular characteristics extracted from eye movements for each observer (fixation duration, saccade amplitude and saccade duration) and compared to the free-exploration condition. Finally, the task effect was

[1] www.robots.ox.ac.uk/~vgg/data/oxbuildings

quantified analyzing the positions of eye fixations of all observers (dispersion).

3.1. Behavioral data

Correct answer rates were computed for each participant for CAT, VS and SO tasks (Table 1). The correct answer rate was significantly higher for CAT (CI: [98.8; 99.6]) compared to SO (CI: [85; 89.6]) and VS (CI: [82; 87]). This result partly confirms that SO and VS tasks was more difficult than CAT task.

3.2. Ocular characteristics

Saccades were automatically detected by the Eyelink software based on three thresholds: velocity (30 degrees/sec), acceleration (8000 degrees/s²) and saccadic motion (0.15 degree). Only eye positions recorded during the scene presentation (4 sec) were analyzed. Note that the first saccade after the scene onset was called *saccade 1* and was followed by *fixation 1*. Parameters were statistically analyzed using a repeated-measure ANOVA with *Condition* (FE, CAT, SO and VS) as within-subjects factors. Multiple comparisons were assessed with Bonferroni post-hoc tests.

3.2.1. First saccade latency

The first saccade latency after the scene onset was computed (i.e. the delay between the scene onset and the beginning of the first saccade) (Table 1). A main effect of the *Condition* was observed ($F(3,114)=13.05; p<.0001$). VS and SO significantly differed from FE (both $p<.01$) having both shorter latency.

3.2.2. Global effect

Distributions of fixation duration, saccade amplitude and saccade duration were plotted according to the four conditions (Fig. 2; all fixations and saccades were extracted independently of the participant and the scene). Note that distributions for the four conditions are very similar.

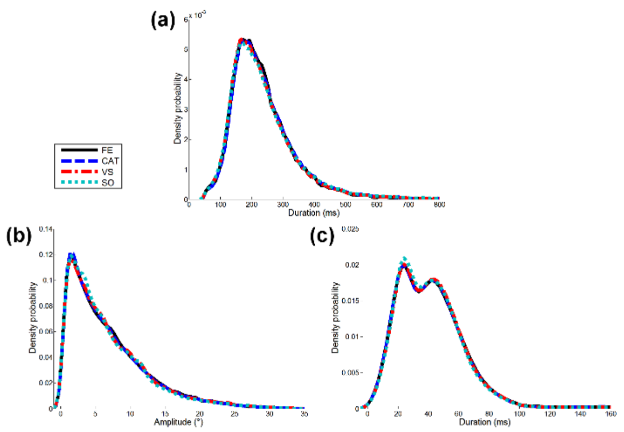

Fig. 2. Distributions of eye movement parameters for the four conditions: (a) fixation duration, (b) saccade amplitude and (c) saccade duration.

The shapes of theses distributions have been classically reported in studies analyzing eye movement recorded when viewing natural scene images [17, 18]. For each condition and each participant, median values of these eye movement parameters were computed; median values were chosen due to the shape of distributions. Table 1 summarizes the mean values over all observers for the different eye movement parameters in function of the condition.

A main effect of the *Condition* was observed for the fixation duration ($F(3,114)=25.9; p<.0001$), the saccade amplitude ($F(3,114)=18.6; p<.0001$) and the saccade duration ($F(3,114)=8.3; p<.0001$). The mean fixation duration was smaller for VS compared to FE (the two other tasks did not differ from FE). The mean saccade amplitude was smaller for SO compared to FE (the two other tasks did not differ from FE). Finally, the saccade duration was also smaller for SO compared to FE (the two other tasks did not differ from FE).

Conditions	FE	CAT	VS	SO
Correct answer (%)	-	99.25 *(0.20)*	84.50 *(0.20)*	87.31 *(0.04)*
First saccade latency (msec)	266.60 *(7.16)*	273.87 *(8.95)*	231.85 *(5.55)*	240.83 *(12.50)*
Fixation duration (msec)	216.05 *(3.46)*	224.42 *(4.21)*	**206.38** ***(3.19)***	223.63 *(3.66)*
Saccade amplitude (°)	5.39 *(0.14)*	5.20 *(0.17)*	5.25 *(0.09)*	**4.67** ***(0.08)***
Saccade duration (msec)	39.51 *(0.70)*	38.84 *(0.79)*	39.68 *(0.56)*	**37.50** ***(0.48)***

Table 1. Mean (and standard error) parameters for the four conditions. Differences with FE were highlighted and underlined.

3.2.3. Task effect across the exploration

Parameters globally analyzed in the Table 1 were also analyzed along the exploration, i.e. according to fixation/saccade rank (Fig. 3; only the first 8 fixations were plotted over 12.7±1.6 fixations made on average).We were only interested in testing the task effect at each fixation/saccade rank and so the evolution of parameters along the exploration was not discussed.

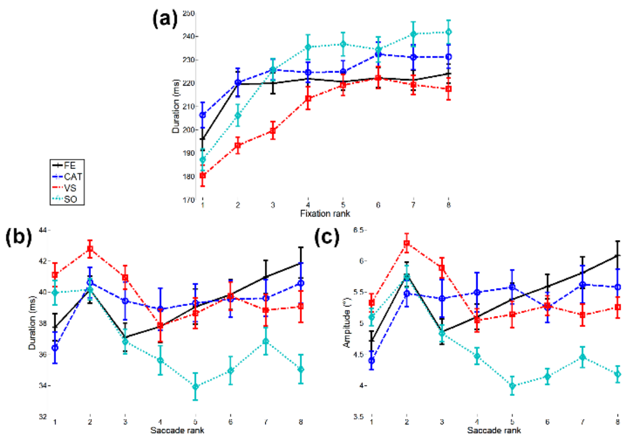

Fig. 3. Mean (and standard error) parameters for the four conditions according to fixation rank: (a) fixation duration, (b) saccade amplitude and (c) saccade duration.

For all ANOVAs (per fixation/saccade rank for the three parameters: fixation duration, saccade amplitude and saccade duration) a significant effect of the *Condition* was observed. The fixation 1 was shorter for VS than FE ($p<.001$), and for VS and SO compared to CAT ($p<.001$). Fixation 2 is shorter for VS and SO compared to FE and CAT (all $p<.01$). Fixation 3 was shorter for VS compared to the three other conditions (all $p<.001$). After fixation 4, no significant difference between VS, FE and CAT was observed whereas SO showed longer fixations than for the three other conditions ($p>.01$).

Results are quite similar for the saccade amplitude and the saccade duration; this might be explained by the relation between these two parameters (also called the main sequence) [19]. A significant difference was observed for SO and VS when comparing to CAT and FE for the saccade 1; SO and VS implied larger saccade 1 with longer duration than FE and CAT. From saccade 4, a difference was observed only for SO; this difference might be explained by the fact that when observers found the two target objects of the question they made saccades from one object to the other more or less until the end of the scene. Indeed, the distance between the two targets was on average 8.34±5.86° (when considering the center of the objects but this distance could be smaller).

3.3. Eye fixation positions

The aim of this second part is to analyze the task effect on the regions gazed at by observers i.e. the fixation positions. The variability of eye positions between observers was analyzed using a metric called "dispersion". Analyses were run using scene as observation.

For each scene and for each fixation rank, the dispersion D is defined between the fixation positions of n observers as shown in (1):

$$D = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \qquad (1)$$

with (x_i, y_i) the fixation coordinates of the observer i.

The dispersion corresponds to the mean Euclidean distance between the eye positions of different observers. Small dispersion values small variability between fixation positions of observers. The dispersion was computed according to fixation rank for each scene and then averaged over all scenes (60 scenes per condition) (Fig. 4). The global evolution of the dispersion along the exploration was not discussed but was already reported in others studies [7, 20]. Only the task effect was analyzed at each fixation rank. At the very beginning of exploration, for fixation 1, CAT and SO differ from FE with smaller dispersions.

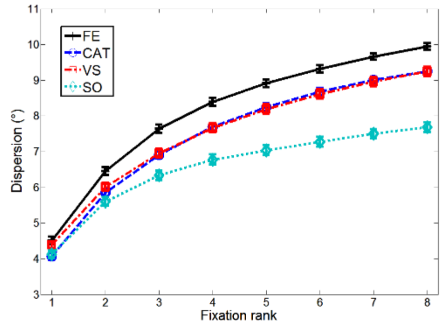

Fig. 4. Evolution of the mean (and standard error) dispersion for the four conditions according to the fixation rank.

No significant difference was observed between VS and FE. From fixation 2, the three cognitive tasks, CAT, SO and VS differ from FE with smaller dispersion. This means that eye positions are less dispersed for the three cognitive tasks compared to FE. Moreover, during the SO task the dispersion is the smallest form fixation 3 to fixation 8. This might be explained by the fact that when observer found the two target objects he made saccade from one object to another. This already explains the results obtained concerning the saccade amplitude and duration that were smaller for SO than VS, CAT and FE.

4. DISCUSSION

All these results show that the task greatly influences eye movement parameters and also the variability between the fixation positions of observers. The first results concerns the fact that, for SO and VS tasks, the first saccade latency (the latency to start the visual exploration) was shorter. Moreover, the first fixations (fixation 1 and 2) were also shorter for SO and VS compared to FE and CAT. This difference is still observed for the fixation 3 for VS task but disappears for the SO task. These two tasks also modify the saccade amplitude and duration with a saccade 1 larger for SO and VS and still a saccade 2 larger for VS. These results showed that the top-down process, measured here with the cognitive task given to participants, influenced eye movement parameters from the beginning of the exploration. The task difficulty level can be indexed on the modification of eye movement parameters compared to a free exploration condition. We observed that the two more difficult tasks (SO and VS) showed the most important influence on fixations and saccades. Fixations are shorter for VS compared to SO task. More the task is difficult more it decreases the fixation duration at least at the very beginning of the exploration.

The results for VS can be interpreted in regards with the solving of the task. Differences were observed for fixation and saccade parameters for fixations/saccades 1 to 3. After fixation 4, no more significant difference was observed compared to FE. This might reflect the fact that participants solved to task and continue the exploration of the scene due to our experimental setup. This result was verified by analyzing fixations on the target: on average, the target was fixated, and the task potentially solved on average after 4.53±0.13 fixations. For SO, the same explanation might be proposed, except for the end of exploration where saccade parameters might be explained by the fact that observers made saccades between the two targets. In this condition, the first target was fixated after 2.37±0.05 fixations on average and the second target at least one fixation after; hence, the task might be solved after 4 fixations. These results shown that eye movement parameters might be used to determine when observer solve the task.

This eye-tracking study also provides useful data to improve existing models of visual attention. Moreover an effort has to be done in modelling to include the duration of fixations which was shown to greatly modified eye movement parameters. In fact many models provide a saliency map for a scene without giving a dynamic scan-path, whereas in many applications it might be important to have access to the temporality of the saliency.

REFERENCES

[1] A. L. Yarbus, Eye-movements and vision, New-York: Plenum Press, 1967.

[2] S. K. Mannan, K. H. Ruddock and D. S. Wooding, "Fixation sequences made during visual examination of briefly presented 2D images," *Spatial Vision,* vol. 11, no. 2, pp. 157-178, 1997.

[3] P. Reinagel and A. M. Zador, "Natural scene statistics at the centre of gaze," *Network: Computation in Neural Systems,* vol. 10, pp. 341-350, 1999.

[4] G. Krieger, I. Rentschler, G. Hauske, K. Schill and C. Zetzsche, "Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics," *Spatial Vision,* vol. 13, no. 2,3, pp. 201-214, 2000.

[5] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology,* vol. 12, pp. 97 - 136, 1980.

[6] A. Torralba, A. Oliva, M. Castelhano and J. Henderson, "Contextual guidance of eye movements and attention in real-world scenes : The role of global features on object search," *Psychological Review,* vol. 113, no. 4, pp. 766-786, 2006.

[7] S. Marat, T. H. Phuoc, L. Granjon, N. Guyader, D. Pellerin and A. Guérin-Dugué, "Modelling spatio-temporal saliency to predict gaze direction for short videos," *International journal of computer vision,* vol. 82, no. 3, pp. 231-243, 2009.

[8] O. Le Meur and J. C. Chevet, "Relevance of a feed-forward model of visual attention for goal-oriented and free-viewing tasks," *Image Processing, IEEE Transactions,* vol. 19, no. 11, pp. 2801-2813, 2010.

[9] J. M. Henderson, P. A. Weeks Jr and A. Hollingworth, "The effects of semantic consistency on eye movements during complex scene viewing," *Journal of experimental psychology: Human perception and performance,* vol. 25, no. 1, p. 210, 1999.

[10] M. S. Castelhano, M. L. Mack and J. M. Henderson, "Viewing task influences eye movement control during active scene perception," *Journal of Vision,* vol. 9, no. 3, pp. 1 - 15, 2009.

[11] J. Henderson, "Human gaze control during real-world scene perception," *TRENDS in Cognitive Sciences,* vol. 7, no. 11, pp. 498-504, 2003.

[12] J. M. Henderson and A. Hollingworth, "Eye movements during scene viewing: An overview," *Eye guidance in reading and scene perception,* vol. 11, pp. 269-293, 1998.

[13] M. S. Castelhano and K. Rayner, "Eye movements during reading, visual search, and scene perception: An overview," *Cognitive and cultural influences on eye movements,* pp. 175-195, 2008.

[14] M. Mills, A. Hollingworth, S. Van der Stigchel and M. D. Dodd, "Examining the influence of task set on eye movements and fixations," *Journal of Vision,* vol. 11, no. 8, pp. 1-15, 2011.

[15] M. Pomplun, T. W. Garaas and M. Carrasco, "The effects of task difficulty on visual search strategy in virtual 3D displays," *Journal of vision,* vol. 13, no. 3, p. 24, 2013.

[16] A. H. Young and J. Hulleman, "Eye movements reveal how task difficulty moulds visual search," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 39, no. 1, p. 168, 2013.

[17] B. Tatler, I. Gilchrist and J. Rusted, "The time course of abstract visual representation," *Perception,* vol. 32, pp. 579-592, 2003.

[18] G. Ionescu, N. Guyader and A. Guérin-Dugué, "SoftEye software.," *IDDN. FR,* vol. 1, no. 000, 2009.

[19] B. Tatler and B. Vincent, "Systematic tendencies in scene viewing," *Journal of Eye Movement Research,* vol. 2, no. 2, pp. 1-18, 2008.

[20] T. Ho-Phuoc, N. Guyader, F. Landragin and A. Guérin-Dugué, "When viewing natural scenes, do abnormal colors impact on spatial or temporal parameters of eye movements?," *Journal of Vision,* vol. 12, no. 2, pp. 1 - 13, 2012.

[21] A. T. Bahill, D. Adler and L. Stark, "Most naturally occurring human saccades have magnitudes of 15 degrees or less," *Investigate Ophthalmology & Vison,* vol. 14, no. 6, pp. 468-469, 1975.

[22] A. Coutrot and N. Guyader, "How saliency, faces, and sound influence gaze in dynamic social scenes," *Journal of vision,* vol. 14, no. 8, p. 5, 2014.