

Eye and Mouse Coordination During Task: From Behaviour to Prediction

Alexandre Milisavljevic^{1,2,3}, Kevin Hamard³, Coralie Petermann³, Bernard Gosselin¹,
Karine Doré-Mazars² and Matei Mancias¹

¹*Numediart institute, University of Mons, Mons, Belgium*

²*Psychology institute, VAC EA7326 team, Paris Descartes University, Paris, France*

³*Research and Development department, Sublime Skinz, Paris, France*

Keywords: Behaviour, Visual Attention, Webpages, Mouse-tracking.

Abstract: The study of web users' behaviour is of crucial importance for understanding people reaction when browsing websites. Eye-tracking is a precise tool for this purpose, but it is hard to scale up when trying to apply it to a wide range of situations and websites. On the other hand, mouse-tracking fulfills these requirements. Unfortunately, mouse data provides a limited approximation of the eye position as it was shown in the literature. In this paper, we investigated the relationship between mouse and eye behaviour on several kind of websites with three different tasks to create models based on these behaviours. Our findings were that 1) saliency Pearson's correlation is not suitable to analyse eye and mouse coordination, 2) this coordination is altered according to the task, 3) scroll speed directly influence where the eyes are during the scroll, 4) amplitude vary according to eyes position before the scroll and 5) by using the X axis variations it is possible to find the moments where it is easier to model eyes location from mouse location.

1 INTRODUCTION

Understanding why a user visits a webpage has been a central question since the beginning of the twenty first century. To answer this question, Eye-tracking has been used as a precise tool to estimate intention impact on user's gaze. However, these kinds of studies are hard to scale up and apply it to a wide user panel is difficult. That is why mouse-tracking emerged as an efficient proxy to determine user's attention. Since then, correlation between mouse movements and eye movements has been found (Mueller and Lockerd, 2001; Chen, 2001; Rodden and Fu, 2007; Rodden et al., 2008; Cooke, 2006; Guo and Agichtein, 2010; Huang and White, 2012; Navalpakkam et al., 2013) and modelling attempts followed (Guo and Agichtein, 2010; Huang and White, 2012; Navalpakkam et al., 2013; Boi et al., 2016).

Nevertheless, a majority of these studies focused on SERP (SEarch Result Pages) from search engines putting aside the rest of the web and tasks. In addition, eye-mouse and eye-task relationships have been studied separately (Yarbus, 1967; Castelhana et al., 2009; Mills et al., 2011) but rarely together. That is why, the goal of this study was to explore the eye-mouse-task relationship in a more diversified environment.

Chen (2001) was the first to show that areas visited by the mouse were also visited by the eye in free-viewing condition. Rodden and Fu (2007) also showed that regions visited by the mouse were also visited by the eye but they were the first to highlight the better correspondence on Y axis between mouse and eye. Unlike previous work, they set-up an experiment with pre-defined search queries on a search engine. Guo and Agichtein (2010) confirmed Rodden and Fu (2007) results about more accurate correlation on Y axis. Their main contribution was the first attempt to automatically infer the user's eye position using mouse movements. They also suggested the presence of images did not have a significant effect on eye-mouse coordination. Huang and White (2012) presented that amount of time spent on a search web page by a participant can affect where they were pointing and looking and then used this finding to enhance their algorithm. They showed that gaze-cursor alignment was distinct for each participant but did not highlight significant difference among women and men. Navalpakkam et al. (2013) updated previous work by investigating more recent SERP which now includes images and more complex content. They showed that this content induced different behaviour. Then they proposed a non-linear model outperforming state-of-the-

Table 1: Generic tasks used in the study.

Category	Description
Free viewing	Browse the website by visiting at least two other pages
	Browse two articles of your choice
Target finding	Browse the following pages: calendar, team and news
	Buy the specific given item
Text reading	Read the two first paragraphs

art models because of its non-linearity.

Our exploratory study aimed to investigate effect(s) of user's goal on eye and mouse coordination in ecological conditions (different categories of web sites, no scroll limitations). Our hypothesis was the following: there is a direct link between the eye, the mouse movements and the task which fluctuates according to the task. Thus, we could enhance the precision of the current models.

This paper was structured as follows: experiment set-up was described in section 2 followed by the results of the static and dynamic analyses in section 3. Finally we discussed and concluded the paper in section 4.

2 METHOD

We recruited five participants with normal or corrected-to-normal vision (4 males and 1 female) aged between 24 and 25 years from the local signal processing department. All participants were right-handed and fluent with computer operations. They were tested on 10 various websites including blogs, e-commerce platforms, etc. From the calibration phase at the beginning of the study to the end, the whole process took about 20 minutes per person.

2.1 Tasks

We used a set of five tasks distributed in three classes as presented in Table 1: free viewing, target finding and text reading. The number of pages that could be visited during Free viewing tasks was limited to two but was not limited in time and participants had full scroll possibilities. The reading task was specific enough to prevent any free interpretation in order to simulate participants' willingness to read a specific paragraph. Finally, we chose two types of target finding tasks: one in which participants were instructed to find and buy an item, the second in which they had to find a given page.

2.2 Set-up

To record eye movements, we used a FaceLAB 5 eye-tracker at 60Hz without head constraint on a 17-inch screen set to a resolution of 1920X1080. Instructions and websites were displayed in Google Chrome maximized with a resolution of 1920X955.

To record Mouse movements we developed a plug-in using WebExtensions¹ standard. It took the form of an ON/OFF button on the browser top bar and has only been used by the operator. The extension monitored the following metrics: time-stamp, event type (click, movement or scroll), mouse's coordinates, offset induced by the scroll, URL and screen size. The plug-in developed in Javascript was uploading all mentioned metrics on the fly or every 50-60ms to a NodeJS² server via a socket connection. The same server inserted in real time the data in a MySQL database without further processing. The server also kept track of the page to deliver to the participant.

2.3 Procedure

Participants started on a homepage describing the context of the study. To visit the next planned page by the study they had to click on a Javascript bookmark situated in the browser's bookmark top bar. The first click on it led them to the first task instruction. All tasks were stored in HTML format locally. After reading the instruction, participants could once again click on the bookmark and begin the task. When the task was completed participants had to click again on the bookmark to read the next instruction and so on. At the end of the study, participants were asked to answer to a survey about their knowledge of the websites.

3 RESULTS

We ran three sets of analyses in order to highlight coordination between eye and mouse movements. First, we used 2D saliency metric *PCC* (Pearson's Correlation Coefficient) to check consistency between overall eye and mouse movements. Then we repeated the same analysis between participants' eyes movements. Second, we applied literature's temporal and distance estimation to our task-related context to bring out tasks' influence on eye and mouse coordination. Third, we analysed the participants' behaviour while scrolling because - at our knowledge - it

¹<https://developer.mozilla.org/Add-ons/WebExtensions>

²<https://nodejs.org/en/>

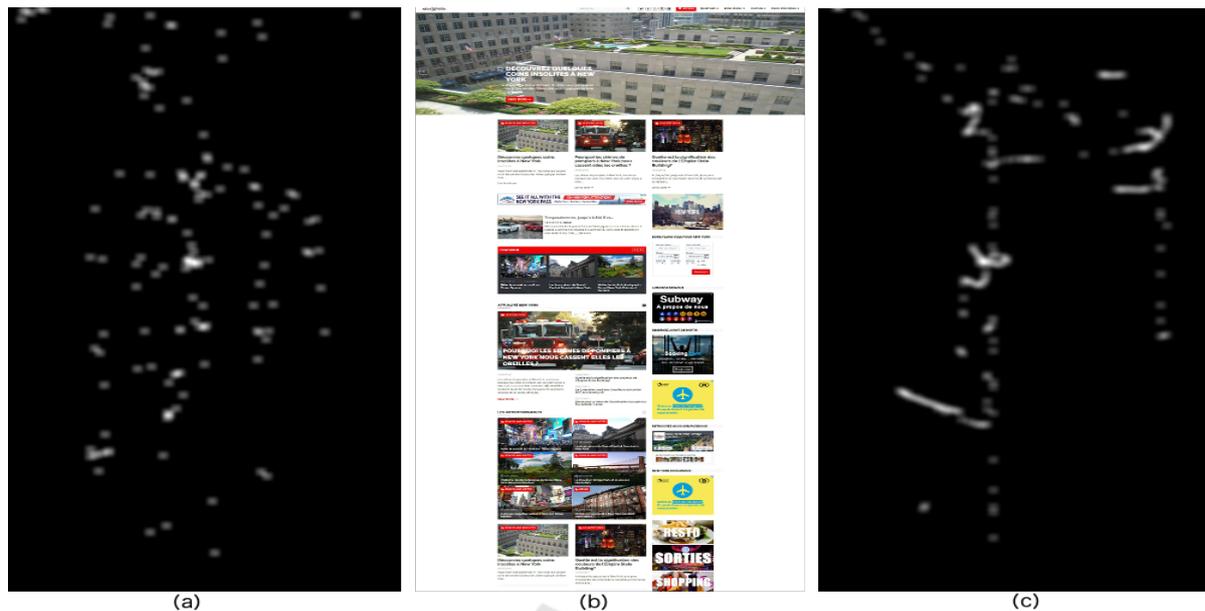


Figure 1: (a) eye fixation density map, (b) original website and (c) mouse fixation density.

has not been treated by the literature whereas it could be an essential information to the understanding of eye and mouse coordination. Finally, we used the results of the two first sets of observations to create two Gaussian-based models to approximate eye position from mouse position.

While the first approach focused on a static and spatial analysis, the second and third aimed for a dynamic analysis taking into account the temporal evolution of both eye and mouse tracks.

3.1 Static Fixation Densities Comparison

Pearson's Correlation Coefficient (PCC) also known as the Pearson Product-Moment Correlation (1), is a metric used in saliency maps comparison by authors like Ouerhani et al. (2004) and Le Meur et al. (2007) and used to compare fixations and mouse movements by Tavakoli et al. (2017). PCC has a value between -1 and 1. When the coefficient is almost equal to 1, there is a strong relationship between the two variables. The goal was to apply this metric to highlight eye and mouse coordination changes between tasks. The originality of this metric lies in the fact that it uses probability densities instead of raw variables values. To do so, we computed PCC between eyes density map and mouse density map. To obtain these maps, fixations from eye-tracking and mouse-tracking were convolved with a Gaussian filter. Thus PCC was computed between images (a) and (c) as shown in Figure 1.

Table 2: Pearson's correlation coefficients for intra and inter analyses.

Task	inter	intra-eye	intra-mouse
Free viewing	0.132	0.036	0.082
Target finding	0.171	0.028	0.107
Text reading	0.176	0.440	0.162

$$P_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \quad (1)$$

We obtained for our three tasks classes (free viewing, target finding and text reading) defined in section 2.1 correlation scores as in Table 2, "inter" column. Both classes correlation and their relative difference remained small which showed that mouse-tracking could not be directly used to model eye movements. For this reason, we decided to refine the investigation based on motion dynamics in the next sections.

Furthermore, when comparing eye-tracking results between different participants on the same stimulus, we obtained results in Table 2, "intra-eye" column which showed a higher correlation for "Text reading" task than for the two others. This result confirmed that if the task and its location were precise, then most of the participants would produce similar eye-gaze patterns. We observed the same behaviour for mouse tracks in Table 2, "intra-mouse" column, but with a lower overall correlation which showed that mouse behaviour remained less consistent than eye behaviour.

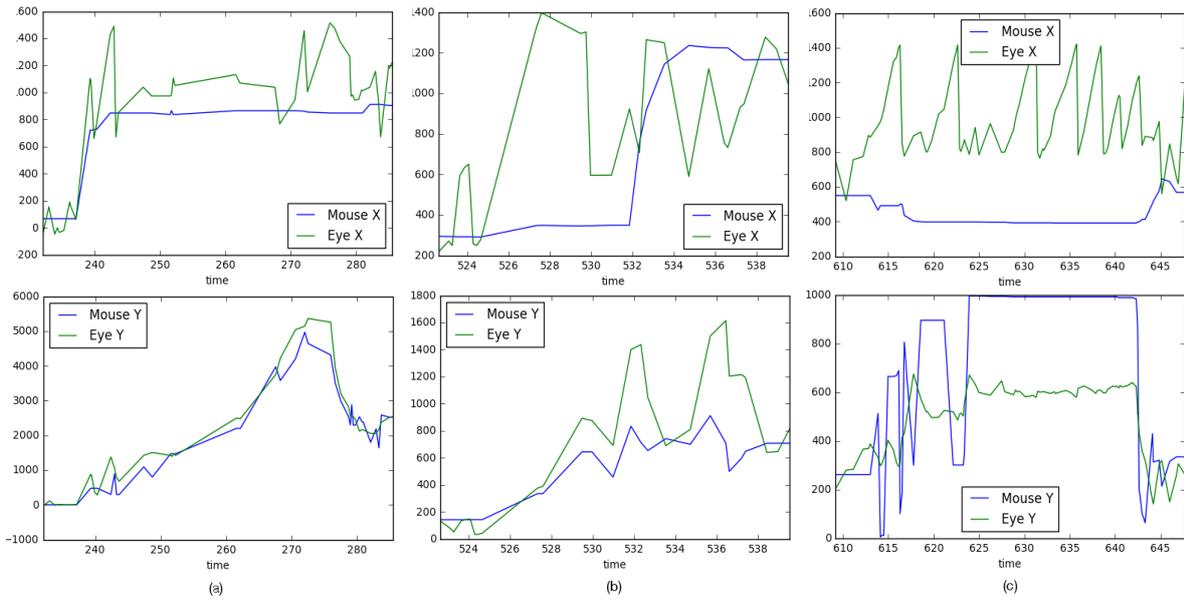


Figure 2: First column (a) is a free viewing task, second column (b) is a target finding task and third column (c) is a text reading task.

3.2 Dynamic Analyses

Considering the dominant use of scroll in our experiment, modern vertically-based designs and the tend of Human eye to be more efficient horizontally, we separated X and Y coordinates to enhance granularity in our dynamic analyses. For each X and Y coordinate we got temporal vectors which were synchronized between mouse and eye. To do so, we matched eye fixations with mouse events and then down-sampled mouse data to fit eye data. We chose to not interpolate as in Deng et al. (2016) because it could have generated non-existing fixations and wrong results.

3.2.1 Temporal and Distance Estimation

We observed for some participants a time shift on Y axis between mouse and eye with the mouse being delayed as in Figure 2 (a) and (c) right columns. This finding joined Huang and White (2012) previous work in which they detected a lag between the mouse and the eye. This could be explained by the fact that, in visual exploration context, the eye is the only mean of perception and leads the hand movements.

We computed euclidean distance (3) and obtained an eye-mouse distance of 554 pixels. This result was not in accordance with the average 229 pixels from state-of-the-art (Rodden and Fu, 2007; Guo and Agichtein, 2010; Huang and White, 2012). We then refined our analysis by separating the two axes. Using formula (2) we got a mean distance of 409 pixels for X axis and 291 pixels for Y axis. With this results

we began to have a better consistency on Y axis as expected. However, Bejan (2009) demonstrated that our eyes scan horizontally faster than in the vertical dimension. Based on our results, we could assumed that participants kept their mouse vertically stationary to scroll down or up and used it as a vertical pointer, allowing them to horizontally browse without difficulties. Thus the participant could easily move his eye on X axis more often. That is why the participant tended to move it's eye on X axis more often.

We then continued with separate axes to compute correlation. Compared to distances, correlation coefficients between mouse and eye were drastically different. Chen (2001) obtained a correlation of 0.58 with more than 50% of the pages associated with correlations larger than 0.8. In our study, we measured a mean correlation of 0.64 on Y axis and 0.18 on X axis.

Difference between axes got even more significant when we examined these correlations coefficients according to their corresponding task. As exposed in Table 3, free viewing task had the best correlation on Y with 0.9. This result reflected a greater trend to use the mouse as a vertical pointer as in other tasks. Coefficients for target finding were more balanced with an increased correlation on X but a decrease on Y . Finally, text reading correlations expressed the fact that participants did not used much the mouse during this task. We could assume that more the cognitive load of the task is important more the correlation drop on both axes.

Table 3: Pearson’s correlation on X and Y axes.

Task type	r_x	r_y
Free viewing	0.176	0.921
Target finding	0.383	0.699
Text reading	0.006	0.32

$$d(i) = |x_m(i) - x_e(i)| \tag{2}$$

$$d(X, Y) = \sqrt{(x_m - x_e)^2 + (y_m - y_e)^2} \tag{3}$$

3.2.2 Scroll’s Speed and Direction Influence

As we previously exposed, mouse and eyes were more correlated on Y axis. In addition, scroll events are a barely studied subject while it is a common behaviour in all webpages browsing. We based all our calculus on scroll sessions which corresponds to a set of continuous scroll events ended with a mouse movement. Scroll is an important feature providing good information about the degree of participants’ interest on a website. Another advantage is that the scroll is measurable on desktop and mobile. Through the following analyses, we highlighted influence of behaviour on scroll’s speed and amplitude.

We collected for each scroll session the direction (up or down) and the absolute speed. After empirical tries and errors and after taking into account the amount of data, we also separated the browser screen into 3 equals categories as in Figure 5 to detect patterns.

For the current analyses we removed text reading tasks because it did not included enough scroll events. For both amplitude and speed influence test, we performed a one-way independent ANOVA (analysis of variance) (4) test. The ANOVA examines if the mean of numeric variables differs across levels of categorical variables. After checking all assumptions (normality of errors, equal error variance across category, independence of errors), we hypothesized:

$$H_0 : \mu_0 = \mu_1 = \mu_2 \tag{4}$$

H_1 : At least one mean is not equal to the others.

As shown in Table 4, we considered that all means were equal to each other. The statistic test we ran was the ratio of the between-category variance and the within-category variance. If this ratio was greater than the critical probability distribution F, we could reject the null hypothesis. After obtaining a p-value below the 0.05 threshold, we could affirm the rejection of the null hypothesis with a confidence rate of 95 %. Thus, we can conclude that there is an effect of scroll speed on eyes category position.

Table 4: Result test ANOVA with significance level (p-value) and F-score.

Indicator	Task	Down	Up
F-test	Free viewing	4.26	7.07
	Target finding	3.76	-
P-value	Free viewing	0.017	0.001
	Target finding	0.031	-

To go further, we had to determine and define this influence. We focused on means for each tasks using a Tuckey’s test. We observed that while scrolling quickly, participants positioned their eyes at the opposite side of the scroll’s direction to be able to detect bottom-up characteristics through peripheral vision as shown in Figure 3. Furthermore, when participants were looking for a specific information (target finding task), they tended to quickly look towards the center of the screen when the scroll speed decreased.

Then we focused on scroll’s amplitude, which is the distance between the start and the end of a scroll session. We wanted to know if participants adapted their eyes position before scrolling. Here again we differentiated the target finding and the free viewing tasks and calculated the means distance for each area before the participant scroll.

Table 5: Result test ANOVA with significance level (p-value) and F-score.

Indicator	Task	Down	Up
F-test	Free viewing	3.08	10.44
	Target finding	0.09	-
P-value	Free viewing	< 0.001	< 0.001
	Target finding	0.9	-

We could conclude using ANOVA test that for the free viewing task, when the eyes were located at the bottom of the screen and before scrolling up, the scroll amplitude increased with p-value < 0.05 as shown in Table 5. As expected, when the eyes were located on the top of the screen before scrolling down the scroll amplitude were much higher, see Figure 4 (a) and (b). About the target finding task, there was no significant impact (c) of amplitude on the screen area before scrolling (p-value > 0.05). However, we noticed that when searching specific information, participants did not have a long scroll amplitude in order to not miss an element (text, blocs, titles, etc) and to differentiate them.

The scroll event could improve the prediction of the localization of the eyes on Y axis using the combination of direction, amplitude and speed variables.

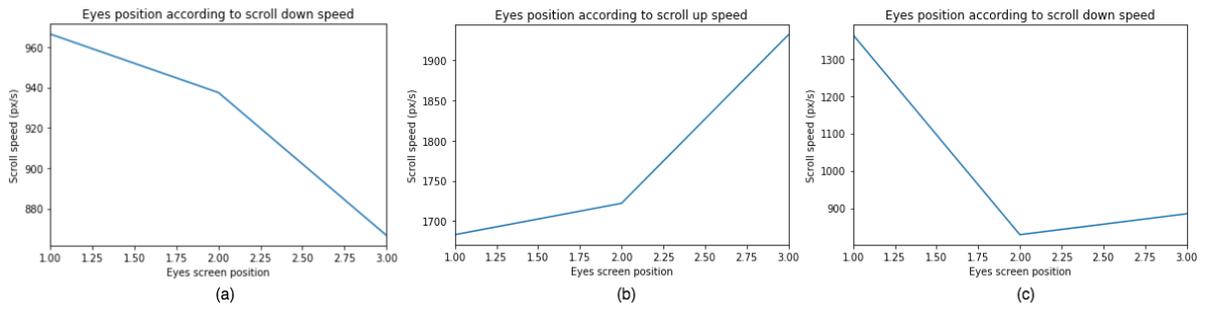


Figure 3: Eyes position according to scroll speed, (a) and (b) corresponds to free viewing task, (c) corresponds to target finding task.

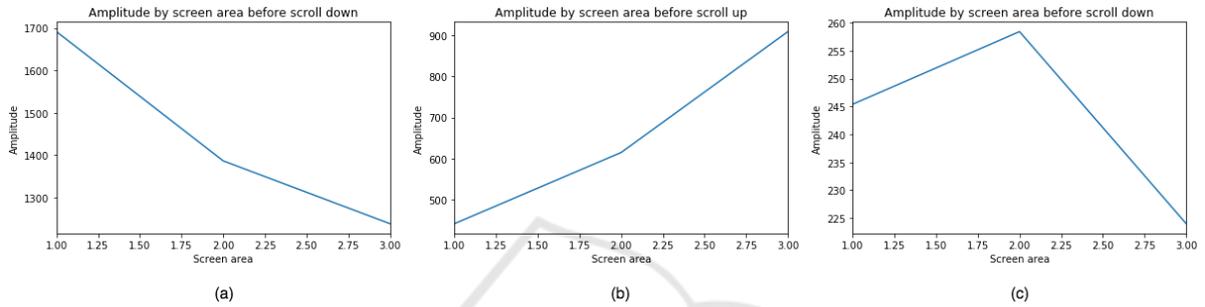


Figure 4: Scroll amplitude according to screen area before scroll, (a) and (b) corresponds to free viewing task, (c) corresponds to target finding task.



Figure 5: Screen’s three areas.

3.3 Model

Previous analyses provided several insights about users behaviours on webpages given more or less specific tasks. We built our models from these, more particularly from the eyes movements standard deviations. As in section 3.2.1, we separated X and Y axes to infer the parameters of a Gaussian model which predicted the eyes position based on the mouse position and cognitive load of the task. From these standard deviations we were able to define a confidence area around the mouse in which the eyes had a 70% probability to be in it. We chose to base our calculus on the 70th percentile because it was the minimum confidence rate we observed in the state of the art. As shown in Table 6, columns “x std.” and “y std.”, the 70th percentile (5) gave a first coarse pixel area around the mouse cursor.

$$percentile = \mu \pm Z\sigma \tag{5}$$

Table 6: Standard deviation (percentile 70%) normal and using only sudden X changes.

Task	x std.	y std.	x std. thrs.	y std. thrs.
Free viewing	558.0	416.4	-	-
Target finding	486.4	403.8	361.3	251.0
Text re-ading	627.8	257.9	-	-

But we were interested in a better model, so we focused on specific behaviours during tasks. As shown in Figure 2 (a) target finding class, we identified sudden changes on X axis. After analysing participants’ videos and comparing with several target finding tasks among them, we found that these sudden changes matched participant’s interest. When the participant had a target finding goal and found his target, he quickly moved his mouse to the point of interest.

Thus, we manually defined a threshold at the beginning of each sudden changes and we computed the standard deviation before and after every final sudden change on X . We obtained better results as shown in Table 6, column “x std. thrs.” and “y std. thrs.”. Area covered by the 70th was reduced by around 150 pixels for both axes. With this second model, we were able to increase the accuracy but only by focusing on a specific event.

4 CONCLUSION AND DISCUSSION

We first compared eye and mouse data with the saliency metric *PCC*. We did not find significant consistency between participants' eyes and mouse positions (inter) and between participants' eyes (intra). However, results showed that participants behaved in a more similar way when they had the same task with the same location (reading task).

Then, we got deeper with dynamic analyses. We showed that using distance and correlation, we were able to highlight more interesting coordinations between eyes and mouse. We had better results on *Y* axis than *X* axis and succeed to demonstrate behaviour differences between tasks. In addition, scroll analyses, clearly showed a relation between eyes position and scroll speed while browsing and amplitude before the scroll.

Finally, we made a model for each task able to predict the area around the mouse's cursor in which the eyes had 70% chances to be located in. However, eyes location uncertainty compared to mouse position remained high, even if we succeed to enhance the model during target finding task by observing brutal changes on *X* axis.

In this paper, we presented results of a preliminary study, used as a validation to conduct a bigger experiment, including more participants. This will allow us to analyse the impact of participants' age on their mouse movements. Moreover, we did not use scroll events analyses to enhance our models. In future work, we think that doing so, could boost the precision of the model by reducing the area around the mouse's cursor. We could also investigate new relations between the scroll and the eyes by analysing scroll in 2D. Then, we could use machine learning models to integrate new features and more user behaviours such as mouse patterns. Finally, our main objective is to propose the most accurate model in order to use it in real time to predict web user behaviours.

ACKNOWLEDGEMENT

We thank French Research and Technology Association (ANRT) and Sublime Skinz for supporting this work.

REFERENCES

Bejan, A. (2009). The goldenratio predicted: Vision, cognition and locomotion as a single design in nature. *International Journal of Design & Nature and Ecodynamics*, 4(2):97–104.

- Boi, P., Fenu, G., Davide Spano, L., and Vargiu, V. (2016). Reconstructing User's Attention on the Web through Mouse Movements and Perception-Based Content Identification. *Transactions on Applied Perception*, 13(3):1–21.
- Castelhana, M. S., Mack, M. L., and Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9(3):1–15.
- Chen, M.-c. (2001). What can a mouse cursor tell us more? Correlation of eye / mouse movements on web browsing. In *Conference on Human Factors in Computing Systems*, pages 281–282.
- Cooke, L. (2006). Is the Mouse a "Poor Man's Eye Tracker"? *Annual Conference-Society for Technical Communication*, 53:252 – 255.
- Deng, S., Chang, J., Kirkby, J. A., and Zhang, J. J. (2016). Gaze–mouse coordinated movements and dependency with coordination demands in tracing. *Behaviour & Information Technology*, 35(8):665–679.
- Guo, Q. and Agichtein, E. (2010). Towards predicting web searcher gaze position from mouse movements. In *Extended Abstracts of Conference on Human Factors in Computing Systems*, pages 3601–3606.
- Huang, J. and White, R. (2012). User See, User Point: Gaze and Cursor Alignment in Web Search. In *Special Interest Group of Conference on Human Factors in Computing Systems*, pages 1341–1350.
- Le Meur, O., Le Callet, P., and Barba, D. (2007). Predicting visual fixations on video based on low-level visual features. *Vision Research*, 47(19):2483–2498.
- Mills, M., Hollingworth, A., and Dodd, M. D. (2011). Examining the influence of task set on eye movements and fixations. *Journal of Vision*, 11(8):1–15.
- Mueller, F. and Lockerd, A. (2001). Cheese: tracking mouse movement activity on websites, a tool for user modeling. In *Extended Abstracts of Conference on Human Factors in Computing Systems*, pages 279–280.
- Navalpakkam, V., Jentzsch, L. L., Sayres, R., Ravi, S., Ahmed, A., and Smola, A. (2013). Measurement and modeling of eye-mouse behavior in the presence of nonlinear page layouts. In *International Conference on World Wide Web*, pages 953–964.
- Ouerhani, N., Wartburg, R. V., and Heinz, H. (2004). Empirical Validation of the Saliency-based Model of Visual Attention. *Electronic Letters on Computer Vision and Image Analysis*, 3(1):13–24.
- Rodden, K. and Fu, X. (2007). Exploring how mouse movements relate to eye movements on web search results pages. In *Special Interest Group on Information Retrieval Workshop on Web Information Seeking and Interaction*, pages 29–32.
- Rodden, K., Fu, X., Aula, A., and Spiro, I. (2008). Eye-mouse coordination patterns on web search results pages. In *Extended Abstracts of Conference on Human Factors in Computing Systems*, pages 2997–3002.

- Tavakoli, H. R., Ahmed, F., Borji, A., and Laaksonen, J. (2017). Saliency Revisited: Analysis of Mouse Movements versus Fixations. In *Conference on Computer Vision and Pattern Recognition*, pages 4321–4329.
- Yarbus, A. L. (1967). *Eye Movements*. New York: Plenum Press.

