Evaluating gaze control on a multi-target sequencing task: the distribution of fixations is evidence of exploration optimization

Giaccomo Veneri\textsuperscript{a,b,c,*}, Francesca Rosini\textsuperscript{b}, Pamela Federighi\textsuperscript{a,b}, Antonio Federico\textsuperscript{b}, Alessandra Rufa\textsuperscript{a,b,*}

\textsuperscript{a}Eye Tracking & Vision Applications Lab - University of Siena
\textsuperscript{b}Department of Neurological Neurosurgical and Behavioral Science - University of Siena
\textsuperscript{c}Etruria Innovazione Spa

Abstract

Many high cognitive applications, such as vision processing and representation and understanding of images, often need to analyse in detail how an ongoing visual search was performed in a representative subset of the image, which may be arranged into sequences of loci, called regions of interest (ROIs). We used the Trial Making Test (TMT) in which subjects are asked to fixate a sequence of letters and numbers in a logical alphanumeric order. The main characteristic of TMT is to force the subject to perform a default and well-known path. The comparison of the expected scan-path with the observed scan-path provides a valuable method to investigate how a task force the subject to maintain a top-down internal representation of execution and how bottom-up influences the performance. We developed a mechanism that analyses the scan path using different algorithms, and we compared it with other methods: we found that fixations outside the ROI are direct influence of exploration strategy. The paper discusses the method in healthy subjects.

Keywords: Eye Tracking, Trial Making Test, Fixations, Saccades, Attention

1. Introduction

Recent studies are focused on methods and models that evaluate how humans explore real scenes in a naturalistic approach and how machine vision should emulate human searching. In the real world, visual search is a common task that enables humans to explore the environment and direct attention towards regions of interest. In experimental settings in which neuro-physiological and cognitive functions are investigated, quantification methods for visual search are used to evaluate the allocation of attention during scene viewing [1, for a short review].

In order to understand how mechanisms drive attention during visual exploration, many studies in which image characteristics are manipulated have been conducted. The main concept is that particular regions of interest in the scene are selected assuming their cognitive relevance or local image’s saliency. When image salience is thought to guide visual search, the mechanism is called bottom-up. Conversely, when mechanisms driving visual search depend more on human intention, they are called top-down. These two exemplifications describe the main theories of visual search, although endogenous and exogenous components presumably work together in normal circumstances. By considering this dichotomy, a variety of formal models have been proposed on the last decade, in order to describe the attentional selection mechanism: Feature Integration Theories [2, FIT], Guided Search [3], Theory of Visual Attention [4, 5, TVA] and a new purely bottom-up model Winner Take All [6, WTA]. It has been suggested that early selection stages are purely driven by the image saliency factors in a bottom-up prevalence, and later selection is due to the combination of top-down and bottom-up factors. A key debate in this literature is whether bottom-up can override top-down and vice versa. For instance Theeuwes [7] and Daniel Schreij [8] found that the appearance of distracters reduced search efficiency, presumably due to the involuntary capture of attention. Conversely, Chen X [9] found that visual search on the real world is dominated by the top-down mechanism.

The focus of recent researches, however, has shifted to how these processes should be combined and their relative contributions to search guidance, how top-down and bottom-up work together to perform an efficient visual exploration and how they interfere with each other (see der Stigchel et al. [10] for a review); therefore, the research is directed towards unified methods of analysis that may better reflect real conditions. In our research, we aimed to investigate how the mechanism of bottom-up and top-down works together on neuro-psychological context during the ongoing visual search. We aimed to use a task which encourages free exploration, avoiding too high saliency features, such as proposed by Theeuwes. We did not want to use a real word image because we needed an easy method to evaluate “bottom-up”-“top-down” competition.
or collaboration; the key idea was to adapt the version of Trial Making Test part B (TMT) [11] and some variants: the test is a neuro-psychological instrument in which numbers and letters have to be connected in numerical and alphabetical order (1 – A – 2 – B – 3 – C – 4 – D – 5 – E). Recent studies [12] proposed the TMT as a powerfully test to evaluate sequencing, symbol classification, memory and searching. Wolwer and Gaebel [13] demonstrated the validity of its application in patients, adapting the “paper pencil version” of test to the “computer version” which uses the cursor and a tone as feedback for achieving the target. In the “eye-tracking version” proposed in our study, the subject is required to connect letters and numbers by moving the gaze (see Method) over each sign of the sequence without any feedback such as tone or guided search.

The main characteristic of TMT is to force the subject to perform a default and “a priori known” path; the key idea was to compare the expected scan-path (1 – A – 2 – . . . – E) with the observed scan-path.

The Scan-path was one of the first methods [14] to identify patterns of eye movement: [15] defined a number of spatial Regions of Interest (ROIs) in the scene being scanned and recording the fixation sequence as a series of letters representing the fixated locations. Brandt and Stark [16] used string-edit analysis to compare the viewing pattern of a scan-path. Recently, Cristino et al. [17] developed an interesting method (ScanMatch) which consists on transition matrix among ROIs and the usage of Levenshtein distance to compare scan path. The Levenshtein distance measures the editing cost of transforming one string into another one using, in its basic form, a set of three operations (insertion, deletion and substitution) with a cost of one for each operation. Related works [18, 19, 20], in which the cognitive processes underlying visual search performance are investigated, measures such as fixation duration, fixations per trial, saccadic latency and saccadic distribution, have been developed. In this context, Region of Interest (ROIs) are predefined over cognitively relevant parts of the image, and features such as time spent in ROI and saccade trajectories [21, for a survey] are calculated. Other authors extract the regions of interest automatically using fixation distribution during visual exploration [22, 23]. In a more complex manner, extraction of saliency maps was developed to predict fixation locations [6].

In our method, due to the distinctive characteristic of TMT, in which subject must follow a predefined sequence (attended scan-path) of symbols, we propose to evaluate the scan-path made by subject respect to the attended exploration, in order to extract some features representing the differences between attended and observed exploration. The method was tested in a group of normal volunteer subjects: we developed a set of computational indicators based on scan-path and fixations analysis to evaluate the visual search performance. We compared the results with the ScanMatch proposed by Cristino et al. [17] in order to provide a valuable reference.

2. Materials and Methods

We enrolled 30 volunteers (17 female and 13 male) aged 25 – 45 years (CTRL) with normal vision and not taking medicines and without refractive defects were enrolled. All participants in the study were trained on the TMT test by a psychologist. Subjects were seated at a viewing distance of 78 cm from a 24” color monitor (51 cm x 33 cm). Eye position was recorded using the ASL 6000 system, which consists of a remote-mounted camera sampling pupil location at 240Hz. Nine point calibration and 3-point validation procedures were repeated several times to ensure all recordings had a mean spatial error of less than 0.1 degree. Data was controlled by a Pentium4 3 GHz computer acquiring signal by a fast UART serial port.

Head movement was restricted using a chin rest and bite. Subjects were arranged in three groups: subjects doing the simplified trial making task part B (TMT), the simplified trial making task part B plus distracter (TMTD) and the E-Masked test (ET). The experiment was repeated in three different sessions mixing the participants among groups: each subject performed TMT, TMTD, ET in three different days in order to avoid any bias due to memory.

2.1. Task

In the TMT, subjects are asked to search for the simple sequence 1 – A – 2 – B – 3 – C – 4 – D – 5 – E with their gaze (See Fig. 1.a); each letter/number was displayed in red (60 cd/m2) on a black background, sub-tending of ~3x4 degrees a visual angle, and arranged in random positions to avoid geometric trajectories.

A common distinctive feature of this task is to force the subject to take a predefined path ordered by sequence [24]. Considering the characteristics of this task, scan path analysis should be the best method to measure the visual search strategy. TMTD: in order to study the influence of distracters and the exploration around targets, we developed a second task in which the image was produced “merging” the TMT with some E-masked distracters (Fig. 1.b). ET: in order to evaluate free exploration (in the absence of a sequencing command) we adapted a common E-masked test [3] to the geometry of the TMT. Subjects were asked to count right-oriented “Es” (Fig. 1.c). To force the subject to direct the gaze over the sign, the right- and left-oriented “Es” were masked. To obtain comparable results among the tasks, the signs were disposed in the same location as for TMT. The TMT, TMTD and ET were also tested with different geometric distributions of signs.

The temporal execution of the tasks was the same for both tasks: 1 second red ball at the centre of the screen, 20 seconds the task, 1 second black screen. The subjects performed the test several times in three different arrangements of signs, but only the data of test depicted in Fig. 1 was used.
Figure 1: (TMT) Trial Making Test Part B: the subject is required to connect 1–A–2–B–3–C–4–D–5–E (letter on top and number on the bottom) with the gaze. Image shown on a 1024 \times 768 px in 20000 ms of time. (TMTD) TMT and distracters (ET) Masked E test: the subject is looking for right oriented E.

Table 1: Indicators were compared with ScanMatch method.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Indicator</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SI</td>
<td>Sequencing index</td>
<td>Ability to perform task correctly. TMT specific</td>
</tr>
<tr>
<td>FDS</td>
<td>Fixation distance (Euclidean)</td>
<td>An expression of fixation distribution around ROI.</td>
</tr>
<tr>
<td>FDN</td>
<td>Fixation distance from target ROI</td>
<td>An expression of fixation planning to reach the next target.</td>
</tr>
<tr>
<td>D</td>
<td>Duration of fixation</td>
<td>Duration of fixation</td>
</tr>
<tr>
<td>DE</td>
<td>Direction error</td>
<td>Gaze direction versus right (expected) direction. It’s an expression of saccades planning in order to reach next target.</td>
</tr>
<tr>
<td>∆DE</td>
<td>Delta direction error</td>
<td>An expression of improves saccade planning to reach the next target in on-going visual search.</td>
</tr>
</tbody>
</table>

3. Calculations

Data was stored in a comma separated file imported into Matlab. A simple blink-removal filter was applied; the filter substituted blink values (pupil diameter 0) or missing data (horizontal or vertical coordinates out of range) by linear interpolation data. Large segments of missing data were marked (duration > 40 ms) in order to exclude them from analysis.

Numbers and letters were sampled as pre-defined rectangular ROIs centered on letters and numbers and having widths and heights merging from 80 \times 80 pixels to 100 \times 100 pixels. The same ROI spatial distribution was pre-defined for all tasks. Fixations were calculated using a dispersion-based algorithm proposed by Salvucci and Goldberg [25] and adapted by Veneri et al. [26]. Saccades were found by Fischer et al. [27] algorithm. The minimum duration of fixation was set at 50 ms, the maximum dispersion to identify fixations was at one degree and the saccade threshold at 30 deg/sec [28].

Peripheral vision of subjects was assessed computing the likelihood to reach the target as the function of distance of current fixation to the target [29, 30].

Scan-path was estimated through six indicators (Table 1 and Appendix Appendix A) and compared with ScanMatch method.

ANOVA was used to find the difference among the groups ET, TMT and TMTD; post-hoc analysis was performed by the Holm-Sidak procedure for multiple Student’s t tests. Kolmogorov-Smirnov test was used to evaluate the assumption of normal distribution. The power of a statistical test and the sample size was checked by t-test.

3.1. Sequencing index

The method discovers the sequence followed by subjects in order to provide an indicator of task performance; it is based on a tree exploration algorithm, sampling only the area of letters and numbers of the TMT. During the first phase, the method marks the segments of raw eye movement in the ROI and purges rapid crossing. Rapid crossing is defined as segments where duration in ROI is less than 10 ms (velocity > 52.6 deg/sec). Sometimes, subjects cross the ROI at low velocity because they have already been visited or the gaze is affected by tremor due to a movement disorder, preventing
distinctive fixations. The proposed procedure was preferred than fixation algorithms, because is easier and avoid any complex parameter definition, which should fail. Only letters and numbers were considered as a target ROIs; distracters were excluded.

During the second phase, the procedure identifies the correct sequence: the algorithm makes a tree of all possible combinations of the sequence (see Fig. 2), and assigns a score for each correct step, such as $1 - A$ or $A - 2$.

Maximum depth is chosen as the correct sequence. Finally, the score is normalized to the maximum score possible (for instance 10 with TMT and TMTD).

3.2. Fixation distance

In order to evaluate scan path with respect to ROIs we developed two indicators to describe the distribution of fixations. The first indicator measures the Euclidean distance of fixation to nearest ROI (FDS). The value aims to provide a measure of dispersion of fixations on the scene.

The second indicator measures the Euclidean distance of fixation to target ROI (FDN). The aim of the indicator was to determine whether fixation was on the target and to evaluate user performance to reach the targets.

3.3. Eight-Pointed Star Model (8PS): direction error

Engel [31] proposed the usage of the eight-pointed star model of the conspicuous area. Ponsoda et al. [32] and later Machner et al. [33], assessed search strategies by attributing the relative number of saccades towards the eight main orthogonal and diagonal directions of space. In the Eight-Pointed Star (8PS) procedure we used the model proposed by Ponsoda et al. [32] and adapted it to ROIs. We calculated the projection of gaze into and out of ROIs in the eight directions (Fig. 3).

Sequencing ability was quantified based on the difference between the observed direction and expected direction of the 8PS model. If a ROI was not entered, a maximum Direction Error was assigned. In the case of multiple crossing, the minimum Direction Error was assigned, assuming other crossings to the pre-visiting or re-visiting. The aim of the method was to evaluate simultaneously motor skills and sequencing ability: a subject coming from the correct ROI and going to the correct ROI had a Direction Error (DE) close to zero. However, Direction Error was affected by TMT geometry (placing of letters and numbers). The symbols near the border are a special case of this artefact because the subject’s gaze is driven by borders and empty areas. In order to avoid this artefact and to develop a method independent of geometry, a spatial correction was applied to the 8PS model. For symbol, Direction Error was multiplied by a weigh proportional to the directions actually available.

We defined a second indicator of ability to improve performance during the task. The indicator was expressed as delta Direction Error step by step: Delta Direction Error ($\Delta$DE). The value provided an estimation of ongoing ability to update the spatial map during visual exploration. $\Delta$DE must be scaled proportionally to mean direction error.

4. Results

4.1. Subjects vision assessment

It has been demonstrated that information acquired from the visual periphery in one fixation (peripheral preview) can influence the subsequent pattern of eye movements [30, 34, 35]. In order to assess that correct execution of the task was not affected by a defect on peripheral vision and scene understanding, we verified the probability to reach the target as a function of FDN. Fig. 4 shows the probability density function of normal subjects depending from distance only; the result showed that ability to reach the target depends strictly from distance: we did not find any significant influence of fixation’s duration to reach the target such as the Findlay et al. [29] research ($t(429) = -1.347519, p = 0.17$); we found a significant difference between fixations preceding the “fixation in target” and the remaining fixations ($t(429) = -3.79, p < 0.001$); similar result was found evaluating the distribution of “fixation in ROI” and the remaining fixations ($t(429) = -4.25, p < 0.001$).
The probability to reach the target depended from the fixation to target distance: likelihood reported that the subject could reach the target with 72% of probability if fixation was performed in a neighbourhood of two degrees. (a) We evaluated fixation distance to the target, and we found a significant difference between fixations preceding the “fixation in target” and the remaining fixations ($t(429) = -3.79, p < 0.001$). (d) Duration of fixations did not affect the ability to reach the target ($t(429) = -1.347519, p = 0.17$). (c) Similarly, we evaluated the distance from fixation to nearest ROI and the probability to “go into the ROI” and we found a significant difference ($t(429) = -4.25, p < 0.001$). (d) The duration of fixations preceding a “fixation in ROI”: we did not find a significant difference ($t(429) = -1.34, p = 0.18$).

<table>
<thead>
<tr>
<th>Distance to Target (deg)</th>
<th>Probability to reach the target</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 4</td>
<td>72%</td>
</tr>
<tr>
<td>4 - 6</td>
<td>51%</td>
</tr>
<tr>
<td>6 - 10</td>
<td>30%</td>
</tr>
<tr>
<td>18 - 25</td>
<td>20%</td>
</tr>
<tr>
<td>≥ 25</td>
<td>8%</td>
</tr>
</tbody>
</table>

Table 2: Peripheral vision was tested as the probability to reach the target conditioned to the probability to enter the nearest ROI when the latest fixation was in the intervals 2-10-18-25 degrees.

We evaluated the probability to reach the target depending from distance to target conditioned to the probability to enter the nearest ROI:

$$P_{go \text{ into Target}}(go \text{ into nearest ROI}) = \frac{P_{go \text{ into Target}} \cap go \text{ into nearest ROI}}{P_{go \text{ into nearest ROI}}}$$

The probability to reach the target was the 72% for $FDN \leq 4$ (Table 2).

4.2. Healthy subjects performance

The Sequencing Indicator (SI) was firstly tested comparing TMT paper pencil version and TMT eye-tracking version on 10 normal subjects. No significant differences were found ($t(18) = -1.24, p = 0.23$) between the two groups. We then evaluated SI on TMT, TMTD and ET on 30 normal subjects. Since no differences were found between TMT and TMTD, SI is a robust indicator of individual ability to perform a logical sequence and is not affected by distracter ($p = 0.89$). Since ET does not ask the user to follow a sequence but just to explore the scene, we found a significant difference between the TMTs and ET ($F(2, 87) = 486.9, p < 0.001$). This result further confirms the effectiveness of our algorithm.

The aim of the FDS indicator was to describe fixation distribution around the target ROIs. We found a significant difference between TMTD and TMT/ET, as expected ($F(2, 87) = 55, p < 0.001$) Holm Sidak confirmed the hypothesis ($\alpha(87) = 0.05, p = 0.0352$); indeed, distracters forced subjects to fixate symbols far from target ROIs. Unexpectedly, significant differences were found also between TMT and ET ($\alpha(87) = 0.0253, p < 0.001$), despite they have the same geometry and therefore, the same saliency (Fig. 7). In order to confirm the robustness of indicator, we analysed fixation duration in ROIs and out of ROIs; fixation duration is intended to be an indicator of fixation purpose: short fixations are interpreted to be for exploring and orienting, on the contrary, long fixation should be regarded for semantic scene acquisition [1]. Thus a link between FDS and short fixation would be a further confirmation of the proposed model. We verified this hypothesis comparing fixation duration in ROIs and outside the ROIs only for TMT and ET. We applied right-tail t-test to verify that fixation duration in ROIs was greater than duration fixation outside of ROIs. A significant difference between the two parameters ($t(772) = 3.187, p < 0.001$) was found.

Comparing FDS with the score ScanMatch proposed by Cristiano et al. [17], we found a correlation between the two indicators for the TMT ($R^2 = 0.19, p = 0.02$) for the TMTD ($R^2 = 0.58, p < 0.001$) and not for ET (Fig. 6).

We evaluated the Receiver Operating Characteristic (ROC) of FDS, FDN and ScanMatch of the tests; test was assumed positive when $SI = 1.0$. Since ET is not a sequencing test, there was no subject performing the sequence, then, we evaluated the ROC curve on a mixed group of ET and TMT data. Tab. 3 reports the area under the curve: the results suggested that ScanMatch was more sensitive to the ability to perform the sequence rather than FDS and FDN, but FDN and FDS are more appropriate to evaluate the change of search strategy, for instance, when the subject performed ET or TMT. Our preliminary con-
Table 3: Area under the Receiver Operating Characteristic (ROC). P-values (reported into the brackets) did not report a significant differences with respect to the line of no-discrimination (area=0.5): FDS and FDN were able to discriminate the execution only on the mixed group ET-TMT. ScanMatch was more sensitive to the ability to perform the sequence correctly; FDS and FDN were more sensitive to the change of strategy.

Figure 6: Distance to nearest ROI (FDS) was correlated to the scan-path evaluated by ScanMatch method both on TMT ($R^2 = 0.19, p = 0.02$) and TMTD ($R^2 = 0.58, p < 0.001$); FDS was not strictly dependent on the salience (bottom-up), as expected, rather the exploration strategies (top-down). Considering the characteristics of peripheral vision and generally of attentive spot, fixations outside the ROI Task TMT/TMTD are due to mechanisms of efficiency rather than the salience of the image exploration.

Conclusion was that FND and FDS provided a different level of information about the scan-path, not strictly related to the sequencing indicator SI.

We tested the FDN comparing TMT/TMTD and ET under the same conditions as above. This indicator uses ROI sequencing to evaluate the intention of subjects to fixate according to an internal representation of the task. No significant differences were found (Fig. 7) between TMT and TMTD ($t(58) = 0.8242, p = 0.4132$).

Analysis of scan-path performance was carried up by the 8PS model studying the overall direction taken by subjects with respect to the ROI centre. The precision of DE was calculated comparing TMT/TMTD and ET. Significant difference was found ($F(87) = 48.3, p < 0.001$) between the three tasks. Post-hoc analysis by the Holm-Sidak procedure confirmed significant differences between TMT-ET ($\alpha(87) = 0.0170, p < 0.001$), TMTD-ET ($\alpha(87) = 0.0243, p < 0.001$) and ($\alpha(87) = 0.05, p = 0.004$). In fact, ET did not require any predefined route respect to the TMT or TMTD, and the direction error DE was in line with this hypothesis.

Moreover, DE of TMT was smaller than DE of TMTD (Fig. 7); the result is a further proof of the precision of the indicator because distracters force the user to explore fake ROIs increasing the direction error. In order to understand how the DE indicator evolves during the on-going visual search, we calcu-


5. Discussion

The FDS indicator confirmed all hypotheses: FDS grew up when the subject was forced by TMTD saliency to make sparser fixations, but we found, also, a strong significant difference between TMT and TMTD. Considering the characteristics of peripheral vision and generally of attentive focus [36] explained in Section 4.1, fixations outside the ROI on the task TMT/TMTD were due to mechanisms of efficiency rather than to the salience of the image exploration. We concluded that FDS is an indicator of visual search and exploration strategies depending on both from the saliency (bottom-up) but, particularly, from top-down intention.

Results found on ΔDE suggested that the indicator should be interpreted as feedforward ability to improve the search mechanism in order to reach the target during visual search [37]. ΔDE is in fact a negative indicator: performance is inversely proportional to the indicator. On TMTD, performance is not significant greater than on TMT, suggesting that the indicators depend strictly from subject’s ability. ΔDE does not make sense applied to ET, and should be 0; we reported it only as proof of correct implementation of the indicator.

We found that DE was significant different between TMT and TMTD; this value is a further proof provided by various authors that visual exploration efficiency degrades proportionally to distractors [38, 7, 8].

SI remained stable on TMT and TMTD; this should be interpreted as the ability of subjects to perform the sequence, with distractors too.

FDN indicator remained unclear: FDN seems independent of top-down command. This would support the hypothesis that visual search does not require memory [39, 40] and fixations outside the target ROIs are planned only by visual search map in order to reach the target efficiently as possible.

6. Conclusion

According to Yarbus [41], Itti and Koch [6], Noton and Stark [15], Privitera and Stark [22], Awh et al. [42], the construction of an accurate visual representation depends crucially on optimal and precise selection of subsequent fixation points. Thus, the study of how fixations are made with respect to pre-defined ROIs is primary for understanding the visual search behavior. We developed three methods. The sequencing algorithm was specific for the TMT and related tasks where the subject is asked to make a sequence. The fixation distance indicators (FDS and FDN) evaluate the distribution of fixation and are useful to study visual search in complex scenes (real or geometric) [43]. The 8PS model evaluates how saccade direction is taken by users to explore the scene. Saccade direction can be interpreted as the instantaneous intention [44] of the user to move the eyes to explore the target portion of a scene according to an internal spatial model (bottom-up) or task command (top-down).

Table 1 summarizes the proposed method: generally speaking four of the six indicators proposed confirmed the hypothesis. We suggest that analysis of fixations should be used to understand visual search exploration and saccades analysis to study how cognitive process plan exploration in order to reach the target. This hypothesis was confirmed by a related study by Pondoda et al. [32], Pomplun [44]; who adapted the “saccade selectivity” method, which describes the distribution of endpoints of first saccades, to the nearest target; in our case, we used fixations as endpoints of saccades because in clinical contexts fixation is more robust than saccades due to movement disorders. Thus we concluded that TMT and the proposed method should be useful to study visual search in a top-down context.

Further experiments based on TMT will be addressed to integrate our visual search model [45] in order to validate the hypothesis that the fixations outside the ROIs are an optimisation mechanism similar to that proposed by Todorov [46] or van Beers [47].

Appendix A. Formal definition

Appendix A.1. ROI definition

The dimension of ROIs was set by the following formula:

If \( R((x_1, y_1), (x_2, y_2)) \) is the smallest rectangle containing the symbol, the ROI was set:

\[
ROI = R((x_1 - K, y_1 - K), (x_2 + K, y_2 + K)) \quad (A.1)
\]

where \( K \) is one degree. If the intersection of two ROIs was not empty, the area was assigned by the Voronoi mechanism (Barber et al. [48]).
Algorithm 1 Pseudo code to identify ROI crossing. The first block marks each sample (x,y,t) at the time t in ROI proportionally to the points near (x,y,t) and in ROI i. The second block reconstructs the sequence avoiding rapid crossing.

<table>
<thead>
<tr>
<th>Block marking the sample proportionally to neighborhood</th>
</tr>
</thead>
<tbody>
<tr>
<td>for all ROI, do</td>
</tr>
<tr>
<td>for all (x,y,t) ∈ ROI, do</td>
</tr>
<tr>
<td>S(x,y,t) ← 0</td>
</tr>
<tr>
<td>for all (x’,y’,t’) ∈ ROI, and t’ ∈ (t-W,t+W) do</td>
</tr>
<tr>
<td>S(x,y,t) ← S(x,y,t) + 1</td>
</tr>
<tr>
<td>end for</td>
</tr>
<tr>
<td>end for</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Block marking sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>QUEUE ← empty</td>
</tr>
<tr>
<td>for all (x,y,t) ∈ ROI, do</td>
</tr>
<tr>
<td>if S(x,y,t) ≥ th and ROI ≠ QUEUE(last) then</td>
</tr>
<tr>
<td>push ROI on QUEUE</td>
</tr>
<tr>
<td>end if</td>
</tr>
<tr>
<td>end for</td>
</tr>
</tbody>
</table>

Appendix A.2. ROI sequencing

If (x,y,t) is the sample of Euclidean coordinates at the time t of the input signal, the Algorithm 1 describes the method. The sequence made by a subject is extracted from the variable QUEUE. The variables th and W are the threshold and time window, respectively. W was set at 100 ms and th at 10 to avoid rapid crossing.

Appendix A.3. Peripheral vision performance

\[
pdf_p(d) = \sum t'(f_s,f_s,d) \quad (A.2)
\]

where

\[
t'(f_s,f_s,d) = \begin{cases} 
1 & : \text{next fixation is in target ROI(r_s,r_y)} \\
0 & : \text{else}
\end{cases}
\]

and

\[
t(f_s,f_s,d) = \begin{cases} 
1 & : \delta((f_s,f_s),(r_s,r_y)) \leq d \\
0 & : \text{else}
\end{cases}
\]

Equation A.2 is the percentage of fixations that precede a proper target, divided by the number of fixations as a function of distance from the target.

Appendix A.4. Fixation distance

The formal definition of fixation distance from nearest ROI:

\[
FDS = f_s(x_f,y_f) \cdot \min_{ROI} \delta((x_f,y_f),(x_o,y_o)) \quad (A.5)
\]

where (x_f,y_f) is the centroid of fixation (x_o,y_o) is the centroid of ROI and f_s(x,y) is a linear-step function (Fig. A.10):

\[
f_s(x,y) = \begin{cases} 
0 & : (x,y) \leq w \\
0.5 & : w \leq (x,y) < w + w \cdot 0.1 \\
1 & : w + w \cdot 0.1 \leq (x,y)
\end{cases}
\]

and \( \delta \) is the Euclidean distance.

Fixation distance from target ROI was defined as:

\[
FDS = f_s(x_f,y_f) \cdot \delta((x_f,y_f),(x_o,y_o)) \quad (A.7)
\]

where \((x_f,y_f)\) is the centroid of fixation at time t and \((x_o,y_o)\) is the centroid of the next ROI according to the sequence followed by the user.

Appendix A.5. Direction error performance

Gaze projection was defined as the direction of the saccade mapped onto the nearest main star direction. We defined “observed direction” \((o)\) as the direction of the subject’s gaze. We defined “expected direction” \((e)\) as the direction from one ROI to the next according to the sequence.

\[
d = (o - e).w \quad (A.8)
\]

For instance, weight \((w)\) was set at 1 for letters in the centre and 1.6 =8/5 for letters near the edge. The final formula for all ROIs was:

\[
d = \{d_1 \ldots d_N\} \quad (A.9)
\]

where

\[
d_i = \{d(x,y) \in ROI, \min_{ ROI} \delta((x,y),d) : \text{else} \} \quad (A.10)
\]

\( N \) = number of ROIs = 10 and maximum error = 4 = 8/2. The scalar error (indicator) was expressed as the mean of equation A.9:

\[
DE = \text{mean}(d) \quad (A.11)
\]

The scalar delta error was expressed as:

\[
\Delta DE = \frac{\sum_{i=1}^{N-1} e_i - e_{i-1}}{\text{mean}(d)} \quad (A.12)
\]

References


