The Contribution of Eye Tracking to the Usability Evaluation of Network Management Tools

Marco C. Pretorius, Andre P. Calitz and Darelle van Greunen
Department of Computer Science & Information Systems
PO BOX 77000, Nelson Mandela Metropolitan University, Port Elizabeth, 6031
Tel: 041-504 2094    Fax: 041-504 2831
Marco.Pretorius@nmmu.ac.za, Andre.Calitz@nmmu.ac.za, Darelle.vanGreunen@nmmu.ac.za.

ABSTRACT
Network management (NM) tools have been developed to analyse the large amount of data generated by network applications and to display the data using information visualisation techniques. The general increase in the use of information visualisation techniques has highlighted the need for principles and methodologies for the evaluation of NM tools. The usability evaluation of NM tools is traditionally conducted by means of task performance measures and subjective measures such as questionnaires. Eye movement data can supplement the data obtained through user testing by providing more specific information about the user’s mental processes. This paper discusses a methodology that combines traditional usability methods and eye tracking methods for the usability evaluation of the visualisation techniques used by NM tools. Preliminary results from a pilot study show that eye tracking does provide additional value to the usability evaluation of NM tools.

Keywords — Eye tracking, network management, usability evaluation, visualisation evaluation.

I. INTRODUCTION
Data communication networks such as the Internet and wide-area networks are rapidly increasing in size and capability. The ability for a network manager to assess the effectiveness of the network infrastructure, is greatly enhanced by visually representing the statistical information associated with network usage and directly associating that information with the network layout. This allows the network manager to plan long range infrastructure management as well as deal with short term and immediate crisis [6].

Network management can be described as the monitoring and controlling of a computer network in order to function efficiently and provide value to users. The availability and ability of NM tools to generate data on which to make informed decisions has increased and the capability to communicate information has become unrestricted. However, the ability to access and analyse information in order to make informed decisions is on the decline [23]. Given this situation, there is a growing need for usable solutions to facilitate and support the visualisation of data and the associated decision-making process.

The main aim of usability evaluation is to identify problems that avoid or interfere with users’ tasks, causing stress or reducing performance. Assessing the role of visual attention with conventional usability methods like click analysis, questionnaires or simply asking users what they paid attention to, is simply not enough when dealing with NM tools. As technology has advanced in recent years, eye tracking has become a promising tool in order to answer questions relating to where the user’s visual attention is on the screen.

Incorporating eye tracking into software usability evaluation can provide knowledge that is not obtainable from traditional usability testing methods [11]. Eye tracking could be a helpful method in usability testing for the assessment of the relevant mental strategies which cannot be measured by means of think-aloud protocols or questionnaires.

II. BACKGROUND
A. Network Management
Networks are critical to modern society, and a detailed understanding of how they work is essential for their operation. Network data is very difficult to display without the use of some form of visualisation techniques. The aim of information visualisation techniques is to present data in methods that accurately communicate information, and need minimal effort for comprehension [12]. Making the effort to understand Internet traffic patterns, network throughput, usage patterns, application delay, downtime and other network management data sets, is made much easier by visual representations of this complex data, rather than looking strictly at tables and statistics.

B. Usability
Usability is the quality of a system with respect to ease of learning, ease of use, and user satisfaction [19]. ISO 9241 defines it as the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments [1].

Certain questions are difficult to answer efficiently with traditional usability evaluation methods [11,15]. For example, if a test participant spends longer than expected looking at a particular screen, without making the appropriate selection to reach a goal, what exactly is s/he doing wrong? Think aloud protocols, questionnaires or interviews afterwards will not always reveal the reason for this failure [11,15]. The test participant might have overlooked the appropriate control; a system component might have distracted him/her; the meaning of the control might not have been apprehended; etc. Different interpretations may lead to different design modifications.

Success rates and completion time metrics can inform designers about when a test participant had difficulties with an
interface, but it can not necessarily inform them what specific areas of the interface caused such problems [3]. The importance of finding objective usability evaluation techniques that can identify trouble areas of interfaces has encouraged researchers to look at how eye movements could be used to comprehend the way that users view, search and process interface information [3].

Traditional usability techniques are fairly effective for evaluating interface usability when tangible tasks are considered [8]. However it is a much more difficult task to evaluate usability when intangible tasks such as “understand data” or “make decision based on information” is considered. This is the type of tasks dealt with when evaluating network management tools.

C. Eye tracking
Eye tracking can be defined as a technique to determine eye movement and eye-fixation patterns of a person [15]. The human eye moves by alternating between saccades and fixations. A saccade is the quick movement of the eye in order to move focus from one area to the next. A fixation is the time spent looking at the newly found area. An eye tracker follows the eye around during its saccades and tracks the location of the fixation points.

Software designers can gain useful information about natural human movements, by tracking eye movements [16]. Eye tracking data can expose response biases of participants resulting from an artificial testing environment. This would be undetected in traditional usability testing techniques and therefore eye tracking data results in a higher validity of usability data [21].

In a study conducted by Stasko et al. the strategies participants used in performing the required tasks, were observed [23]. They state that the task performance was clearly influenced by the strategy employed. If an eye tracker was used in collecting this information, the observers would clearly see where the participant was looking when performing a task. Goldberg et al. states that variables that are derived from eye tracking methods can provide insight into users’ decision making while searching and navigating interfaces [10].

Morrison et al. studied the effects of eye tracking in tactical decision making environments [14]. In order to evaluate the utility of a display, a researcher needs to know what information a test subject is looking for and where he is looking to obtain that information. They claim that this is very difficult to achieve using traditional evaluation methods and that suitable measurement tools are required.

Scan paths, the time spent looking at various areas of interest on the screen and the use of visual attention are just some of the benefits that eye tracking can add to the usability testing of NM tools.

III. FOCUS OF RESEARCH

A. Objectives
Traditional usability methodologies exist as well as several eye tracking methodologies. The question arises if these methodologies can be combined for the evaluation of the user interface of NM tools. Established guidelines exist for the effective design of graphical information, as well as theoretical proposals for how visual information is processed. Traditional techniques can be used to establish the relative usability of graphical information, but eye tracking can add new and interesting insights [18].

The goal of this research is to investigate the added value of eye tracking data combined with usability evaluation data when evaluating the interface of a NM tool.

This research will attempt to answer the following questions:

- How can eye tracking and usability evaluation metrics be combined to evaluate NM tools?
- Will eye tracking give an added value to usability evaluation data when evaluating NM tools?

The hypotheses for this case study are as follow:

- $H_0$: Eye tracking will not give added value to the usability evaluation of NM tools.
- $H_1$: Eye tracking will give added value to the usability evaluation of NM tools.

B. Pilot Study: AppVis
AppVis 1.0 is a NM prototype tool that allows network managers at the Nelson Mandela Metropolitan University (NMMU) to analyse and explore application performance on the network. Application performance management entails the comprehension of how a network application performs from a user perspective [17]. NMMU has an extensive network infrastructure that supports several application services. This prototype system, AppVis 1.0, uses novel visualisation techniques to visualise the application delay performance of the Integrated Tertiary Software (ITS) application implemented at NMMU [17].

![Figure 1: AppVis 1.0 Network Overview screen](image)

AppVis 1.0 contains four different types of information visualisation graphs. This prototype system has not been evaluated extensively [17]. As AppVis was developed at NMMU, it increases the feasibility of such an evaluation. Figure 1 illustrates the Network Overview screen of AppVis 1.0.
The user interface is divided into three co-ordinated views: a graphical view; a textual view; and a data filtering view. The results of the data analysis and exploration are displayed in a graphical and textual form simultaneously based on the filtering criteria set within the data filtering view [17].

The prototype system supports the following tasks [17]:

- Displaying a network overview of application delay;
- Displaying a subnet view of application delay metrics for an individual VLAN;
- Performing trend analysis of application delay metrics for an individual VLAN;
- Filtering application delay metrics based on time period and VLAN and delay statistics; and
- Allows analysis and exploration of application delay metrics with zooming, rotating, panning and details-on-demand features.

IV. METHODOLOGY

A. The Evaluation Methodology

Usability evaluation methodologies suggested by Barnum [1], Dumas and Redish [5], Rosson and Carroll [19], Rubin [20] and Faulkner [7] were studied for this research. Eye tracking techniques by Xu [25], Gao [9], Goldberg et al. [10], Cowen [3], Renshaw et al. [18] and Bennel and Otten [2] were used to combine with appropriate usability evaluation methods. The result was a methodology proposed for the formal usability testing of a NM tool incorporating eye tracking. Table 1 lists the basic steps involved in planning and effectively implementing a formal usability test of NM tools.

<table>
<thead>
<tr>
<th>Step</th>
<th>Step description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Establish the team.</td>
</tr>
<tr>
<td>2</td>
<td>Define the product issues and audience.</td>
</tr>
<tr>
<td>3</td>
<td>Formulate the research hypothesis.</td>
</tr>
<tr>
<td>4</td>
<td>Set goals and define usability measurements.</td>
</tr>
<tr>
<td>5</td>
<td>Define eye tracking metrics.</td>
</tr>
<tr>
<td>6</td>
<td>Establish the user profile.</td>
</tr>
<tr>
<td>7</td>
<td>Select the tasks to include in the test.</td>
</tr>
<tr>
<td>8</td>
<td>Determine how to categorise / analyse results.</td>
</tr>
<tr>
<td>9</td>
<td>Develop and write the test plan.</td>
</tr>
<tr>
<td>10</td>
<td>Prepare the test materials, environment and team.</td>
</tr>
<tr>
<td>11</td>
<td>Recruit the test participants.</td>
</tr>
<tr>
<td>12</td>
<td>Conduct a pilot test.</td>
</tr>
<tr>
<td>13</td>
<td>Conduct the usability test.</td>
</tr>
<tr>
<td>14</td>
<td>Tabulate and analyse the data.</td>
</tr>
<tr>
<td>15</td>
<td>Recommend changes.</td>
</tr>
<tr>
<td>16</td>
<td>Report the results.</td>
</tr>
</tbody>
</table>

Table 1: The proposed methodology

B. Measures

The following standard usability metrics were used in the pilot study:

i) Effectiveness:

- **Task completion rate.** This metric consists of determining the percentage of tasks each participant completes successfully in the task list. This will include the percentage of tasks completed per participant and per task.
- **Number and percentage of tasks completed with and without assistance.** This metric will show if the participant completed a task with or without the assistance of the team members.

- **Error rate recovery.** This metric consists of monitoring the number of errors made by the user, as well as the total errors from which the user could not recover.

ii) Efficiency:

- **Task completion time.** This metric involves measuring the total time that participants spend performing the assigned tasks.
- **Real-time events.** This metric involves monitoring and filtering events such as the click of a mouse, the push of a key or the participant writing down his answer to a task list question.

iii) Satisfaction:

A subjective ratings scale and a modified version of Questionnaire for User Interaction Satisfaction (QUIS™) were used to measure the satisfaction. This post-test questionnaire contains questions concerning the overall satisfaction, screen design, terminology, learnability and system capabilities of the NM tool.

The eye tracking metrics used for this case study are discussed briefly below:

- **Number of fixations.** The number of fixations is negatively correlated with search efficiency. Large numbers of fixations point to less efficient search perhaps resulting from poor display element arrangements.

- **Fixation duration.** Relatively long fixation duration is an indication of the complexity and difficulty of a display.

- **Number of fixations on each area of interest (AOI).** This metric is an indication of the importance of a system element.

- **Number of gazes on each AOI.** The eyes are drawn to informative areas. This metric also reflects the importance of a system element.

- **Scanpath.** The strategy that a participant uses to complete a task can be obtained from this metric. We can get an indication of the efficiency of the arrangement of elements in the user interface.

- **Time to the 1st fixation on the target AOI.** This metric is useful for tasks where a precise search target exist.

The task list for this case study was designed in such a way as to allow these metrics to be captured.

C. Participants

The main criterion for the users of this product is to have a sound knowledge in the domain of network performance management. A background questionnaire was used to screen the participants for this evaluation. This questionnaire reflects the possible participant’s NM tool experience, computer experience, age and gender. Questions regarding the participants’ eye sight were also asked in the questionnaire. The pilot study showed that a participant wearing glasses or contact lenses does not affect the accuracy of the eye movement data. Six participants were tested for the pilot study. Of the six participants, one wore contact lenses and one wore glasses. Their right eye was calibrated for the eye
tracking purposes. All participants claimed to have worked with between one and three NM tools. The participant profile is included in Table 2. Six to Eight participants will be used in the case study for this research.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Total</th>
<th>Male</th>
<th>Female</th>
<th>20-29</th>
<th>30+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Years Experience</th>
<th>Total</th>
<th>&lt;1</th>
<th>1-3</th>
<th>4+</th>
<th>NM tools Used</th>
<th>1-3</th>
<th>4+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2: Profile of the participant population

D. Apparatus
The SMI iView X RED eye tracker was used for this experiment. This remote eye tracker was developed for absolutely contact-free measurement of eye movements including automatic head-movement compensation. The eye tracker was placed directly in front of the participant just below the display screen. The video files and data files were recorded using the iView X software. The data was saved for later analysis. Fixations were defined as being at least 250 ms in duration in a radius of 50 pixels.

Participants were seated approximately 60cms from the screen in a comfortable chair allowing minimum movement. Tasks were read out loud to the participant as to eliminate the participant from looking down on a piece of paper. Several tasks required an answer from the participant. These answers, the tasks read, as well as any comments made by the participants were recorded.

E. Task Scenarios
A training task was given to a participant in the form of a system brief and a PowerPoint presentation. The participants were briefed about the system goals and objectives. This was followed by a PowerPoint presentation displaying the type of graphs of the system so that the participant could obtain a general look and feel for the system but not giving away too much information to be evaluated.

The experimental tasks are divided into two scenarios each containing several tasks. These tasks represent typical activities that would be conducted using the software. The following scenarios were evaluated for the purpose of the pilot study:
- Import application delay metrics; and
- Display a network overview of application total delay.

F. Procedure During Test
The participants were welcomed and briefed about the experiment, which was followed by an explanation of the equipment to be used. It was explained that only the eye, voice and stimulus display would be recorded. The participant was required to also complete an informed consent form. The think aloud protocol was explained to the participant and was encouraged to use it.

After the training tasks the participants were given time to make themselves comfortable in front of the PC before the eye tracking calibration commenced. A 9-point calibration with corner correction was used at all times. The participants were asked to keep their head as still as possible during the experiment as to minimise inaccuracy caused by head movements. Participants were offered the opportunity to stand up and relax half-way through the experiment. After every three or four tasks, depending on the task length, the accuracy of the eye movements was checked. If the accuracy would appear to be out, the participant’s eye would be recalibrated. Data recording commenced with the test administrator reading the task, and ended with the participant either answering or completing the task. The duration of the experiment was between 40 minutes and one hour. Following the tasks, a post-test questionnaire was administered. Finally an experiment debriefing was given followed by the participants being thanked for their time.

V. ANALYSIS AND RESULTS

A. Data Collection
Data was collected and calculated by means of:
- Video recordings, one file per task, were captured live and evaluated at a later stage. The video recording included a cursor which indicates the participant’s eye movements. Audio in the form of the participant or the test administer speaking were also included in the video files.
- Eye tracking data files.
- A preference questionnaire was used to gather feedback using a modified version of the Questionnaire for User Interaction Satisfaction (QUIS™).
- Monitoring of tasks.

B. Usability and Eye Tracking Measures
The usability measures results obtained from the preliminary analysis are summarised in Table 3. All tasks were performed correctly, except for task 2.6 where one participant did not complete it correctly and task 2.7 where one participant did not know how to complete the task and asked for assistance. Task completion times were also recorded for each task. Each task together with useful eye tracking measures obtained are explained next.

<table>
<thead>
<tr>
<th>Task</th>
<th>Mean Task Completion Rate</th>
<th>Mean Task Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Task 2.1</td>
<td>100%</td>
<td>n/a</td>
</tr>
<tr>
<td>Task 2.2</td>
<td>100%</td>
<td>n/a</td>
</tr>
<tr>
<td>Task 2.3</td>
<td>100%</td>
<td>n/a</td>
</tr>
<tr>
<td>Task 2.4</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Task 2.5</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>Task 2.6</td>
<td>83%</td>
<td>83%</td>
</tr>
<tr>
<td>Task 2.7</td>
<td>83%</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Table 3: Usability metrics

Task 1
The eye tracking data obtained from the first task allowed the researcher to explore information otherwise not available through traditional usability techniques. The participant was required to start the system and to determine whether new
delay metrics have been collected. Participants completed the tasks correctly and in time and also provided the correct answers. Usability tests would usually suggest that not much can be done to improve this interface. However, Figure 2 illustrates the fixations of a participant on the different AOIs on the screen. The System Information AOI is much smaller than the System Logo and Name AOI, but it has more fixations than the last one. The eye tracking data would suggest that the system designers should place emphasis on the System Information part of the screen by making it bigger and presenting the information in a more clear and uncluttered way.

Figure 2: Fixations on AppVis 1.0 Introduction Screen

Task 2.1
Task 2.1 required the participants to view the interface for a fixed amount of time. Participants showed the main interest in the area of the screen containing the graph, as it had the most fixations, biggest fixation percentage and biggest gaze percentage.

Task 2.2
This task was completed successfully by selecting the threshold option from the menu and then by adding a new threshold. For example, one participant took 26 seconds before he fixated on the target AOI (menu) and another four seconds before selecting the option. Another participant took only five seconds for his first fixation on the target AOI but failed to realise that it was correct. After 16 seconds, another fixation was made on the target and another two seconds before selecting the option. The scanpaths from the different participants indicated that they searched in the filtering area for the completion of this task. This data suggests that the option of adding a threshold should be available from the filtering area as well.

Task 2.3
Task 2.3 requested the participant to change the graph to a 2-Dimensional view and was completed with relevant ease. The scanpath data showed that the participants went to the target AOI in a short amount of time.

Task 2.4
Task 2.4 required the participants to determine which VLAN had the highest maximum total delay. This information could be extracted from the graph or from the textual view. The number of fixations, fixation percentage and gaze percentage favored the graphical AOI. All participants fixated several times on the graph and not once on the textual view. All participants gave the correct answers.

Task 2.5
All participants could not give the correct answer for task 2.5, where they needed to determine which VLAN had the smallest total delay. Figure 3 illustrates the fixations of a participant for the completion of task 2.5. The participants searched for the VLAN with the smallest mean total delay indicated by the blue circles. However, the VLANs with no delay are the smallest and are not indicated as blue circles on the graph. Participants could not tell this from the graph. The participants had a mean total of 21 fixations on the graph and still gave the incorrect answer. Thus, it was difficult to extract the information from this graph.

Figure 3: Fixations for task 2.5

Task 2.6
Task 2.6 required the participants to display “details-on-demand” for a specific VLAN, which is an information box that pops up on the screen if the mouse is focused on a specific VLAN. This term was explained to the participants during training, but they still took exceptionally long to complete this task. The mean time for the first fixation on the target AOI was 27 seconds.

Task 2.7
Participants had to drill down into a specific VLAN which resulted in a new graph being displayed to complete this task. Double clicking on the VLAN in the filtering area is one of several ways in which to achieve this task. One participant searched the filtering area several times but could not achieve the task. After searching through the graphical and textual AOIs the participant asked for assistance. The task could also be completed by means of the toolbar, but only one fixation was made on the toolbar. Figure 4 shows the AOI order of this participant. The coloured bars indicate the fixations made. Another participant completed the task easily once he first fixated, after seven seconds, on the filtering area.

Figure 4: AOI fixation order for participant
C. Questionnaires
The results of the modified version of Questionnaire for User Interaction Satisfaction (QUIS™) were analysed by calculating the mean across all participants for the overall rating that was given for each usability criteria. The mean ratings calculated for each usability criteria indicated a positive response by participants towards the prototype system. A mean rating of 4.1 was calculated for the screen design indicating a high degree of satisfaction.

D. Limitations of Pilot Study
The pilot study conducted had the following limitations:

- Only a limited data set was analysed. The case study will have eight to ten participants completing a comprehensive set of tasks. The analysis of that data should yield more significant results.
- The participants used in the pilot study did not have prior knowledge of the NM tool used. The case study will use participants who have used the system before.
- Excessive head movements during an eye tracking experiment can cause the eye tracking data to be incorrect. However, participants’ eye movement accuracy was tested against a fix set of points after every three or four tasks. None of the participants had this problem.

VI. CONCLUSIONS AND FUTURE WORK
There has been an increase in the use of NM tools by network managers to effectively manage large networks. The need to make informed and quick decisions has increased, yet the ability to assess and analyse information has declined. This is as a result of the use of sophisticated information visualisation techniques in such NM tools. Traditional usability evaluation methods have been employed to evaluate these graphical reports. With the increased use of information visualisation techniques, a need has arisen to combine other evaluation techniques with traditional usability evaluation methods. This study proposes that by adding eye tracking evaluation to traditional usability evaluation methods, value will be added to the evaluation of NM tools. Not only will such a combined method offer the opportunity to measure user actions, but to also record eye movements which are critically important when dealing with network management.

AppVis 1.0 was selected as the NM tool to be evaluated in the case study. Pilot studies have been conducted. The proposed methodology has proved to work successfully. Analyses and results of the pilot studies indicate that eye tracking does give added value to the usability evaluation of NM tools. Future work will include further tests and more detailed analysis of the results.

VII. ACKNOWLEDGEMENT
The authors would like to thank the Telkom Centre of Excellence Programme and the Department of Computer Science and Information Systems at the NMMU for making this research possible.

REFERENCES

Marco C. Pretorius received his BSc and BSc(Hons) degrees at the University of Port Elizabeth now known as the Nelson Mandela Metropolitan University. He is presently completing a Masters degree in Computer Science at the Nelson Mandela Metropolitan University. His current research involves the investigation of eye tracking as an added value to the usability evaluation of Network Management tools.