Regressions Re-visited: a New Definition for the Visual Display Paradigm

James A Renshaw
Leeds Metropolitan University
Beckett Park
Leeds LS6 3QS,UK
t.renshaw@leedsmet.ac.uk

Janet E Finlay
Leeds Metropolitan University
Beckett Park
Leeds LS6 3QS,UK
j.finlay@leedsmet.ac.uk

David Tyfa
University of Huddersfield
Queensgate
HD1 3DH,UK
d.tyfa@hud.ac.uk

Robert D Ward
University of Huddersfield
Queensgate
HD1 3DH,UK
r.d.ward@hud.ac.uk

Abstract
We revisit the definition of regressions in eye tracking, having found existing definitions, formulated within a reading paradigm, unsuitable for visual display assessment. The new definition is tested using eye movement data recorded during a usability evaluation of two series of graph designs. The new definition gave a stronger result in the usability assessment than was obtained using the previous measure, and gave a result consistent with other eye movement usability metrics derived for the two series of graph designs. The newly defined measure is easily computed and is sensitive enough to reveal a correlation between improved performance and a decrease in regressions.

Categories & Subject Descriptors:
H.5.2 [Information Interfaces and Presentation (e.g. HCI)]; User Interfaces—Evaluation/methodology, screen design; H.1.2 [Models and Principles]: User/Machine Systems—Human information processing.

General Terms:
Human Factors; Measurement; Performance.

Keywords:
Regression; eye tracking; usability; evaluation; design.

INTRODUCTION
The reading process is complex and not as smooth and continuous a process as one's perception leads one to believe. Reading is undertaken by means of a series of fixations; instances when the eye remains relatively still within a particular location; and saccades, jerky ballistic movement of the eye to the next fixation location. Sometimes whole words can be skipped [4]. On occasions the reader reverses direction to re-fixate on a portion of the text in order to re-read it, only then to resume reading again and it is to these instances of reversal of direction that the term regression is applied. There are few studies in non-reading environments involving the use of regressions as a means of measuring usability. However, Yamamoto and Kuto [6] studied the eye movements of participants as they viewed displays of rows and columns of data, in order to determine an optimum arrangement for the display. They measured the proportion of regressions made and were able to demonstrate a relationship between an increase in the proportion of regressions and a decrease in task performance.

Later Goldberg and Kotval [3] describe an experiment, in which the usability of various arrangements of tool icons on a screen, were assessed by several eye movement metrics, of which regression counting was one, and compared with usability ratings made by a panel of designers. They offered the following as a definition of a regression:

> ".. a backtrack [regression] can be described by any saccadic motion that deviates more than 90° in angle from its immediately preceding saccade. These acute angles indicate rapid changes in direction due to changes in goals and mismatch[es] between users' expectation and the observed interface layout."
>
> [3]

They found that the number of back tracks correlated well with the usability ratings of the designs as judged by the panel. However, it is not clear how the back tracks were detected, or how they were counted.

It is our contention that for visual displays, and particularly for non textual displays, the definition above is not adequate and that a definition of more precision is needed in which the relative position of the first and last fixation of the regression is specified. For example, in the case of long saccades, under the current definition, the location of the final fixation could differ widely from the first, despite there being an acute angle between the saccades linking them, and there would be uncertainty as to whether this was a mere shift in interest or a revisit to an area of uncertainty similar to the regressions experienced in reading.

As for the cause of regressions, other workers have suggested that the incidence of regressions in reading may relate to reading ability and/or the semantic difficulties of the text or its textual and typographical values [6, 4,] or participant fatigue [1]. It is also likely that many regressions could be attempts to correct a saccade that overshot a target or oculomotor errors [4]. These problems may also be present in the viewing of visual displays and their influence will be manifest in eye movement based usability assessments. Consequently their measurement is important.

The purpose of this paper is to propose a tighter definition of regressions for adoption in a non reading context and test its validity as a means of evaluating usability. The sensitivity of the new measure is further assessed by examining the pattern of regressions over time and their relationship to task performance.

FORMULATION OF AN IMPROVED DEFINITION OF REGRESSION

A regression may be considered as being made up of three consecutive fixations, (see Figure1a-1b), one occurring at the beginning (point A), one at the point at which the backward path of the regression starts (point B) and the final fixation at the point (point C) to which the regression has been made. In reading; regressions occur on the same line so their start and end points are more or less defined; in scene perception it is not quite as easy.

In formulating the new definition certain factors had to be considered:

1. The context of viewing a visual display is one in which formal rules of reading do not always apply particularly when the use of words is sparse or non existent.
2. The effective area of high visual acuity is between 2-3^{o} of visual angle [2].

As a result of the first factor (1) it is important to state, within the definition, that there must be a return to an area within reasonable proximity to the source of the uncertainty causing the regression. The consequences of failing to do so are demonstrated in figures 1a and 1b, where, under the current definition, the third fixation (point C) could be incorrectly counted as a regression. Figure 1b shows that under the proposed definition such instances would be excluded from such a count. Factor 2 provides the rationale behind specifying that the distance between the first and last fixations should be the equivalent of 2^{o} of visual angle.

Consequently the proposed definition of a regression is:

> That the final fixation of a regression, is within a specified radius (up to 2^{o}) of the first fixation and that the angle between the first saccade and the second saccade must be less than 90^{o}.

Not only is this revised definition more precise but it also enables the number of regressions in data sets of contiguous fixations to be computed easily.

ASSESSMENT OF NEW DEFINITION

Data from an experiment, in which eye movements had been recorded whilst participants looked at two forms of line graph was used to test the new definition and these results compared to those obtained using the previous definition.

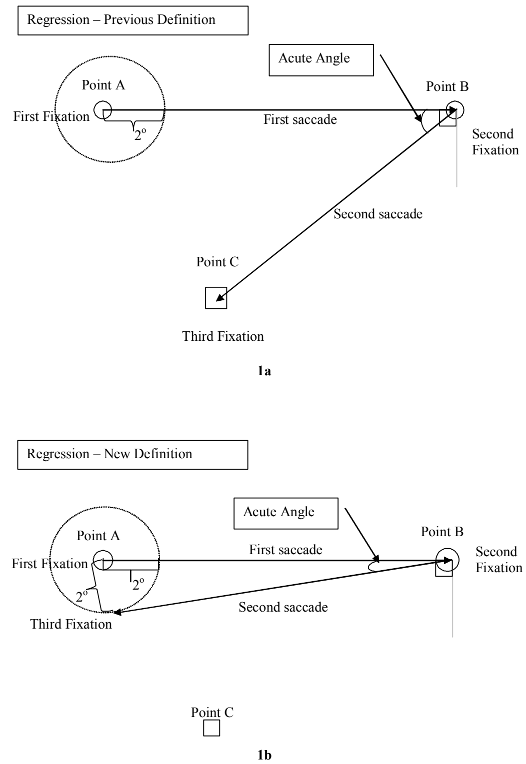

Figure 1a: A Regression as per Current Definition

Figure 1b: A Regression as per the New Definition

The data derived from an experiment in which participants viewed two series of eight graphs. One series had been designed to conform to graphic design guidelines (graph 1) whilst the other had flouted them (graph 2) [5].

The graphs were created using standard Microsoft Excel graph facilities. Both displayed plots of the same variables: sales, costs and profit. But the design and the proximity of the legend area to the data differed considerably; with design compliant graph (graph 1) having a simpler legend, in close proximity to a two dimensional display of the data, whereas the second graph design (graph 2), incorporated a superfluous third dimension and a more complex legend, remote from the variables to which it related.

In previously reported results [5] it was found that graph 1 design was the more usable evidenced by the fact that the set task was completed in less time, more accurately and with greater user satisfaction. These properties were evaluated through conventional means: timed tasks and questionnaires, supported by eye metrics after [3] but the results excluded a count of regressions.

In a further analysis, a temporal examination of the data was undertaken, to establish whether learning or familiarity effects could be detected using the regression measure.

METHOD

Experimental Design

In the experiment the eye movements of participants were tracked whilst they carried out tasks using the two contrasting styles of graph. Each graph was designed using standard Excel XP graphic facilities, and featured the same variables: sales, costs and profit on the vertical axis, with time on the horizontal axis.

The experiment was a repeated measures design with each participant viewing 8 graphs of each type of design format. Graphs were viewed one at a time in a random sequence until all 16 graphs had been displayed. The participants were asked to answer one question on each graph displayed, by means of clicking on the relevant area of the graph with a mouse. This signalled the end of the eye tracking data collection for that particular graph. After having viewed all the graphs they rated the usability of the two types of graph on a scale of 1 (low) to 5 (high) to complete the experiment.

Participants

Volunteer participants (15 male, 9 female) took part in the experiment. Their ages ranged from 20 to 55 and they were drawn as an opportunity sample from the mathematics, computing, multimedia and health faculties of two universities. Sixteen wore no corrective lenses, five wore glasses and three wore contact lenses. Two declared themselves to be novices in the use of graphs; the remainder were quite familiar: three stated they used graphs very frequently, thirteen frequently, eight hardly ever.

Apparatus and Materials

The experiment used an ASL504 pan/tilt eye tracker system capable of detecting a bright back lit pupil image caused by the retinal reflection of a near infrared beam of light emanating from the eye tracker, which was placed immediately in front of the participant just below the display screen. The resolution of the eye tracker is better than one degree and it has a sampling rate of 60 Hz. The data was recorded using GazeTracker software, capable of recording fixation data and mouse clicks. ASL Eyepos software was used as a back up method of recording information on a second PC, in addition to the video recording of eye movements by means of a conventional video recorder. Participants were seated approximately 60 cm. from the screen in a high backed chair – no other restraining mechanism was used. The graphs were shown on a 17 inch PC colour monitor with a screen area of 1024 x768 pixels. The images were displayed at random. Time stamped fixation and mouse click data was exported to spreadsheets for subsequent analysis.

RESULTS AND DISCUSSION

The question to be addressed is whether the newly defined regression metric would yield results comparable to those derived from the previous definition.

The proportion of regressions was determined, under the old definition, by computing the angle between each successive saccade and counting the incidences of acute angles. An almost identical number of pairs of saccades were counted (3729 for graph 1 and 3732 for graph type 2) and a t-test established no significant difference between the graph types ($t(23) = 0.179$, $p>0.05$).

However, using the new, more precise, basis of computing regressions, a significant difference was detected (mean proportion of regressions for graph 1 = 14.61%, graph 2 = 17.25%, $t(20)=2.48$, $p<0.05$), which may be interpreted as indicating graph 1 to be more usable than graph 2: a result compatible with the other eye measure metrics reported previously [5].

The next task was to emulate the work of [6] and test whether there was evidence of an improvement in the performance over time. A Pearson correlation test on the completion times and the display sequence of both graph types was undertaken. It showed no significant correlation for either graph type (graph 1: $r = -0.2$, $p= .44$, $n=16$, graph 2: $r = -0.475$, $p=0.63$, $n=16$). This may indicate that the differences in performance times were not sensitive enough to detect performance improvements.

An examination of regressions however, yielded a stronger result. The new basis of defining regressions was used to count the number of regressions in the legend area: selected because its design remains the same, in terms of content and position, through out the experiment for both graph types. A Pearson correlation test between completion times and the proportion of regressions in the legend area showed a strong and significant correlation for graph 2 but not for graph 1(graph 1 $r = 0.267$, p 0.318, $n=16$, graph 2: Pearson correlation $r = 0.6$, $p = 0.013$, $n=16$,).

A plot of the relationship is shown in Figure 2.

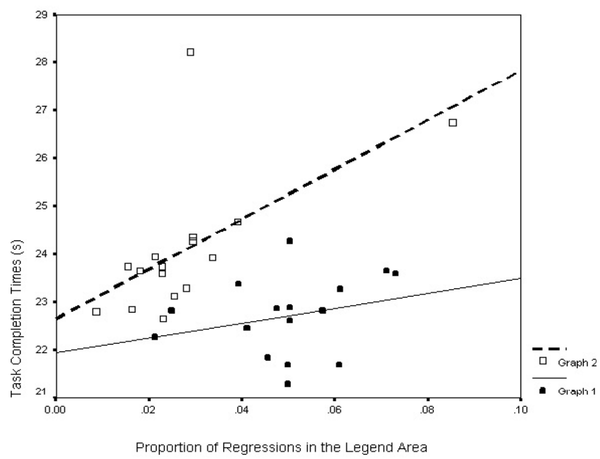

Figure 2: Plot of task Completion Time vs Proportion of Regressions in the Legend Area

Despite the lack of a significant correlation for graph 1 it is still interesting to note the differences between the two designs. Figure 2 shows that, for graph 2, performance is improved as the number of regressions decreases. There is no similar correlation for graph 1. Furthermore it will be observed that the mean proportion of regressions generally for graph 2 is lower that that for graph type 1. A repeated measure t-test confirms this difference to be significant (graph 1 mean proportion = 0.05, graph 2 mean proportion = 0.03 , $t(15) = 4$, $p<0.05$). At first glance this is a surprising result, for more difficult scenes were anticipated to induce more regressions. However, a transaction matrix analysis of the fixations established that participants abandoned the use of the legend area in graph 2 as time went on, and, as a consequence, lowered the probability of there being regressions, hence the lower proportion of regressions for graph 2.

CONCLUSION

In this paper a more precise definition of regressions has been compared with a previous definition and found to be more sensitive to differences in usability; yielding results similar to those obtained using other established eye movement metrics. The previous definition was found wanting because it failed to be sensitive enough to enable the detection of the difference in usability of the two graph types. Additionally the new measure enabled rapid computing of the number of regressions, avoiding lengthy visual inspections. An analysis of the regression data has indicated that the higher proportion of regressions in the legend area of a graph was related to a lower performance in terms of task completion times confirming findings by other researchers.

The tighter definition suggested here will enable further work to be undertaken into the causes and effects of regressions providing greater insights into the effects of designs and environments on human computer interactions.

REFERENCES

1. Bruneau, D., Sasse, M.A., McCarthy,J. (2002) The Eyes Never Lie: The Use of Eye Tracking Data in HCI Research *Proceedings of the CHI2002 Workshop on Physiological Computing*.

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

3. Goldberg, J. H. and Kotval, X. P. (1999) Computer Interface Evaluation Using Eye Movements: Methods and Constructs *International Journal Of Industrial Ergonomics,* **24,** 631-645.

4. Rayner, K. (1998) Eye Movements in Reading and Information Processing: 20 Years of Research *Pychological Bulletin,* **124,** 372-422.

5. Renshaw, J. A., Finlay, J.E., Ward, R.D.,and Tyfa,D (2003) Designing for Visual Influence: An Eye Tracking Study of the Usability of Graphical Management Information *Human-Computer Interaction INTERACT '03,* **1,** 144-151.

6. Yamamoto, S., Kuto,Y. (1992) A Method of Evaluating VDT Screen Layout by Eye Movement Analysis *Ergonomics,* **35,** 591-606.