Combining Keystroke Logging with Eye Tracking

Andersson, Bodil; Dahl, Johan; Holmqvist, Kenneth; Holsanova, Jana; Johansson, Victoria; Karlsson, Henrik; Strömqvist, Sven; Tufvesson, Sylvia; Wengelin, Åsa

Published in:
Writing and Digital Media

2006

Document Version:
Early version, also known as pre-print

Link to publication

Citation for published version (APA):
Andersson, B., Dahl, J., Holmqvist, K., Holsanova, J., Johansson, V., Karlsson, H., Strömqvist, S., Tufvesson, S., & Wengelin, Å. (2006). Combining Keystroke Logging with Eye Tracking. In L. van Waes, M. Leijten, & C. M. Neuwirth (Eds.), *Writing and Digital Media* (pp. 166-172). (Studies In Writing; Vol. 17). Elsevier.

Total number of authors:
9

COMBINING KEYSTROKE LOGGING WITH EYE TRACKING

BODIL ANDERSSON*, JOHAN DAHL**, KENNETH HOLMQVIST***, JANA HOLSANOVA***, VICTORIA JOHANSSON**, HENRIK KARLSSON*, SVEN STRÖMQVIST**, SYLVIA TUFVESSON** AND ÅSA WENGELIN**[1]

**Oribi AB Lund*
***Department of Linguistics, Lund University*
****Department of Cognitive Science, Lund University*

INTRODUCTION

The aim of this paper is to show how keystroke logging can be combined with eye-tracking, to enhance the study of writing processes in subjects of different ages and different writing skills.

Producing a text involves an interplay between several mental processes such as planning, encoding, monitoring and revision (e.g., Hayes and Flower 1980). Production-rate data derived from key-stroke logging offer a window on these processes. In our previous research (see, e.g., Wengelin, 2002; Strömqvist et al., 2004), we have explored this window by means of ScriptLog (Strömqvist & Karlsson, 2002), a tool for research on the online process of writing which we are continuously developing. ScriptLog keeps a record of all events on the keyboard (i.e., the pressing of alphabetical and numerical keys, cursor keys, the delete key, space bar etc, and mouse clicks), the screen position of these events and their temporal distribution (the time that elapses from one key-board event to the next). From a ScriptLog recording of writing activity you can analyse not only the final edited text with its lexical and grammatical aspects and global content structure, but also the "linear" text with its temporal patterning, pauses and editing operations. Many of these patterns, however, are ambiguous with respect to different interpretations. For example, keyboard inactivity, just like silences in speech, can be indicative of the process of planning the continuation of the text. But keyboard inactivity may also be indicative of reflexion

[1] *Authors in alphabetical order*

Luuk Van Waes, Mariëlle Leijten & Chris Neuwirth (Editors), Writing and Digital Media.

and the monitoring of text already produced. In this context, additional data on the distribution of visual attention in the writer can serve a disambiguating function (Holmqvist et al., 2002).

In order to forward this line of investigation, we have combined the keystroke logging program ScriptLog with the eye movement technology iView X HED + HT. Our proposed methodology provides a powerful yet non-intrusive way of getting closer to the textwriting subject. The methodology provides a particularly valuable window on those phases of the text-writing process which necessitate visual feedback, namely, instances of monitoring and revision demanding that relevant parts of the emerging text actually be read. The present paper explains the technical aspects of our method in greater detail as well as an analysis tool that assists the analyst in analysing the vast amount of data produced by keystroke logging and eye tracking. We also present selected examples of our analyses of writers from different age groups, some of whom have reading and writing difficulties.

OUR SYSTEM: SCRIPTLOG + SMI IVIEW

An eye-tracker is simply a device to measure where people are looking. There are several eyetracking technologies, but this text will only discuss infra-red video systems. In order to use an eye-tracker in combination with ScriptLog, several parameters need to be considered.

First, you need to consider which kind of data you are interested in. *Video data* show the field of view of the writer with a small circle indicating the point of gaze. The video may also contain sound. Video is typically used for the presentation of the recording situation. Furthermore it can be used for retrospective interviews; i.e. the video is shown to the writer immediately after the data collection and the writer is interviewed about his writing and visual behaviour. However, videos are very time comsuming to analyse and code for quantitative analyses.

Data coordinates are much easier to work with, but more difficult to collect. Basically, coordinates presume a coordinate system with dimensions. The writer's computer monitor is a good candidate for a coordinate system. It must be noted, however, that the eye-tracking coordinates of the monitor will differ from the pixel coordinates of the monitor itself. Assuming that a translation between the two coordinate systems can be made, a second problem arises: The eye-tracking data refer to the screen as it looked at the time of writing. If we want to know which word the writer looked at at a specific time, ScriptLog must be able to reproduce the screen *exactly* as it looked at the time of writing.

Writers do not look at exclusively at the screen. Most writers spend a considerable time looking at the keyboard. Eye-tracking data should include fixations on the keyboard, which could be part of the coordinate system of the monitor, or be a plane of its own (as in our set-up). When the task is to write about a picture, it is interesting to know when and where the writer looks at the stimulus, and that calls for a third plane.

To record data coordinates in a coordinate system of the stimulus, either a remote eyetracker or a headmounted eye tracker with headtracking can be used. A remote system requires the writer to sit very still while writing and therefore we chose to use a headmounted eye tracker with head tracking (SMI iView X HT). An eye camera and a scene camera are placed on a bicycle helmet worn by the writer. On top of the bicycle helmet, there is a magnetic sensor that keeps track of the head in 6D:

position and direction. The eye tracker calculates a vector for the gaze direction that eminates from the eye (in the head). The head position and direction together with the eye direction allow for a real time calculation of the position where the combined eye-head vector hits the monitor plane, or the keyboard plane, or the stimulus picture plane.

In order for this to work, the environment in which the recording takes place must be measured. This needs to be done onlye once, in advance of a series of recording sessions. A simple type of virtual reality model of the monitor, keyboard and picture is created, which tells the eye-tracker their positions in space, and what extension they have. Since data are recorded in the co-ordinate system of this virtual model, the keyboard, monitor and stimulus picture may not move away from the measured position. Consequently, we have fixed all these planes to the table, and indirectly to the floor.

It is important to place all planes within as small a visual angle as possible. An eye tracker covers approximately 60 degrees of visual angle for a writer who does not turn his/her head. If the writer looks further away, data will be partially lost.

When all the coordinates systems are in place, we are able to get precise data on where the writer looks. But we also want to know *when* s/he looks at the monitor, in order to, for instance, find out whether s/he looks up at the monitor more often when s/he has just concluded a clause than when s/he has concluded a word. Synchronisation of data in our system is achieved by letting ScriptLog send a start signal to the eye tracker when a ScriptLog recording is started.

When the system is set up, a recording typically starts by placing the helmet on the writer and adjusting the camera. The writer is then asked to sit down by the keyboard in the recording environment. On the monitor, 13 calibration points are shown in a certain sequence, and the writer is asked to look at the points in that order. The eye tracker analyzes the video signal of the eye in each position and builds an interpolation matrix that allows us to get precise data also between points. The interpolation matrix resulting from the calibration is specific to the writer's individual eye; and calibration should be made at least once for each recording.

A calibration of the scene camera follows immediately. This procedure allows the eye-tracker to output video in parallel with data coordinates. The recording then starts. In order to achieve high data quality, careful positioning of the eye camera is necessary before calibration and recording. If the writer uses glasses or lenses, the reflexes from infra-red light in these may cause disturbances. Lighting conditions must be chosen with care. It is also advantageous to have filtering options on the eye video processing in the eye tracker. Settings for the eye video processing may have to be monitored continuously for some subjects.

For each writer and each text, the eye tracking part of our set-up outputs

- An MPEG-2 video of the visual field with overlaid gaze cursor, timestamp and sound
- A data file which gives the following data for every 20 milliseconds: Plane number, gaze coordinates in the coordinate system of that plane, head position and head orientation, eye position, and time.

DATA ANALYSIS: AN EXAMPLE

As was mentioned earlier one type of output from a writing session with ScriptLog is the *linear file*. The linear file presents every event during a writing session, and includes pauses of any minimum duration defined by the analyst. An example of a linear file produced during a picture elicitation experiment is shown in Fig 1.

Insert fig 1 here

This linear file shows that the writer paused four times while writing this sentence, but it doesnt't tell us anything about *what* the writer did in those pauses. We will now proceed to show how additional information about gaze behaviour can help us refute hypotheses about what was going on during the pauses in the production of this text fragment.

For the analysis of the interaction between writing and gaze behaviour, we have developed an analysis tool, inspired by the so-called multimodal time-coded score sheets developed by Holsanova (2001). The tool is used for the analysis of temporal and semantic synchrony in picture viewing and picture description. A merged time-coded data file produced by ScriptLog and iView provides a visual representation of the verbal and visual flow. This representation gives the analyst an enhanced picture of the writer's attention processes: Which objects or areas were scanned visually and which objects or areas were described verbally at a certain point of time? The tool provides an overview of how the writers' attention was distributed between the stimulus picture, the keyboard, the computer monitor and elsewhere during the writing session. In addition we have implemented what we call a writing filter which specifies the subject's writing activities. Let us discuss an example.

The text fragment in the example was produced by a female university student who participated in an experiment where the task was to describe a detailed picture from a childrens' book in writing. During the first 12 minutes of the writing session the writer has given an overview over the picture's main characters and their activities and described the animals and plants present in the picture. She is just about to start the last quarter of her picture description where she will focus on minor details.

The linear file (Fig. 1) shows that she has just made some revisions to her text and started a new paragraph (<RETURN>). The first sentence in the new paragraph is "Mitt på bilden finns det också kor som är bruna och vita." ('*In the middle of the picture, there are also cows that are brown and white'.)* A first glance at the linear file reveals that her writing is interrupted by pauses. Let us turn to the visual presentation format in Fig. 2 to find out how her visual attention was distributed.

Insert fig 2 here

The visual presentation format consists of five horisontal synchronized tiers. The tiers enable us to analyse how the visual attention is distributed over the areas in the writer's experimental environment during the writing session.

In the presentation format time progresses from left to right. The bottom tier contains the stream of keystroke-logged text-writing where the distance between the keystrokes shows the writing speed (the shorter distance between the letters the faster the writing). The second (blue) tier from the bottom defines writing activities (based on a changeable predefined pause criterium; here 1 second). The third (green), forth (pink) and fifth (yellow) tiers represent the distribution of the subjects' visual attention during writing. The green slots mark time when the subject was looking at the keybord, the pink when she was lookin at the screen, and the yellow finally when she was looking at the stimulus picture. We can also observe the flow of simultaneous perception and production, e.g. when the subject is looking at the keyboard while writing (intersection of the blue and green slots) or when the subject is looking at the monitor while writing (intersection of the blue and pink slots). Another advantage is that we can determine with greater accuracy what happens in a pause, e.g. whether the person is scanning the picture, re-reading her own text, or looking for a key on the keyboard.

The fragment shown in Fig. 2 reveals that the subject - after a pause - begins a new sentence with quite constant speed: “Mitt på bilden finns det också [...]” 'In the middle of the picture there is also [...]' In the beginning of this sentence, the subject tends to look at the keyboard whereas later on, she mainly looks at the monitor when writing. Before naming the objects and writing the word “kor” 'cows', this person stops and makes a short pause (2.02 sec) in order to revisit the stimulus picture. She may, for example, be checking whether the cattle in the picture are cows, bulls, calves or perhaps sheep. After the visual information gathering is accomplished, she finally writes “kor” 'cows'. After writing this, she distributes her visual attention between the screen and the keyboard, before she finishes the sentence by writing '”som är bruna och vita'” ‘that are brown and white”.

Our analysis tool can be used as a point of departure for various kinds of analyses, such as

- correlation studies between for example pause durations or pause locations and gaze behaviour
- comparative studies of visual attention between different groups, such as experienced-unexperienced writers, or good-poor writers;
- qualitative case studies, for example of how writers integrate semantic information gathered from text and from picture during writing, text revisions and editing

To sum up, the combination of keystroke-logging and eye-tracking provides a powerful means for analysing the dynamic interplay on-line between production and perception during writing. This complex interplay has hitherto been little studied. Further research along the lines presented here will help determine the nature of this interplay and how it is shaped by different kinds of subjects and different kinds of writing tasks.

REFERENCES

Hayes, J. R. and Flower, L. (1980) Identifying the organisation of the writing process. In L. W. Gregg, and E. R Steinberg,. (eds) *Cognitive processes in writing*, p. 3-30. Hillsdale, NJ: Larence Erlbaum Associates,

Holmqvist, K., Johansson, V., Strömqvist, S. and Wengelin Å (2002) Studying reading and writing online, in S. Strömqvist (ed). *The diversity of languages and language learning*, Lund University, Centre for languages and literature

Holsanova, J. 2001: *Picture Viewing and Picture Description: Two Windows on the Mind*. Doctoral dissertation. Lund University CognitiveStudies 83.

Strömqvist S. & Karlsson, H. (2000) *SciptLog for Windows – user's manual*. Technical Report, Lund University: Department of Linguistics and University College of Stavanger: Centre for Reading Research

Strömqvist, S., Nordqvist, Å. & Wengelin, Å. 2004. Writing the frog-story: developmental and cross-modal perspectives. In S. Strömqvist and L. Verhoeven (eds.). *Relating events in narrative - typologiocal and contextual perspectives*. Mahwah, NewJersey: Lawrence Erlbaum Associates Publishers, 359-394.

Wengelin, Å (2002) Text Production in Adults with Reading and Writing Difficulties, *Gothenburg Monographs in Linguistics 20*, Göteborg University, Department of linguistics

Fig 1. LINEAR FILE

```
<RETURN>Mitt på bilden finns det också <2.021>kor<3.031>
<4.210>som är bruna och vita<14.351>.
```

(uc52fb)

Eng:
In the middle of the picture there are also <2.021>cows<3.031> <4.210>that are brown and white<14.351>.

Fig 2. Multimodal time coded score sheet of keystroke data and eye tracking data combined