



Universität Potsdam

Reinhold Kliegl

Automated and interactive analysis of eye fixation data in reading

first published in:
Behavior research methods and instrumentation 13 (1981), S. 115-120,
ISSN 0005-7878

Postprint published at the Institutional Repository of the Potsdam University:
In: Postprints der Universität Potsdam
Humanwissenschaftliche Reihe ; 137
<http://opus.kobv.de/ubp/volltexte/2009/3986/>
<http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-39861>

Postprints der Universität Potsdam
Humanwissenschaftliche Reihe ; 137

Automated and interactive analysis of eye fixation data in reading

REINHOLD KLIEGL

University of Colorado, Boulder, Colorado 80309

A package of five FORTRAN programs that provides for fast user-controlled analyses of reading eye fixations is described. The package requires the data to be in a fixation format and to be rescaled to screen dimensions. OLDEYE identifies six types of fixations and calculates descriptive statistics on each of them, on their associated saccades, and on their average pupil diameter. CONVRT represents the text as a string of words that can be coded according to experimentally relevant variables. PLTFIX prints fixation durations by letter position and sequence of occurrence. MODDAT is an interactive program for marking parts of the text in which the data quality is below acceptable standards. It also allows the correction of systematic errors due to calibration or drift. MATCH combines the outputs from OLDEYE, CONVRT, and MODDAT and calculates 11 dependent measures for every word. The output of MATCH is suitable for input to conventional multivariate statistical programs.

Historically, three levels of analysis for eye fixations in reading can be distinguished. Characteristic of early eye movement research was the analysis at the overall text level. Typical statistics calculated were average fixation duration, number of fixations, number of regressions, and average interfixation distance. Primarily for technological reasons, eye fixations were not matched onto their corresponding text parts. Recent advances in eye monitoring technology and interest in psycholinguistic processes have resulted in a shift from the overall text statistics to analyses at the word-by-word level. Contemporary eye movement research in reading analyzes the fixation data by relating them to linguistic and perceptual properties of the word fixated (e.g., Just & Carpenter, 1980; Kliegl, 1980). In parallel with this psycholinguistic research, questions of eye guidance and peripheral sensitivity have become more and more prominent. Experiments in which the position of the fixation causes a change of the text in foveal or peripheral vision have given insight into the extent of peripheral preprocessing (e.g., McConkie, 1979; O'Regan, 1979; Rayner, 1979) and the conditions under which two fixations are likely to occur on a single word (O'Regan, 1980, 1981). Of course, any theory about the process of reading and eye fixations should take into account both the linguistic characteristics of the text and the more peripheral influences on eye guidance. However, relatively little "cross talk" between these two branches of eye movement research appears to have occurred. It seems that data are collected and analyzed only with regard to a very specific question, and, although the data

might also speak to several other issues, normally this information is not reported. To facilitate a more integrated approach a package of five FORTRAN programs has been developed to allow a fast analysis of eye fixation data in a format in which reading task manipulations, linguistic aspects, and eye guidance issues can be tested simultaneously. The input requirements are that the eye movement data be in a reduced and rescaled format (see Davidson, 1981; Kliegl & Olson, 1981). Every record in the data file gives the duration of one fixation, its character position on the line, its line number, and its average pupil diameter. The record is flagged if the fixation is a blink. Further, the experimental text material should be available in the form in which it appeared on the screen. Output of the programs is on a screen-by-screen basis.

PROGRAMS

OLDEYE: Calculate Overall Text Statistics

The first program fulfills three functions: (1) identification of saccades and blinks or loss of stable measurement in the data base, (2) identification of six types of fixations, and (3) calculation of descriptive statistics separately for each type of fixation, its associated saccade, and its average pupil diameter.

As a preliminary step in separating fixations from saccades, the successive reduced data sample clusters are checked to see if they are closer than a parameter (XDELTA). For example, if XDELTA is one, fixations separated by one character position are combined. This step is necessary because noise events sometimes result in the breakup of true fixations. Data resulting from saccades are separated from fixations by setting another parameter (TDELTA) in the program. This parameter specifies the maximum number of samples in a data cluster that would be called a saccade. In our research,

This research was supported by USPHS Grant HDMH11681-01A1. I thank Richard K. Olson and Brian J. Davidson for helpful suggestions and advice. Requests for reprints and program listings should be sent to Reinhold Kliegl, Department of Psychology, University of Colorado, Boulder, Colorado 80309.

this parameter is set at four, since almost all true fixations are longer than 67 msec (four samples). Although saccade data are excluded from further analyses, their times are summed to allow calculation of percentage of time spent in saccades.

For identification of blinks, the flags output by the reduction programs are used. There are two types of blinks: one has an average pupil diameter below a critical value, and the other has an average pupil diameter within the tolerance limit (but one or several samples contributing to this fixation have pupil diameters below the tolerance limit; see Kliegl & Olson, 1981). Obviously, the second type is more stringent than the first. The user must decide which option to choose. Note, however, that in the Applied Sciences Corporation system the pupil diameter also reflects loss of stable measurement. The time spent in blinks is summed and expressed as a percentage of total reading time. These data points are not included in any other statistical calculations, but they are kept in the data base as indicators for text parts with bad measurement in the plotting program. The percentages of time spent in saccades and in blinks can be taken as indicators of the overall quality of the data. For example, values higher than expected according to

other research are very often indicative of unstable measurement.

The second function of this program is to identify different types of fixations. These include fixations at the beginning and at the end of a line, fixations preceded and succeeded by a left-right saccade (i.e., the "standard" reading fixation), fixations preceded by a right-left and followed by a left-right saccade (i.e., the "standard" regressive fixation), fixations preceded by a left-right and followed by a right-left saccade (i.e., fixations preceding a regressive saccade), and fixations preceded and succeeded by right-left saccades (i.e., a series of regressive fixations). This classification was motivated by Goltz (1975), who used it as a basis for distinguishing between skilled and less skilled readers. During the process of flagging the fixations according to their type, the length of the saccade following the fixation is calculated and included in the data base. Length is measured in number of characters.

The third function of this program is to calculate mean and standard error for all fixations and for each of the six types of fixations, the saccades following them, and the pupil diameters associated with them. The program also determines extreme values for fixation

Table 1
Output of Program OLDEYE

STATISTICS OF TEXT: 1						CODE:s101				
MEANS AND SE OF FIXATION PAUSES										
ALL: 16.7	BLINE: 20.4	ELINE: 13.8	REGR: 15.8	PROGR: 16.8	PREC. REGR.: 16.9	SEQ. REGR.: 23.3				
(SE= 0.74)	(SE= 3.24)	(SE= 0.94)	(SE= 1.57)	(SE= 1.03)	(SE= 0.82)	(SE= 8.95)				
MEANS AND SE OF SACCADES										
ALL: 9.3	BLINE: 15.7	ELINE: -52.7	REGR: 16.8	PROGR: 8.9	PREC. REGR.: -4.4	SEQ. REGR.: -14.3				
(SE= 1.00)	(SE= 6.56)	(SE= 3.16)	(SE= 5.41)	(SE= 0.50)	(SE= 0.75)	(SE= 3.48)				
NUMBER OF OBSERVATIONS										
ALL: 93	BLINE: 11	ELINE: 10	REGR: 13	PROGR: 46	PREC. REGR.: 9	SEQ. REGR.: 3				
82 (FOR SACC.)										
MEANS AND SE OF PUPIL DIAMETERS										
ALL: 94.8	BLINE: 92.4	ELINE: 95.0	REGR: 95.0	PROGR: 95.2	PREC. REGR.: 94.1	SEQ. REGR.: 95.7				
(SE= 0.49)	(SE= 2.17)	(SE= 1.53)	(SE= 0.86)	(SE= 0.60)	(SE= 1.87)	(SE= 0.33)				
MAX. FIX.: 45. MIN. FIX: 5. MAX.PUP.: 101. MIN.PUP.: 72. MAX.SAC.: 79. MIN.SAC.: -78.										
TOTAL NUMBER OF FIXATIONS: 93										
N OF INITIAL REGRES.: 0 DOUBLE REGR.: 0 TRIPLE REGR.: 0 QUADR. REGR.: 1 QUINT. REGR.: 0										
READING TIME (IN SEC): 29.12										
FIXATION TIME: 0.90 MOVE TIME: 0.09 BLINK TIME: 0.01 (PROP. OF READ. TIME)										
FREQUENCIES:	80-129	130-179	180-229	230-279	280-329	330-379	380-429	430-479	480-529	530 AND LONGER
	2	10	22	21	16	9	6	3	0	4
	0.02	0.11	0.24	0.23	0.17	0.10	0.06	0.03	0.00	0.04

durations, pupil diameters, and saccade sizes. Further, it determines the number of initial regressions (i.e., regressions that occur immediately following a return sweep and serve to position the eye at the beginning of the line) and the number of double, triple, quadruple, and quintuple regressions. It calculates the reading time in seconds and breaks this reading time up into proportion of time spent in fixations, spent in saccades, and allocated to loss of delimiters or blinks. Finally, an absolute and relative frequency distribution of fixation durations is derived. Table 1 is an example output of OLDEYE's summary statistics.

Note that fixation durations are given in a number of samples. For the frequency distribution, the durations are given in milliseconds.

These summary statistics, with the exception of the frequency distributions and the regression counts, are also output to a file suitable as input for statistical packages (e.g., SPSS). The third output generated by OLDEYE is the original data base with moves removed, but enhanced with information about the type of the fixation and the length and direction of the saccade following a fixation.

PLTFIX: Plot Fixations and Text

The data base generated by OLDEYE and the experimental text material in its display format are the input for PLTFIX. The program matches the fixations with corresponding letters, using the character position and line number of each fixation. It also keeps track of the

Table 2
Output of Program FIXPLT

```

TEXT exp.txt          READ BY s101  SCREEN: 1 IOUT: 13 0 0 0 0 0 0 0 0 0 0 0 0 0
                        The unusual events described in this chronicle occurred
                          4 3          5          6          7 9          8          10
                          8 18         27          15         1113         19          15

at Oran.  Everyone agreed that, considering their somewhat
1214 11          15          16          17 18          20          19
1345 10          11          16          14 12          15          14

extraordinary character, they were out of place there.  For
22 21          23          24          25          26          27
11 15          44          16          20          16          20

its ordinariness is what strikes one first about the town of
29 28 30          31          32          34          35
18 10 35          17          18          18          15

Oran, which is merely a large French port on the Algerian
37 36 2 38 1          39          40          41          33 42          43
16 2222 18 25          11          17          9          18 14          13

coast, headquarters of the Prefect of a French Department.
44 45          46          54 49 48          5350          52 51
7 14          20          12 24 20          4125          17 13

The town itself, let us admit, is ugly.  It has a smug,
55          56 57          4759          6058          62 61          64 63
10          17 12          1019          1319          15 16          18 15
                                                65
                                                15

placid air and you need time to discover what it is that makes
67 66          68 69          70          71 72 73          74 75
28 12          21 24          15          5 11 14          25 12

it different from so many business centers in other parts
77 76          78 79 80          81
12 10          22 15 888          13

```

sequence in which the fixations occur. This value is referred to as the series number of the fixation. Table 2 is an example output of PLTFIX.

Line 1 contains the text, Line 2 displays the series number, Line 3 displays the duration in number of samples, and Lines 4 and 5 are used when a second fixation falls on a letter position already occupied by a previous fixation. If more than two fixations have identical coordinates, the first and the last fixations on that position are plotted. The array IOUT in the output header contains fixation series numbers that fell outside the text boundaries. If a fixation is marked as blink, "BBB" is output instead of the fixation duration. Note, however, that the position of a blink does not always correspond to the text part at which it occurred. The reason for this is that the loss of delimiters causes a loss of accuracy in coordinates. The series numbers may be used to infer the correct position.

The output of PLTFIX allows the user to judge the quality of the calibration and of the obtained data. If a consistent shift, especially of line numbers, is observed, the user might consider rerunning the reduction and calibration programs with readjusted calibration flags for the quadrants in which the drift is observed (see Kliegl & Olson, 1981, for details).

CONVRT: Convert and Code the Experimental Text

So far, we have not attempted to relate fixations and text for purposes of a quantitative analysis. The statistics of OLDEYE can be calculated without too much concern for an exact fixation to text correspondence. Position information is relevant only in a relative way, (i.e., whether a fixation occurs after a positive or a negative saccade, before or after a return sweep, etc.). Most contemporary research in eye movements strives for a fine-grained analysis level. To achieve this, the experimental text has to be converted to a string of words that are coded according to variables considered important. CONVRT is a partially interactive program that serves these needs. Its input is the experimental text. The program converts the text to a data base in which every word forms a record, removing punctuation marks in the process. Every record is supplied with the following information about the word: its coordinates on the screen (character position of first letter, line number), length of the word in number of letters, whether it is the first or last word in a sentence. The program then tries to locate the word in a dictionary that has word frequency norms associated with it (Kučera & Francis, 1967). If the word is available, the frequency

Table 3
Output of Program MATCH

a	b	c	d	e	f	g	h	TFT	TNF	IFT	INT	DFE	PF	DLF	PLF	LS	PDF	PDL
30	7	500	1	19081	what			17	1	17	1	17	1	0	0	9	98	98
35	7	700	2	501	strikes			18	1	18	1	18	5	0	0	9	101	101
43	7	300	2	32921	one			0	0	0	0	0	0	0	0	0	0	0
47	7	500	2	13601	first			18	1	18	1	18	2	0	0	6	99	99
53	7	500	1	18152	about			18	1	18	1	18	2	0	0	9	101	101
59	7	300	16997	11	the			0	0	0	0	0	0	0	0	0	0	0
63	7	400	2	2121	town			15	1	15	1	15	1	0	0	-51	101	101
68	7	220	1364	11	of			0	0	0	0	0	0	0	0	0	0	0
10	8	410	2	02	oran			60	3	38	2	22	4	16	1	7	72	93
30	9	200	1364	11	of			0	0	0	0	0	0	0	0	0	0	0
33	9	300	16997	11	the			10	1	10	1	10	0	0	0	9	100	100
37	9	700	2	12	prefect			56	3	44	2	20	5	24	3	14	99	95
45	9	200	1364	11	of			0	0	0	0	0	0	0	0	0	0	0
48	9	100	12323	71	a			0	0	0	0	0	0	0	0	0	0	0
50	9	600	2	1391	french			66	2	25	1	25	4	0	0	12	95	95
57	9	1022	2	2253	departmen			30	2	30	2	13	9	17	4	-8	93	96
1510	311	16997	11	the				10	1	10	1	10	3	0	0	6	97	97
1910	400	2	2121	town				0	0	0	0	0	0	0	0	0	0	0
2410	600	2	3042	itself				29	2	29	2	17	0	12	4	12	96	92
3210	300	2	3841	let				19	1	19	1	19	2	0	0	5	98	98
3610	200	2	6721	us				0	0	0	0	0	0	0	0	0	0	0
3910	500	2	442	admit				32	2	19	1	19	1	0	0	-6	93	93
3112	400	2	10302	many				22	1	22	1	22	4	0	0	7	89	89
3612	800	2	3922	business				15	1	15	1	15	6	0	0	5	90	90
4512	700	2	2242	centers				999	0	0	0	0	0	0	0	0	0	0
5312	200	12134	11	in				0	0	0	0	0	0	0	0	0	0	0

Note—(a) position on line, (b) line number, (c) length of word (characters), (d) beginning and end of line, (e) beginning and end of sentence, (f) code (function vs. content), (g) word frequency, (h) number of syllables. For explanation of other labels, see text.

value is returned to the program. If the word cannot be located, the user is prompted for the value. This word and its values are new entries in the dictionary, making it available for a later search. For every word, the program prompts the user to enter the grammatical case of the word. The interactive part of this program is easily expandable to provide other information (e.g., the number of syllables, level of proposition, and type of case, in the sense of Fillmore's, 1968, case grammar). The output of this program is a template of the experimental text that is the same for every subject and has coded any qualities of the text of interest to the user. The columns to the left of the word in Table 3 show the output of CONVRT.

MODDAT: Mark Areas of Bad Measurement

In its implementation, MODDAT is a part of the final program, MATCH. It is discussed as a separate program for conceptual clarity. Before running MODDAT, the output of PLTFIX should be scrutinized carefully. Do the series numbers follow an expected course or do they jump between lines for parts of them? Do the "BBB" series numbers (i.e., fixations at which no stable measurement was obtained) fall between the preceding and succeeding series number? Are there parts of lines that do not show fixations? Questions like these should guide the user's judgment about the quality of reduction and calibration for this text. If there are too many discrepancies between the output and the values, this text might be excluded. If there appears to be a systematic drift across several lines explainable in principle by an off-calibration mark, one might consider rerunning the calibration program with a changed value for this mark. We also advise marking the intended changes on the output of PLTFIX. As noted, under conditions of loss of pupil delimiters, the position information does not necessarily indicate the area of text for which no accurate measurement could be obtained. The loss might be indicated on a single word, although it involved several words. Further, the "BBB" mark might be a line above, below, or outside the text boundaries. In this case, it is necessary to infer the area of missing data with help of the series number of the "BBB" fixation and to tell the program to ignore these text parts for further analyses. The program prompts the user in a line mode and in a word mode. Line mode is used when a whole line of text is to be excluded. In word mode, it is possible to mark a single word or a string of words beginning anywhere in the text. The program prompts the user for the word coordinates of the first word of the string and for the number of words to be marked. In the MATCH program to be described below, words that receive a fixation marked "BBB" are marked the same way as words marked during the interactive MODDAT program. Therefore, in the case of misplaced "BBB" fixations, it is necessary

to delete these fixations from the data base prior to running MATCH to prevent marking of text parts with stable measurement conditions. MODDAT prompts the user for the series number of fixations to be excluded from the analysis.

Regardless of the quality of calibration and the cooperativeness of the subject, there is always a chance that some drift will occur during reading. It might be introduced by a head movement or an off-calibration flag. As mentioned above, in some cases, it might be advisable to rerun the sequence of programs with adjusted calibration flags. If there is drift only for a few fixations, MODDAT can be used to reassign line numbers to fixations. Applied Sciences eye trackers show more fluctuations from machine noise along the vertical dimension. This, together with a slightly too low calibration flag, can cause fixations for a part of the line to be systematically allocated a line number too low or too high. Identification of these conditions should be guided by series numbers and by a consistency of direction of misplacements across a few lines at about the same horizontal positions.

An example of how MODDAT works is given in Tables 2 and 3. Fixations 33 and 47 were assigned the preceding line number. Therefore, in Table 3 "first" and "the" (preceding "Prefect") are shown with the durations of the corresponding fixations. "Center" has a "BBB" mark in Table 2. In Table 3, a value of 999 in the first column after the word indicates that this word should be ignored during analysis.

MODDAT creates a protocol of the interactive modification dialogue. At a later date, the user can simply specify the name of the MOD-file and the program will execute all modifications automatically. Of course, it is possible to create any number of MOD-files with varying criteria, for example, for reassigning line numbers. Storage of the OLDEYE data base and a MOD-file allows fast recreation of the data base used in final analyses.

MATCH: Match Fixations and Words

MATCH is the final program in this sequence. Its input is the OLDEYE data base, the corresponding MOD-file, and the corresponding converted text from CONVRT. The converted text file is used as a template into which the fixations are filed. During this filing process, the program calculates the following 11 measures for every word of the template: total fixation time (TFT), total number of fixations (TNF), initial fixation time (IFT), initial number of fixations (INF), duration of first fixation (DFF), absolute letter position of first fixation (PFF), duration of last fixation (DLF), absolute letter position of last fixation (PLF), length of saccade following the last fixation on the word (LS), pupil diameter during first fixation (PDF), and pupil diameter during last fixation (PDL). With the exception

of TFT and TNF, all measures are calculated only for the initial reading of the word. A sample output of MATCH is given in Table 3.

For a fixation to be allocated to a word, its position has to match the space preceding the word or a position occupied by a letter of the word. This allows a convenient calculation of relative fixation position by dividing the absolute fixation position (PFD or PLD) by the length of the word. A fixation on the space preceding the word will have a relative position of 0, and a fixation on the last letter of the word will have a relative position of 1.

TFT is the sum of all fixation durations falling on a word, IFT is the sum of all fixation durations during the initial reading only. Therefore, if a fixation strikes a word, then jumps to a different one and then back onto it, TFT will have the sum of both fixations, whereas IFT, DFF, and DLF all show the value of the first fixation only. Note that IFT is very similar to Just and Carpenter's (1980) gaze duration, except that gaze durations also contain time spent in saccades and blinks within the word boundaries; gaze duration is more like reading time per word. Words to be excluded from the analysis (see MODDAT) have a value of 999 in TFT and 0 in all others.

The data base generated by MATCH is suitable for input to conventional multivariate statistical analysis programs. There are also possibilities for reexpressing variables' values (e.g., calculating a log transform for word frequency values) and for filtering the data base according to any logical condition of word information and dependent measures. For example, one can extract all words of a certain length that received two fixations during the initial reading and compare them with words of the same length that received only one fixation. It is this flexibility that we think can facilitate third-level type analyses for data obtained in reading of continuous paragraphs.

CONCLUSION

A package of five programs has been developed for analyzing eye movement data of reading experiments. The programming language is VAX-11 FORTRAN. Only minor modifications, mostly at the input and output side, are required to make the programs work on other systems. Further, RT-11 Version 2B equivalents of OLDEYE, CONVRT, and FIXPLT are available.¹

Collection of eye movements during reading usually generates a great amount of data. There is a great temptation to "stuff" these data into the computer and allow some general algorithm to decide their fate. The

design of this package pursues a philosophy of exploratory data analysis (Tukey, 1977). The output at every stage of analysis is transparent to the user and, therefore, allows for user intervention. This is especially true for MODDAT. The programs do not force interactivity, but they strongly encourage it. We do think that "staying in touch" with the data base is a potentially very valuable heuristic strategy and puts the command of analysis more into the user's hands than does a completely automated analysis program.

REFERENCES

- DAVIDSON, B. J. A rotating buffer system for on-line collection of eye monitor data. *Behavior Research Methods & Instrumentation*, 1981, **13**, 112-114.
- FILLMORE, C. J. The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory*. New York: Holt, Rinehart, & Winston, 1968.
- GOLTZ, R. T. *Comparison of the eye movements of skilled and less skilled readers*. Unpublished doctoral dissertation, Washington University, 1975.
- JUST, M. A., & CARPENTER, P. A. A theory of reading: From eye fixations to comprehension. *Psychological Review*, 1980, **87**, 329-354.
- KLIEGL, R. *Eye movements during reading: An attempt to interface psycholinguistic and eye-guidance research*. Unpublished master's thesis, University of Colorado at Boulder, 1980.
- KLIEGL, R., & OLSON, R. K. Reduction and calibration of eye monitor data. *Behavior Research Methods & Instrumentation*, 1981, **13**, 107-111.
- KUČERA, H., & FRANCIS, W. N. *Computational analysis of present-day American English*. Providence, R.I: Brown University Press, 1967.
- MCCONKIE, G. W. On the role and control of eye movements in reading. In P. A. Kolars, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language*. New York: Plenum, 1979.
- O'REGAN, J. K. Moment to moment control of eye saccades as a function of textual parameters in reading. In P. A. Kolars, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language*. New York: Plenum, 1979.
- O'REGAN, J. K. The control of saccade size and fixation duration in reading: The limits of linguistic control. *Perception & Psychophysics*, 1980, **28**, 112-117.
- O'REGAN, J. K. The "convenient viewing position" hypothesis. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception*. Hillsdale, N.J: Erlbaum, 1981.
- RAYNER, K. Eye movements in reading: Eye guidance and integration. In P. A. Kolars, M. E. Wrolstad, & H. Bouma, *Processing of visible language*. New York: Plenum, 1979.
- TUKEY, J. W. *Exploratory data analysis*. Reading, Mass: Addison Wesley, 1977.

NOTE

1. All programs are available upon request. Programs will be sent with an example data set for easy verification and conversion.