Journal of Eye Movement Research
1(2):2, 1-12

A pragmatic approach to multi-modality and non-normality in fixation duration studies of cognitive processes

Chie Nakatani and Cees van Leeuwen
Laboratory for Perceptual Dynamics
RIKEN Brain Science Institute

Interpreting eye-fixation durations in terms of cognitive processing load is complicated by the multimodality of their distribution. An important source of multimodality is the distinction between single and multiple fixations to the same object. Based on the distinction, we separated a log-transformed distribution made to an object in non-reading task. We could reasonably conclude that the separated distributions belong to the same, general logistic distribution, which has a finite population mean and variance. This allowed us to use the sample means as dependent variables in a parametric analysis. Six tasks were compared, which required different levels of post-perceptual processing. A no-task control condition was added to test for perceptual processing. Fixation durations differentiated task-specific perceptual, but not post-perceptual processing demands.

Keywords: fixation duration distribution, task effects, multi-modality, non-normality

Introduction

When a text is read or an image is scanned, the eyes jump from one location to the next in a sequence of saccades. Between two saccades, the eyes usually fixate on what is being processed at that point in time (Yarbus, 1967). As processing takes longer, fixations will, on average, last longer; hence fixation duration was espoused for measuring information processes during reading, visual search, object identification, scene viewing and other tasks. (See reviews by Henderson & Hollingworth, 1998; Inhoff & Radach, 1998; Rayner, 1998).

To assess the difference in fixation durations between different processing conditions, preferably parametric statistical tests are applied, that require the data to be distributed approximately normally. Fixation durations are generally distributed asymmetrically. They could be understood as following either a Gamma distribution, if we regard them as intervals between random, stochastically independent saccades, an exponential distribution if saccades are considered stochastic transitions between discrete internal states, or yet another long-tailed distribution if the state transition depends on the time spent in the state itself (Engbert & Kliegl, 2001). In all these cases, a logarithmic transformation of the data would yield an approximately normal distribution. This transformation has been practiced widely in order to apply analyses of variance (ANOVA) or the General Linear Model (GLM).

Such analyses are meaningful only if fixation duration distributions are homogeneous. Homogeneity of the distribution, however, cannot be taken for granted in fixation duration studies. Often the number of fixations to an object differs and the distribution is a mixture of single and multiple fixation cases. In reading text, for example, whereas most words receive a single fixation only, 10 to 30 % receive two or more subsequent fixations. The probability of such refixations depends on physical and cognitive factors, such as word length and frequency (Rayner, Sereno, & Rany, 1996; Yang & McConkie, 2001). The duration of a first fixation is generally longer than that of a second one (O'Regan & Levy-Schoen, 1987; Rayner, et al., 1996; Sereno, 1992); therefore the overall fixation duration distribution will be multimodal (Feng, 2006; Velichkovsky, 1999).

Multi-modality is not always an obstacle to comparison between conditions; when conditions yield consistent effects across first and subsequent fixations, aggregate measures such as gaze duration could be used; gaze duration is the sum of fixation durations made to a target region in succession, including single fixation cases. In reading/word recognition, cognitive factors such as word frequency consistently affected first fixations and refixa-

DOI 10.16910/jemr.1.2.2
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

tions; for instance, low frequency words were fixated longer than high-frequency ones in single fixation cases as well as in both the first and subsequent refixations of multiple fixation trials (Kennison & Clifton, 1995; Rayner & Duffy, 1986; Raney & Rayner, 1995; Rayner, Sereno, Morris, Schmauder & Clifton, 1989; Sereno, 1992).

Beyond reading tasks, it is presently unknown, however, if cognitive factors consistently influence first and subsequent fixation durations. De Graef, Christiaens, & d'Ydewalle (1990) propose to distinguish first and subsequent fixations to a target. They argue that whereas the first fixation duration reflects perceptual processing, later fixations predominantly reflect post-perceptual cognitive processes. Aggregate measures such as gaze duration, therefore, are non-preferred and, as a result, we need to face the multiple fixations problem. When the number of re-fixations is small, they could simply be excluded from analysis. When their number is large, each type of fixation needs to be analyzed separately. In scene recognition, multi-fixations occurred about 30 % of the time when eyes fixated on an object (Nakatani & Pollatsek, 2002), which is too significant a proportion to discard.

In the present article we propose a simple method to deal with multi-modality in fixation durations. First we apply a logarithmic transformation to reduce the skewness of the fixation duration distribution. Next, we separate the transformed distribution according to fixation order within an object and investigate whether the bulk of the separated distributions offers a sufficiently good fit to the normal distribution, and, if so, with sufficiently similar variances to apply GLM directly to the transformed data. If this is not the case, we proceed with means, assuming that these are normally distributed under the central limit theorem. To warrant this, we need to investigate the separated distributions to see if they have finite means and variances. To this purpose probability density functions are fitted.

A consequence of fitting probability density functions is that it allows us to determine whether the separated distributions could reasonably be assumed to follow the same theoretical distribution, with well-defined sample means and variances. If this is the case, experimental factors are used along with fixation order as predictors in fitting a GLM to averaged data. If no common probability density function could be found, averaged data from each fixation order are fed separately to a GLM, of which the predictors are the experimental factors.

To demonstrate the proposed method, fixation durations under six cognitive task conditions were recorded. The same set of stimuli, (sequences of letters and digits) was used for each task to minimize variation due to physical differences between stimuli. As a result, they will share the same mandatory perceptual processing components. In a control condition, the stimuli were presented without a task. A comparison between task-no task conditions will allow us to evaluate the consistency of cognitive, i.e. non-mandatory perceptual effects on fixation duration for different fixation types, distinguished according to fixation order. The tasks differ on whether they invoke categorical or spatial information, as well as on whether they require an immediate or deferred choice response. In other words, they differ in their post-perceptual cognitive processing requirements. We examined whether task effects are consistent amongst the separated fixation types.

Methods

Participants

Eight undergraduate and graduate students with normal, or corrected to normal vision (five women and three men, mean age: 20.0 years) from the greater Tokyo area were paid one thousand yen per hour for participation.

Stimuli

Eight letters (A, B, C, D, E, F, G, H) and eight digits (1, 2, 3, 4, 5, 6, 7, 8) constitute the stimuli, sequences of which were presented in each task. Each stimulus fit within a 2º x 2º area, rendered in blue on a gray background. Each was presented at one of 16 positions of an invisible 4 by 4 square grid in the display. Centers of each position were approximately 3.5º apart. Starting from an arbitrary point on the grid, the next stimulus was always presented at one of four adjacent locations, above, below, right or left of the current one, while on the edges of the grid the number of possible 'next' locations was reduced to three, and to two in the corners. For each individual participant, a new random sequence of 1024 letters and digits was generated, which was then uniformly used for presentation throughout all the different tasks (Figure 1, top row).

Design

There were six task variations and a control condition. The task conditions involve category (letter vs. digit) or

DOI 10.16910/jemr.1.2.2
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

spatial location of the stimuli. There were three category tasks, a category-judgment task, a category-count task and a category-pattern task. In the category-judgment task, participants were instructed to press the right button of the button box each time a letter stimulus appeared and the left button for each digit. After the button press, the next stimulus appeared. This is done for all the 1024 stimuli in the sequence. Thus, this task is expected to reflect the capacity to judge item category. In the category-count task, participants were asked to count the number of letters in the sequence. They pressed an arbitrary response button for continuation to the next item. After the presentation of the whole sequence, participants were asked if the actual count was greater or less than an arbitrary chosen number, for instance 675, by pressing the right or left button (e.g., "Press right button, if the number of letters >= 675. Press left button, if the number of letters < 675"). This task in addition to judgment of item category reflects on the capacity to update memory. In the category-pattern task, participants were asked to detect whether the sequence contained a repeating pattern of 16 letters and digits (e.g., `letter - letter - digit - digit - letter - letter - digit - digit - letter - letter - letter - letter - digit - digit - digit - digit'). After the sequence, two patterns were presented and one of them was chosen as the presented pattern. (e.g., "Press right button, if repeated pattern was 'digit-letter-digit-letter-digit-letter-digit-letter-digit-letter-digit-letter-digit-letter-digit-letter'; press left if `letter - letter - digit - digit - letter - letter - digit - digit - letter - letter - letter - letter - digit - digit - digit - digit'"). This task, in addition to category judgment, and memory update for serial pattern matching was expected. In sum, judgment of item category was needed in all tasks. Each of these tasks is more complex than the previous one.

Location tasks comprised the following: a location-judgment task, a location-count task and a location-pattern task, which were similarly structured to those in the category tasks. In the location-judgment task, participants were asked to press the right button if a stimulus is presented in Rows 1 and 3. If the stimulus was presented in Rows 2 and 4, the left button was pressed. This involves a spatial judgment. In the location-count task, a participant counted the number of the stimuli presented in Rows 1 and 3. In the end of 1024 stimulus presentations, participants were asked to answer if the final count was equal or greater than an arbitrarily chosen number by pressing a button (e.g., "Press right button, if the number of letters >= 675. Press left button, if the number of letters < 675"). In this task, in addition to a spatial judgment, each stimulus requires an update of memory. In the location-pattern task, participants were asked to register how often, relatively speaking an item was presented at each of the matrix locations. After the sequence, two sets of 16 square tiles (4 x 4) were presented in which a color code ranging with three intermediate steps from dark gray to bright blue indicated the relative item frequency at each location – the dark gray represents the lowest presentation frequency and the bright blue indicated the highest presentation frequency. Participants responded to the correct pattern by pressing the right or left button. This task, in addition to spatial judgment and memory update, required the detection of a spatial sampling distribution.

All sequence presented had a 16-stimulus pattern corresponding to one of the response alternatives of the category pattern task. This pattern was repeated 64 times in the sequence. In addition, all the pattern had a spatial distribution corresponding to one of the response alternatives in the location pattern task.

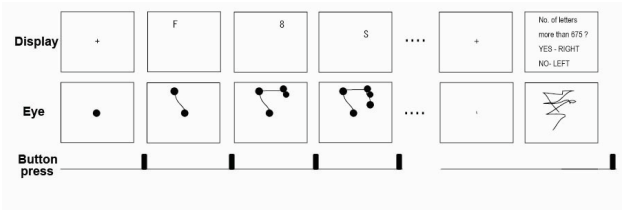


Figure 1. Schematic illustration of the tasks.

Top row: After an initial fixation cross, a letter or a digit was presented in one of 16 display locations. In the end of the stimulus sequence, a question display was presented. Different questions were used depending on task conditions. In the figure, the question belonging to the category-count task is shown. Middle row: Participants were asked to follow each stimulus with eyes (fixations are represented as black circle), then pressed right or left buttons. Bottom row: The button presses were required to start the next stimulus; in the category-judgment and location-judgment conditions, the button presses were categorization responses (See text for details of task conditions).

A no-task viewing condition was added as control condition. Participants viewed the sequence of 1024 letters or digits were asked to fixate on each item and press an arbitrary button in order to continue to the next one.

Procedure

The experiment was performed in a light-attenuated experimental room, using an image-based eye-tracking system (EyeLinkI, SR Technologies, Ontario, Canada).

DOI 10.16910/jemr.1.2.2
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

Participants wore a light-weighted headset with two CCD cameras and a head movement sensor, and held a button box in both hands while seated in front of a 21-inch CRT display. Participants were instructed to follow the stimulus with their eyes while keeping the head still. Special instructions and practice sessions (32 stimuli) were given prior to each task condition. Feedback on performance was given at the end of the practice session in the category-count, category-pattern, location-count and location-pattern conditions. After the practice session, two experimental sessions were run with eye movement recording. Both eyes were tracked with sampling rates of 250 Hz, but only right eye data were analyzed. In each of the experimental sessions, a sequence of 1024 items was presented. The next item appeared immediately after participants had pressed a response button. Recalibration and drift correction were performed whenever needed to secure measurement accuracy. Each participant performed all the tasks in counter-balanced order: half started with object conditions, the other half with spatial conditions. Within object and spatial conditions, one third started with the item, another third started with the count and one other third started with the pattern task. Due to the large number of trials, the experiment was performed over two subsequent days.

Results

Bad segments of eye movement record were removed prior to data analysis. The remaining data were analyzed using an eye-event filter to extract fixations. The three parameters of the filter, velocity, acceleration and motion of saccades were set to detect most of the saccades larger than 0.6 degrees and fixations longer than 100 ms, the regular range for cognitive psychological experiments. Out of the fixations detected, those started before the presentation of a stimulus were excluded. Those that ended after a button press was adjusted by truncating the period after the button press. Since a button press initiates next stimulus presentation, latency of saccade from one stimulus to another was excluded. As a result, a total 76142 of fixation durations were analyzed.

A natural-logarithm distribution was applied here to all fixation durations. However, a Kolmogorov-Smirnov test showed the transformed distribution to differ significantly from the normal distribution ($p < .001$). Figure 2 shows that the transformed distribtuion is right skewed (skewness = -0.82, SE skewness = 0.01) and leptokurtic (kurtosis = 1.21, SE kurtosis= 0.02)[1]. Multiple peaks are in evidence in Figure 2: around 3.30 (27 ms), 5.50 (245 ms) and 6.10 (446 ms). The first peak occurs for times faster than express saccades (80-100 ms, Fischer & Ramsperger, 1984; Fischer & Weber, 1993). The second peak is in the normal range of fixation durations. The third one is slightly out of the ordinal range

.

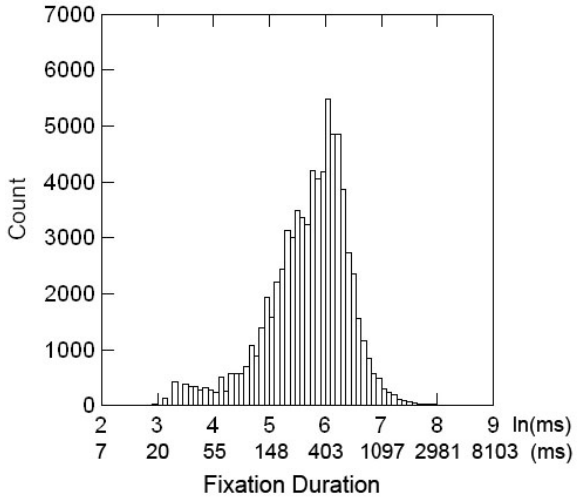


Figure 2. Distribution of fixation durations

Fixation duration distribution across all conditions in natural-logarithmic time units (upper x-axis) and ms (lower x-axis). The median of the distribution is 5.83 in ln ms, or 340 ms.

One might wish to argue that the logarithmic transform has been sufficient to enable a parametric analysis, even though there may be multi-modality in Figure 2 as a consequence fixation type, since adding this factor as a predictor to a GLM may take care of that. To illustrate what could happen if we thus ignore the distribution issue, we fitted a Linear Mixed Effect model (LME, Pinheiro & Bates, 2000) to the data, in which we included as predictors the most likely sources of multi-modality. These are: single fixation trials vs. multiple fixation trials,

[1] *The criterion used here for non-normal skewness and leptokurticness was absolute value of skewness divided by the SE of the skewness, and absolute value of kurtosis divided by the SE of the kurtosis. When these ratios are more than two, it is judged that kurtosis or skewness differ significantly from those of normal distribution.*

DOI 10.16910/jemr.1.2.2
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

task conditions, location errors, individual differences and their interactions. Task effects may be another source of multi-modality in the pooled distribution. Location error was chosen because bad landing positions could invoke a correction saccade (Vitu, McConkie, Kerr, & O'Regan, 2001). The location error was computed as distance between the current fixation position and the center of the stimulus, and was dealt as random variable. Participants were also a random variable, and as the other variables were fixed, a mixed-effect model resulted. The model was implemented as part of the non-linear mixed-effect (nlme) library of R (CRAN, http://www.R-project.org). Effects of random variables were estimated by restricted likelihood for the log-transformed fixation durations. The results are summarized in Table 1; Single vs. multiple fixation, task, and location error factors all showed highly significant effects on fixation duration. However, also several interactions were highly significant, some of which involve the single vs multiple fixation factor. In particular, the triple interaction of Task, Fixation, and Location error causes worries. Should we take this analysis seriously, this analysis leads, possibly misleadingly, to the conclusion that task effects are not consistent across different types of fixations.

One robust effect that cannot easily be ignored in Table 1 is the effect of single vs multi-fixation trials. This confirms that in our data the first fixations on a word in multi-fixation cases tend to be shorter than those in single fixation cases (Rayner et al.,1996; O'Regan, Vitu, Radach, & Kerr, 1994; O'Regan, 1992). Notwithstanding our concerns about the above analysis we will, therefore, proceed by separating the data according to fixation type. The number of single fixation trials was 37576. The fixations in the single-fixation cases are denoted as 1st/1-fixation, hereafter. In most of multiple fixation trials, two to three fixations were made; the number of two-fixation trials was 13499, yielding 26998 fixations. The first and second fixations in the two-fixation trials are termed 1st/2-fixation, and 2nd/2-fixation, hereafter. The number of the three fixation trials was 2501, yielding 7503 fixations, and the first, second and third fixations are called 1st/3-fixation, 2nd/3-fixation, and 3rd/3-fixation, hereafter. Trials with more than three fixations were excluded from further analysis due to small number of occurrences, leaving 72077 fixations. Multiple fixation trials occurred about 30% of trials in single-character stimulus presentation[2]

Table 1. Fitting results of the LME model

	DF		F-value	p- value
(Intercept)	1,	76107	54617.53	<.0001
Task Conditions	6,	76107	58.05	<.0001
Single-fixation vs. Multi-fixation trials	1,	76107	22088.75	<.0001
Location Error	1,	76107	15.26	<.0001
Task Conditions * Single-Multi	6,	76107	23.86	<.0001
Task Conditions * Location Error	6,	76107	15.12	<.0001
Single-Multi * Location Error	1,	76107	139.23	<.0001
Task Conditions * Single-Multi * Location Error	6,	76107	32.51	<.0001

Note. Degrees of freedom, especially denominator DF, differ from those in a general linear model (GLM), because the mixed models use individual data points. The interactions imply that cognitive effects are inconsistent across different fixation types. However, this result may be a misleading consequence of ignoring the distribution.

Figure 3 shows that the distributions of the different fixation types nicely separate the peaks observed previously in Figure 2. The third peak in Figure 2 corresponds to the main peak of the1st/1-fixation, 2nd/2-fixation, and 3rd/3-fixation; the second peak in Figure 2 to the main peak of the 1st/2, 1st/3 and 2nd/3 fixations. The first, early peak in Figure 2 corresponds to a minor peak in 1st/2, 1st/3 and 2nd/3 fixations, around 3.50 in natural logarithmic units (about 33ms). Descriptive statistics of each distribution are listed in Table 2. In all the separated distributions, skewness and kurtosis differ from those of the normal distribution. This precludes application of the GLM directly to these data.

A slight skew and leptokurtic tendency can be observed in all fixation categories. The non-normality poses a practical problem for analyzing cognitive effects across fixation types using parametric tests such as ANOVA. The problem can be circumvented by applying the analysis to the mean fixation durations, of which the distribu-

[2] *The percentage of multiple fixation trials was computed by (Number of 2-fixation trials + Number of 3-fixation trials) /(Number of Single fixation trials + Number of 2-fixation trials + Number of 3-fixation trials) *100, i.e., (13499 + 2501)/(37576 + 13499 + 2501) *100 = 28.86.*

DOI 10.16910/jemr.1.2.2
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

tion may be assumed to be normal based on the central limit theorem. This is possible if the data follow theoretical distributions that have defined means and variances. – This precaution is necessary as some distributions (e.g. the Cauchy distribution) can be leptokurtic without having a well-defined mean. A set of 40 probability density functions were fit to log-transformed fixation duration using a commercial package (EasyFit 4.0, MathWorks). Each of the models was fit to each type of fixation of each participant in each condition, except for cases with less than 50 fixations. As a result, the number of fit in 1st/3, 2nd/3 and 3rd/3 fixations was 40 % less than that of others. Probability density functions fit were ranked based on Kolmogorov-Smirnov test results. Average rank was computed for each fixation type over conditions and participants.

A typical example of model fits is shown in Figure 4. Fitting results in all fixation types are summarized in Table 3. In all but 2nd/2 fixation, the Wakeby distribution fits the best. This distribution, which is rarely encountered in psychology, has five parameters: location, shape and scale parameters for left tail, and separate shape and scale parameters for the right tail. Separate parameters for each tail made the distribution flexible enough to accommodate the very short (~30 ms) fixations, which our method had failed to separate off. Because of the loss of uniqueness resulting from the use of five parameters, we forgo on a process interpretation of this fit. Let it be said that this distribution has well-defined, finite, population means and variance, which enables the envisaged application of the GLM to sample means.

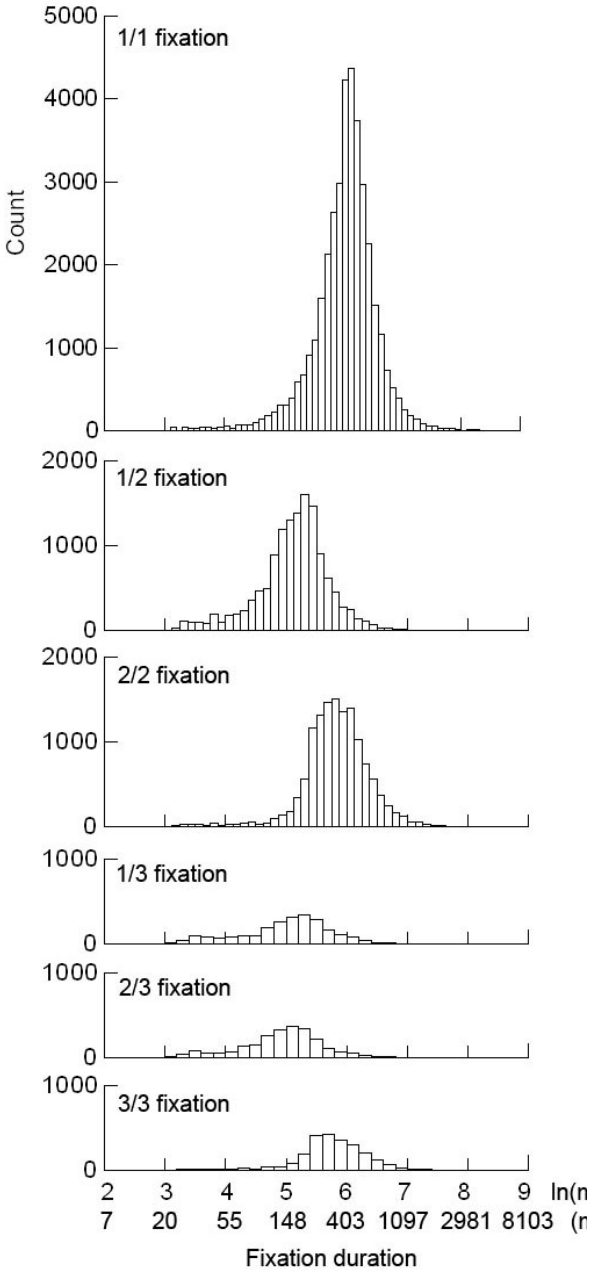


Figure 3. Fixation durations distributions in ln ms units plotted for six types of fixations

From top to bottom: 1st/1-fixation, 1st/2-fixation, 2nd/2-fixation, 1st/3-fixation, 2nd/3-fixation and 3rd/3-fixation distributions, where nth/m-fixation means the n-th fixation from those trials in which m fixations to an object were made.

DOI 10.16910/jemr.1.2.2
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

Table 2. Statistics of six fixation types

Type of fixation	1st / 1-fixation	1st / 2-fixation	2nd /2-fixation	1st / 3-fixation	2nd / 3-fixtion	3rd / 3-fixation
No. of fixation	37576	13499	13499	2501	2501	2501
Median	6.09, ~444 ms	5.19, ~180 ms	5.84, ~344 ms	5.02, ~152 ms	5.12, ~168 ms	5.76, ~318 ms
Mean	6.04 , ~420 ms	5.12, ~167 ms	5.83, ~339 ms	4.92, ~137 ms	4.99, ~147 ms	5.74, ~310 ms
SD	0.54	0.59	0.55	0.69	0.76	0.62
Variance	0.29	0.35	0.31	0.47	0.58	0.38
Skewness (SE)	-1.00 (0.01)#	-0.61 (0.02)#	-0.81 (0.02)#	-0.38 (0.05)#	-0.43 (0.05)#	-0.91 (0.05)#
Kurtosis (SE)	4.16(0.03)#	1.25 (0.04)#	3.84 (0.04)#	0.36 (0.10)#	-0.04 (0.10)	2.75 (0.10)#

Note. Median and mean durations are also presented in ms. The symbol '#' indicates that |skewness/SE skweness| >2 and | Kurtosis /SE kutosis| > 2, which is the criterion used for a non-normality.

The generalized logistic distribution (three parameters; location, scale and shape) was ranked as second or third best in all fixation types. For the 2nd/2 fixation the log-logistic distribution was the best fitting. The log-logistic, which belongs to the generalized logistic family, also mostly comes second or third, and first in the 2nd/2 fixation type. Like the Wakeby distribution, ithe generalized (log) logistic distribution has finite population means and variances, which enables application of the GLM to the sample means.

The generalized logistic distribution has only a single shape parameter and is therefore less flexible than the Wakeby distribution, of which two shape parameters can be adjusted for right and left tails independently. Based on this observation, we may conclude that the generalized logistic distribution is the most plausible candidate as the theoretical distribution of all the separated fixation durations. The theoretical implications of this will be mentioned in the discussion; for now it is of importance that the conclusion of a uniform distribution means that the data are homogeneous and can jointly be evaluated by a single statistical model.

Mean fixation duration was computed for each fixation type for condition in each participant. Mean and SD (in parenthesis) of the means in each fixation over all conditions were, 6.08 (0.24), 5.08 (0.24), 5.81 (0.17), 4.82 (0.30), 4.85 (0.32), and 5.70 (0.22) for 1st/1, 1st/2, 2nd/2, 1st/3, 2nd/3 and 3rd/3 fixations, respectively. On the other hand, mean (and SD) of the means in each conditions over the fixation types were; 5.47 (0.52), 5.43 (0.54), 5.48 (0.50), 5.48 (0.54), 5.48 (0.51), 5.09 (0.60) and 5.26 (0.60) for category-judgment, category-count, category-pattern, location-judgment, location-count, location-pattern task and control conditions, respectively. Condition means in each fixation type are shown in Figure 5. We tested whether task effects were significant and consistent across fixation types by fitting a linear mixed effect (LME) model. Fixation types and task conditions were within-subject factors, thus within-subject covariance is a potential cause of Type - I error. The LME

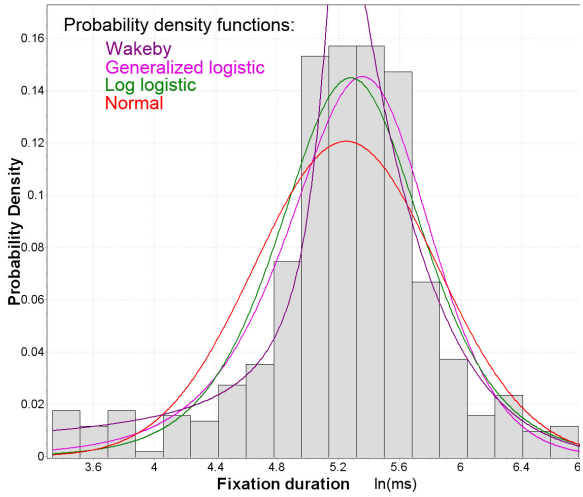


Figure 4. Fitting probability density functions to an empirical fixation-duration distribution. Distribution histogram of the 1st/2 fixations (N=509) of Participant 2 in the category-count condition is given as a representative example. The top-three best-fitting probability density functions are given: Wakeby, generalized logistic, and log-logistic, plus the normal distribution. The empirical data were statistically indistinguishable from each of the top three distributions (p > .1 on Kolmogorov-Smirnov tests), but different from the normal distribution (p < .01)

DOI 10.16910/jemr.1.2.2
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

Table 3. Average rank order of fitted distributions

Note. Top three distributions are listed for each fixation type. Their average rank is shown in parenthesis.

Rank	1st / 1-fixation	1st / 2-fixation	2nd /2-fixation	1st / 3-fixation	2nd / 3-fixtion	3rd / 3-fixation
1	Wakeby (1.69)	Wakeby (2.23)	Log logistic (3.23)	Wakeby (1.29)	Wakeby (2.76)	Wakeby (2.00)
2	Generalized logistic (3.49)	Generalized logistic (4.73)	Wakeby (3.48)	Generalized logistic (3.88)	Generalized logistic (6.70)	Log logistic (3.88)
3	Log Logistic (3.97)	Log Logistic (6.38)	Generalized logistic (3.60)	Log Logistic (5.65)	Generalized extreme value (6.76)	Generalized logistic (4.64)

model is able to estimate unknown covariance, such as individual differences (Pinheiro & Bates, 2000). The model also accommodates differences in the number of means per cell. These occurred since some participants did not show the triplet of fixations, 1st/3, 2nd/3 and 3rd/3, in each of the conditions. In addition, some segments of eye movement data were omitted due to bad quality of recording. The best fitting MSE model was tested using the F-statistic

The fixed factors of the LME model were fixation types (6) and conditions (7; the six task conditions and the control condition), while participants constituted a random variable. Effects of random variables were estimated by restricted likelihood method. Main effects of both fixation types and tasks are highly significant, $F(5, 185) = 290.73$, $p < .0001$ and $F(6, 185) = 20.26$, $p < .0001$, respectively. But the interaction of the two was not significant, $F(30, 185) = 0.98$, $p > .1$. The absence of an interaction indicates the consistency of the task effects on fixation duration across fixation types.

Post-hoc tests were performed, in which participants were the only random variable. Thus, we mention only the fixed variables of each LME model, hereafter. The number in parenthesis after each variable indicates its number of levels in the model; p-values are uncorrected. The difference between the task and control conditions was evaluated by a model including fixation types (6) by six pooled task conditions vs. the control condition (2). The main effect of task vs. control condition was significant, $F(1, 215) = 14.08$, $p < 0.001$; as shown in Figure 5 longer in the task than in the control condition. This result signifies that fixation durations differ according to non-mandatory perceptual, i.e. cognitive processes. Effects of fixation type also occurred in this analysis, but no interaction with task effects occurred. This means that perceptual processes consistently affected the fixation durations, irrespective of fixation type. Visual inspection of Figure 5 suggests, moreover, that fixation durations in the location-pattern condition were shortest amongst the task conditions. Secondly, therefore, a post-hoc analysis compared this condition with the other five task conditions; a model was fit including fixation types (6) by location-pattern vs. the other five pooled task conditions (2). The main effect of task was significant, $F(1, 179) = 102.34$, $p < .0001$; the location pattern task yielded shorter fixations than all the other task conditions. Interestingly, the location pattern task yielded fixation durations at the level of the control condition, or even marginally shorter $F(1,48) = 3.67$, $p < .1$. Most likely, therefore, in this task conditions, visual inspection is minimized. On the other hand, a model including fixation types (6) by the remaining five task conditions (5) failed to yield a significant effect, $F(4, 131) = 1.64$, $p > .1$. This, despite the difference in post-perceptual processing demands between these tasks. Together, these analyses indicate that fixation durations differentiate according to (non-mandatory) perceptual processing demands, but not according to post-perceptual demands. In all these post-hoc tests, fixation type effects occurred, but none of them yielded an interaction with task conditions; all $Fs < 1.4$, $p > .1$. The results showed that task-related, i.e. cognitive, perceptual processing demands influenced fixation duration, but that post-perceptual cognitive factors did not. These results were consistent in log-transformed units of duration across all fixation types.

Post-hoc testing of the differences among fixation types started off by comparing the duration of the last fixation (i.e., 1st/1, 2nd/2 and 3rd/3) with that of the earlier ones (1st/2, 1st/3 and 2nd/3). A model with the last vs.

DOI 10.16910/jemr.1.2.2
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

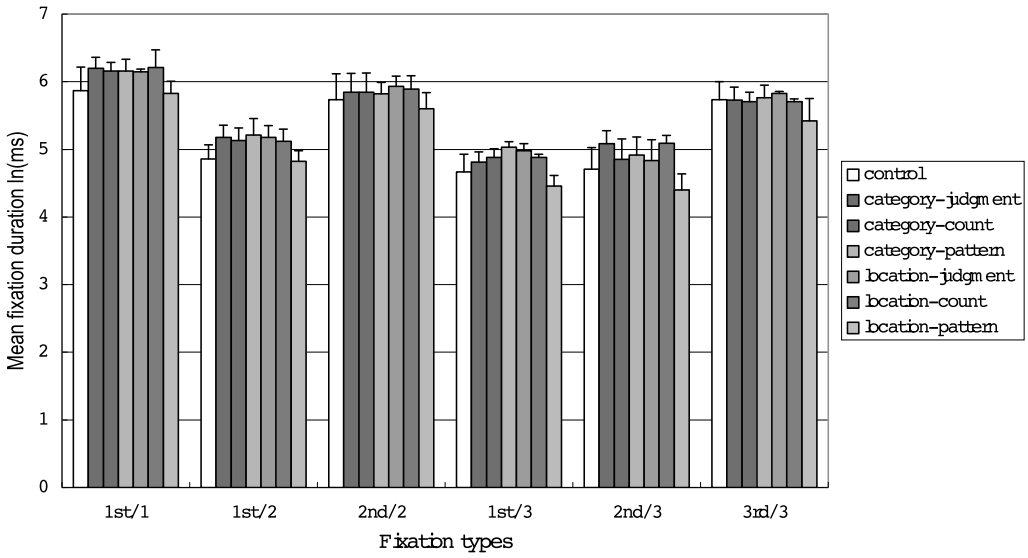


Figure 5. Mean fixation duration in task and control conditions in six fixation types. Mean of log-transformed fixation durations were plotted for each fixation type; from the left ,1st/1-fixation, 1st/2-fixation, 2nd/2-fixation, 1st/3-fixation, 2nd/3-fixation and 3rd/3-fixation, where nth/m-fixation means the n-th fixation from those trials in which m fixations to an object were made. Vertical lines indicate SD.

earlier fixations (2) by conditions (7) was fit. As expected, the last fixation was significantly longer than the earlier fixations, the main effect of the last vs. the earlier yielded $F(1, 213) = 879.69$, $p < .0001$. To test within the last fixations: one model tested 1st/1 vs. 2nd/2 (2) by conditions (7); the main effect of fixation type was $F(1, 57) = 58.83$, $p < .0001$; 1st/1 was longer than 2nd/2 fixations. Analogously, a second model tested 1st/1 vs. 3rd/3 (2) by conditions (7), yielding that 1st/1 were significantly longer than. 3rd/3 fixations, $F(1, 57) = 72.60$, $p < .0001$, and a third model tested 2nd /2 vs. 3rd/3 fixations (2) by conditions (7), showing that 2nd /2 were longer than 3rd/3 fixations, $F(1, 57) = 7.84$, $p < 0.01$. To test for differences among the first fixations, a model included 1st/2 vs. 1st/3 fixations (2) by conditions (7), 1st/2 fixation was longer than 1st/3 fixations, $F(1, 57) = 31.33$, $p < .0001$. Analogously, between second fixations, 1st/2 and 2nd/3 fixations (2) by conditions (7), and former were longer than the latter, $F(1, 57) = 23.00$, $p < .0001$. However, duration of 1st/3 and 2nd/3 fixations were not significantly different; the F value for the main effect of fixation type reached only $F(1, 57) = 0.31$, $p > .1$. Notice that in all these post-hoc tests, there are effects of task; however, interactions between fixation type and task where nowhere obtained. All fixation types, therefore, are equally affected by task conditions. This, even though some fixation durations may include a manual response component (the last fixations: 1st/1, 2nd/2 and 3rd/3), whereas the others did not. This indicates that even though manual responses may have an effect on fixation durations, this does not affect the differences between task (and control) conditions.

Discussion

We compared six different task conditions (and a control condition) all using the same visual stimuli, and studied the patterns of eye-fixation, to see if measures of fixation duration could meaningfully be applied across these tasks. The visual stimuli in the current experiments were simple ones, yet fixation patterns were not simple. It was confirmed that the order and number of fixations given to an object was of predominant importance to fixation duration. This observation is in line with a common finding in reading tasks. Here, the effect of order and number of fixations outweighs that of task differences, and therefore that of cognition. It is therefore important to see if the effects of task, c.q. cognition are consistent across fixation types.

To test the consistency, fixation duration distributions had to be separated according to fixation type, natural-log

DOI 10.16910/jemr.1.2.2
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

transformed, and investigated for their theoretical distribution, before the effect of task could be examined. Effects of task-related perceptual demands on fixation duration, in log units, were observed to be consistent for all fixation types. Fixations were generally longer in task than in control conditions, but in the location-pattern task they were equal, or even shorter, than in the control condition. The location task does not require inspection of individual targets, as only their locations matter, so it is likely that the non-mandatory aspects of the perceptual process are minimized in this condition. There was no difference in duration amongst the remaining tasks. We may conclude that the fixation durations are insensitive to post-perceptual processing demands imposed by our tasks. Interestingly, this was equally the case for all fixation types. This conclusion runs counter to de Graef et al. (1990), who argued that perceptual and post-perceptual processes differentially affect first and subsequent fixations. Difference in method could be one of the reasons for this discrepancy, as the analysis in our Table 1 might suggest. Alternatively, we may consider a difference in stimuli: In De Graef's et al (1990) study, semantically rich scenes were used. Nevertheless, the consistency across fixation types observed in the present study underpins the use of fixation duration to estimate perceptual processing demands, at least for a category of relatively simple stimuli.

Number and order of fixation are large sources of variance; the proposed procedure to factor them out will, therefore, yield greater power to detect effects of cognitive process of interest. The proposed analysis failed, however, to separate short fixations (latency around 30 ms); they may be held responsible for the right-skewness of the separated log-fixation durations. This might correspond to the duration of micro-saccades (cf. Engle, 2008 for review), of which investigation requires eye movement record with higher spatial precision than our current one. We were able to monitor, and factor out, the fixations where a manual response occurred. This happened, by definition, in the last fixation. On the other hand, the consistency of the task effects across all fixations types, including those which contained manual responses, means that these did not interfere with task-related effects on fixation durations.

The Wakeby distribution renders overall the best fitting probability function. The Wakeby distribution is widely used in hydro-engineering to model flood level/ extreme rainfall, which is highly variable (Houghton, 1978). The ability of these models to accommodate asymmetry due to the short fixations is offset by the need to have two extra parameters, compared to the next-best distribution. Therefore, from the point of model uniqueness, and because of the insignificance, from a cognitive point of view, of the short fixation durations, we prefer a distribution, which was ranked as the second: the generalized logistic distribution. This distribution has an interesting theoretical implication. It is known as the distribution of the maxima of finite samples, with variable size, taken from an exponential distribution (Galambos, 1978; Gnedenko, 1982; Voorn, 1987). This would imply that fixation duration reflects the waiting time for a set of N simultaneous processes to complete, each of which has a fixed probability of terminating per unit of time. Note that these are logarithmic units; in linear units, these processes would follow a power-law. Such an interpretation of fixation durations is intuitively plausible, and a natural extension of the idea that they are based on a single stochastic transition between discrete internal states (Engbert & Kliegl, 2001). Instead of a single transition, the eye-movement waits for the slowest of N transitions to be completed. Instead of stochasticity, the power-law distribution might suggest some form of determinism. We cannot tell at this point, whether this "waiting" is under the executive control, but this would certainly be an interesting issue to pursue.

For studies in which the relative effect of visuo-cognitive factors on fixation duration is the main interest, our procedure provides a pragmatic solution; fixations should be separated by its number/order in a region of interest, log transformed, and their distributions investigated. Depending on the outcome of these tests, parametric models can be applied directly to the data (if they are normally distributed), to their means (if they are not, but can reasonably be assumed to follow the same theoretical distribution, with well-defined means and variances), or be analyzed separately for different fixation types.

Beyond such procedural issues, identification of fixation duration distribution has a more fundamental significance as well. In the current study, we propose that the duration reflects waiting time for completion of multiple processes, of which the processing time follows a power-low distribution. Power-low distributions have a long right tail. Thus it is natural for long fixation durations to occur and such large variability does not constitute a reason against using fixation duration as a measure of cognitive processes (cf. Viviani, 1990). Adequate models of fixation duration distributions will help us to further to understand the underlying processes behind.

DOI 10.16910/jemr.1.2.2
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

References

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52,* 317-329.

Engbert, R. (2006). Microsaccades: a microcosm for research on oculomotor control, attention, and visual perception. *Progress in Brain Research, 154*, 177-192.

Engbert, R. & Kliegl, R. (2001). Mathematical models of eye movements in reading: a possible role for autonomous saccades. *Biological Cybernetics, 85*, 77 -87.

Feng, G. (2006). Eye movements as time-series random variables: A stochastic model of eye movement control in reading. *Cognitive Systems Research, 7*, 70-95.

Fischer, B. & Ramsperger, E. (1984). Human express saccades: extremely short reaction times of goal directed eye movements. *Experimental Brain Research, 57*, 191-195.

Fischer, B. & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences, 16*, 553-567.

Galambos, J. (1978). *The asymptotic theory of extreme order statistics.* New York, NY: Wiley.

Gnedenko, B. V. (1982). On limit theorems for a random number for random variables. *In K. Ito and J. V. Prokhorov (Eds.). Probability theory and mathematical statistics. Proceedings for the 4th USSR-Japan Symposium*, *Tbilisi, USSR, Lecture Notes in Mathmatics,* 167-176. Berline, Germany: Springer-Verlag.

Henderson, J. M. & Hollongworth, A. (1998). Eye movements during scene viewing: an overview. *In G. Underwood (Ed.). Eye Guidance in Reading and Scene Perception.* Amsterdam, The Netherlands: Elsevier.

Houghton, J. C., (1978). Birth of a parent: the Wakeby distribution for modeling flood flows, *Water Resources Research, 14,* 1105-1110.

Inhoff, A. W. & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. *In G. Underwood (Ed.). Eye Guidance in Reading and Scene Perception.* Amsterdam, The Netherlands: Elsevier.

Kennison, S. M. & Clifton, C. (1995). Determinanys of parafpveal preview benefit in high and low working memorycapacity readers: Implications for eye movement control. *Journal of Experimental Psychology: Memory and Cognition, 21*, 68-81.

O'Regan, J. K. & Levy-Schoen, A. (1987). Eye movement strategy and tactics in word recognition and reading. *In M. Coltheart (Ed.). Attention and performance XII: The psychology of reading,* 363-383. Oxford, UK: Erlbaum.

O'Regan, J. K., Vitu, F., Radach, R., & Kerr, P. W. (1994). Effects of Local processing and occulomotor factors in eye movement guidance in reading. *In J. Yagge & G. Lennerstand (Eds.), Eye Movements and Reading*, 329-348. Oxford, UK: Pergamon.

O'Regan, K. (1992). Optimal viewing position in words and strategy-tactics theory of eye movements in reading. *In K. Rayner (Ed.), Eye Movements and Visual Cognition: Scene Perception and Reading.* New York. NY: Springer.

Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS.* Berlin, Germany: Springer-Verlag.

Raney, G. E. & Rayner, K (1995). Word frequency effects and eye movements during two readings of a text. *Canadian Journal of Experimental Psychology, 49,* 151-172.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin,124,* 372-422.

Rayner, K. & Duffy, S. A. (1986). Lexcal complexity and fixation times in reading: Effects of word frequency, verb complexity and lexical ambiguity. *Memory & Cognition, 14*, 191-201.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye Movement Control in Reading: A Comparison of Two Types of Models. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1188-1200.

Rayner, K., Sereno, S. C., Morris, R.K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognition Processes, 4* (Special issue), 21-49.

Sereno, S. C. (1992). Early lexical effects when fixating a word in reading. *In K. Rayner (ed.). Eye movements and visual cognition: Scene perception and reading,* 304-316. New York, NY: Springer-Verlag.

DOI 10.16910/jemr.1.2.2
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692

Velichkovsky, B. M. (1999). From levels of processing to stratification of cognition: Converging evidence from three domains of research. *In B. H. Challis, & Velichkovsky, B. M. (Eds.). Stratification in cognition and consciousness. Advances in consciousness research,* 203-235. Amsterdam, The Netherlands: John Benjamins.

Vitu, F., McConkie, G. W., Kerr, P., & O'Regan, J. K. (2001). Fixation location effects on fixation durations during reading: An inverted optimal viewing position effect. *Vision Research, 41, Special Issue: Eye movements and vision in the natural world: In honour of Han Collewijn,* 3513-3533.

Viviani, P. (1990). Eye movements in visual search: cognitive, perceptual and motor aspects. *In E. Kowler (Ed). Eye movements and their role in visual and cognitive processes,* 353-393. Amsterdam, The Netherlands: Elsevier.

Voorn, W. J. (1987). Characterization of the logistic and loglogistic distributions by extreme value related stability with random sample size. *Journal of Applied Probability, 24*, 838-851.

Yang, S. N. & McConkie, G. W. (2001). Eye movements during reading: a theory of saccade initiation times. *Vision Research, 41, Special Issue: Eye movements and vision in the natural world: In honour of Han Collewijn,* 3567-3585.

Yarbus, A. (1967). *Eye movements and vision.* New York, NY: Plenum Press.

DOI 10.16910/jemr.1.2.2
This article is licensed under a
Creative Commons Attribution 4.0 International license.
ISSN 1995-8692