Distributions of fixation durations and visual acquisition rates

Arthur Lugtigheid

Supervisors
Prof. Dr. M.A. Frens
(Department of Neuroscience, Erasmus MC Rotterdam)

Dr. R. Zeelenberg
(Department of Psychology, Erasmus University Rotterdam)

Studentnumber
277528

Abstract

In this paper, we investigated fixation durations and distributions of visual acquisition times in two experiments. In the first experiment, we reanalyzed fixation data obtained from previous experiments involving visual search and showed that: (a) the distribution of fixation durations is best described by the lognormal distribution and (b) there is a strong linear relationship between the mean fixation duration and the standard deviation of fixation duration. In the second experiment, we investigated how much of the variability in fixation durations can be explained by variability in visual acquisition rates (which we defined as the combined processes of the sampling of the retinal image and the processing of visual information). We designed a two-alternative forced choice discrimination task that mimics a visual search paradigm, but in which the maximal duration of visual acquisition was manipulated. We found a linear relationship between the mean required stimulus duration and the standard deviation of required stimulus duration that was almost identical to the relationship found in experiment 1 and conclude that variability of fixation durations can be explained by the variability in visual acquisition rates.

Introduction

In observing the world around us, we make many eye movements. These eye movements are necessitated by the structure of the human retina. Only a small part of the retina, the fovea, has a high density of photoreceptors and makes up the high resolution central part of the retina that is responsible for high-acuity vision. Consequently, if we want to inspect interesting or important details of the visual world, they have to be projected onto the fovea. This is accomplished by the execution of saccades, which are fast eye movements that bring the fovea to a new portion of the visual scene. Acquisition of visual information occurs during fixations, which are the inter-saccadic intervals in which the eye is relatively stable. The location of the fixation determines what information is available for processing, whereas fixation duration mainly determines how much of the available information can be acquired. The importance for repeated sequences of saccades and fixations is especially demonstrated in every day tasks like reading, or visual search: in both tasks, a series of fixations is required to acquire information from the visual scene—information that can not be acquired from a single fixation.

Defining the processes that take place during a fixation has been the subject of many studies of the past decades. Viviani (1990), for instance, argued that during a fixation at least three processes may occur. These processes are (1) sampling of the visual field; (2) analysis of the foveal part of the visual field; and (3) planning of the next saccadic eye movement. These processes do not
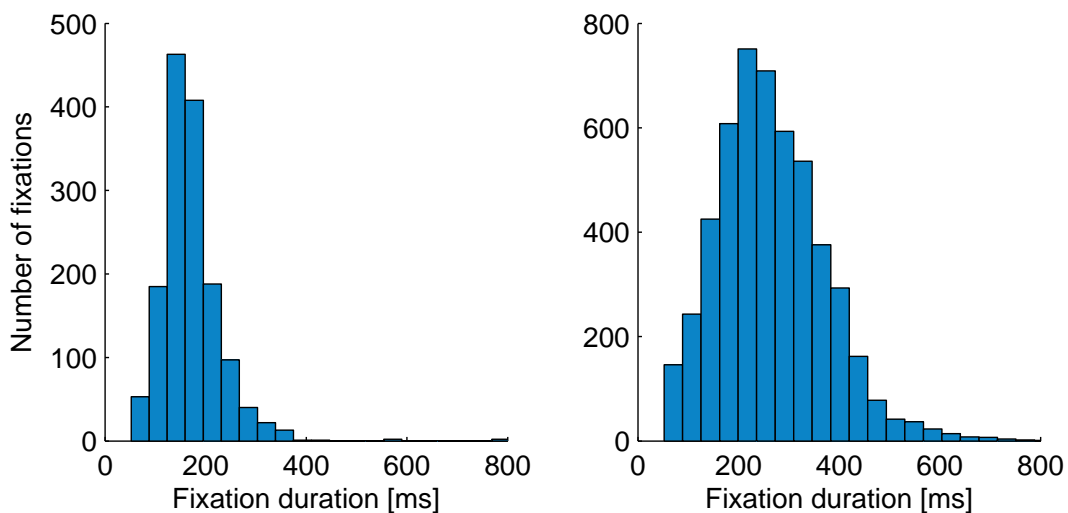

Figure 1: Two representative frequency distributions of fixation durations. Data were obtained from the fixation dataset of experiments 1 and 2 in Vlaskamp et al. (2005).

necessarily occur serially, since evidence suggests that at least the planning of the next eye movement may take place during the analysis of the foveal part of the visual field (Hooge & Erkelens, 1996). Since little or no information can be a aquired during a saccade, it is reasonable to assume that the processes described by Viviani (1990) should be reflected in the duration of a fixation.

Empirical evidence points to two general characteristics of fixation durations. First, the duration of fixations are highly variable on a trial-by-trial basis (eg. Buswell, 1935). Second, distributions of fixation durations are almost always reported to be non-normal with a characteristic skew to the right (Harris, Hainline, Abramov, Lemerise, & Camenzuli, 1988; Suppes, 1990), suggesting that fixation durations are stochastic by nature (see figure 1). Thus, distributions of fixation times are highly predictable, but the durations of individual fixations are not. The properties (eg. the skewness and kurtosis) of distributions of fixation durations are of considerable interest since they can provide information about the process that underlies the control of fixation duration— information that is not readily represented by first order statistics like the mean and standard deviation. As a consequence, models aiming to quantify this process produce distinct predictions about the distributions of fixation times.

The main purpose of the study reported here is to examine fixation durations and their distributions under conditions where involvement of (complex) cognitive processes is minimal. In the first section (which we will refer to as experiment 1) we will discuss two attempts at quantitative modeling of distributions of fixation duration and evaluate their assumptions by reanalyzing data from previous experiments involving visual search and crowding. In the second section, we will present and discuss the results of a second experiment in which we manipulated the time that is available for processing.

Experiment 1

In the past decades several attempts have been made to model the shape of the distribution of fixation durations (e.g. Suppes, 1990; Harris et al., 1988; McConkie & Dyre, 2000). In this section, we will review two attempts at quantitative modeling of distributions of fixation durations. We do not intend to provide a comprehensive description of the models, but instead will focus on the predictions made by these models regarding the shape of the distribution of fixation durations. We will then evaluate these models by retrospectively reanalyzing data from previous experiments involving visual search.

Models of fixation durations

Harris et al. (1988) reported that distributions of fixation durations for infants and naive adults who were engaged in free viewing, could best be described by an exponential distribution with a randomly distributed onset time. The exponential distribution is a waiting-time distribution and it was chosen for this model because it represented a memoryless system and it has a constant termination rate. They suggested that the duration of a fixation could be divided into two sequential stages, called the alpha- and beta-periods. The alpha-period was thought to represent the time between the onset of a fixation until the start of the processing of information made available from that fixation. It was conceptualized as a refractory period in which fixations were not terminated. The beta-period was thought to represent a period in which processing of information took place, the duration of which was represented by a waiting time for the execution of the next saccade that would be triggered by some random terminating impulse. This impulse would occur with a constant probability, described by an exponential function. In this sense, the model proposed by Harris et al. is equivalent to the convolution of two distributions; one based on an exponential function, representing the random waiting time for a termination signal in the beta-period, and the other based on some function representing the random onset times for when the beta period begins.

There are two main problems concerning this model; First, the data sets on which Harris et al. (1988) based their analysis were very small. They reported sample sizes ranging from ten to a few hundred fixations, therefore introducing quite a lot of sampling error. Second, the distribution that described the alpha-period was never specified by Harris et al. (1988) and so it remains unknown exactly which distributions the model is a convolution of, making it nearly impossible to truly test the model.

The second model evaluated here is a model proposed by Carpenter (Carpenter & Williams, 1995; Reddi & Carpenter, 2000; Reddi, Asrress, & Carpenter, 2003), who showed that the reciprocal of saccadic latency times (defined as the time from the onset of a suddenly appearing target until a saccade is initiated) obeys a normal distribution. The resulting distribution is called a recinormal distribution. Their LATER model (Linear Approach to Threshold with Ergodic Rate) provides a simple explanation for the recinormality of saccadic latency times. In their model, incoming information causes a decision signal S to rise linearly with a rate r from an initial value S_0 until it reaches a threshold level S_T at which point a response (e.g. a saccade) is initiated. The time taken to reach the threshold is equal to the fixation duration. The resulting recinormal distribution is explained by supposing that the rate of rise r varies randomly between trials in a normal fashion (see figure 2).
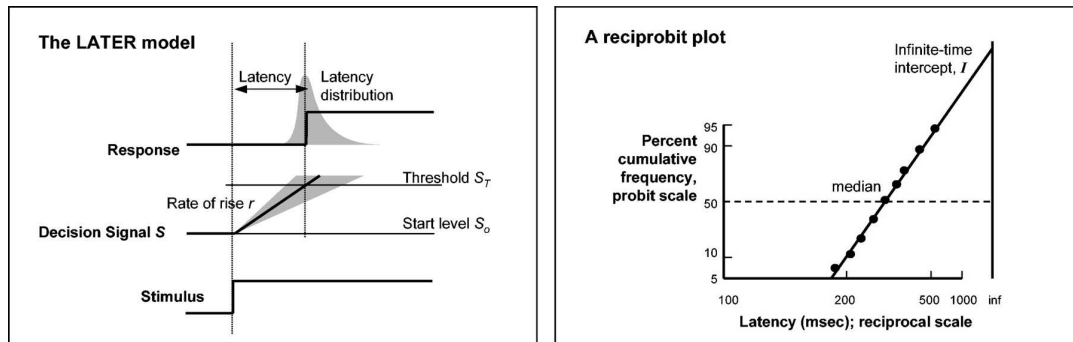
Figure 2: The LATER model (Reddi & Carpenter, 2000). The left panel shows a schematic representation of the model; upon the presentation of a stimulus, decision signal S rises from start level S_0 with rate r to threshold S_T, resulting in a recinormal latency distribution. The right panel shows a reciprobit plot, which shows that the inverse transformation of latency is normal: if the cumulative frequency (on a z-scale) is plotted against $\frac{1}{RT}$ it yields a straight line.

The model has been shown to account for manipulation of prior probability (Carpenter & Williams, 1995), speed versus accuracy instructions (Reddi & Carpenter, 2000) and the supply of information (Reddi et al., 2003). More recently, Carpenter and McDonald (2007) showed that the LATER model also predicts distributions of fixation duration in reading. There is also strong evidence for this model from neurophysiology, where it is shown that populations of visuo-motor neurons in the frontal eye fields of monkeys begin firing in advance of a saccade, with the activity rising linearly as a reaction on a stimulus. The rate of rise varies randomly from trial to trial and the saccade itself is initiated when this activity reaches a fixed threshold (e.g. Schall & Hanes, 1998).

One problem of the model, however, is the assumption that the variability of the rate of rise r is normally distributed. If this variability is not normally distributed, the distribution of saccadic latency times can produce a distribution that is not recinormal. Furthermore, Ratcliff (2001), using simulations, showed that the results presented by Carpenter could also be explained with a diffusion model. This model assumes the same gradual accumulation of evidence as the LATER model, but adds noise to this accumulation of information, which also allows the model to explain how incorrect responses are generated.

Method

Distributions of fixation duration predicted by the models of Harris et al. (1988) and Carpenter (2000) were compared to eye fixation data obtained from previous visual search experiments carried out by Vlaskamp, Over, and Hooge (2005) in which the effect of element spacing on search performance was investigated. In each trial of these experiments, a search display was presented and observers

5

were instructed to find a target (a closed square) in a display of distractors (squares with gaps in them) with different element spacings. The distance between elements was varied by varying the number of elements in the display while keeping the display size constant. Furthermore, the target-distracter similarity was varied by varying the gap size in the distractors. The experiment consisted of 12 conditions; four element spacings in combination with three distractor gap sizes (for a full description of the experiments, see Vlaskamp et al., 2005). In total, ten observers participated, yielding a dataset of 338141 fixation durations from 157 combinations of displays.

Predictions of each model were fit to the distributions of fixation duration obtained from the 157 combinations described above, using the method of moments. This method is a straightforward test of distributions and involves tests of higher order moments of a distribution. Although not conclusive for all distributions, the third and fourth central moment typically provide an accurate and unique description of the shape of statistical distributions. The third central moment of a distribution is the skewness, which is a measure of the degree of asymmetry of a distribution. If the left tail is more pronounced than the right tail, the distribution is said to have a negative skew. If the reverse is true, it is said to have a positive skew. If both tails are equal, it has zero skewness. The fourth central moment of a distribution is the kurtosis, which is a measure of 'peakedness' of a distribution. A distribution with positive kurtosis (leptokurtic) has heavier tails and a higher peak than the normal distribution, whereas a distribution with negative kurtosis (platykurtic) has higher tails and is flatter.

Results

The gamma distribution

To test the model of Harris, we used an exponential distribution, which is a gamma distribution of the first order. The order of a gamma function is denoted by shape parameter α. If α is an integer then the distribution represents the sum of α exponentially distributed random variables. One property of the exponential is that the coefficient of variance (CV, defined by the formula $\frac{\sigma}{\mu}$) is always equal to 1, thus $\sigma = \mu$ (see Harris et al., 1988). Figure 3a shows σ as a function of μ for our distributions. There is a clear linear relationship, with a slope of 0.51 and an R^2 of 0.738 ($p < 0.01$), but because the slope was not equal to 1 (i.e. the data points do not follow the unity line) we rejected the exponential distribution as a model for our data. This conclusion was confirmed by subsequent analysis of shape parameter α and the higher order moments of our distributions.
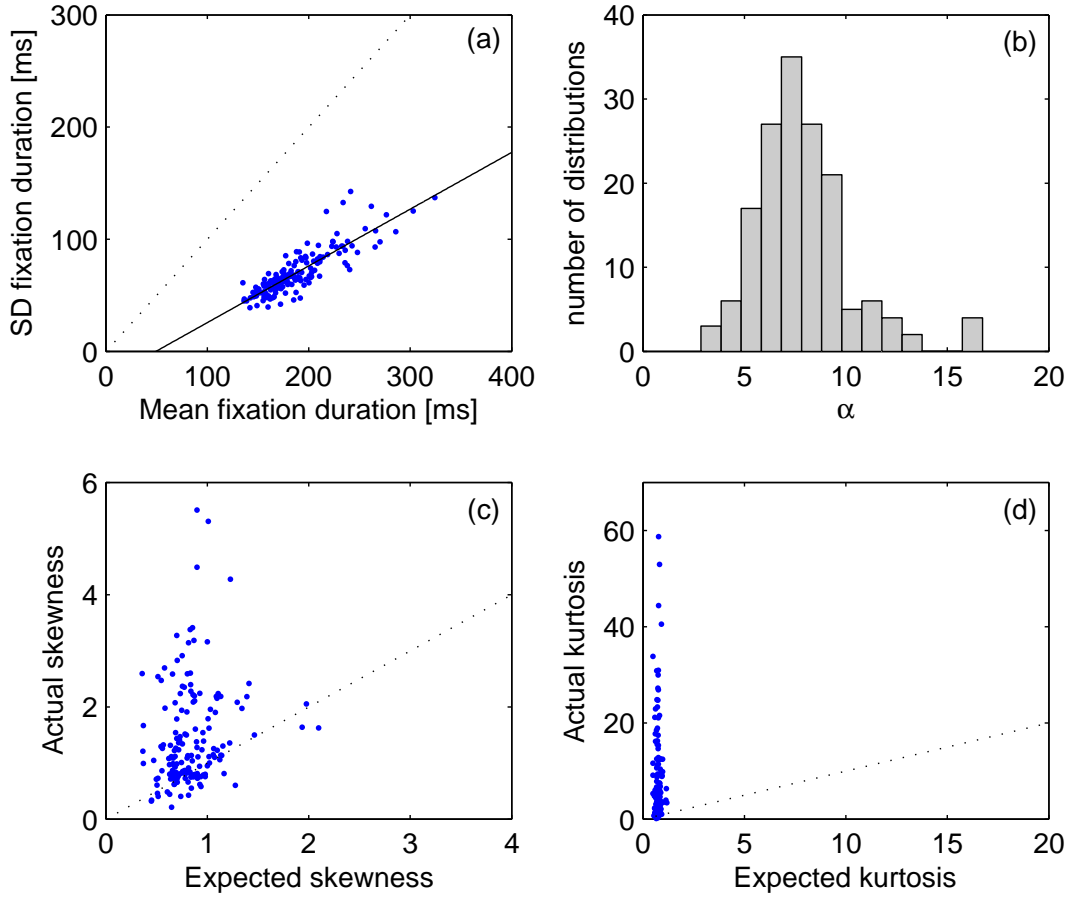
Figure 3: Our analysis of the gamma distribution. (a) σ as a function of μ. The linear relationship observed here follows the formula $y = 0.51x - 25$. (b) Histogram of the best fitting α for our data. (c) The actual skewness as a function of the predicted skewness. (d) the actual kurtosis as a function of the predicted kurtosis. Dotted lines in panels (a), (c) and (d) represent unity. Data were obtained from the fixation data set of experiments 1 and 2 in Vlaskamp et al. (2005).

The order of the best fitting gamma distribution for each of our distributions was calculated using the following formula (see appendix A for derivations):

$$\alpha = \frac{1}{CV^2} \tag{1}$$

Figure 3b shows a histogram of the results. As can be seen, our data is best described by a gamma distribution of the seventh order ($\alpha = 7$), but it highly unlikely that the duration of a fixation is the result of seven independent processes, each of which are exponentially distributed.

To calculate the predicted skewness (S) and kurtosis (K), we used the following formula's (see appendix A for derivations):

$$K = 6CV^2 \qquad (2)$$
$$S = 2CV \qquad (3)$$

The results are shown in figures 3c and d. If our data could be modeled by a gamma distribution, all data points for both the skewness (figure 3c) and the kurtosis (figure 3d) would be located on or around the unity line. As can be seen from the figures, this is not the case for our data.

The Recinormal distribution

The LATER model (Reddi & Carpenter, 2000) assumes a recinormal distribution of fixaiton durations. Because the recinormal distribution is derived from the normal, we expected the distribution to be normal after reciprocal (1/RT) transformation. The normal distribution has a skewness and (excess) kurtosis of zero. Consequently, if our distributions can be described by a recinormal distribution, then the reciprocal of our distributions would yield data points that are clustered around zero skewness and kurtosis (in figure 4 represented by the
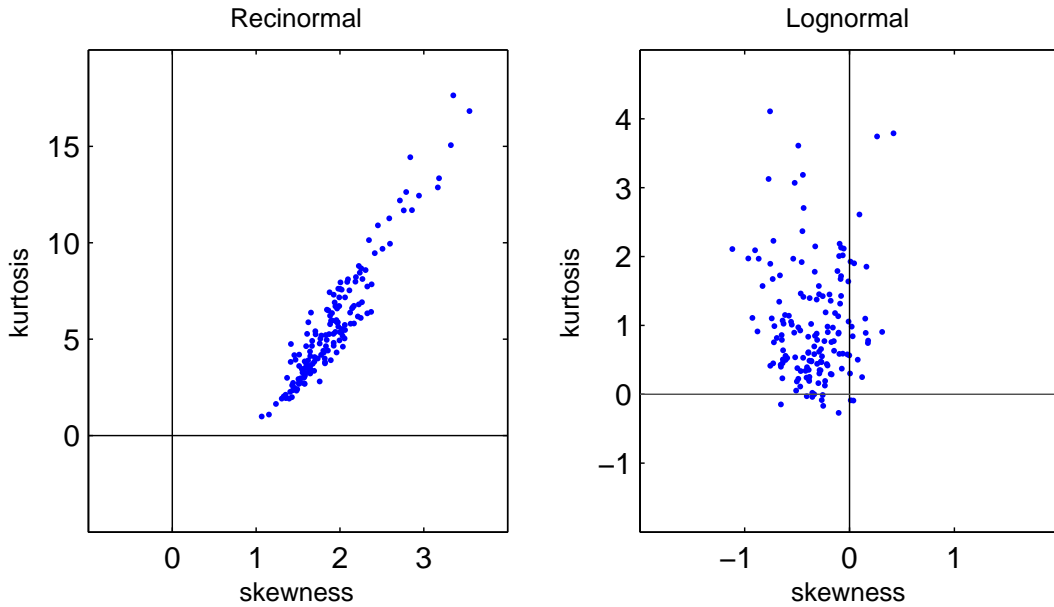

Figure 4: (a) Skewness as a function of kurtosis for the recinormal distribution. (b) Skewness as a function of kurtosis for the lognormal distribution. Data were obtained from the fixation data set of experiments 1 and 2 in Vlaskamp et al. (2005).

intersection of the horizontal and vertical line). As can be seen from figure 4, this is not the case for our data, and thus the recinormal distribution was rejected as a model for our fixation data. There is however an apparent linear relationship, a property which we can not readily explain and might require further investigating.

Other possible distributions

So far we rejected both the exponential and the recinormal distributions as possible models for our data. The typical skewed shape of the distributions, however, gives rise to several other candidate distributions, such as the lognormal distribution. The evaluation of the lognormal distribution was carried out in the same manner as the recinormal distribution, because if a distribution is distributed lognormally, then the logarithm of this distribution is normally distributed. If our distributions follow a lognormal, then the observed skewness and kurtosis should be located on or around the point where both skewness and kurtosis are equal to zero as is the case for a normal distribution. As can be seen from figure 4b, the lognormal provides a good fit to our distributions.

A summary of the analysis of the higher order moments of our distributions is depicted in figure 5. It shows the mean and median difference (here refered to as the error) between the observed and the expected skewness and kurtosis. An error of zero would mean that our distributions fit the model perfectly. From this figure it is clear that the lognormal distribution provides the best fit.
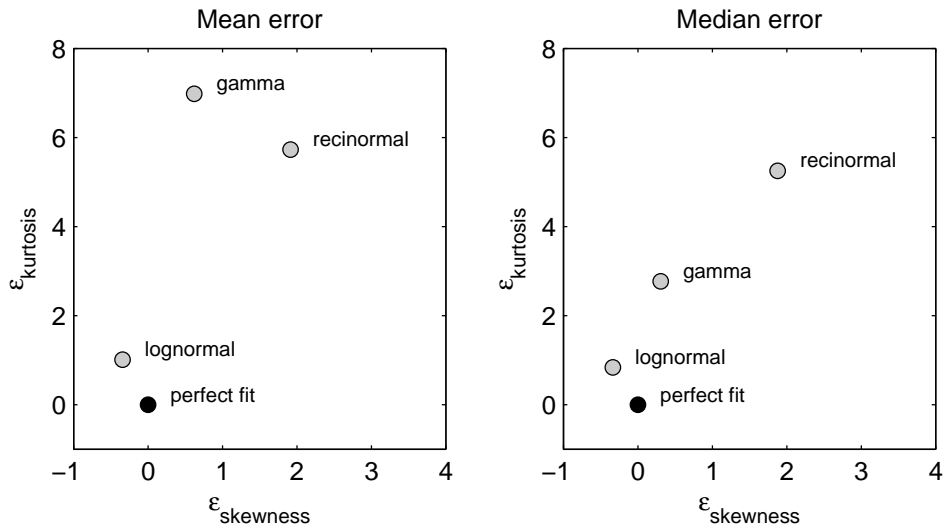

Figure 5: Error plots of the skewness and kurtosis for the gamma, recinormal and lognormal distributions.

Discussion

In the previous section we evaluated two models of fixation duration by retrospectively reanalyzing distributions of fixation durations that were obtained from the fixation data collected by Vlaskamp et al. (2005). By analyzing the skewness and kurtosis of these distributions, we showed that they are best described by a lognormal distribution, rather than the exponential distribution, proposed by Harris et al. (1988), and the recinormal distribution, proposed by Reddi and Carpenter (2000). Futhermore, we showed that there is a linear relationship between the mean fixation duration and the fixation duration standard deviation that has a slope of 0.51 and an R^2 of 0.738 ($p < 0.01$).

What does lognormality of distributions of fixation durations imply for the processes that take place during a fixation? Ulrich and Miller (1993) described a number of simulation studies in reaction time (RT) modeling, showing that lognormal distributions may be the product of several conceptually different mechanisms of information processing. Three possible causes of lognormality were found: exponentially transformed normal random variables, a product of independent random variables, or both. Although there is no direct proof for this, when applied to durations of fixations, one possible interpretation might be that the duration of a fixation is governed by a number of processes, the duration of each may be normally distributed. Ulrich and Miller (1993) also briefly touched on the possibility that a lognormal could be the result of a convolution of a finite number of gamma distributions. They stated that "a serial model could provide a lognormal distribution of RTs if the individual stages had an appropriate combination of gamma distributions" (p. 522-523). However, they conclude that it would be unlikely that an appropriate combination of gamma distributions would arise by chance.

Experiment 2

In experiment 1, we showed that the distributions of fixation durations are positively skewed with a linear relationship between the mean fixation duration and the standard deviation of fixation duration. Furthermore, we concluded that the shape of the distributions is not accurately be described by the model of Harris et al. (1988) or the LATER model (Reddi & Carpenter, 2000). The lognormal distribution showed the best fit to our data, but the choice for the lognormal was data-driven, rather than based on putative mechanisms. Clearly, this result is unsatisfactory and additional research is required.

Positively skewed distributions are also consistently reported for reaction times (e.g. Wagenmakers & Brown, 2007; Ratcliff & Smith, 2004) and it has been

suggested that the lognormal distribution provides a good fit (Ulrich & Miller, 1993). One model that has been especially successful in modeling the skewed shape of distributions of reaction times is the diffusion model (Ratcliff & Smith, 2004). The diffusion model is a sequential sampling model and has been applied to a wide range of two-choice tasks. Like the LATER model, it assumes that evidence is accumulated gradually over time. As evidence is accumulated, the process wanders about until it eventually reaches a threshold, triggering a response. For the remainder of this paper we will refer to this process as the 'visual acquisition', which is thought to represent two processes: the sampling of the retinal image and the processing of visual information. The speed at which visual acquisition occurs is the visual aquisition rate. In contrast to the LATER model, the evidence accumulation process of the diffusion model is stochastic: there are random moment-to-moment fluctuations (noise) in the evidence. This randomness in fluctuations explains the variability in RTs (e.g. different response times are observed when an identical stimulus is presented).

The diffusion model also predicts that when task difficulty increases, RT mean (μ) and RT standard deviation (σ) increase at the same rate (Wagenmakers, Grasman, & Molenaar, 2005). Wagenmakers and Brown (2007) showed that the relation between RT standard deviation and RT mean is close to linear in a large number of experiments, independent of the underlying descriptive distribution that was assumed. They also showed that the mean of the underlying normal distribution increases with the difficulty of the task. Assuming a constant standard deviation of the underlying normal (S), this predicts a linear increase of σ with μ, where the slope is proportional to the value of S. The slope of this relationship is the single free parameter that gives an indication of how the variability is dependent on the mean, and hence task difficulty.

To investigate whether the variability in visual acquisition rates can explain the variability in fixation durations, we developed an experiment in which we mimic a visual search paradigm. While in a typical visual search paradigm an observer is allowed to freely scan a display, in this experiment we cued stimulus locations and manipulated the maximum duration of visual acquisition. In short, we engaged observers in a two-alternative forced-choice discrimination task and measured psychometric functions between duration of visual acquisition and correct discriminations. We then calculated the mean and standard deviation of these psychometric functions and asserted their relationship.

To compare the variability of fixation duration to the variability of visual acquisition rates, we used the linear relationship we found between the mean fixation duration and the standard deviation of fixation duration in experiment 1. If variability in fixation durations is due to variability in visual acquisition rates, we expect equal slopes for both relationships of μ/σ. That is, a slope of

0.51 was expected. If variability of the visual acquisition rates only mildly influences the variability of fixation durations, we expect the slope to be smaller.

Methods

Observers

Eleven observers (eight male, three female) participated in the study. All observers had normal or corrected-to-normal vision. Observers were allowed to wear glasses or contact lenses, but we carefully checked that the frames of the glasses did not affect the accuracy of our tracking device. All gave their informed consent to participate in the experiment.

Apparatus

Stimuli were generated in MATLAB (the Mathworks) using the Cogent Graphics toolbox (see http://www.vislab.ucl.ac.uk/Cogent/) and presented on a Dell Optiplex GX620 (3GHz Intel processor; 1 GB RAM; 256mb ATI Radeon X600 videocard) with a 22-inch Viewsonic FB227B CRT monitor (refresh rate 100Hz; resolution 1280 by 1024 pixels). Observers sat in a dimly lit room with their head movements constrained by a bite board. Viewing was binocular at a distance of 80 cm. Two-dimensional orientation of the left eye was measured using a head-mounted Eyelink II system (SR-research) at a sampling rate of 500Hz.
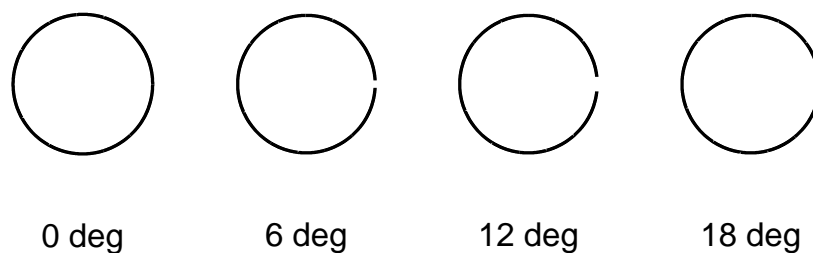

Figure 6: The three conditions of stimulus discriminability used in the experiment, defined as C's with gap size 6 (hard), 12 (medium) and 18 deg.

Stimuli

The stimuli used for the experiment were either a Landolt C (which we refer to as a C) or a closed circle (which we refer to as an O). Both the O's and C's had a diameter of 3 deg visual angle and had a line width of 1 screen pixel. Orientation of each C was randomly chosen from the directions: 0, 90, 180 and 270 deg. Stimulus location was randomly chosen from one of four corners of an invisible virtual square that was centered on the screen and measured approximately 9 deg by 9 deg visual angle. A target, however, would never appear at the same location in any two consecutive trials, therefore forcing the observer to make a saccade to the new stimulus location in each trial. All stimuli (and the Eyelink screens for calibration and drift correction) were displayed in black (rgb: 0,0,0) on a grey (rgb: 120, 120, 120) background. We defined task difficulty in terms of gap size of the C's; the smaller the gap, the more difficult the task. We defined three difficulty levels: 18 deg for easy, 12 deg for medium and 6 deg for hard (see figure 6). The masks we used were composed of multiple superimposed semi-circular patterns. Five different patterns were generated and for each trial one of them was chosen randomly.

Design

In our experiment, observers engaged in a single-interval two-alternative forced choice (2AFC) discrimination task by reporting whether they had seen a C or an O. The experiment consisted of three experimental sessions, each of which measured one condition of stimulus discriminability. Each session consisted of two consecutive stages. In the first stage, psychometric threshold was estimated using an adaptive staircase procedure. The reasons for the implementation of a staircase procedure were twofold: First, in advance of testing, the threshold is usually unknown and a lot of data would have to be collected at points on the psychometric function that provide little information about its shape (e.g. at the tails). Adaptive staircases can be used to cluster the points that are sampled around the threshold. This can prove especially valuable in experiments where a large between-subject variability as a function of task difficulty is anticipated, as is the case in the current experiment. Second, an adaptive staircase would allow the observer to start at an easy pace to get used to the task.

The particular type of staircase used in this experiment was the 'transformed up-down' procedure (Levitt, 1971). This procedure does not require strong assumptions regarding the underlying psychometric function (its shape and the type of distribution) and is robust to changes in sensitivity during testing. We chose to use a 1-up/3-down staircase, in which the stimulus intensity increases by one after the observer makes one error and decreases by one after the ob-

server gives three consecutive correct answers. This method approximately targets the 79% criterion level (Levitt, 1971). The termination rule for the staircase was either reaching a maximum of 300 trials, or having reached 40 reversals—a reversal being a change of direction in which the stimuli are chosen. Threshold was then estimated by calculating the average of the stimulus durations at the reveral points. The first 20 reversals were not included to ensure that only reversals around the threshold level were included in the estimation.

A disadvantage of an adaptive staircase, however, is the lack of information about the exact shape of the psychometric function, which is not as accurate as when a method of constant stimuli is used. Because we wanted to obtain an accurate estimate of the threshold and slope of the psychometric function, our staircase procedure was followed by a method of constant stimuli. Instead of being presented in a descending or ascending order, this method presents stimulus intensities randomly, drawn from a sampling plan. This sampling plan determines which, and how many, points are measured of the psychometric curve. See appendix B for a motivation of the sampling plans we used for this experiment. The offset of the sampling plan (the position on the abscissa) was derived from the threshold value obtained from the staircase procedure. For both the staircase and the method of constant stimuli, trials were identical (with the exception that stimulus durations in the staircase were dependent of responses in the previous trials).

Each observer performed a total of 2700 experimental trials (150 trials per 6 sampling points and 3 stimulus conditions) spread over three sessions. The duration of each session was about 45 minutes. Each of the 3 conditions was presented in a different random order for each observer. In each trial for both the staircase procedure and the constant stimuli, the possibility of presenting a C or an O was 50%. The order of C's and O's for both the adaptive staircase and the method of constant stimuli was randomized.

Procedure

In total, there were 4 sessions, separated by one or more days. At the start of each session, the Eyelink II system was placed on the head of the observer and calibrated using the Eyelink innate nine-point calibration routine. Prior to the first session, observers received instructions and their visual acuity was measured. They then participated in a training session to get acquainted with the task. During this session, observers engaged in a short staircase procedure, which consisted of 200 trials and tested for an easy gap size of 24 deg. The staircase 'path' was visually inspected by the experimenter and when an observer did not converge to a threshold value, they were asked to take part in another training session.
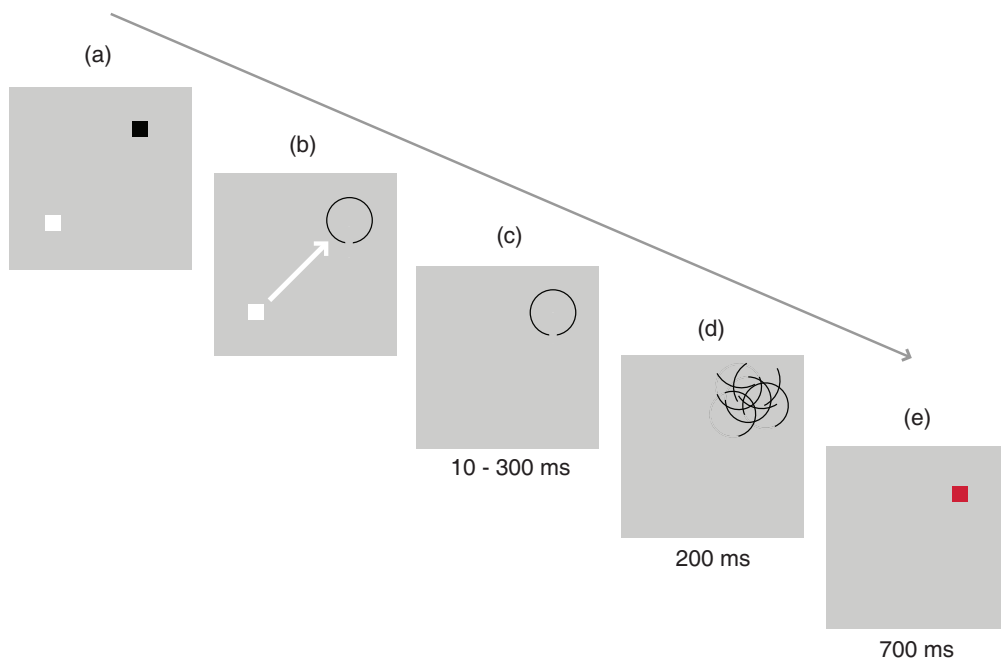
Figure 7: A schematic representation of a single trial as used in experiment 2. (a) A trial started at the stimulus location of the last trial (white square). First, a cue (black square) appeared to guide a saccade to the next stimulus location. (b) When a saccade (shown as the white arrow) was detected, this cue was replaced by a C or an O. (c) The observer was allowed to fixate the stimulus for a very brief period (10 to 300 ms). (d) The stimulus was followed by the mask, which was shown for 200 ms. (e) The mask was replaced by a red fixation point. The red fixation square was shown for 700 ms, during which the observer was allowed to respond.

Each trial started with the presentation of a peripheral cue (a small white square) that could appear in one of the four possible locations. The purpose of the cue was to guide the saccade to the next stimulus location. Subjects made a saccade to this cue and during this saccade, the cue was replaced by the stimulus, which would be either a C or an O. After the observer had fixated the stimulus for a brief period (ranging from 10 - 300 ms), it was removed and immediately followed by a mask to prevent after images. The mask was shown for 200 ms and replaced by a red fixation point, which was shown for 700 ms. The purpose of this fixation point was twofold. First, it was meant to hold the gaze of the observer on this location, so that the saccadic eye movement in the next trial would start from this location. Second, while the fixation point was visible, the observer could respond by pressing the space bar if a C was seen. No response was required if an O was seen. A trial always ended 700ms after the mask had been removed.

Data analysis

Response data from the experiment was saved for offline processing and later analyzed using MATLAB (The Mathworks). Psychometric functions were fit through our response data (on the abscissa measuring display duration, and on the ordinate the fraction of good answers) using the psignifit toolbox version 2.5.6 for MATLAB (see http://bootstrap-software.org/psignifit/) which implements the maximum-likelihood method described by Wichmann and Hill (2001a). A frequently used general form of the psychometric function is:

$$\psi(x; \alpha, \beta, \gamma, \lambda) = \gamma + (1 - \gamma - \lambda)F(x; \alpha, \beta) \tag{4}$$

Parameter $\gamma = \Psi(0)$ is the lower asymptote, which represents the guessing rate (or chance level). Since our experiment is designed as a 2AFC experiment, we assume γ to be fixed at .5. The upper asymptote is given by $1 - \lambda$, where λ represents the observer lapse rate (responding incorrectly regardless of the stimulus intensity). The shape of the curve is determined by its parameters $(\alpha, \beta, \gamma, \lambda)$ and the choice of a two-parameter function F, which is typically a sigmoid function. Parameters α and β represent its displacement on the abscissa and its slope, respectively. In our analysis we used a cumulative normal distribution to model the psychometric function.

Results

A typical example of data obtained from our experiment for a single observer and one condition of gap size (in this case 12 deg) is shown in figure 8. The top panel shows the results from the staircase procedure and the resulting estimated threshold value. All observers showed a similar initial decrease of stimulus intensity for each condition of gap size and most reached their threshold values within 60 to 80 trials. The threshold for this session was estimated to be equal to 50 ms. The bottom panel shows the results of the method of constant stimuli for the same observer. Here, each point the fraction of correct responses is plotted as a function of presentation time. Each data point represents a presentation duration that was measured and the sigmoid represents the psychometric function that was fit through these points. As can be seen from figure 8, all sampling points follow the psychometric function quite well and especially the threshold and the slope (the parameters we were interested in) could be estimated reliably.

The threshold value was defined as the stimulus intensity at the 75% criterion level (represented by the horizontal dotted line), which is equal to the stimulus intensity in the inflection point of the psychometric function. For the result shown, the actual threshold was equal to 43.31 ms.
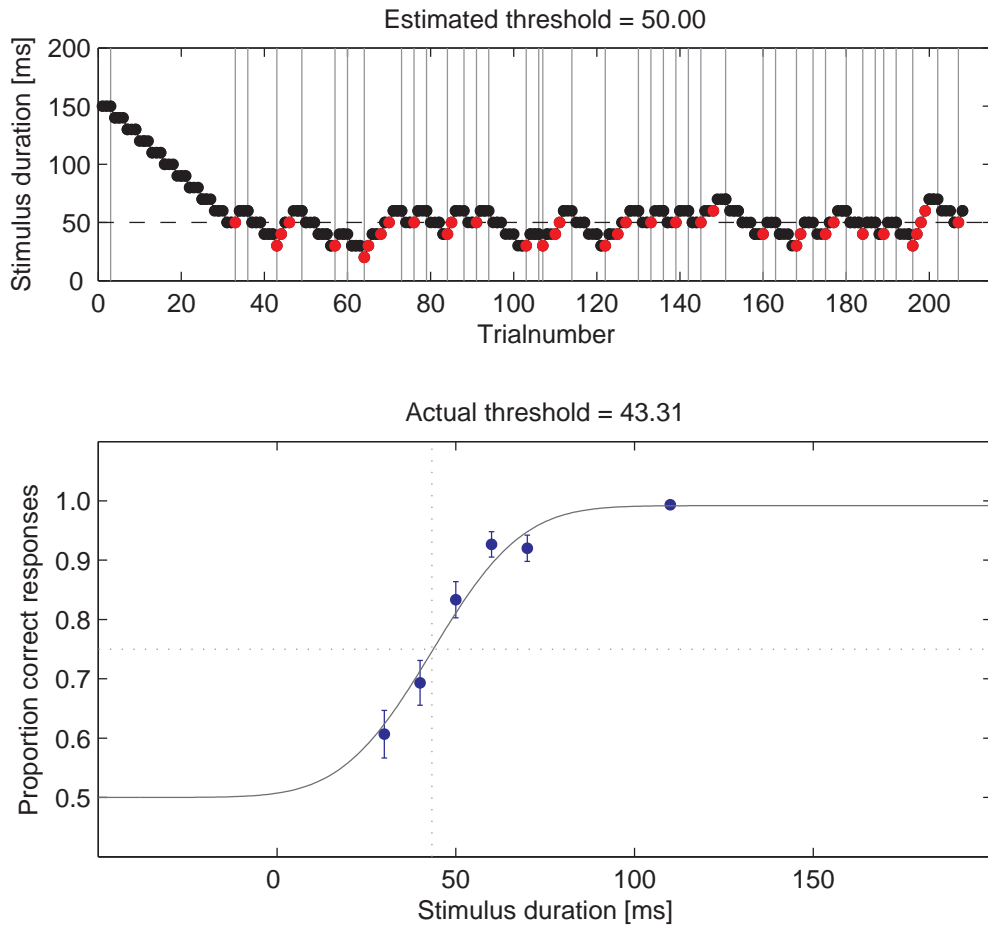
16

Figure 8: Representative results of one session for a single observer. (a) The results of the adaptive staircase procedure. Black data points represent trials in which a correct response was given, red data points represent trials in which an incorrect response was given. Stimulus duration decreases after three consecutive correct answers and decrease after each incorrect answer. Each reversal is marked by a vertical grey line (b) The results from the method of constant stimuli. It shows the fraction of correct responses for each stimulus duration (with the respective standard deviations shown as error bars) and the psychometric function that was fit.
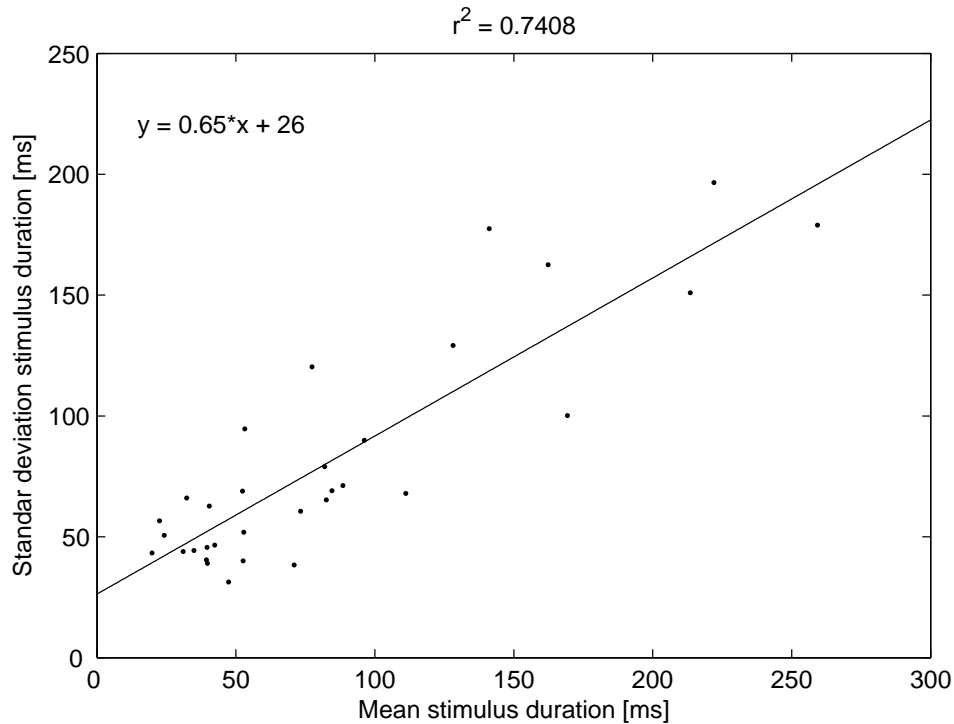
Figure 9: The mean stimulus duration as a function of stimulus duration standard deviation for each observer and condition of gap size. The linear relationship shown here follows the formula $y = 0.65x + 26$. The observed R^2 is equal to 0.7408.

Because the psychometric function used for our experiment was a cumulative normal distribution, the mean of its underlying non-cumulative distribution is equal to the threshold value, whereas its standard deviation is equal to the β-parameter of the psychometric function (see formula 4), and is proportional to the slope at threshold. Figure 9 shows the mean as a function of the standard deviation for each distribution. It shows a clear linear relationship, that follows the formula $y = 0.65x + 26$ and has an R^2 of 0.878 ($p < 0.01$).

Discussion

The main objective of this experiment was to examine the relationship between the mean fixation duration and the standard deviation of fixation durations. We assumed that the variability in fixation durations was caused by variability in visual acquisition rates. As a consequence, we predicted that we would find a linear relationship between the mean and standard deviation of the psychometric functions that were fit in the current experiment, with a slope of 0.51 (as was shown in experiment 1).
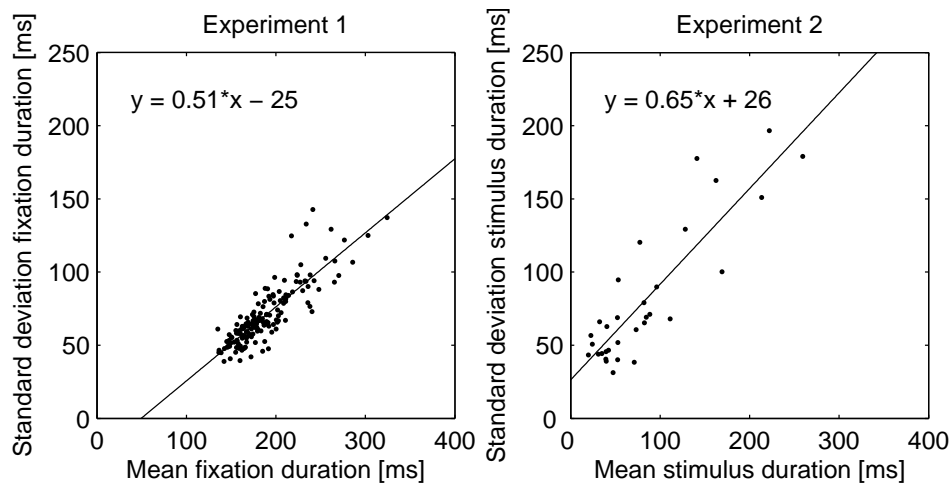
Figure 10: A comparison of the results of experiments 1 and 2. The left panel shows the linear relationship between the mean fixation duration and the fixation duration standard deviation, found in experiment 1. The right panel shows the linear relationship between the mean required stimulus duration and the standard deviation of the required stimulus duration, found in experiment 2.

The results of our experiment show that there is a clear linear relationship between the mean stimulus duration and the standard deviation of stimulus duration. If we compare this relationship found in our experiment to the relationship found for the mean fixation duration and the fixation duration standard deviation found in experiment 1, it is clear that the results are very similar (see figure 10). First, there is a clear linear relationship with a slope that is approximately equal (0.51 for the data of Vlaskamp et al. and 0.65 for our experiment). Furthermore, the R^2 of both relationships is almost identical. From these two results we conclude that variability of the fixation durations can be explained by the variability of visual acquisition times.

There is however an offset between the two linear relationships of approximately 100 ms (see figure 10). This offset can be explained in terms of saccade preparation. In our experiment, fixation locations were cued whereas in the experiment of Vlaskamp et al. (2005) the observers actively searched through the display and therefore planned a saccade to the next fixation location.

General discussion

In this paper, we investigated distributions of fixation duration and variability in fixation durations. In experiment 1, we evaluated two models of fixation duration by retrospectively reanalyzing a fixation data set collected by (Vlaskamp et al., 2005). Our results show that the lognormal distribution provided the best fit for these distributions. Furthermore, we showed that there is a linear relationship between the mean fixation duration and the standard deviation of fixation duration, that has a slope of 0.51 and an R^2 of 0.738 ($p < 0.01$). In experiment 2, we showed that variability in fixation durations can be explained by the variability in visual acquisition rates.

Our results suggest that current models of fixation duration, like the exponential model proposed by Harris et al. (1988) and the LATER model proposed by Reddi and Carpenter (2000), are unsuitable to model distributions of fixation durations and consequently the variability in fixation durations. Here, we propose that a model like the diffusion model could be a suitable alternative to model fixation durations: for reaction times, the model gives a good account of the skewed distributions (Ratcliff & Smith, 2004) and the linear relationship between the mean and standard deviation (Wagenmakers et al., 2005). To our knowledge, however, the model has only been successfully applied to two-choice tasks in which reaction times were the dependent variable and further research is needed to investigate how it can be applied to fixation duration.

References

Buswell, G. (1935). *How people look at pictures*. Chicago: Univ of Chicago Press.

Carpenter, R. H., & McDonald, S. A. (2007). LATER predicts saccade latency distributions in reading. *Exp Brain Res*, *177*(2), 176–183.

Carpenter, R. H., & Williams, M. L. (1995). Neural computation of log likelihood in control of saccadic eye movements. *Nature*, *377*(6544), 59–62.

Harris, C. M., Hainline, L., Abramov, I., Lemerise, E., & Camenzuli, C. (1988). The distribution of fixation durations in infants and naive adults. *Vision Res*, *28*(3), 419–432.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*, *49*(2), Suppl 2:467+.

McConkie, G., & Dyre, B. (2000). Eye fixation durations in reading: Models of frequency distributions. In D. H. . J. P. A. Kennedy R. Radach (Ed.), *Reading as a perceptual process.* Oxford, England: Elsevier.

Ratcliff, R. (2001). Putting noise into neurophysiological models of simple decision making. *Nat Neurosci*, *4*(4), 336–337. (Comment)

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychol Rev*, *111*(2), 333–367.

Reddi, B. A., Asrress, K. N., & Carpenter, R. H. (2003). Accuracy, information, and response time in a saccadic decision task. *J Neurophysiol*, *90*(5), 3538–3546.

Reddi, B. A., & Carpenter, R. H. (2000). The influence of urgency on decision time. *Nat Neurosci*, *3*(8), 827–830.

Schall, J., & Hanes, D. (1998). Neural mechanisms of selection and control of visually guided eye movements. *Neural Netw*, *11*(7-8), 1241–1251.

Suppes, P. (1990). Eye-movement models for arithmetic and reading performance. *Rev Oculomot Res*, *4*, 455–477.

Ulrich, R., & Miller, J. (1993). Information processing models generating lognormally distributed reaction times. *Journal of Mathematical Psychology*, *37*, 513–525.

Viviani, P. (1990). Eye movements in visual search: cognitive, perceptual and motor control aspects. *Rev Oculomot Res*, *4*, 353–393.

Vlaskamp, B. N. S., Over, E. A. B., & Hooge, I. T. C. (2005). Saccadic search performance: the effect of element spacing. *Exp Brain Res*, *167*(2), 246–259.

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychol Rev*, *114*(3), 830–841.

Wagenmakers, E.-J., Grasman, R., & Molenaar, P. (2005). On the relation between the mean and the variance of a diffusion model response time distribution. *Journal of Mathematical Psychology*, *49*, 195–204.

Appendix A   Equations

Skewness and kurtosis of a gamma distribution

From the coefficient of variance (*CV*) of a gamma distribution, one can deduce the equations for skewness (*S*) and kurtosis (*K*) as shown here. The equations for the mean and the standard deviation of a gamma function are[1]:

$$\mu = \alpha\Theta \tag{A-1}$$

$$\sigma^2 = \alpha\Theta^2 \tag{A-2}$$

where θ is the scale parameter and α is the shape parameter. We can use these formulas to define α in terms of the CV:

$$CV = \frac{\sigma}{\mu} = \frac{\sqrt{\alpha}\Theta}{\alpha\Theta} = \frac{1}{\sqrt{\alpha}} \tag{A-3}$$

$$\alpha = \frac{1}{CV^2} \tag{A-4}$$

The equations for Skewness and Kurtosis of a gamma distribution are:

$$S = \frac{2}{\sqrt{\alpha}} \tag{A-5}$$

$$K = \frac{6}{\alpha} \tag{A-6}$$

Therefore, if we substitute α with equation (A-4) we get:

$$S = \frac{2}{\sqrt{\frac{1}{CV^2}}} = 2CV \tag{A-7}$$

$$K = \frac{6}{\frac{1}{CV^2}} = 6CV^2 \tag{A-8}$$

---

[1]Weisstein, Eric W. "Gamma Distribution." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/GammaDistribution.html

Appendix B   Sampling plans

One of the questions we wanted to address before starting our experiments was which sampling plan we should use for the method of constant stimuli to ensure optimal deployment for our trials. In other words; should we capitalize on the amount of sampling points or should we capitalize on the amount of trials per sampling point? Generally, the best would be to capitalize on both, but if a maximum amount of trials per session is set, there is a trade-off between the amount of trials per sampling point and the amount of sampling points measured: more sampling points result in less trials per sampling point and more trials per sampling point results in less sampling points measured. For our sessions, we set a maximum of 900 trials, resulting in sessions of approximately 45 minutes.

To test which was the optimal sampling plan for our experiment, we conducted several Monte Carlo simulations; six-point, eight-point and ten-point data sets were generated binomially assuming a 2AFC design. We used a cumulative log-normal as the underlying performance function, with parameters $\sigma = 1.5$ and $\mu = 90$. The lower asymptote γ (guessing rate) was fixed at the 50% performance level, the upper asymptote λ was fixed at 100% performance level.

For each data set, seven sampling schemes were simulated, each dictating a different distribution of sampling points along the stimulus intensity axis. One thousand iterations were run per condition. A condition was defined as a unique combination of sampling scheme, number of trials (ranging from 10 to 400 with increments of 10) and number of sampling points. A maximum likelihood fit was performed on each data set to obtain parameters α (mean) and β (standard deviation) for each PF. For each of the conditions, the simulated values for α and β were calculated, yielding 1000 values of α and β for each condition. Then, the standard deviation (S) for α of each condition was calculated. The optimal sampling plan was defined as that those with the lowest S for α, or, in other words, where there was the least variation in the simulated means.

Initially, we tested several sampling schemes, including those skewed to the low or high performance values, those that were clustered around the threshold, those that were spread out away from the threshold and combinations of the aforementioned, i.e. with some values clustered around the threshold and some values at the tails. After running these simulations, we selected the two best performing schemes from each data set and compared them (see figure 11a). The results of our simulations were to some extent similar to the results of simulations conducted by Wichmann & Hill (2001a); For each of the data sets a combination of sample points clustered around the threshold combined with one or more sample points above 80% correct yielded the best results.
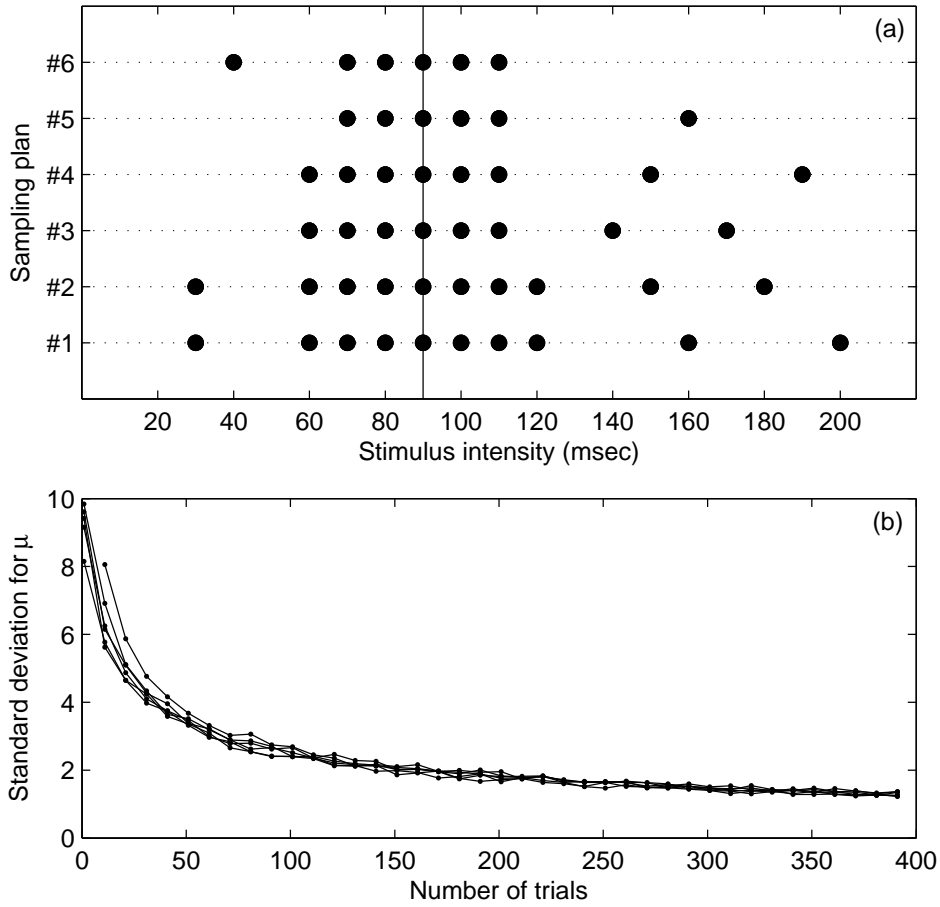
Figure 11: Sampling plan simulation results. (a) The two best performing sampling plans from each of the data sets used; the 10-point set (1 and 2), 8-point set (3 and 4) and the 6-point set (5 and 6). (b) simulation results for the sampling plans defined in panel (a)

The results (see figure 11b) from the selected best-performance six-point, the eight-point and ten-point sampling schemes are very similar. For each sampling scheme there is a clear near-exponential downward trend: the more trials per sampling point, the smaller the §. From this simulation study we concluded that, bearing in mind the synergy between the amount of trials per sampling point and amount of sampling points, a sampling plan with six sample points and 150 trials (a total of 900 trials) would be optimal. An example will illustrate this choice: If we were to measure 1200 trials per session using a ten-point sampling plan, we can measure 120 trials per sampling point. If we were to use a six-point sampling scheme, we can measure 200 trials per sampling point. Since the S of α decreases with increasing number of trials, it can be concluded that the method with less sampling points, but more samples is superior over

the methods with more sampling points, but less sampling points. For our experiments, we therefore used sampling plan number 5: a 6-point sampling plan with one point above 80% performance and 150 trials per sampling point.