

Gaze Control in Dynamic Broadband (1/f) Noise Sequences

Christoph Rasche, Karl R. Gegenfurtner

Abteilung Allgemeine Psychologie, Justus-Liebig-Universität
Otto-Behaghel-Str. 10F, 35394 Giessen, Germany
Email: rasche15@gmail.com, gegenfurtner@uni-giessen.de

Abstract

Research on gaze control has been typically characterized for stationary targets on homogenous backgrounds. In this study, gaze behavior is investigated with a dynamic broadband (1/f) noise display, which offers much more the kind of natural background distraction as when moving through natural scenes. To elucidate the possible bottom-up component of saccadic target selection, we analyzed the fixation locations during free viewing and find that looming dark spots are preferred and the ROC area value for a target/random patch distinction was 0.57. For a luminance-target search, this bottom-up component was overwritten but the target/random patch distinction was not larger (ROC area value also 0.57). Exploiting the principle of the classification image, it can be shown that saccadic decision time may not be fixed but rather depend on target properties. Saccadic orienting properties are in qualitative agreement with measurements obtained from a stationary, homogeneous background, but saccadic constant and variable error increased. A cued search reveals that despite the presence of the distracting background, attentional shifts take place to obtain identification judgments about targets appearing in the parafovea.

Introduction

Gaze control has been predominantly investigated on a homogeneous background with well defined targets (e.g. Findlay and Gilchrist, 2003; Roos et al 2008). Such highly controlled conditions are suitable to accurately characterize parameters and dependencies such as saccadic latency toward a target, the dependence of latency on target parameters, target detectability in the periphery and so on. How those parameters and dependencies change when more complex visual input is presented has only been marginally investigated. In analogy, Roos et al. make even a distinction between *evoked* saccades, the ones triggered in laboratory conditions, and *spontaneous* saccades, the ones triggered when freely viewing real images. Others have approached the investigation of spontaneous saccades using broadband noise to mimic the background structure of natural images, which show a decline in the frequency amplitude spectrum (Field, 1987;

Simoncelli, Olshausen, 2001). For instance, Geisler et al (2006) have investigated the search behavior for Gabor targets placed in a stationary, 1/f-noise background using gaze-contingent methodology. Rajashekar characterized the influence of structural cues in visual search in such 1/f displays (Rajashekar et al., 2006). White et al. (2008) found that saccadic latencies toward Gabor patches are shorter in 1/f displays than on a homogenous background lacking any structure, when target visibility is matched between background conditions. Here, we extend the broadband-noise analysis to dynamic displays, specifically a flickering bar-code movie whose frames are taken row-wise from a 1/f-noise image (see figure 1 for one frame). The noise may be circumscribed as consisting of a fast fizzling noise placed on a slower, rolling noise, caused by the high and low frequencies respectively. The choice of spatial one-dimensionality is done for reason of technical simplicity but is justifiable by the fact that a large number of saccades are made along the cardinal axes (Einhäuser et al 2007).

Because a 1/f noise display lacks any specific contour structure, it offers the possibility to determine whether the bare 1/f background contains features attracting gaze, which could be part of the bottom-up component of saccadic target selection in natural scenes. This is pursued in a **first** experiment, in which subjects freely viewed the noise movies without any specific instruction. We analyzed the movie snippets of those locations where saccades landed, in a manner equivalent to the analysis of fixation and non-fixation patches in scene studies (e.g. Reinagel and Zador, 1999, Tatler et al 2006). In those studies, the image patches at the fixation locations were compared to randomly chosen patches to determine the bottom-up component of saccadic target selection. The distinctness between the fixated and random patches is determined using classifiers. In our study, the patches correspond to space-time patches due to the use of a dynamic display: for each saccadic onset, the preceding space-time sequence at the spatial location of saccadic landing is extracted (a subimage of the 1/f stimulus image). This space-time patch presumably has triggered the visual system to place its focus on it: it is therefore called a *trigger* patch. The set of trigger patches is compared to a set of *random* patches, which were chosen randomly from the movie sequences. A trigger/random patch classification proves that the patches differ, meaning that there exists a bottom-up feature in the dynamic 1/f noise display.

Is this bottom-up feature still being present during a target search? To investigate this, we carried out a **second** set of experiments, during which subjects search the movie display for a temporarily appearing, just-noticeable target (see figure 1), which appears only occasionally. It turns out that the target search overwrites the bottom-up component, which one could have expected, but it surprisingly does *not* lead to a better trigger/random patch distinction. Are the detection dynamics for those targets different from the detection dynamics of the bottom-up features? We analyze this using the

method of classification images (Ahumada, 2002; Shimozaki, 2007). A classification image is the average of those patches, which had been chosen erroneously (miss). Such an image therefore reveals how the actual target stimulus is represented. In our study, the majority of saccades can be regarded as misses and the classification image corresponds to the average of the space-time trigger patches. Its temporal cross-section can give hints about the saccadic decision dynamics. We find that the dynamics are different between the free-viewing targets and those luminance search targets. But for a systematic manipulation of the target parameters, we measure target dynamics which we have not completely understood. The target search also allows us to determine saccadic orienting parameters such as saccadic latency, undershoot, constant error and variable error, which we can compare to laboratory experiments. The amount of undershoot and variable error is larger than in typical laboratory experiments.

In a **third** set of experiments, we increased the complexity of visual search by introducing an identification task, specifically the detection and identification of letters embedded in the noise movie. The letters appear only transiently and are therefore difficult to detect and to identify, requiring thus full attention and possibly foveation. A comparable real-world scenario would be the detection and recognition of road signs while driving in dense fog. We would expect that this difficult search task can be facilitated by cueing.

Methods

Subjects. Male and female students (age 23-30) served as subjects. All subjects had normal or corrected to normal vision. All subjects were naive with respect to the experiment.

Equipment. Subjects were seated in a dimly lit room facing a 21-inch CRT monitor (ELO Touchsystems, Fremont, CA, USA) driven by an ASUS V8170 (Geforce 4MX 440) graphics board with a refresh rate of 100 Hz non-interlaced. At a viewing distance of 47 cm, the active screen area subtended 45 degrees of visual angle in the horizontal direction, and 36 deg vertical on the subject's retina. With a spatial resolution of 1280 x 1024 pixels this results in 28 pixels/deg. The subject's head was stabilized in place using a chin rest. Eye position signals were recorded with a head-mounted, video-based eye tracker (EyeLink II; SR Research Ltd., Osgoode, Ontario, Canada) and were sampled at 250 Hz. Subjects viewed the display binocularly through natural pupils. Stimulus display and data collection were controlled by a PC.

Noise movies. The gray-scale values for the bar-code are taken line-wise from a 2D noise image $I[x,t]$ of size 1200*1000 pixels presented as 8-bit (see figure 1). Such a noise

image was generated with normally distributed random values, whose frequency spectrum was then altered to describe a 1/frequency decline. A single bar code (= one column = one frame) was stretched vertically to a height of 100 pixels (100*1200 pixels) and placed into a gray background (40 cd/m² luminance). A frame was shown for 10ms, a movie (=one trial) thus lasted 10 seconds. Each movie $I(x,t)$ was different to avoid potential learning effects.

Search Targets. The target stimulus consisted of a finite-pulse function added temporarily to the noise stimulus (see figure 1), a luminance increment by a value of $a=0.2$ (for a given range of gray-scale values from 0 to 1) with a duration of $d=300$ ms, presented randomly every 3 seconds (0.333Hz). This constituted the *fixed-amplitude* target condition in which distal targets are less well detected due to the peripheral decline in visual acuity (Figure 4, dotted curves). In a *scaled-amplitude* target condition, the target amplitude a_{trg} was set in proportion to eccentricity e by the following equation: $a_{trg} = a_{min} + a_{max} \cdot \exp(-e) \cdot a_{max}$, whereby a_{min} is a minimal amplitude and a_{max} is the maximal amplitude (the function starts at a_{min} and saturates at $a_{min} + a_{max}$).

Procedure. Subjects performed blocks with 50 trials, generally 3 blocks per day, 6 blocks per experiment. Each block was preceded a calibration. Before each trial, subjects fixated a central spot and a drift correction was carried out to minimize errors of headband slippage or other factors. Average, spatial measurement accuracy was 0.28 ± 0.55 degrees, which is as accurate as other studies paying attention to accuracy (e.g. 0.4 ± 0.1 in Tatler et al 2005).

Analysis of target search: Subjects were instructed to move their focus toward the targets and press a button when seeing a conspicuity. Target detection is defined as the temporal coincidence of a ‘saccadic hit’ and a button press. A saccadic hit required a saccadic flight toward the target and a landing within 5 degrees of target eccentricity. The temporal tolerances for saccadic latency and button response were 400 and 1200ms respectively. Given this condition, chance level for an accidental saccadic landing near a target is very small but can not be calculated. Each search condition was carried out by 4 to 5 persons. A subject typically performed 3 to 6 blocks of one experiment during which the target was presented 850-900 times (ca. 3 target presentations per trial).

Analysis of saccade trigger-patches: For each saccadic landing position, a neighborhood of its preceding spatio-temporal movie sequence (patch or *trigger patch*) was extracted. A single patch is denoted as $P_i[x,t] (\subset I[x,t])$, with the center x value corresponding to the (spatial) saccadic landing position.

The classification image is the average of trigger patches (Ahumada, 2002; Shimozaki, 2007):

$$C(x, t) = 1 / N \sum_{i=1}^N P_i (x, t), \quad [1]$$

with N the number of patches per condition. Trigger patches were sorted according to their saccadic arrival direction (left/right); left ones were mirrored along the spatial dimension. An equal number of *random* patches were collected for statistical comparison, which were drawn (spatially and temporally) from random locations. An average of 2990 patches was collected per person. Patches toward a target were omitted.

Trigger/random patch classification: A classification between trigger versus random patches was carried out using a support-vector machine (Noble 2006). We thereby followed the exact same implementation as in Kienzle et al (2006) [see Chang and Liu 2001 for software]. Trigger patches of varying width (1 to 8 degrees) were selected. The use of a large number of patches is necessary to obtain reliable classification estimates, but also requires down sampling of the patches due to memory limitations of the classifier methodology. The spatial dimension was subsampled to 14 pixels and only 10 time frames were selected (100ms before saccadic onset), thus using in total a 140-dimensional vector.

Letter task. During the noise movie, letters of an approximate size of 1x1 degree (taken from a 64 x 64 bitmap) appear transiently with low contrast (see figure 3 for a schematic). They are added to the luminance profile like the above luminance targets in order to make them appear very faint: the letters are hard to detect and identify. The viewer's goal is to identify as many letters as possible. The viewer is supported by cues which appear at the spatial location of letters just before the letters appear. Cues have always 100% validity but appear with varying frequency. Thus, the viewer ideally follows two kinds of signals, the randomly appearing letters and also the occasional cueing markers which point out the locations of upcoming letters and facilitate detection. Although the experiment may appear very simple, the marker's appearance has a number of parameters, which may influence the detection and identification rate of the letter task.

Results

Exp1: Free Viewing

Asking subjects to perform free viewing is not as easy as it seems because some subjects immediately ask what they were supposed to look for. To obtain 'unbiased' viewing, observers were simply asked to 'be inspired'. Subjects who participated in multiple experiments, started with the free-viewing task first. A first step was to analyze the

classification image of the saccade trigger patches (Ahumada, 2002; Shimozaki, 2007), i.e., the average of the movie snippets that immediately preceded and followed saccade onsets. The classification image of all subjects showed a depression (trough) in the luminance level hinting that subjects' saccades were directed toward 'dark spots' in the movie. Figure 2 is a representative example for one subject (time axis running from bottom to top). To highlight the area which is significantly different from the random patches, those values were set to the mean grayscale value, which were within the range of values given by the classification image for the random patches. This highlighted area is now called a *salient area*. The salient area's spatial minimum is located approximately 1 degree to the right of the landing point (0 deg eccentricity) reflecting systematic saccadic undershoot; the minimum occurs approximately 130ms before saccadic onset. The width of the salient area extends approximately 5-7 degrees at the time of the minimum; its duration lasts more than a second before saccadic onset. This large size is primarily due to the 1/f correlation along both dimensions (which bears slowly changing low frequencies).

Because we use a dynamic display, there exists the possibility that subjects follow a dynamic motion caused by the slowly rolling background. In fact, the stimulus can be used to trigger smooth pursuit movements if one intentionally follows such a slow background wave (eye velocity traces show typical smooth pursuit behavior). If subjects pursued any such motion stimuli systematically, then this should express itself as a rotation of the salient area's main axis away from the vertical orientation (left/right direction taken into account). No such tilt could be detected in any classification image. To determine the gaze-attracting strength of the bottom-up feature, we employ the trigger/random patch classification. The use of support vector machines yielded ROC area values ranging from 0.55 to 0.61 for different observers (mean=0.57). The analysis was done for patches of different spatial width (1-8 degrees). The ROC area values differed only slightly and the optimal patch width varied from subject to subject. It was 5 degrees in average, although this was not significant.

In summary, the free viewing showed the presence of a bottom-up feature, which is stationary (non-moving) dark spot in the noise movie. This is a non-trivial finding because the search for similar features in Gaussian noise movie did not yield nearly the same classification image even after 6000 fixations. This bottom-up feature can be distinguished from a random background stimulus and its distinctness is characterized with the ROC area value (mean=0.57).

Exp 2: Target Search

The target search was carried out with just-noticeable targets. When a subject performed the search task for the first time, it would take the subject a few trials until s/he had

discovered the conspicuity. Classification images were again generated for the (non-target) trigger patches. They look similar to the one for free-viewing but show a positive salient area (positive luminance elevation from background – not shown for lack of novelty). Put differently, it is inverted as compared to the free-viewing classification image and clearly evidences that the bottom-up feature was completely overwritten by the top-down search.

The classification analysis for trigger and random patches (omitting the few patches with targets) yielded ROC values ranging from 0.54 to 0.63 (mean=0.57), which are approximately equal to those measures for the free-viewing trials (also mean=0.57). This was surprising as we expected that they would be higher given the circumstance that subjects were given a specific search target.

Target detection dynamics:

We now study the temporal cross-section of the classification images in order to gain hints about the target detection dynamics. Figure 3 top graph shows the one for free-viewing and for luminance search, averaged across subjects. For free-viewing (curve inverted for comparison) a long-lasting, gradual integration takes place: subjects apparently follow the slow dynamics of the dark spots. For the target search in contrast, the integration occurs later and shows a higher amplitude: subjects seem to search for rather abrupt onsets as opposed to the free-viewing condition. To relate to a study by Neri and Heeger, we also analyzed the variance of the trigger patches (the variance is taken instead of the mean, see equation 1 in methods section), but no salient areas could be determined.

Orienting properties:

Figure 4 shows the manual reaction times, saccadic latencies and detection rates in dependence of eccentricity averaged across subjects for the fixed-amplitude and scaled-amplitude target condition (dotted and solid respectively). For the scaled-amplitude condition - the gaze-contingent increment in target amplitude - the following parameters were chosen: $a_{min}=0.2$, $a_{max}=0.5$. The manual reaction time for the fixed-amplitude condition increased from ca. 500ms to ca. 550ms for 1 to 15 degrees eccentricity (top graph, black). For the scaled-amplitude condition reaction times remained roughly constant across eccentricity. Saccadic latencies in contrast decreased initially for eccentricities up to about 12 degrees, from ca. 240 down to 200 milliseconds (top graph, blue). For larger eccentricities, latencies show signs of increase. For the scaled-amplitude condition latencies were slightly lower by about 10 to 30 milliseconds.

XXX BW: saccadic latencies: relation to CB paper?

The detection rate for fixed-amplitude targets (lower graph, figure 4, blue dotted) clearly decreased for large eccentricities as expected given the peripheral decline in visual acuity, from a value of ca. 0.5 at 4 degrees to less than 0.3 for 20 degrees and more. The low detection rate for small saccadic amplitudes (1-3 degrees) is due to the difficulty to determine a saccadic jump toward the target given the noisiness of the background and probably due to the subject's low propensity to saccade toward proximal targets. The detection rate for scaled-amplitude targets increases from 4 to 8 degrees eccentricity to saturate at a value of 0.6. It then slowly declines for large eccentricities. There was also a substantial amount of saccades toward the targets which were not followed by a button-press response. They are shown in red and range between 13 and 18 percent. An eccentricity-dependent saccadic hit tolerance was tested as well (as opposed to the fixed 5-degree tolerance), but resulted only in a scaling of the curve's amplitude. The luminance level for detected and not detected targets (hits and misses) was analyzed as well: targets which were added to a low-valued range of the luminance profile were less likely detected than targets added to a high-valued range of the luminance profile. Even some of the subjects sensed that they could not properly detect targets of low-luminance level.

Figure 5 shows the landing precision of saccades for the fixed-amplitude condition. The upper plot depicts the amount of undershoot (mean saccadic amplitude; constant error) in dependence of target eccentricity for the 1st and 2nd (corrective) saccade (black and blue). The landing variability (variable error) was expressed as the standard deviation of the saccadic amplitudes (cyan). All three relationships can be fitted with a linear equation and yield significant fits ($p < 0.05$). Expressed differently, the average saccadic amplitude toward a target is by ca. 18 percent too short and its landing position varies with ca. 10 percent of target eccentricity. A corrective saccade occurs with a probability of 10% and is preferably triggered for target eccentricities in a range of 10 to 20 degrees (bottom graph). For the scaled-amplitude condition these relationships were virtually the same (not shown), hinting that landing precision is independent of target parameters.

For the fixed-amplitude condition two types of parameter changes were tested: a) a higher target luminance, 0.3 and 0.4, whereby duration remained constant at 300ms; b) a longer target duration, 600 and 900 milliseconds, whereby amplitude remained constant at a value of 0.2. Changes were carried out blockwise. Figure 6 shows the average detection rate, the average latency and the occurrence of the luminance maximum for a given parameter change. The detection rate increases with longer duration and higher amplitude (upper two graphs), but changes in luminance have a larger effect than equal changes in duration, e.g. note the different increase for doubling a parameter value (300

to 600 ms as compared to 0.2 to 0.4). The red curve labeled 'eye only' shows the saccadic hits. With increasing detection rate the latencies decrease (lower two graphs), so do the luminance maxima of the classification image (Figure 3). The temporal cross section of the classification images reveals more clues about the integration dynamics (center and bottom graph, figure 3). The amplitude of the saliency area increases with increasing target duration and also becomes steeper in its onset indicating that subjects search for sharper onsets (center graph). For the increase in target amplitude however (bottom graph), the amplitude perplexingly decreases.

Exp 3: Cued Target Search

The letter task was carried out with varying degrees of cueing. As cues the scaled-amplitude target was used in an effort to cue letters of varying eccentricity with equal efficiency. Cues appeared 550 ms before onset of a letter (always 100% validity) and lasted 500ms, leaving a temporal gap of 50ms between cue and letter to avoid potential masking effects. Cues appeared with varying frequency per condition: 0, 25, 50, 75 and 100%. The conditions with 0% and 100% cueing represent the control conditions for which no supporting cues appeared at all (0%), or for each letter appearance one (100%). Figure 7 summarizes the performance, with the left and right plot representing the foveation and identification rate respectively. The foveated rate is the proportion of letters to which the gaze was moved to. Two tolerances for a 'foveation hit' are used: a 1 degree tolerance representing the central fovea and a 5 degree tolerance representing the parafovea (central and para foveation rate respectively). For 0% cueing, the central and para foveation rate was at 0.09 and 0.32 respectively. Both increase with higher cueing frequency but only by a few percent, leaving the impression that cueing is of little help. The foveation rate for cued and not-cued letters ('guided' and 'not' respectively) however proves that cueing does have an effect: the cued foveation rate increases, whereas the not-cued foveation rate decreases. Although the total foveation rate shows little increase, the proportion of menu-selected letters quadruples from ca. 0.09 to 0.37: it starts at the center foveation proportion and ends by just exceeding the para foveation rate (magenta curve in left graph of figure 6). This evidences that the subject does not centrally foveate a letter to make an identification response, but that its parafoveal presence suffices. The subject must therefore perform some covert attentional shifts toward the letter stimulus to obtain an identification judgment. The total identification rate, determined as the proportion correct of the selected letters, increases steadily from 0.02 to 0.09. The absolute identification level is small yet irrelevant to the goal of this experiment. The cued responses increase steadily as well. The not-cued responses increase slightly, remain steady and then drop. The initial increase may be explained by an increased propensity to

respond when cueing is present. Chance level is indicated by the dashed line and is taken as the proportion of manual selections (see left graph) multiplied by 0.1. Summarizing the letter experiment, foveation performance increased only slightly by a few percent, whereas identification rate increased by a factor of 3.5.

Discussion

The classification image obtained from the free-viewing experiment showed that part of the bottom-up analysis of the saccadic target-selection process is a spatially stationary dark spot (figure 2), which slowly appears and is foveated when it is about to disappear (figure 3, top graph). Its distinctness was determined by a trigger/random patch classification and is expressed as an ROC area value (mean=0.57; maximum=0.61). The optimal patch width was about 5 degrees, although this was not significant. The determined ROC area values are not dependent on the method: we also used the reverse-correlation method developed for the analysis of visual neurons (Simoncelli, Paninski, Pillow, Schwartz, 2004) as well as the classification method by Tatler et al (2006), but the computed ROC values were about the same. The ROC area values are almost as high as the ROC values obtained in studies characterizing fixations in scenes, e.g. 0.63 (Tatler et al 2005) and 0.64 (Kienzle et al., 2006) and it is therefore tempting to conclude that our bottom-up feature reveals a major component of target-selection analysis of static scenes. However, a direct comparison between static and dynamic scenes is difficult because motion attracts gaze exceptionally well (Franconeri and Simons, 2003). It may well be that the bottom-up feature represents just a motion stimulus, after all it appears to be a looming stimulus.

The target search revealed that the bottom-up component can be completely overwritten by the top-down component, which had already been shown for complex scenes (Einhäuser et al 2008). But interestingly, the ROC area values for the target search were not higher despite the clearer objective of the search task: although subjects had to learn the target from the first few trials, they all sensed that they had to look for a luminance increase. Thus, an average value of 0.57 may already be an upper boundary and the characterized bottom-up feature may indeed represent the optimal motion stimulus.

Target detection dynamics:

The temporal cross-sections of the classification images (figure 3) reveal a few issues about the target detection dynamics. For the luminance target search, the subjects adjusted to the target dynamics by searching for steeper onsets as opposed to the gradual increase for the free-viewing patches. Subjects were never told about the parameter changes, though many sensed that the targets varied occasionally between sessions

(blocks). The very early rise of the trigger-patch luminance level, at 400ms or earlier, may be interpreted as evidence that the subject follows multiple potential targets simultaneously. This can not be clearly determined from these data due to the 1/f broadband correlation of the pixel luminance values.

For all conditions, the peak of the classification images occurs between 120 and 150ms (figure 6 bottom graph, asterisks). The peak occurs before the saccadic decision time (time to saccadic onset), which is generally assumed to occur between 75ms and 100ms before saccadic onset (e.g. Nazir et Jacobs 1991, Caspi et al 2004). It therefore seems that the visual system awaits the onset of a signal and triggers a saccade when the signal appears to turn off. Even though this seems to be a general strategy for all target conditions, the temporal location of the peak systematically moves with target 'strength'. This can be taken as a hint that saccadic decision time is dynamic on a scale of 20 to 30 milliseconds, which may explain the varying estimates in the previous studies.

The peak amplitude of the classification images changes with the parameter modifications in a way, that is rather puzzling to us. For the conditions with increased target durations we would have expected an earlier start of integration – similar to the free viewing integration dynamics –, but it is the peak amplitude which rises (figure 3, center graph) and onset becomes slightly steeper. For the conditions with increased target amplitudes, the peak amplitude decreases. The former could be explained by the possibility that subjects 'observe' sharper onsets. This explanation however would be in contradiction with the amplitude decrease for the increased target-amplitude condition.

The detection of salient targets in noisy bar code movies has already been investigated by Neri & Heeger (2002). Their observers were asked to foveally detect and identify a target in a short movie sequence (9 frames, 243ms); their choice of noise was Gaussian, the target was a bright or dark bar. Their primary finding was that this discrimination process consisted of a 'detection' and an 'identification' stage, with the former being indicated by a high level of variance preceding the identification stage by ca. 100ms. Translated to our experiments, this would mean that we should find a saliency field in the variance plots 100ms preceding saccadic landing. No such high levels could be detected, hinting that this type of 2-stage discrimination does not take place for saccadic targets in the periphery. But the lack of such variance may also be due to the different use of exact methodology. Their variance plots are made from a task with all types of responses (hits, misses, false alarms, correct rejections), as opposed to only 'false alarm' responses as in our study. The use of different types of noise and the short target duration may also be reasons for differences. All these methodological differences make it difficult to draw a specific comparison.

Orienting properties:

We determined manual and saccadic reaction times for fixed-amplitude targets and scaled-amplitude targets, the latter being an eccentricity-dependent (gaze-contingent) increase in target amplitude to compensate for the peripheral decline in visual acuity (figure 4 top graph). In both reaction types, the switch to the scaled-amplitude condition yielded lower reaction times. For the manual reaction times a reversal from slightly increasing (fixed-amplitude) to slightly decreasing (scaled-amplitude) took place. In case of saccadic latencies, the decrease seemed to be simply downscaled.

The saccadic latencies are further compared to a study by Kalesnykas and Hallett's (1994), who measured latencies using gaze shifts starting from the display center and moving to the outside. Targets simply turned on (step function) and are displayed on a blank background. For eccentricities of 1 to 3 degrees latencies slightly decrease, remain constant from ca. 4 to 12 degrees and then started to increase for larger eccentricities. The function was therefore given the name the bowl-shaped function. In comparison, in our study saccades were made from arbitrary (horizontal) gaze positions on the display with distracting background motion. The latencies for the fixed-scale condition look like a smeared or dampened version of the bowl function (dotted blue, upper graph figure 4). Latencies gradually decline for the first 7 to 10 degrees and then remain steady up to ca. 20 degrees and then show signs of increase. For the scaled-amplitude targets the initial decrease seems to extend to even 13 degrees. Thus, the bowl-function seems to hold for more natural viewing conditions but has broadened. Although, there are attempts to disprove the presence of a constant-latency range - the flat bottom of a bowl function (e.g. Hodgson 2002), those data were not carried out for a broad range of eccentricities, which makes it difficult to judge about the overall shape of the function. Kalesnykas & Hallett (1994) have also noted a slight deviation for landing precision for eccentricities of up to 6 degrees (Figure 10b), specifically a (constant) error of 8.3% (undershoot of 0.5 degrees for a 6 degree eccentric target). Earlier studies measured an error of 10% (Becker, 1972; Henson, 1997). The constant error in our experiments was about 18% (undershoot of 3.8 degrees for 20 degrees eccentricity), thus the use of a distracting noise movie roughly doubles the amount of undershoot. XXX BW: more connection to CB paper.

The variability of saccadic landing position (variable error) was already quantified by Aitsebaomo and Bedell (1992) but primarily for different target durations. Given the increase in saccadic undershoot with increasing eccentricity, it can be expected that the variability increases correspondingly. Van Beer's study (2007) shows also an increase in

variability (figure 2, increasing scatter for landing positions). The scaled-amplitude condition yielded the same results, indicating that target parameters affect its detection and hence saccadic latency but not saccadic amplitude programming.

The presence of a constant and variable error can be interpreted as an inaccuracy of the saccadic spatial-orienting mechanism and hence a weakness, since it apparently does not allow to precisely place the focus on objects with the primary saccade. But for recognition of structure this primary-saccade precision is not necessary: scenes and letters can be recognized translation invariant to a large extent (Rayner 1998; Kirchner and Thorpe 2006; present cueing experiment). This may explain why we had difficulties to obtain a more precise description of the classification image. Together with the tracker measurement error of ca. 0.5 degrees, saccadic landing precision can vary by more than 1 degree for a 5-degree saccadic jump ($0.5 + 0.18 \cdot 5 = 1.3$), which smears out fine structure in a classification image.

The detection rate decreased with increasing eccentricities for fixed-amplitude targets (figure 4, bottom graph), but for the scaled-amplitude targets the detection rate was reasonably steady in the range of 4-16 degrees, proving that the choice of exponential compensation for the decline in visual acuity was a good first guess of the chosen parameters. However, subjects did not react well to targets of very different luminance level within the same block. For instance, a luminance increase for a black patch resulting in a grayish patch was not as well detected as a luminance increase for a gray patch resulting in an almost white target. Subjects can therefore not deal well with just-noticeable targets of varying luminance level in the same block.

The cued search revealed that the majority of identification judgments are made when the letter appeared only in the parafovea. This may not be surprising given the evidence that recognition is translation-invariant (see above). Still, we would have expected that the noisiness of our display would require a much more stringent foveation for proper letter identification. It seems that even in the presence of a very distracting background stimulus (the 1/f noise display) the parafoveal target presence suffices for recognition. The presence of cueing also encouraged manual selections (identification responses), which is evidenced by the not-cued responses from 0.02 to 0.04 (0 to 25% viewing).

Acknowledgements:

We would like to thank the following members of the Schölkopf department at MPI Tübingen in guidance with some computational methods: Wolf Kienzle with help of the support-vector machines; Jakob Macke with guidance of the reverse correlation technique; Matthias Franz for overview of both techniques. We

also thank Brian White for comments on the manuscript and Nadine Hartig for experimental assistance. This work was funded by the Gaze-based Communication Project (contract no. IST-C-033816, European Commission within the Information Society Technologies).

References

- [1] Ahumada, A. J. (2002). Classification image weights and internal noise level estimation. *Journal of Vision*, 2, 121-131.
- [2] Aitsebaomo, A.P. & Bedell, H.E. (1992). Psychophysical and saccadic information about direction for briefly presented visual targets. *Vision Research*, 9, 1729-1737.
- [3] Becker, W. (1972). The control of eye movements in the saccadic system. *Bibliotheca Ophthalmologica*, 82, 233-43.
- [4] Beers, van Robert J. (2007). The sources of variability in saccadic eye movements. *The Journal of Neuroscience*, 27, 8757-8770.
- [5] Caspi, A., Beutter, B. R. & Eckstein, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. *Proceedings of the National Academy of Sciences of the USA*, 101 (35), 13086-13090.
- [6] Chang, C. & Lin, C. (2001). LIBSVM: A library for support vector machines. Software available at [\url{http://www.csie.ntu.edu.tw/~cjlin/libsvm}](http://www.csie.ntu.edu.tw/~cjlin/libsvm).
- [7] Einhäuser, W., Schumann, F., Bardins, S., Bartl, K., Böning, G., Schneider, E. & König, P. (2007). Human eye-head co-ordination in natural exploration. *Network: Computation in Neural Systems*, 18, 267-297.
- [8] Einhäuser, W., Rutishauser, U. & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, 8, 1-19.
- [9] Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4 (12), 2379-2394
- [10] Findlay, J.M. and Gilchrist, I.D. *Active Vision*. New York: Oxford University Press, 2003.
- [11] Franconeri, S.L. & Simons, D.J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics*, 65, 999-1010.
- [12] Geisler, W. S., Perry, J. S. & Najemnik, J. (2006). Visual search: The role of peripheral information measured using gaze-contingent displays. *Journal of Vision*, 6, 858-873.
- [13] Hodgson, T.L. (2002). The location marker effect: saccade latency increases with target eccentricity. *Experimental Brain Research*, 145, 539-542.
- [14] Kienzle, W., Wichmann, F. A., Schölkopf, B. & Franz, M. O. (2006). A nonparametric approach to bottom-up visual saliency. *Neural Information Processing Systems (08/2006)*.
- [15] Nazir, T.A. & Jacobs, A.M. (1991). The effects of target discriminability and retinal eccentricity on saccade latencies: An analysis in terms of variable-criterion theory. *Psychological Research*, 53, 281-289.
- [16] Neri, P. & Heeger, D. J. (2002). Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nature Neuroscience*, 5, 812-816.

- [17] Noble, W. S. (2006). What is a support vector machine? *Nature Biotechnology*, 24 (12), 1565-1567.
- [18] Rajashekar, U., Bovik, A. C. & Cormack, L. K. (2006). Visual search in noise: Revealing the influence of structural cues by gaze-contingent classification image analysis. *Journal of Vision*, 6, 379-386.
- [19] Rayner, K. (1998). Eye movements in reading and information processing. 20 years of research. *Psychological Bulletin*, 124, 372-422.
- [20] Reinagel, P. & Zador, A. M. (1999). Natural Scene statistics at the centre of gaze. *Network: Computation in Neural Systems*, 10, 341-350.
- [21] Roos, J.C.P., Calandrini, D.M. & Carpenter, R.H.S. (2008). A single mechanism for the timing of spontaneous and evoked saccades. *Experimental Brain Research*, 187, 283-293.
- [22] Shimozaki, S.S., Chen, K.Y., Abbey, C.K. & Eckstein, M.P. (2007). The temporal dynamics of selective attention of the visual periphery as measured by classification images. *Journal of Vision*, 7, 1-20.
- [23] Simoncelli, E. P. & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24, 1193-1216.
- [24] Simoncelli, E. P., Paninski, L., Pillow, J. & Schwartz, O. (2004). Characterization of neural responses with stochastic stimuli. In Gazzaniga, M. S. (Ed.): *The Cognitive Neurosciences III* (pp. 327-338). MIT Press.
- [25] Tatler, B. W., Baddeley, R. J. & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, 45, 643-659.
- [26] Tatler, B. J., Baddeley, R. J. & Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research*, 46, 1857-1862.
- [27] White, B.J., Stritzke, M. & Gegenfurtner, K.R. (2008). Saccadic facilitation in natural backgrounds. *Current Biology*, 18, 124-128.
- [28]

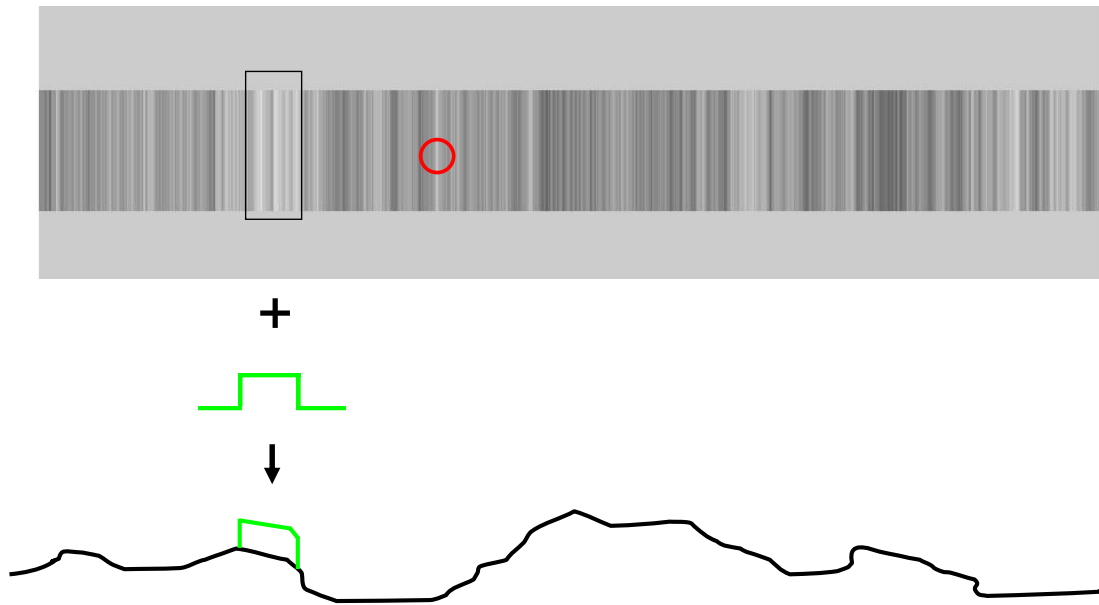


Figure 1. Noise stimulus and target generation. The shown bar code represents one frame (10ms, width = 1200pixels [=40degs], height=100pixels) of the noise movie, which is taken row-wise from a 1/f, 2D noise image). In an oddity search, a target was generated by adding a finite-pulse function (green) to the luminance profile of the bar code (bottom). This is done for a fixed pulse amplitude (fixed-amplitude condition), as well as for a pulse amplitude, which depended on fixation eccentricity (scaled-amplitude condition) [red: gaze at present point in time].

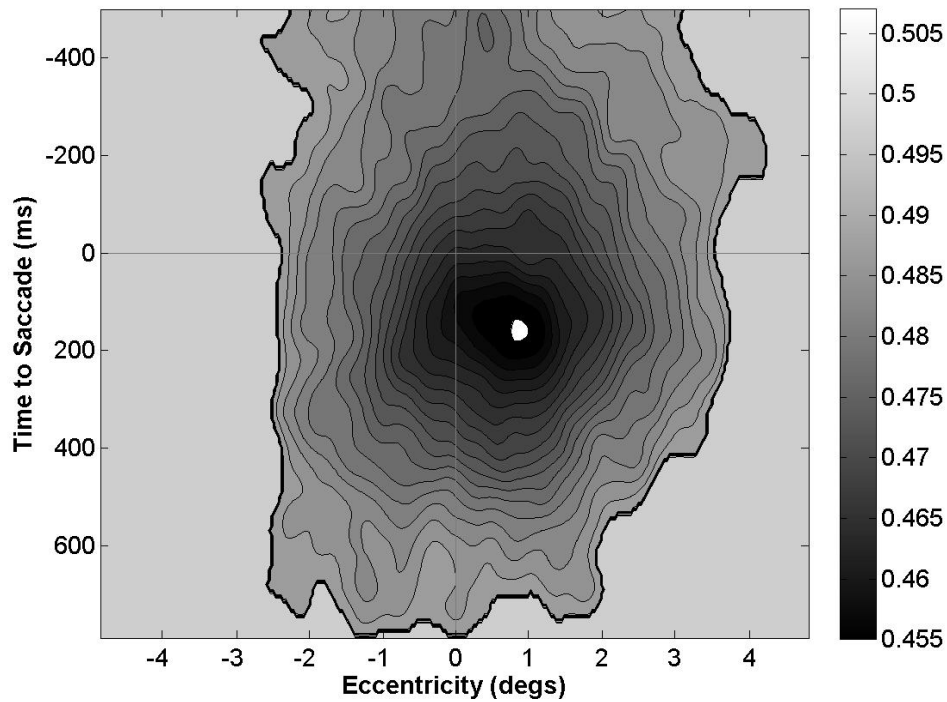


Figure 2. A typical classification image (one observer) for the free viewing experiment (3823 fixations). Y-Axis: Countdown to saccade in milliseconds (0=saccadic onset). X-Axis: Eccentricity relative to saccadic landing position in degrees (0=saccadic landing position). The luminance level is from a range of [0,1]. Trigger patches were flipped according to direction of flight. Values with one standard deviation taken from the random classification image (non-trigger) patches were set to average gray-scale (the surround). The white dot represents the minimum luminance value of the entire field.

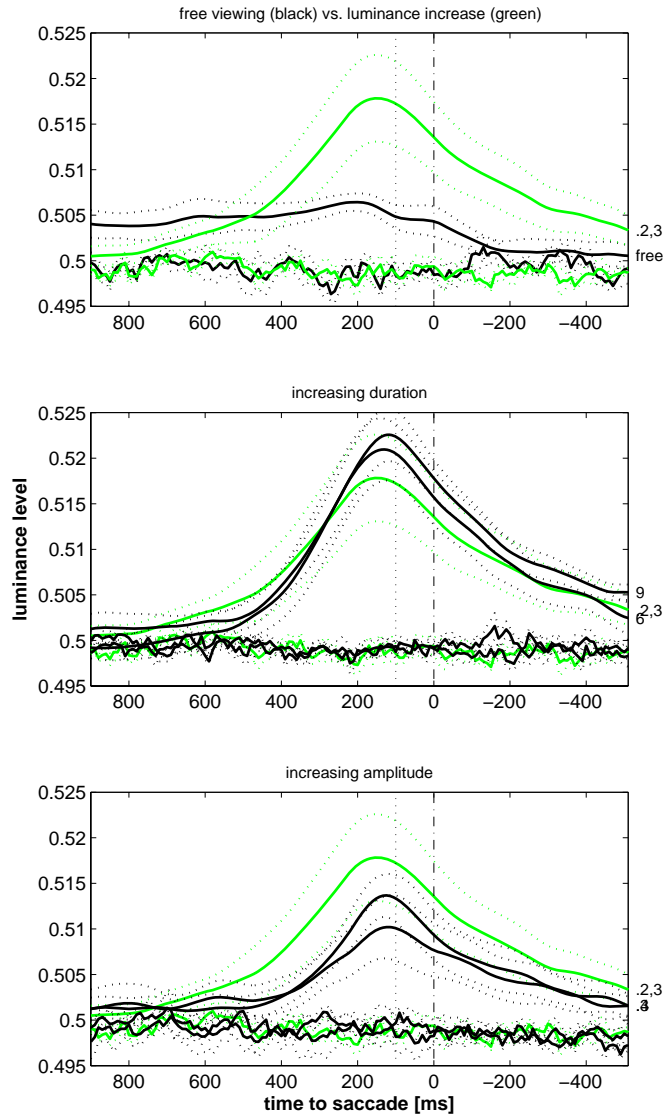


Figure 3. Temporal cross-sections of the subject-averaged classification image (4-5 observers). X-Axis: Countdown to saccade in milliseconds (0=saccadic onset). Y-Axis: Luminance level [0,1]. Top plot: Free viewing versus luminance oddity search (black vs. green). The free-viewing values were inverted for reason of comparison. Center plot: increased target duration [600 and 900ms – denoted as 6 and 9]. Bottom plot: increasing target amplitude [0.3 and 0.4]. The green curve is the ‘standard’ (fixed-amplitude) target condition ($a=0.2$, $d=300$ ms). The dotted curves denote standard error of inter-subject performance. The vertically dotted straight line is placed at 100ms.

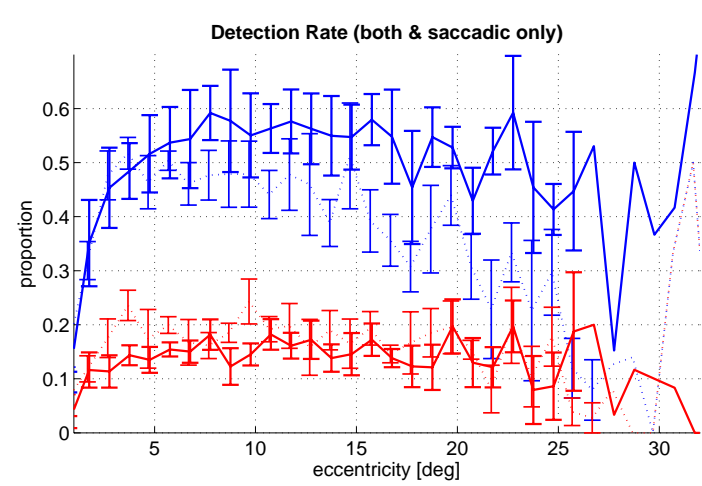
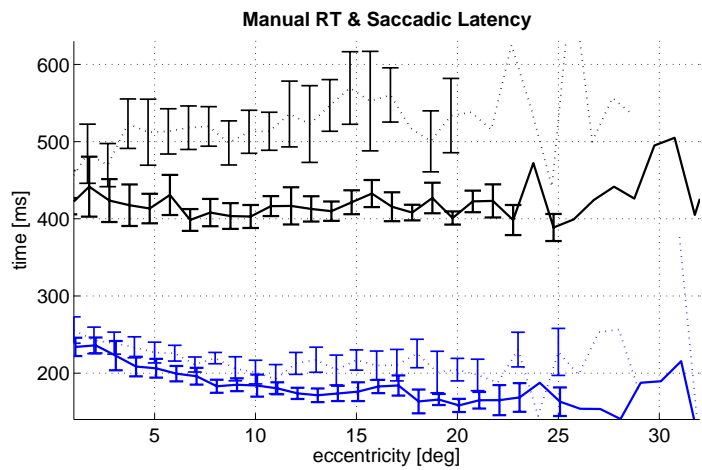


Figure 4: Target search for fixed-amplitude (dotted) and scaled-amplitude (solid) target condition. Top: manual reaction times (black) & saccadic latencies (blue) in dependence of target eccentricity. Bottom: detection rate for targets which captured gaze and were signaled by a button press (blue) and detection rate for targets to which only a saccade was made (red). Error bars = standard error of inter-subject performance.

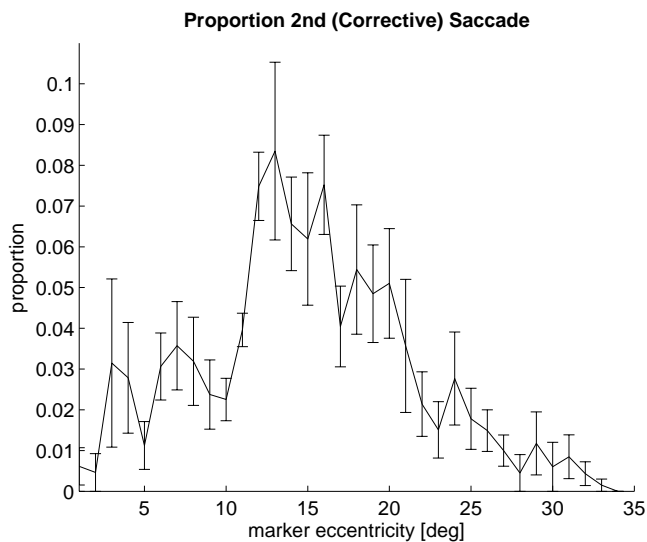
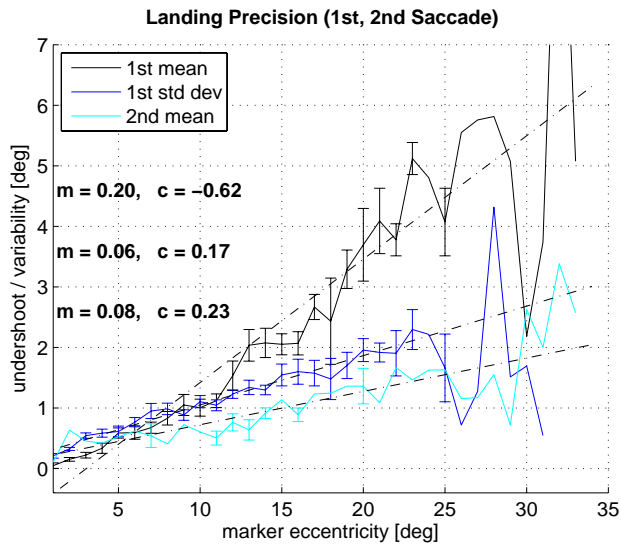


Figure 5: Saccadic landing precision in dependence of eccentricity for visual search. Top plot: Average undershoot (constant error) and variability (variable error [standard deviation]) for primary saccades (black and blue respectively); average undershoot for secondary saccades (cyan) [m =slope; c =intercept]. Bottom plot: Proportion of corrective (2^{nd}) saccades. Error bars = standard error of inter-subject performance.

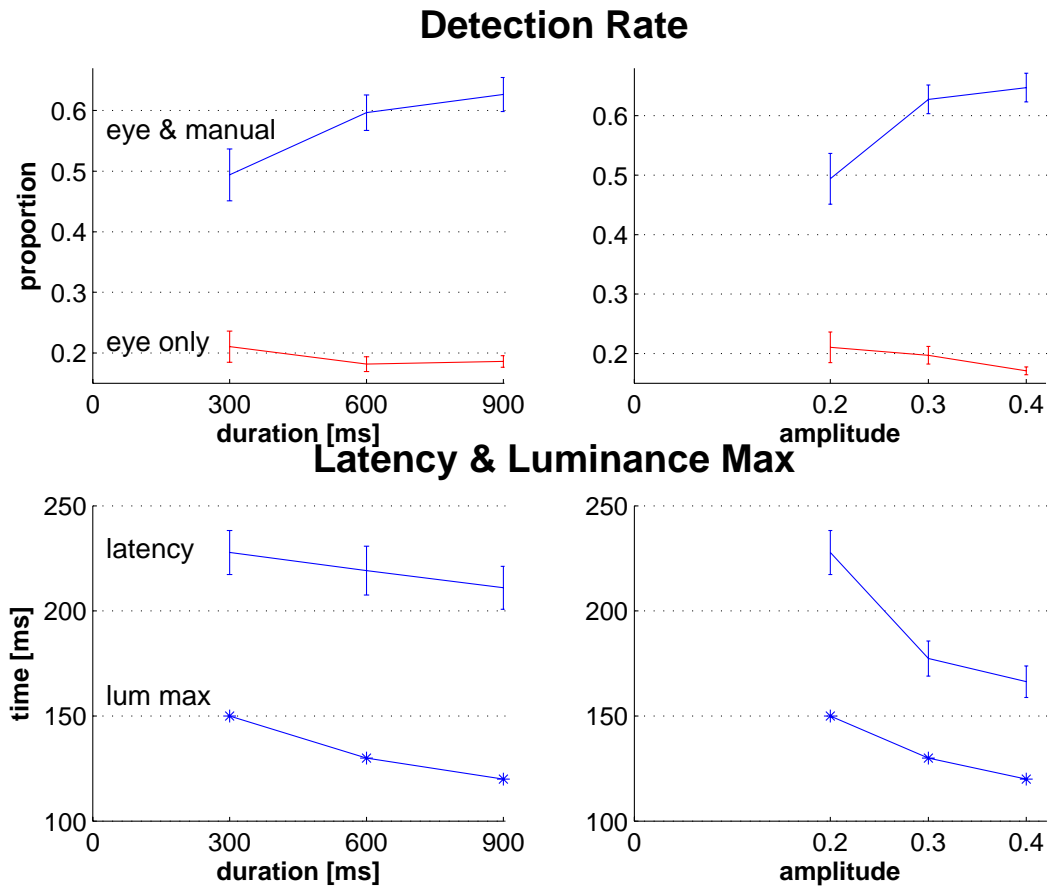


Figure 6. Parameter changes for the fixed-amplitude condition (duration=300ms, amplitude=0.2).
 Upper left: Longer durations: 600, 900ms (red=saccade-only detection).
 Upper right: Higher amplitudes: 0.3 and 0.4.
 Lower left: Saccadic latency and luminance maximum (*) for duration changes. The luminance maximum is obtained from a classification image.
 Lower right: Latency and luminance maximum for amplitude changes.
 Error bars = standard error of inter-subject performance.



Figure 7. The letter task. 10 letters were shown with a frequency of 0.06 Hz each (ca. 6 letters per 10-second trial), for a duration of 500ms at a contrast of 0.1 (contrast not to scale in figure). A letter had a size of ca. 1 x 1 degree. Below the noise display, the set of letters was shown as a menu (letter in display and menu are exact to scale). During visual search, the viewer would select those letters, which s/he had detected and identified.

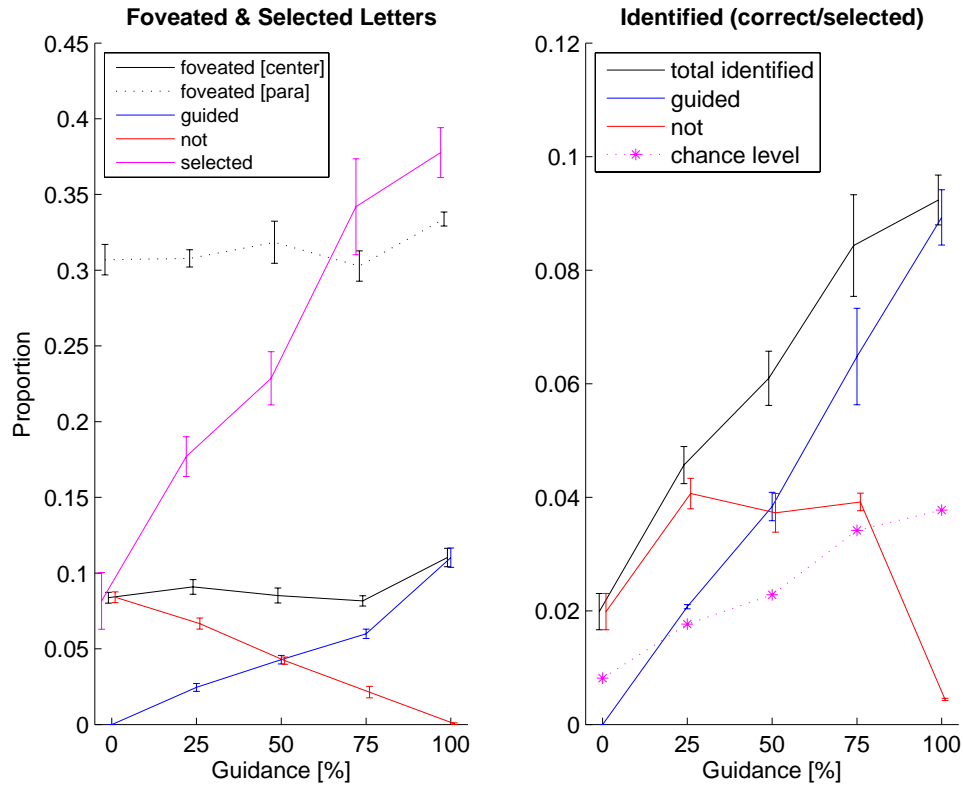


Figure 8: Letter foveation and identification in dependence of guidance (0, 25, 50, 75, 100%). Left: Foveated letters: letters to which the gaze moved to (foveated [center]: central foveation with 1 deg tolerance; foveated [para]: parafoveal foveation with 5 degs tolerance). *guided* (cued): proportion foveation for guided letters (or cues); *not*: proportion foveation for not-guided letters. *selected*: proportion of identification responses (letter selections using menu). Right: Proportion of identified letters. Chance level is taken as the proportion of identification responses (from left) multiplied by 0.1 (letter identification chance level). Error bars = standard error of inter-subject performance.