A254RA

Visual Orienting in Dynamic Broadband (1/f) Noise Sequences

Accepted with the journal ‚Attention, Perception & Psychophysics'

Christoph Rasche, Karl R. Gegenfurtner
Abteilung Allgemeine Psychologie, Justus-Liebig-Universität
Otto-Behaghel-Str. 10F, 35394 Giessen, Germany

Corresponding author: Christoph Rasche, email: rasche15{at}gmail.com

Suggested running head: Visual Orienting in 1/f Noise Movies

Abstract

Visual orienting has been typically characterized using simple displays, e.g. displays with a static target placed on a homogenous background. In this study, visual orienting was investigated using a dynamic broadband (1/f) noise display that should mimic a more naturalistic setting and that should allow performing saccadic orienting experiments with less constraints. In a 1st experiment, it is shown that the noise movie contains gaze-attracting features almost as distinct as the ones measured for (static) real-word scenes. The movie can therefore serve as a strong distracter. In a 2nd experiment, observers carried out a luminance target search, which showed that saccadic amplitude errors were substantially higher (18%) than the ones measured in simple displays. That error is certainly one of the primary factors making gaze-fixation prediction in complex scenes difficult.

Introduction

Visual orienting has been predominantly investigated on homogeneous backgrounds with well defined targets (e.g. Findlay and Gilchrist, 2003; Roos et al 2008), e.g. a saccade is made from a central fixation cross toward a peripheral target of sufficiently high contrast. Such simple, highly controlled conditions are suitable to accurately determine 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. Their analysis concentrated primarily on intersaccadic intervals. The orienting experiments mentioned so far were carried out with a 'static' observer. Visual orienting is also investigated with 'dynamic' observers to determine visual selection during daily activities such as tennis playing or walking through the forest (e.g. Land et al., 1999; Einhäuser et al, 2008). Such experiments are beneficial to gain a more detailed understanding of (spontaneous) saccadic target selection during actions, but they do not allow for the control of the visual environment to

investigate precise orienting behavior in response to targets, for instance, how accurately a target is foveated during search. Our approach to the analysis of visual orienting aimed in-between the static and dynamic condition: on the one hand we desired control over the stimulus and for that the observer is preferably static; on the other hand we wanted a stimulus that could possibly mimic the same degree of attraction (or distraction depending on task) as experienced during 'free-moving' – a simulation of the dynamic condition. For this purpose we generated a movie based on dynamic broadband noise, which was displayed on a typical experimental CRT monitor. The broadband noise had an inverse frequency amplitude spectrum (1/f), similar to the amplitude spectrum of natural images (Field et al 1987). This allows us investigating saccadic orienting properties under close-to-naturalistic conditions.

Our first goal was to characterize visual selection in those noise movies, specifically to elucidate whether, and to what degree, the dynamic 1/f movie stimulus provided preferred features of fixation selection. From scene-fixation studies, it is known that there exists a preference for high contrasts (e.g. Reinagel and Zador, 1999; Tatler et al. 2006; Parkhurst and Niebur, 2003). One may therefore wonder whether this contrast preference also exists in a mere 1/f display (in our case a movie). To investigate this, observers merely perform free-viewing in a first experiment. Analyzing the image properties of the fixation locations, we found that some of them were indeed similar to the image properties of fixation locations of natural scenes. From this, we concluded that the movie provides a good 'natural' attractor. It makes therefore sense to investigate orienting behavior in those movies by carrying out a luminance search task, which is done in a second experiment. It turns out that saccadic landing precision decreased as compared to measurements reported in other studies made with evoked saccades.

The use of 1/f noise has become increasingly popular over the past years, because natural images show the same amplitude spectrum (Field, 1987; Simoncelli & Olshausen, 2001). For instance, Geisler et al. (2006) have investigated the search behavior for Gabor targets placed in a static, 1/f-noise background using gaze-contingent methodology. They found that the search time and number of fixations both increase with spatial frequency and noise contrast. Our experiment rather addresses saccadic orienting parameters, such as saccadic latency, constant (mean) error and variable error.
Cormack and collaborators investigated the detection and identification of shapes (2006, 2007). For instance, using the method of classification images, Rajashekar et al. (2006) found that saccadic target selection was based on structural cues despite the noisiness of the 1/f displays. Tavassoli et al. (2007) continued that line of research but additionally proposed a variant of the classification image paradigm, with which one can obtain useful classification images in as few as 200 trials. Their method consisted of dividing the display into tiles. Although we use the method of classification images as well, their tiling method is not straightforward to apply as we use targets which can appear anywhere in the display and not only at selected locations. 1/f displays were also used by White et al. (2008) to determine the latencies of evoked saccades. Employing a gap

paradigm, they measured that latencies toward Gabor patches were smaller in 1/f displays than on a homogenous background lacking any structure.
Here, we extended the use of 1/f-noise to dynamic displays in order to create a moving input which potentially is as attracting as the input perceived during natural activities. The noise movie was created from a two-dimensional 1/f image (figure 1), the same type of image as used in other studies (e.g. Rajashekar et al. 2006, Geisler et al. 2006). In those studies, the image was presented statically, but in our study it was used as the source for a one-dimensional movie by iterating row-wise through the source image: each row was stretched vertically in space and presented as a frame (figure 2) - the movie thus appears as a flickering bar code. The choice of spatial one-dimensionality was 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).

Experiment 1: Free Viewing

Observers were asked to freely view the noise movie – that is viewing without any specific instruction. They did so naturally on the horizontal axis only (because this is where the dynamic stimulus appeared). An example of a scan path of the horizontal eye-position is plotted into the source image (figure 1), vertical lines represent fixation periods, horizontal lines describe saccadic jumps. To determine whether gaze was drawn toward specific locations/features of the stimulus, we compared the fixated movie patches with a set of randomly collected movie patches. This analysis was done in analogy to the analysis carried out for scene-fixation studies, in which a gaze prediction for real-world scenes is sought (e.g. Reinagel and Zador, 1999; Tatler et al. 2006; Parkhurst and Niebur, 2003; Carmi and Itti, 2006; Vig et al, 2009). They typically used (static) images of real-world scenes while observers performed free viewing. The fixated image patches were then compared to a set of randomly collected image patches in an effort to characterize the bottom-up component of saccadic target selection in natural scenes. In our study, the fixated image patches correspond to spatio-temporal image patches before saccadic landing. Specifically, for a given fixated spatial location, its preceding space-time patch was extracted (see white rectangles in figure 1). This space-time patch presumably has triggered the visual system to place its gaze on it. It is therefore called a *trigger patch*.

To investigate whether there exists a feature in a trigger patch, a first simple step was to look at the average of the collected trigger patches, a method known as the classification image (Ahumada, 2002; Shimozaki, 2007; Tavassoli, 2007). The method is typically applied to image patches of ‘false alarms’ in experimental tasks containing the four response outcomes hits, misses, false alarms and correct rejections, or by linear combination over classification images for all response outcomes. The classification images typically reveal aspects of stimulus representation and recognition mechanisms, when Gaussian noise is used, but recently the methodology has been applied to 1/f noise (e.g. Rajashekar et al. 2006, Geisler et al. 2006). In our study, saccadic landing positions can be loosely regarded as false alarms, because no actual target has been placed at those

locations and the classification image may therefore contain hints about the presence of a possible trigger feature for saccadic target selection in 1/f displays. In our study, the classification image corresponds to the average of the space-time trigger patches, whose temporal cross-section can also give us clues about the saccadic decision dynamics.
To characterize the degree of distinctness of the trigger patches, we apply Tatler et al.'s method (2005, 2006) that was developed for a discrimination of fixation and non-fixation patches. The method determines the ROC area value in a very simple manner and we refer to this as the *fixation prediction*.

Methods

Observers. Two male and three female students (age 23-30) served as observers. All observers had normal or corrected to normal vision. All observers were naive with respect to the aim of the experiment.
Equipment. Observers 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 by 36 degrees of visual angle, in the horizontal and vertical direction respectively. With a spatial resolution of 1280 x 1024 pixels this results in 28 pixels/deg. The observer'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. Observers viewed the display binocularly through natural pupils. Stimulus display and data collection were controlled by a PC using modified C routines from SR Research Ltd and SDL routines.
Stimulus. The two-dimensional 1/f image *I[x,t]* was generated using a 2D image of normally distributed random pixel-intensity values, whose frequency spectrum was then transformed to describe a 1/f-frequency decline. The image size was 1000*1200 pixels, which defined the temporal and spatial characteristics of the stimuli, respectively. Each row was the source for a single frame: the row was stretched vertically to a height of 100 pixels and placed into a gray background presented as 8-bit luminance resolution (40 cd/m^2 luminance). A frame was shown for 10ms, a movie thus lasted 10ms*1000 (pixels) = 10s and constitutes one trial. Each movie *I(x,t)* was different to avoid potential learning effects.
Procedure. Observers performed blocks of 50 trials, generally 3 blocks per day and 6 blocks per experiment. Each block was preceded a 9-point calibration and validation. Before each trial, observers 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 (e.g. 0.4 ± 0.1 in Tatler et al 2005). Asking observers to perform free viewing was not as easy as it seems because some observers immediately ask what they were supposed to look for. To obtain 'unbiased' viewing, observers were simply asked to 'be inspired'. This unusual instruction was done to avoid any mentioning of the overall goal of the experiments.

Observers who participated in multiple experiments, started with the free-viewing task first.

The classification image was generated for trigger patches of sufficiently large size to understand the dynamics. A trigger patch was centered spatially on the point of saccadic landing and temporally at the time of saccadic onset. Trigger patches exceeding the image border were omitted as well as the very first one after a trial was started. Trigger patches were aligned according to the direction of saccadic flight: patches resulting from a leftward-flight were mirrored along the spatial dimension. An equal number of *random* patches were collected for statistical comparison, which was chosen as follows: given the spatial location of a trigger patch, a corresponding random patch was taken, whose spatial location was mirrored along the spatial axis with the half image width as the line of reflection. In other words, the random patches were chosen from the scan path, which was flipped along the spatial axis. In total, an average of 2990 patches was collected per person. In addition to the mean, also the variance of the trigger patches was computed.

Fixation prediction. To determine the degree of distinctness of the trigger patches, the average luminance values of the individual patches were compared to the average values of the random patches. Specifically, a decision threshold is slid through the two luminance distributions and true positives and false positives were determined, from which the ROC area value was calculated (see Tatler's et al. method (2005, 2006) for an elaborate justification of this method). The average luminance value for an individual patch was taken from a small area centered around the minimum of the classification image, for instance 1 degree wide and 60ms long.

Results

Because the use of our movie sequences was novel, we ensured that saccadic search behavior roughly corresponded to regular search behavior. Histograms for fixation durations show a peak at around 320ms (in average) and an 1/time decay (see supplementary material figure 1); the ones for saccadic amplitudes showed a peak between 2 and 8 degrees with long-tailed distributions for the majority of observers. The classification image was generated for patches starting from several hundreds of milliseconds before saccadic onset and extending some hundreds of milliseconds after onset in order to observe the entire dynamics (the patches in figure 1 are smaller for purposes of illustration). Figure 3 is an example for one observer. It shows a depression (trough) in the luminance level. To highlight the area which was significantly different from chance, those values were set to the mean grayscale value (0.5), which were within the range of values given by the classification image for the random patches (see figure 4 for a cross section). This highlighted area is now called a *salient area*. The salient area's spatial minimum is located 0.61 degrees on average beyond the landing point, which was set at 0 degrees eccentricity (standard deviation = 0.67 degrees, one person showed overshoot, not shown); 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 emphasizes slowly

changing low frequencies). The classification image of 4 other observers is shown in figure 4, left column: they all clearly show a large depression of similar size with a minimum at about the same spatial and temporal location. The right column displays the variance of the trigger patches. Only small salient areas could be found but which were not consistent across observers.
The analysis of fixation prediction yielded ROC area values ranging from 0.54 to 0.60 for different observers (mean=0.56).

Discussion

The classification image of trigger patches clearly shows the presence of a large depression for all observers tested and distinguishes itself clearly from the random background. The depression indicates that the visual system preferred to direct its gaze toward a static, looming dark spot in this type of movie stimulus.
The classification image reveals two more aspects of the spot. 1) The location of the spatial minimum - 1 degree beyond saccadic landing - can be interpreted in two ways: the visual system selected a location of large contrast, e.g. the edge of a dark vertical bar, or it selects the center of a dark spot and just lands short of it. Thus, there is the possibility that the bare 1/f 'background' structure of a visual image may contain gaze-triggering features. 2) The temporal minimum occurred at ca. 130ms before saccadic onset, which indicates that the saccadic system decided to jump when the dark spot started to disappear – given a saccadic decision time of 75-100ms (e.g. Nazir et Jacobs 1991, Caspi et al 2004).
With the fixation prediction analysis we determined that this gaze-attracting feature can be distinguished from a random background stimulus by an ROC area value of up to 0.60 (mean=0.56). To ensure that the values are approximately comparable to other methods, we also used the reverse-correlation method developed for the analysis of visual neurons (Simoncelli et al., 2004) as well as the Support-Vector Machine classification method by Kienzle et al (2009). With all these methods similar ROC area values were obtained hinting that the values are independent of the method. In the next section, it will be reported that the 'distinctness' of false alarms for a luminance target was of similar magnitude. Thus, even if the movie consists of only 1/f noise, it may provide the degree of attraction as experienced during natural input.

Experiment 2: Target Search

In this experiment, a temporarily appearing luminance target was inserted into the dynamic noise movie. The target search allows to determine orienting properties such as saccadic latency and saccadic precision which then are compared to measurements reported in studies with evoked saccades under single static stimulus conditions. It is assumed that in particular saccadic variability will increase due to the presence of the dynamic background noise, which has been shown to be a strong gaze attractor in the first experiment and which now serves as a strong distracter.

The target was a rectangular function added to the luminance profile and was therefore not constant in amplitude (see figure 2). The target was shown for a few hundred milliseconds and so appears as a bright bar in the display. It appeared occasionally only, ca. 3 times a trial. The target was of very low contrast and initially difficult to detect and can therefore be considered just-noticeable.
In preparation for a cueing experiment (to be reported elsewhere), we attempted to create a luminance target, which was equally visible in the periphery. To compensate for the peripheral decline in visual acuity, we increased the target amplitude with increasing target eccentricity in a gaze-contingent manner. This is called the *scaled-amplitude* target condition, whereas the measurements made with unmodified target amplitude are called the fixed-target amplitude (fixed means that the added rectangular function was fixed).

Methods

Observers. Two male and four female students (age 23-30) served as observers. All observers had normal or corrected to normal vision. All observers were naive with respect to the aim of the experiment.
Target: The movie stimuli were the same as for the free-viewing condition. Targets were added as a rectangular function to the luminance profile of the source image with an increment of value a_{trg} =0.2 (total intensity range from 0 to 1), for a duration of d=300ms (30 frames) and a spatial width of 1 degree at the same location, see figure 2. Targets were presented spatially and temporally randomly with an average frequency of 0.333Hz. This constituted the fixed-amplitude condition.
To generate the scale-amplitude target, it required a function which would compensate for the decline in visual acuity. The decline has been described by various functions, e.g. a logarithmic decay (Schwartz, 1980), or a one over eccentricity function (Wilson et al, 1990), see also (Rovamo and Virsu, 1979). We therefore chose an exponentially saturating function in which the target amplitude a_{trg} depended on eccentricity e with reference to the present eye-position by: $a_{trg} = a_{min} + a_{max}\text{-}exp(\text{-}e)a_{max}$, whereby a_{min} is a minimal amplitude and a_{max} is a maximal amplitude; the function starts at a_{min} and saturates at $a_{min} + a_{max}$. The parameter values were a_{min} =0.2, matching the amplitude of the fixed-amplitude condition, and a_{max} =0.5, chosen heuristically after a few initial tests.
Analysis of saccade orienting properties: Observers were instructed to move their gaze toward the targets and press a button when seeing a conspicuity. Target detection was 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 spatial landing within 5 degrees of target eccentricity. The temporal tolerances for saccadic latency and button response were 400 and 1200ms respectively. Each search condition was carried out by 4 to 5 persons. A observer typically performed 6 blocks of one experiment during which the target was presented 850-900 times (ca. 3 target presentations per trial).
Classification image: Trigger patches resulting from saccades made toward a target were omitted. As the number of targets was small (ca. 900) compared to the number of saccades, there were still more than 2000 patches in average for each person.

Results

Observers were not given any specific details about the target and were asked to detect the target by themselves. It would take the observer a few trials (max ca. 20) until s/he had discovered the conspicuity on the first block. Over time, the observer developed a clear understanding about the properties of the target, meaning all observers realized that it was a temporarily appearing, vertical oriented, bright bar as indicated by the classification image.
Fixation durations and saccadic amplitudes were slightly lower as compared to the first experiment (decrease of 40ms and 1 deg respectively; see supplementary material figure 2). This difference is equally minor as measured in another study for static images, which compares search versus free-viewing behavior (Tatler et al's 2006). (Because of the same small difference, we considered the free-viewing behavior in our noise movies, as carried out as in experiment 1, as analogous to free-viewing behavior in static scenes).
The classification images of the trigger patches show a salient area, which can be described as a mound or elevation (figure 5). The salient area was of similar size as the salient area of the free-viewing experiment (compare left columns of figures 4 and 5). The spatial location of the elevation's maximum was at 1.34 degrees in average (standard deviation = 1.07 degrees) and roughly corresponds to the minimum of the free-viewing's salient area – the spatial minima and maxima were statistically not different. Some salient areas also contained large, flanking depressions (e.g. observer 9 and 12), which probably reflects individual differences in target representation and were partly caused by the spatial 1/f correlation or luminance values.
We can discern more differences between the two types of salient areas by looking at the temporal cross section of the classification images. Figure 6 shows the one for free-viewing and the one for the luminance target search, averaged across observers. The one for free viewing was inverted for comparison. The cross-section for the target search evidences that integration occurred later and at a higher amplitude, reflecting that observers searched for rather abrupt onsets. In comparison, the cross-section for free viewing evidences that a long-lasting, gradual integration took place: observers apparently followed the slow dynamics of the dark spots.
The fixation prediction analysis for trigger and random patches (omitting the few patches with targets) yielded ROC values ranging from 0.53 to 0.62 (mean=0.57), which were approximately equal to those measured for the free-viewing trials (also mean=0.56).

Orienting properties. Figure 7 shows the manual reaction times, saccadic latencies and detection rates as a function of eccentricity, averaged across observers for the fixed-amplitude and scaled-amplitude target condition (dashed-dotted and solid respectively). 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; $p<0.05$ for t-test comparing ranges from 1-10 and 11-20 degrees). For the scaled-amplitude condition - the gaze-contingent increment in target amplitude - 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 ms (top graph) and then remained constant

at around 210 ms. For the scaled-amplitude condition latencies were slightly lower by about 10 to 30 milliseconds.
The detection rate for fixed-amplitude targets (lower graph, figure 7) 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) was due to the difficulty to determine a saccadic jump toward the target given the noisiness of the background and probably due to the observer's low propensity to saccade toward near targets. The detection rate for button presses as a function of eccentricity was roughly constant at a level of ca. 0.6.
The detection rate for scaled-amplitude targets increased from 4 to 8 degrees eccentricity and saturated at a value of 0.6. It then slowly declined for large eccentricities. 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.
Figure 8 shows the landing precision of saccades for the fixed-amplitude condition. The upper left plot depicts the amount of undershoot (mean saccadic amplitude; constant error) as a function of target eccentricity for the 1st saccade. The landing variability (variable error) was expressed as the standard deviation of the saccadic amplitudes (upper right plot). All three relationships can be fit with a linear equation and yielded significant results ($p<0.05$). Expressed differently, the average saccadic amplitude toward a target undershot by 18 percent and its landing position varied with ca. 10 percent of target eccentricity. A corrective saccade occurred with a probability of 10% and was preferably triggered for target eccentricities in a range of 10 to 20 degrees (lower left plot). The 2nd (corrective) saccade still showed some undershoot (ca. 5%; see lower right plot).
For the scaled-amplitude condition the relationships for landing precision showed surprisingly no notable differences (figure 9). However, the proportion of 2nd saccades seemed a bit altered, with prominently fewer saccades in the range of 1-10 degrees and a sharp incline around 11 degrees.

Discussion

At a first glance, the classification image for the false alarms of the luminance target looked like a mere inversion of the classification image for the free-viewing task. But the temporal cross-section of the classification images revealed that the visual system followed slower changing dynamics when detecting dark spots during free viewing as compared to detecting the sudden onset of a luminance target.
The fixation prediction analysis returned a ROC value of 0.57 on average, which was equally high as the discrimination for the patches obtained from the free-viewing task. This was surprising because we expected that they would be higher given that observers had a specific search target.
The detection of salient targets in noisy one-dimensional movies has been investigated by Neri & Heeger (2002). Their observers were asked to foveally detect and identify a target in a short movie sequence (9 frames, total trial duration: 243ms); their choice of noise was Gaussian. Their primary finding was that this discrimination process consisted of a 'detection' and an 'identification' stage, where the former was 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 but no such high levels were detected (right columns in figures 4 and 5). The lack of such variance may be due to the many methodological differences between our and their study, but the primary reason may be that our results are based on saccadic and theirs on perceptual decisions.

Orienting properties. The scaled-amplitude condition yielded shorter reaction times for both the manual response and the saccadic latency: the manual reaction times for the fixed-amplitude condition were slightly increasing across eccentricities but seemed to decrease for the scaled-amplitude condition. The saccadic latencies for the scaled-amplitude condition seemed to be merely downscaled.

We further compare the saccadic latencies to those measured for evoked saccades in a study by Kalesnykas and Hallett (1994). They determined latencies using gaze shifts starting from the display center and moving to the periphery. Targets simply turned on (step function) and were 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 eccentricites. The function was therefore given the name ‘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 7). 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, although this can not be statistically tested due to the lack of data points. 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 (their figure 10b), specifically a (constant) error of 8.3% (undershoot of 0.5 degrees for a target eccentricity of 6 degrees). Earlier studies measured an error of 10% (Becker, 1972; Henson, 1979). The constant error in our experiments was about 18% (undershoot of 3.8 degrees for 20 degrees eccentricity). It is not clear though, whether the increased variability derives from the search behavior or from the noise background, or from both.

Variability in saccadic landing position (error) was quantified by Aitsebaomo and Bedell (1992) but primarily for different target durations, showing a decrease with increasing target duration. Saccadic landing variability also grows with increasing eccentricity (figure 8), which was also measured in (Van) Beer’s study (2007; figure 2, increasing scatter for landing positions). The scaled-amplitude condition yielded the same degree of

variability, indicating that although target parameters do affect target detection, and hence saccadic latency, they appeared to have little effect on saccadic amplitude.
The detection rate decreased with increasing eccentricity for fixed-amplitude targets (figure 7, 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 (a_{min} =0.2, a_{max} =0.5).

General Discussion and Summary

Visual orienting was studied with a broadband (1/f) noise stimulus, which served as an intermediate form of the typical stimuli used in laboratory experiments (e.g. the gap paradigm), and the stimuli as perceived during natural activities. In a first experiment, the free-viewing task, we demonstrated that the noise stimulus contained a strong gaze-attracting feature, namely a static dark, looming spot (figure 3). This was determined using the method of classification images (Ahumada, 2002; Shimozaki, 2007), whose temporal cross-section also revealed that the spot slowly appears and was foveated when it had started disappearing (figure 6).
To determine the 'strength' of this gaze-attracting feature, we employed a trigger/random patch classification – in analogy to the analysis done for scene fixations (e.g. Reinagel and Zador, 1999; Tatler et al. 2006; Parkhurst and Niebur, 2003; Carmi and Itti, 2006). The ROC area values (mean=0.56; maximum=0.60) did not quite reach the ROC values as the ones obtained in scene-fixation studies, e.g. 0.63 (Tatler et al 2005) and 0.64 (Kienzle et al, 2009). 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 dark spot represents just a motion stimulus, after all it appears to be a looming stimulus.
Concluding the first experiment, it was shown that our noise display provided a strong degree of visual attraction as possibly experienced during free-viewing or moving, yet presenting it in the laboratory allowed us to place well defined targets and to measure orienting aspects as we had done in the 2nd experiment.
In a second experiment, a luminance target search was carried out. It revealed that the ROC area values for the target search were not higher despite the clearer objective of the search task: although observers had to learn the target by themselves from the first few trials, they all sensed that the target consisted of a bright, vertical bar. Thus, in this noise display, an average value of 0.57 may constitute an upper 'distinctness' boundary.
We determined the detection performance for fixed-amplitude 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 7 top graph). Peripheral detection performance for scaled-amplitude targets nearly reached the detection performance of fixed-amplitude targets; saccadic latencies as well as manual reaction times were even slightly shorter. The amount of saccadic undershoot and variable error was roughly twice as large as measured for simple displays (e.g. Findlay

and Gilchrist, 2003; Roos et al 2008). Only 10 percent corrective saccades were carried out, which in turn were preferably performed for target eccentricities ranging between 10 and 20 degrees. We suspect that during natural activities (free moving), this variability is even larger and that the number of corrective saccades is higher.
The presence of a constant and variable error can be interpreted as an inaccuracy or insufficiency of the saccadic spatial-orienting mechanism and hence a weakness, since it apparently does not allow to precisely place the gaze 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; Pelli et al. 2006). If any such translation independent 'recognition' took place in our experiments, then this may explain why the classification image is relatively blurred. Together with the tracker measurement error of ca. 0.5 degrees, the measured saccadic landing precision can vary by more than 1 degree for a 5-degree saccadic jump (0.5 + 0.18*5 = 1.3), which smears out fine structure in a classification image. This is in fact an issue which – to our knowledge - is never really discussed in scene fixation studies. In those studies, it is implicitly assumed that a fixation has landed precisely. Although saccadic orienting is likely more precise in static scenes than in movies, targets in static scenes are typically crowded and that may add to the uncertainty of saccadic target selection. But even if one assumed a very low error of 8% (the error for single saccades made in simple displays), then this adds up to an error of at least 0.9 degrees for a 5-degree amplitude. This inaccuracy makes it certainly difficult to specify what types of structures are fixated. But we believe that this orienting issue can be further discerned, by systematic testing on different image types (line drawings, grayscale scenes, movies) for different types of tasks (detection, recognition).

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 and Erik Groenewold for comments on an earlier version of 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] Carmi, R. & Itti, L. (2006). Visual causes versus correlates of attentional selection selection in dynamic scenes. *Vision Research, 46,* 4333-4345.

[6] 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.

[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] Henson, DB. (1979). Investigation into corrective saccadic eye movements for refixation amplitudes of 10 degrees and below. *Vision Res.* 19(1):57-61

[14] Hodgson, T.L. (2002). The location marker effect: saccade latency increases with target eccentricity. Experimental Brain Research, 145, 539-542.

[15] Itti, L. and Koch, C. (2001). Computational modelling of visual attention. Nature Reviews Neuroscience, 2(3):194--203.

[16] Kalesnykas, R.P. & Hallett, P.E. (1994). Retinal eccentricity and the latency of eye saccades. Vision Research, 34, 517-531.

[17] Kienzle, W., M. O. Franz, et al. (2009). "Center-surround patterns emerge as optimal predictors for human saccade targets." Journal of Vision **9(5):7**: 1-15.

[18] Land, M.F., Mennie, N. & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of everyday living. Perception, 28, 1311-1328.

[19] 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.
[20] Neri, P. & Heeger, D. J. (2002). Spatiotemporal mechanisms for detecting and identifying image features in human vision. Nature Neuroscience, 5, 812-816
[21] Parkhurst, D. J. & Niebur, E. (2003). Scene content selected by active vision. Spatial Vision, 16 (2), 125-154.
[22] Pelli, D.G., Burns, C.W., Farell, B., Moore-Page D.C. (2006). Feature detection and letter identification. *Vision Research*, 46, 4646-4674.
[23] 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.
[24] Rayner, K. (1998). Eye movements in reading and information processing. 20 years of research. *Psychological Bulletin, 124,* 372-422.
[25] Reinagel, P. & Zador, A. M. (1999). Natural Scene statistics at the centre of gaze. Network: Computation in Neural Systems, 10, 341-350.
[26] Rovamo, J. & Virsu, V. (1979). An estimation and application of the human cortical magnification factor. Exp Brain Research. 37(3); 495-510.
[27] 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.
[28] Schwartz, E.L. (1980). Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. *Vision Research, 20,* 645-669.
[29] 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.
[30] Simoncelli, E. P. & Olshausen, B. A. (2001). Natural image statistics and neural representation. Annual Review of Neuroscience, 24, 1193-1216.
[31] 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.
[32] 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.
[33] 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.
[34] Tavassoli, A., van der Linde, I., Bovik, A.C. and Cormack, L.K. (2007). An efficient technique for revealing visual search strategies with classification images. Perception & Psychophysics. 69 (1), 103-112.

[35] Vig E., Dorr, M. and Barth, E. (2009). Efficient visual coding and the predictability of eye movements on natural movies. Spatial Vision. In press.
[36] White, B.J., Stritzke, M. & Gegenfurtner, K.R. (2008). Saccadic facilitation in natural backgrounds. Current Biology, 18, 124-128.
[37] Wilson, H.R., Levi, D., Maffei, L., Rovamo, J. and Devalois, R. (1990). The perception of form: retina to striate cortex. In visual perception : the neurophysiological foundations (ed L. Spillman and J.S. Werner), pp. 231-72, Acadmeic Press, San Diego.

Figure 1. Image source for a noise movie with an example of a horizontal eye-position scan-path (blue) overlaid (fixations run vertical, saccades are the horizontal jumps). The source is a two-dimensional 1/f image *I(x,t)* and represents a space-time plot (width = 1200 pixels = 42.3 deg; 1 deg = 28.4 pixel). Each row is used as a source for a movie frame (10ms) and is stretched vertically (see figure 2). The white rectangles outline the selected *trigger patches* ($P_i(x,t)$; shown schematically only), centered around saccadic landing position (trigger patches close to the image border were omitted).

Figure 2. A sample frame of the noise movie and target generation. The 'bar code' represents one frame (duration = 10ms), whose luminance distribution is taken row-wise from the two-dimensional 1/f image as in figure 1. In a luminance-target search (experiment 2), a target was generated by adding a rectangular function (green) to the luminance profile of the bar code. This was done for a fixed pulse amplitude (*fixed-amplitude* condition), as well as for a scaled pulse amplitude, which depended on fixation eccentricity (*scaled-amplitude* condition) [red: eccentricity at present point in time].

Figure 3. The classification image for one observer for the free-viewing experiment (3823 fixations), the average of the white outlined trigger patches in figure 1. The white dot represents the minimum luminance value of the entire field. Y-Axis: Countdown to saccadic onset in milliseconds (0=saccadic onset). X-Axis: Eccentricity relative to saccadic landing position in degrees (0=saccadic landing position; minus and positive values represent less and more eccentricity). The total luminance level ranges from 0 to 1. Values that were within the range of the values from the random classification image (non-trigger) patches were set to 0.5.

Figure 4. Classification images of trigger patches for the free-viewing experiment for four observers (N=2839, 2823, 4527, 5037 fixations per observer respectively. **Left column**: Mean of trigger patches. The white dot represents the minimum (observer identification number given above left graph). **Right column**: The variance of trigger patches. Axes as in figure 3.

Figure 5. Classification images of trigger patches for a luminance target search. **Left column**: average. The plus sign denotes the location of the maximum. **Right column**: the variance of the trigger patches. Labeling as in previous figures.

Figure 6. Temporal cross-sections of the observer-averaged classification image (4-5 observers). X-Axis: Countdown to saccade in milliseconds (0=saccadic onset). Y-Axis: Luminance level (total range [0,1]). Free: free viewing (black). Lum: luminance target search (green). The free-viewing values were inverted for reason of comparison. The

functions near 0.5 are from the random patches. The dotted curves denote standard error of inter-observer performance. The vertically dotted straight line is placed at 100ms.

Figure 7. Target search for fixed-amplitude (dashed-dotted) and scaled-amplitude (solid) target condition. **Top plot**: manual reaction times (upper two traces) & saccadic latencies (lower two traces) in dependence of target eccentricity. **Bottom plot**: detection rate for targets which captured gaze and were signaled by a button press. Error bars = standard error of inter-observer performance (lacking error bars = 1 observer only).

Figure 8. Saccadic landing precision in dependence of target eccentricity (for the fixed-amplitude condition). **Upper left**: Average undershoot (constant error) for primary saccade [m=slope; c=intercept; r=correlation coefficient]. **Upper right**: Variability (variable error [standard deviation]) for primary saccades. **Lower right**: Average undershoot for secondary saccades. **Lower left**: Distribution of corrective (2^{nd}) saccades (ca. 10% in total). Error bars = standard error of inter-observer performance (lacking error bars = 1 observer only).

Figure 9: Saccadic landing precision in dependence of target eccentricity for the *scaled-amplitude* condition. Plots and labeling as in previous figure.

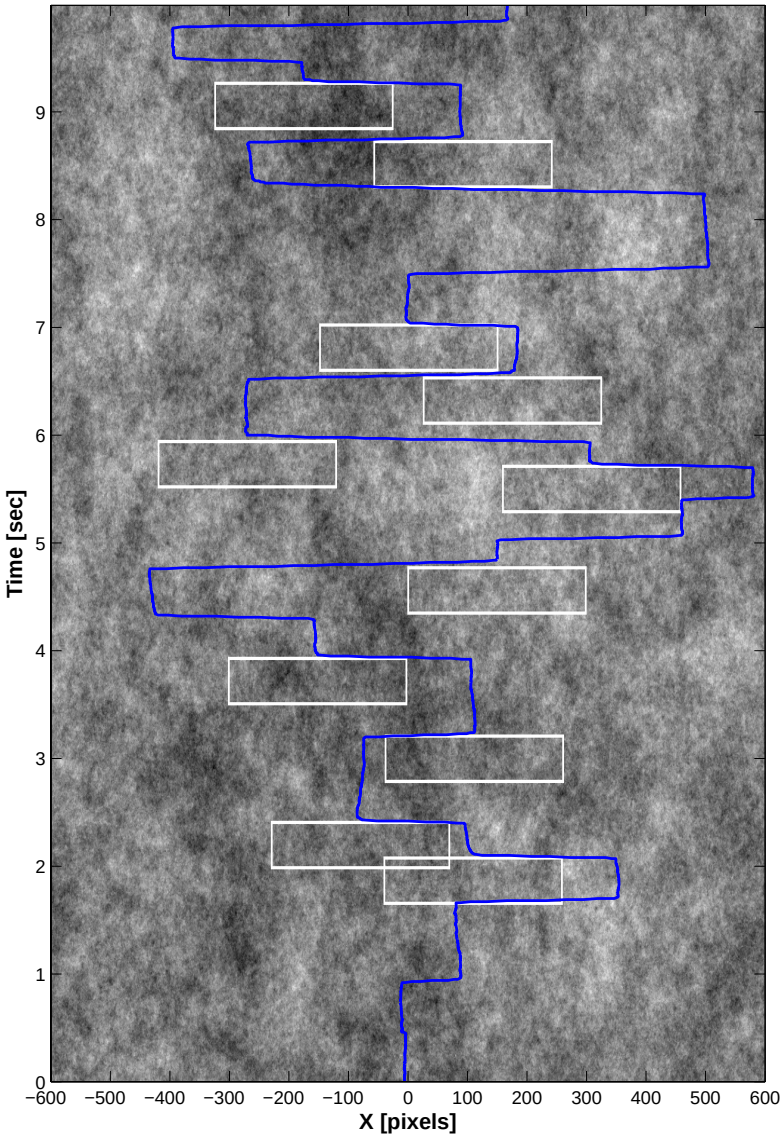
Time [sec]
X [pixels]
0
1
2
3
4
5
6
7
8
9
−600
−500
−400
−300
−200
−100
0
100
200
300
400
500
600

Figure 1

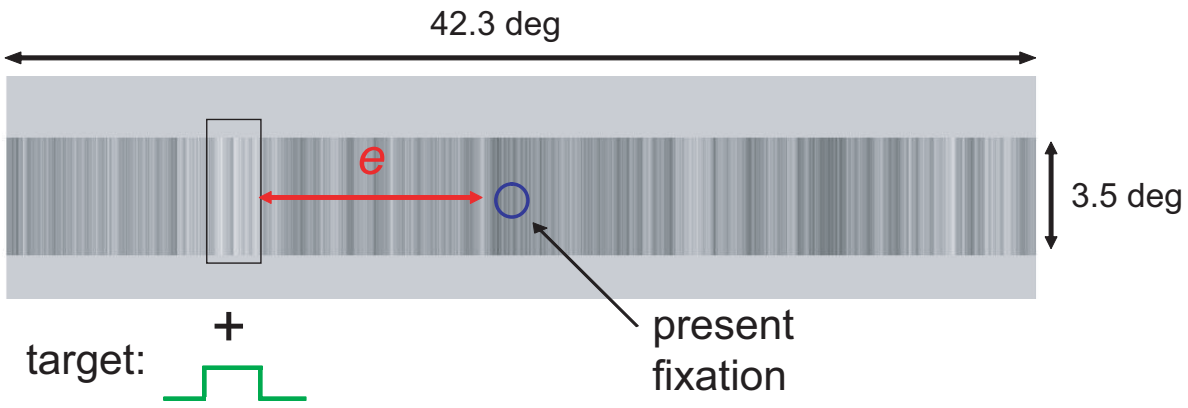

Figure 2

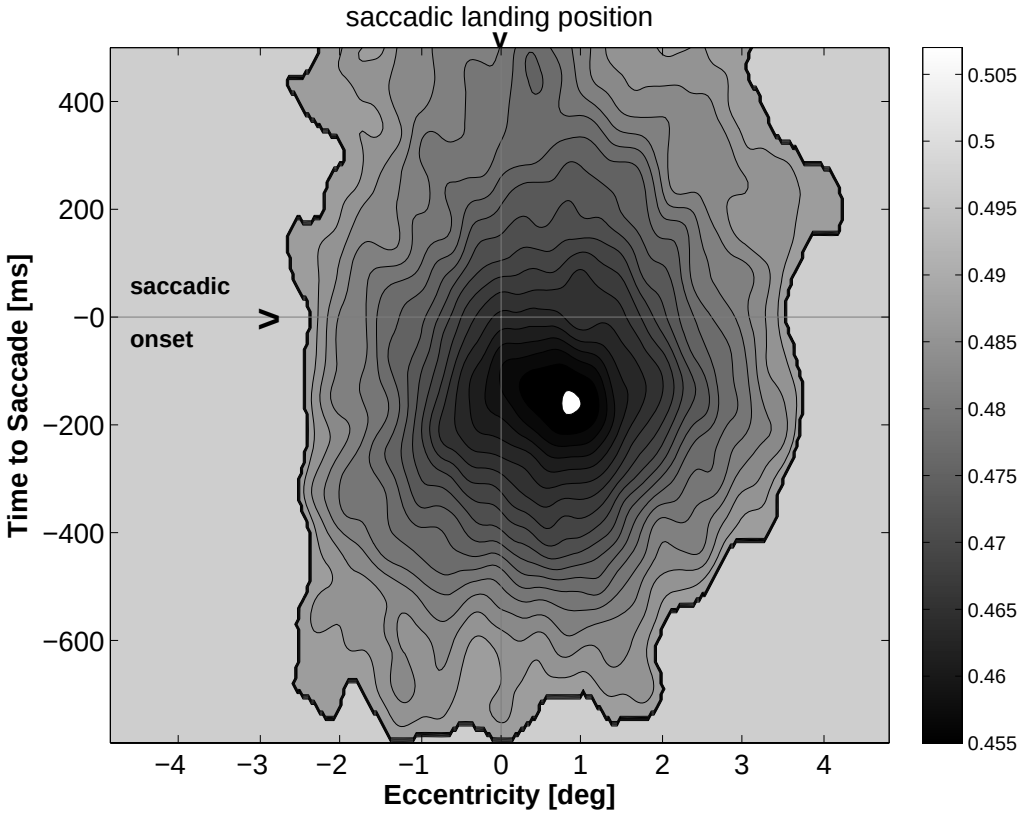
saccadic landing position
saccadic
onset
Time to Saccade [ms]
Eccentricity [deg]
400
200
−0
−200
−400
−600
−4
−3
−2
−1
0
1
2
3
4
0.505
0.5
0.495
0.49
0.485
0.48
0.475
0.47
0.465
0.46
0.455

Figure 3

A254RA

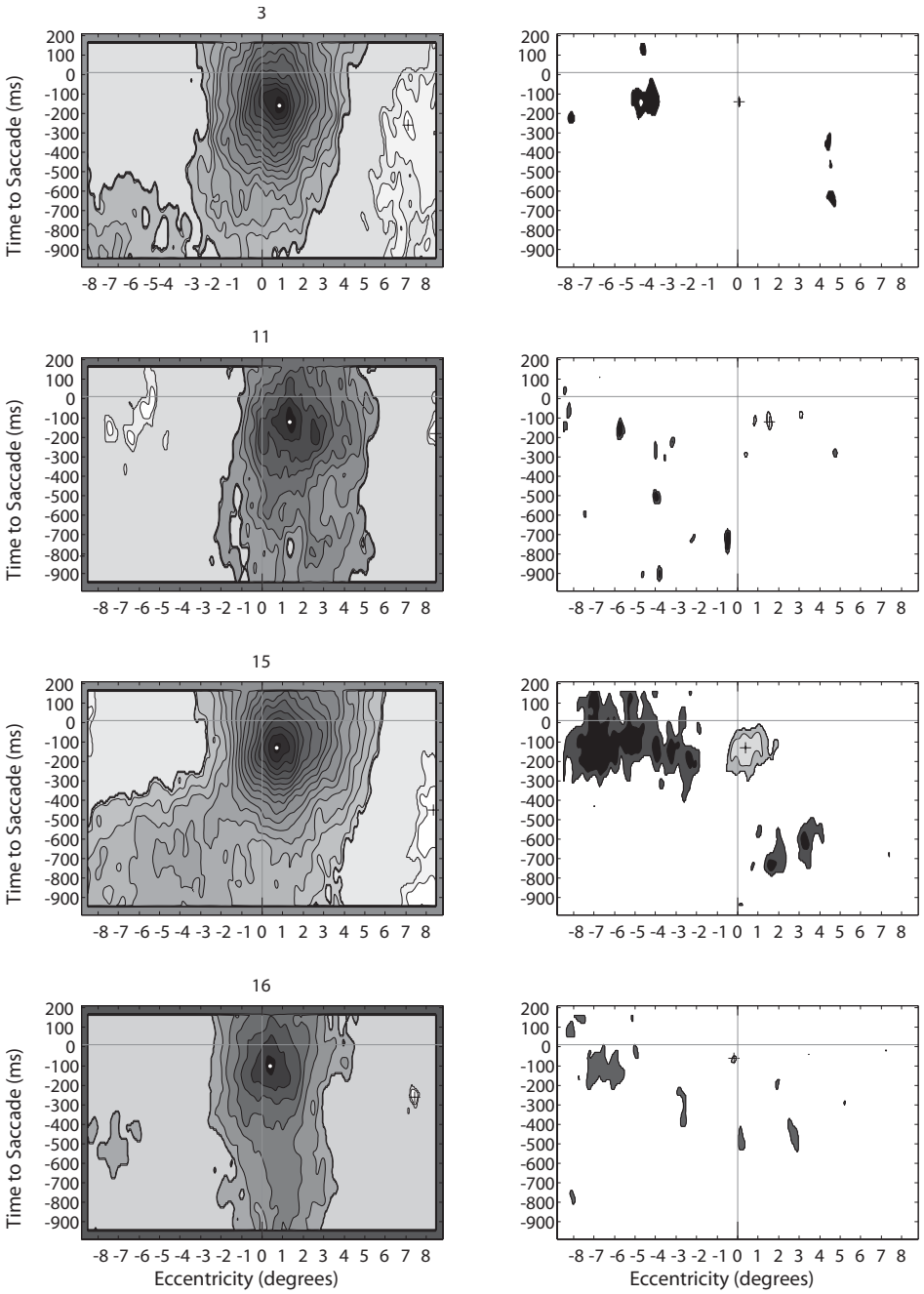

Figure 4

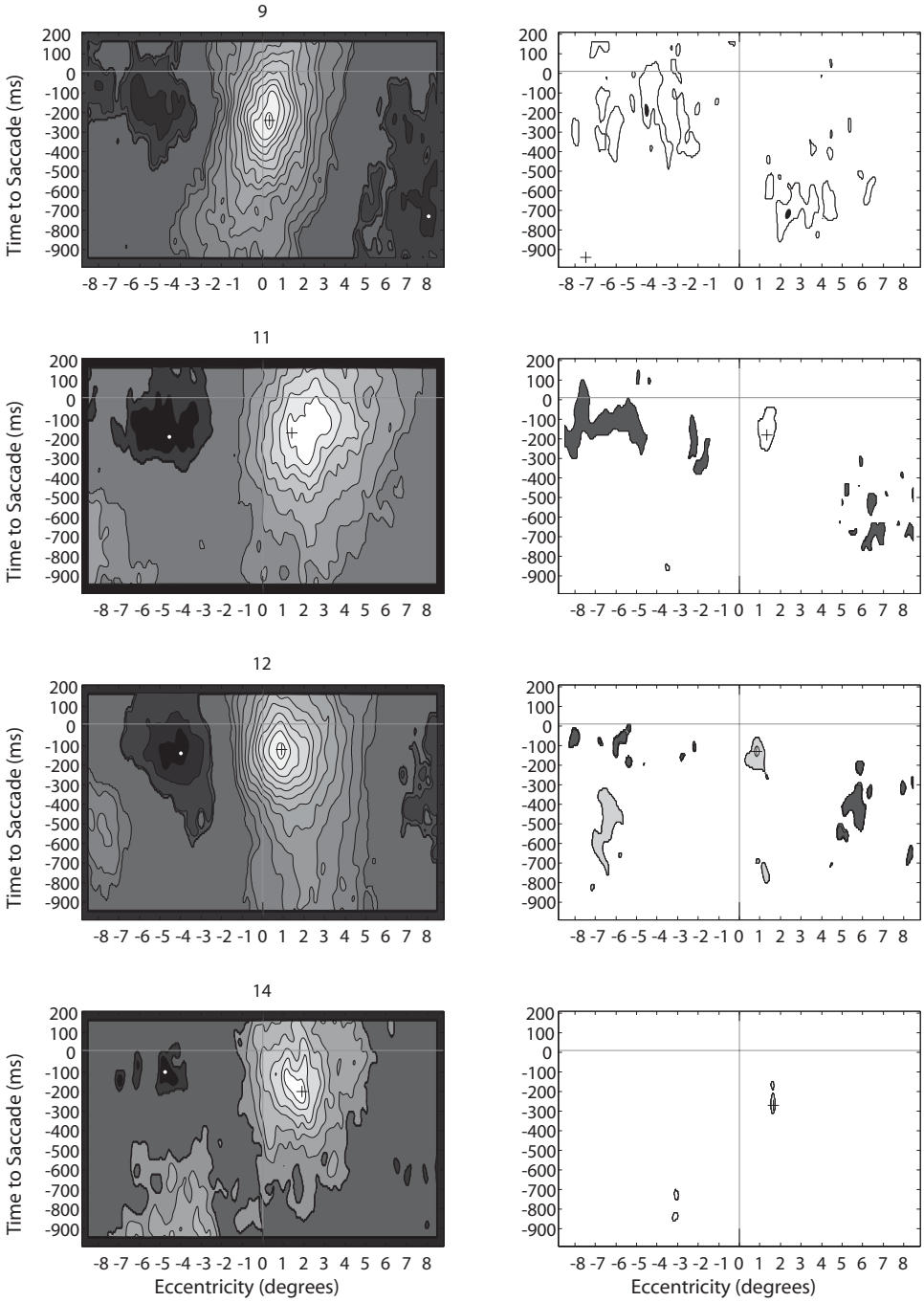
9
11
12
14
Time to Saccade (ms)
200
100
0
-100
-200
-300
-400
-500
-600
-700
-800
-900
-8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8
Eccentricity (degrees)

Figure 5

A254RA

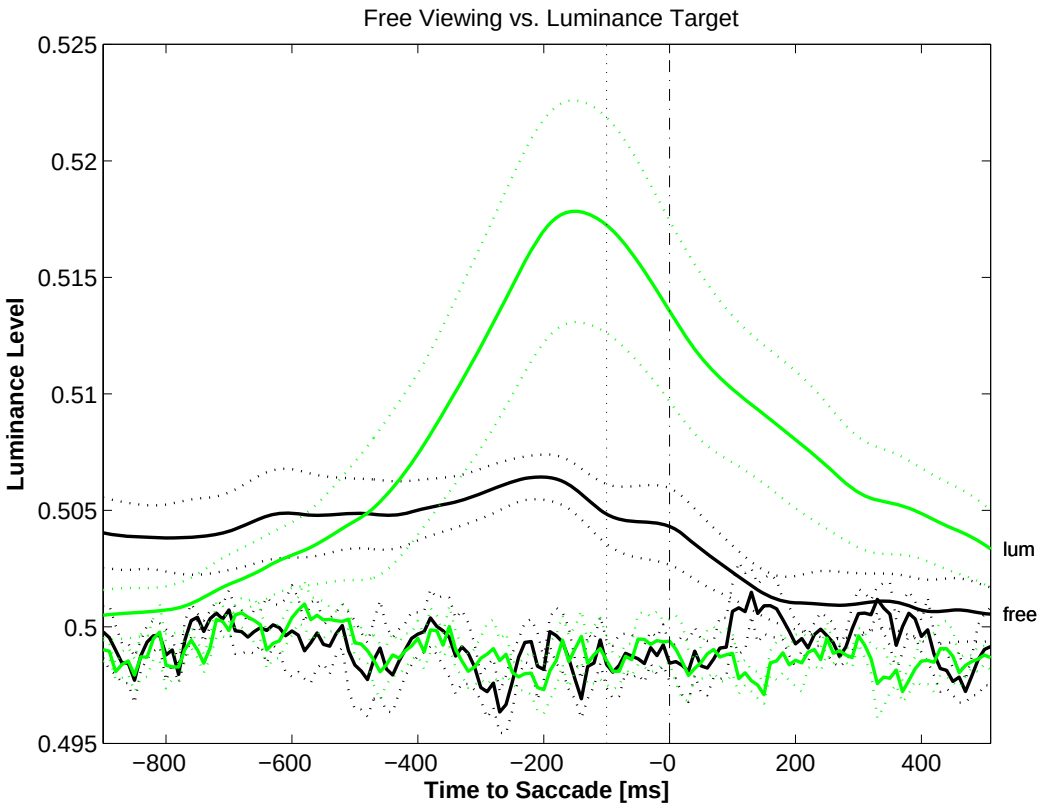

Figure 6

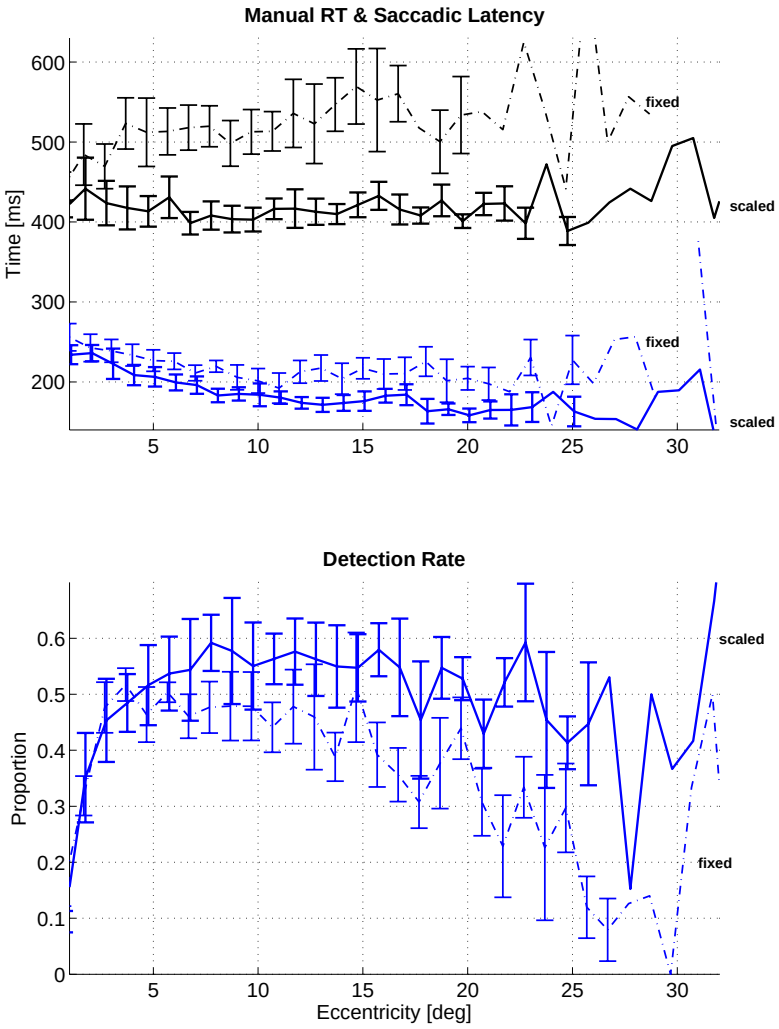
Manual RT & Saccadic Latency
Time [ms]
600
500
400
300
200
fixed
scaled
fixed
scaled
5
10
15
20
25
30
Detection Rate
Proportion
0
0.1
0.2
0.3
0.4
0.5
0.6
scaled
fixed
5
10
15
20
25
30
Eccentricity [deg]

Figure 7

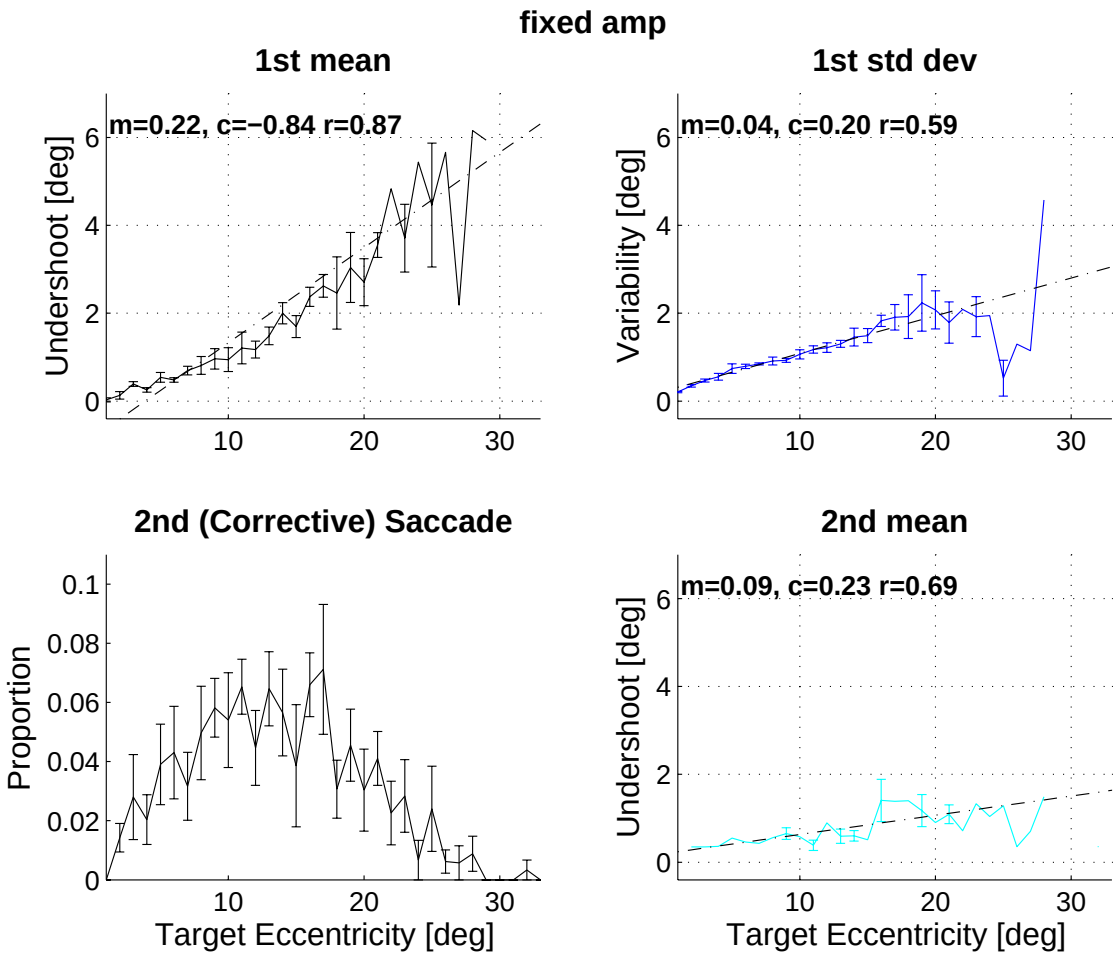
fixed amp
1st mean
m=0.22, c=−0.84 r=0.87
Undershoot [deg]
1st std dev
m=0.04, c=0.20 r=0.59
Variability [deg]
2nd (Corrective) Saccade
Proportion
Target Eccentricity [deg]
2nd mean
m=0.09, c=0.23 r=0.69
Undershoot [deg]
Target Eccentricity [deg]

Figure 8

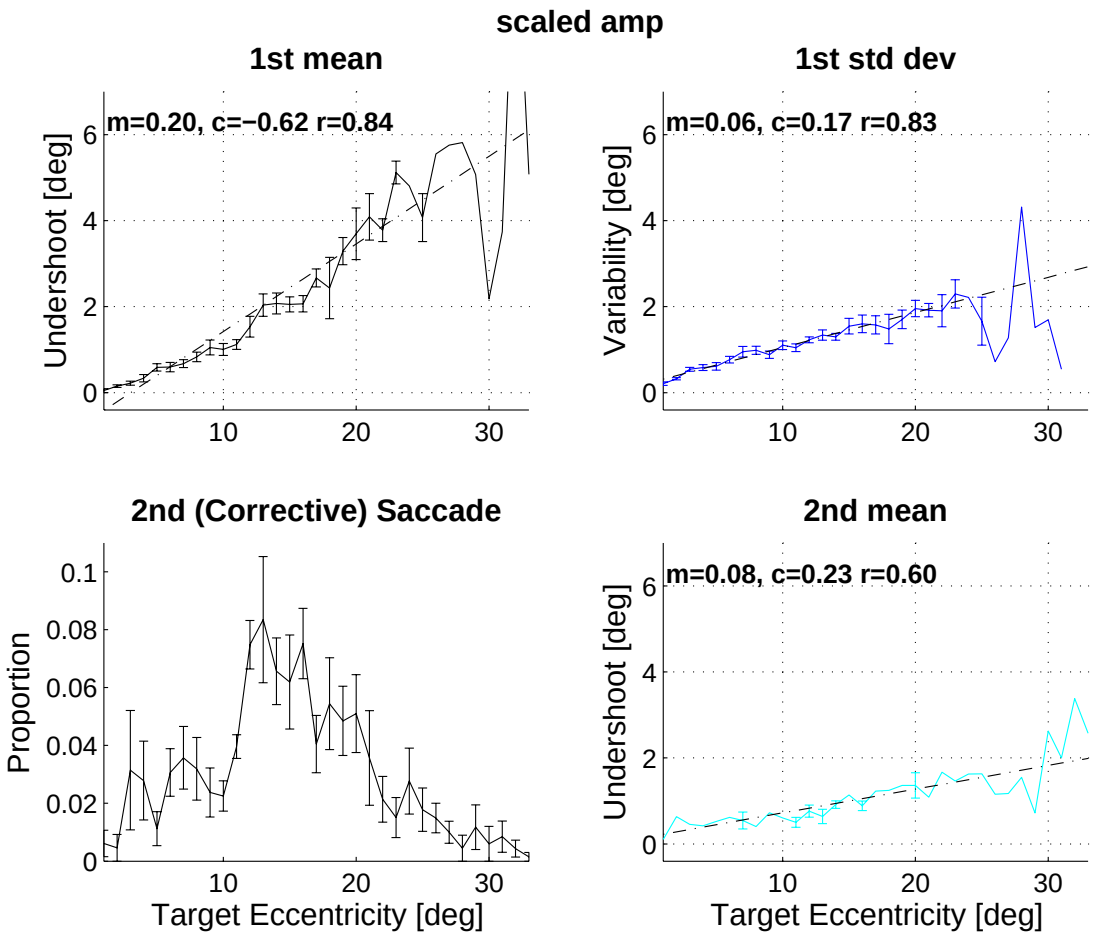
scaled amp
1st mean
m=0.20, c=−0.62 r=0.84
Undershoot [deg]
1st std dev
m=0.06, c=0.17 r=0.83
Variability [deg]
2nd (Corrective) Saccade
Proportion
Target Eccentricity [deg]
2nd mean
m=0.08, c=0.23 r=0.60
Undershoot [deg]
Target Eccentricity [deg]

Figure 9