VISUAL SEARCH FOR A TARGET AGAINST A $1/f^{\beta}$ CONTINOUS TEXTURED BACKGROUND

A.D.F. Clarke, P.R. Green, M.J. Chantler and K. Emrith

Corresponding author: Alasdair.clarke@macs.hw.ac.uk

ABSTRACT

We present synthetic surface textures as a novel class of stimuli for use in visual search experiments. Surface textures have certain advantages over both the arrays of abstract discrete items commonly used in search studies and photographs of natural scenes. In this study we investigate how changing the properties of the surface and target influence the difficulty of a search task. We present a comparison with Itti and Koch's saliency model and find that it fails to model human behaviour on these surfaces. In particular it does not respond to changes in orientation in the same manner as human observers.

INTRODUCTION

The mechanisms governing the deployment of attention to specific regions of the human visual field are an active area of research. One important method for analysing these mechanisms is the use of visual search tasks, in which observers search for a known target in a display. Until recently most research has investigated visual searches among sets of discrete geometric items (see Wolfe (1998) for a review). Theoretical models accounting for visual search performance have assumed that attention is controlled in two main ways; through bottom-up processes that operate on image data, and through top-down processes that draw on higher-level factors such as knowledge of target characteristics or learned search strategies. An example is Wolfe's (1994) influential Guided Search Model, in which image data are used to create basic feature maps (size, orientation, colour, etc), which are then combined to create an activation map. However, before the feature maps are combined they are modulated by top down information about the target: for example if the observer knows that the target is red then the red feature map is emphasised over other colours. A number of variations of this approach have been proposed (see Navalpakkam & Itti, 2005, 2007; Rutishauser & Koch, 2007 for some recent examples).

A limitation to this approach is that image features have to be defined in terms of the properties that distinguish individual items in search displays and it cannot be easily extended to images of natural scenes. In these, even simple low level features such as local contrast, colour and orientation can be measured in many different ways. A recent study by Pomplun (2006) investigated how top down knowledge about the target might influence visual search

in photographs of natural scenes. On each trial, participants were shown a different target region to find in a photograph. Intensity, contrast, orientation and spatial frequency features were constructed and the target's features were compared to the image region fixated. In the case of intensity, they found evidence for top-down guidance; image regions with similar intensity to the target attracted more fixations. For contrast, however, target contrast did not have a significant effect on fixations while display contrast did, providing evidence for a strong bottom-up effect, with more fixations attracted to high contrast locations regardless of the target's contrast. A similar bottom-up effect was also found for the spatial frequency and orientation features.

Models of bottom-up processes involved in controlling human fixation patterns have been influenced by the concept of a saliency map, introduced by Koch and Ullman (1985) and further developed by Itti, Koch & Niebur (1998) and Itti & Koch (2000). The model assumes that image regions with high local contrast, and with local orientations and colours which differ from their surrounding area, attract our attention. Feature maps are generated for a range of resolutions and are weighted in such a way that maps with a small number of strong responses are favoured over maps with a large number of small responses. The resulting *conspicuity maps* are normalised, weighted again, and combined resulting in a *saliency map*.

A number of studies have tested the performance of the saliency model on photographic images of natural scenes and have found correlations with human fixation patterns (Itti & Koch, 2000; Parkhurst, Law, & Niebur, 2002; Peters, Iyer, Itti & Koch, 2005; Parkhurst & Neibur, 2004). However, other work have found a poor match between human fixation and image statistics such as contrast (Einhauser & Konig, 2003; Tatler, Baddeley & Gilchrist, 2005; Henderson, Brockmole, Castelhano & Mack, 2007; Tatler, 2007). The general conclusion is that saliency maps often do not provide a good match to human gaze behaviour. Even where there is a correlation between regions with high salience (as defined by the model) and regions which attact fixations from human observers, this is not neccessarily due to salience, as there is frequently a correlation between regions of high salience and regions containing semantic information likely to require top-down processes for their identification. Furthermore, gaze patterns are known to be very task dependant (see Neider & Zelinsky, 2006b for an example). Theoretical criticisms of saliency models (Baddeley & Tatler, 2006; Vincent, Troscianko & Gilchrist, 2007) have argued that their internal structure is only loosely based on biological evidence and that most of the design choices are fairly arbitary.

In this paper we will present results from a novel method of investigating the control of attention, by using a task that eliminates the possibility of top-down influences on visual search, isolating bottom-up processes so that they can be tested against theoretical models. We achieve this by using synthetic *surface textures*, which have a natural appearance and yet have precisely controlled properties. These surface textures are produced by using $1/f^{\beta}$ -noise to model *the height function* of surfaces. The process of $1/f^{\beta}$-noise occurs frequently in nature and provides a good approximation to the power spectra of many images of natural scenes (Field, 1987; Voss, 1988; van der Schaaf & van Hateren, 1996; Balboa & Grzywacz, 2003). It is important to emphasise that we do not create textured images directly from $1/f^{\beta}$ -noise, as has been done in other studies (e.g. Rajashekar, Cormack, & Bovil, 2002; Kayser, Nielsen & Logothetis, 2006) but instead create a height function which is then rendered

using a lighting model that implements Lambert's Cosine Law (Chantler, 1995, Padilla et al, in press). See Figures 1a and 1b for an example of a height map and the corresponding rendered image. This technique also differs significantly from the methods used by Einhäuser, Rutishauser, Frady, Nadler, König & Koch (2006) and Wichmann, Braun & Gegenfurtner (2006), where random noise is added to the phase spectrum to obscure the contents of a photograph. Our stimuli are constructed with random phase, which is one of their defining characteristics.

A target for a visual search task can be made in these images of synthetic surfaces by introducing an anomaly such as a small indent to the surface. The task of identifying this target can be made easier or harder by varying its size and shape or changing parameters of the underlying surface that control its perceived roughness. These surfaces have several advantages over the types of stimuli that have previously been used in search experiments. Unlike a typical visual search display, they look like natural surfaces such as dressed stone or rough plaster (see Figures 1b and 1c). Unlike photographs, their statistical properties are fully controllable, and the absence of semantic cues makes it possible to test bottom-up processes in isolation. In general, the stimuli can be thought of as bridging the gap between the controllable yet abstract stimuli used in conventional visual search displays, and realistic but uncontrollable photographs of natural scenes. A study by Henderson, Larson, Zhu & David (2007) used three categories of photographic stimuli (objects, close up scenes and finally full scenes where the 3D geometry of the surrounding space can be determined) and the surfaces introduced in this study could be considered as a fourth class in which the surfaces of objects are depicted at a finer-grained level.

While the visual search task used here involves search against a continuous background texture, it contrasts with two other recent studies that use complex backgrounds in visual search tasks (Wolfe, Oliva, Horowitz, Butcher & Bompas, 2002; Neider & Zelinsky, 2006a). Wolfe et al (2002) investigated how the addition of a complex background affected reaction time vs. set size slopes. They concluded that a complex background might slow the accumulation of information in the object identification stage, perhaps because the search items were not cleanly segmented from their surrounding backgrounds in the initial object segmentation phases. Only in Wolfe et al's final experiment, when the search items and background were designed to be very similar to each other was an increase in search slopes observed. Neider & Zelinsky carried on this line of work with a series of experiments using more complex stimuli designed to investigate the effect of target-background similarity (TBS) (Neider & Zelinsky, 2006a). They used photographs of children's toys as search items and constructed "camouflage backgrounds" from the target item by tiling an $n \times n$ pixel patch from the target item. By increasing n, the TBS can be modified while leaving the distracter-background similarity constant. They carried out a series of eye tracking experiments but failed to find any conclusive results or pattern behind the gaze patterns. In the experiments described here, the lack of any distracter items in the stimuli creates a major difference from these two studies, which prevents any straightforward comparisons of the results.

In the present study, we carried out three experiments designed to examine how observers perform in a simple search task where the presence or absence of a target against a continuously textured background must be reported. In Experiment 1, we investigate

effects on search performance of manipulating parameters of the background surface, while in Experiments 2 and 3 we vary the depth and the orientation of the target. Finally we compare the experimental data with the performance of a saliency model of bottom-up control of attention (Itti & Koch 2000).

METHODS

STIMULI

Much work has been done in computer graphics investigating different methods of generating realistic looking synthetic textures. For the experiments in this paper a very simple model referred to as *1/f*$^{\beta}$-noise will be used[1]. We can represent a surface by an *n* x *n* matrix *h*. This matrix is referred to as a height map and for any $(x, y) \in \{Z \times Z \mid 0 < x, y < n\}$, $h(x, y) = z$ gives us the height of the surface. The 1/f$^{\beta}$ noise has only two parameters: the frequency roll-off, β, and the RMS-roughness, σ_{RMS}. The RMS-roughness, σ_{RMS}, is the standard deviation of the surface's height map and acts as a scaling factor in the z axis. The roll off factor β controls the amount of high frequency information in the surface: increasing β causes the high frequencies to drop off more quickly, so the texture appears smoother (Padilla et al, in press). Note that we use β here to denote the roll-off of the magnitude of the inverse discrete Fourier transform of the height map. The same term and symbol β also sometimes refers to the roll off factor in the power spectrum of an image. See Chantler, Petrou, Penirschke, Schmidt & McGunnigle (2005) for a model relating these two parameters. In this paper, we use the word "roughness" to refer to both β and RMS-roughness, σ_{RMS}.

The surface is generated in the Fourier domain with the magnitude spectrum given by:

$$S(u,v) = \frac{k}{\left(\sqrt{u^2 + v^2}\right)^{\beta}}$$

where *k* is the scaling factor required to give us the desired σ_{RMS}, the RMS roughness of the zero mean surface. The phase spectrum is randomised and by using different values to seed the random number generator we can create many different surfaces with the same properties. Taking the inverse discrete Fourier transform of the magnitude and phase spectra gives us our height map h.

The two dimensional height map that represents our surface texture is then rendered to generate an image of the surface under specified illumination. This stage is important, as a surface can have drastically different appearances under different light conditions (Chantler, 1995). We will use one of the simplest models, known as Lambert's Cosine Law. This treats the surface as a perfectly diffuse reflector, i.e. it reflects the same amount of light in all directions. It is easily modelled by the dot product:

[1] A Matlab .m file containing the texture synthesis model can be found at [insert URL here].

$$i(x,y) = \lambda \cdot \rho(x,y)\underline{n}(x,y) \cdot \underline{l}$$

Where *i* is the image we are creating, $\underline{n}$ is the unit surface normal to the height map and $\underline{l}$ is the unit illumination vector. The albedo, ρ, determines how much light is reflected by the surface. The strength of the source light is denoted by λ.

The surface normal $\underline{n}$ is estimated by taking:

$$p(x,y) = h(x,y) - h(x-1,y) \qquad q(x,y) = h(x,y) - h(x,y-1)$$

$$\underline{n} = \frac{1}{\sqrt{1+p^2+q^2}}[p,q,1]^T$$

Self shadowing occurs when $i(x,y) \leq 0$ i.e. the surface is orientated such that its normal makes an obtuse angle with the illumination vector. Self shadowing is implemented by setting all negative values to 0. Cast shadows are not implemented in this model.

The illumination conditions will be kept constant throughout this paper with elevation = 60°, azimuth = 90° and the strength of the source light being set at 150 cd/m^2. The albedo value will be kept constant at 0.8 throughout all the experiments, which is approximately the value of matte white paint.

To use rendered surfaces in visual search experiments, we need to choose a target. Rather than using a target with an artificial appearance, we create an anomaly in the surface texture, in the form of a small pothole in the surface. We create these targets by subtracting the lower half of a small three dimensional ellipsoid from the surface.

Our ellipsoid defect is defined by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

To make the indent appear more realistic it is convolved with a two dimensional smoothing filter to soften the hard vertical edges (see Figure 2 for some examples).

OBSERVERS

Five observers were used for each experiment: all had normal or corrected to normal vision and were between 21 and 30 years old. Observers were given several practice trials on which the target was present, and were told that the target would be present on half the trials and would always be an indent in the surface of the same size and shape. They were instructed to decide whether the target was present or not and to respond with a key press for target present or absent as quickly and accurately as they could. No time limit was imposed on the task. The first two experiments comprised of 300 trials while Experiment 3 had 240 trials. Observers were allowed to take a short rest every hundred trials (120 trials for Experiment 3).

EXPERIMENTAL SET-UP

Stimulus presentation was controlled by Clearview (Tobii Technology Inc). All stimuli were 1024x1024 pixels in size and displayed on a calibrated NEC LCD2090UXi monitor. The pixel dimensions were 0.255mm by 0.255mm resulting in images with physical dimensions 261.12mm by 262.12mm. The monitor was linearly calibrated, gamma=1, with a Gretag-MacBeth Eye-One with the maximum luminance set at 120cd/m^2. This results in the rendered images appearing as if they were being lit under bright room lighting conditions.

A Tobii x50 eye-tracker was used to record observers' gaze patterns. The fixation filter was set to count only those fixations lasting longer than 100ms within an area of 30 pixels. The accuracy of the eye-tracker was 0.5^0-0.7^0 and the spatial resolution was 0.35^o.

The viewing distance was controlled by use of a chin rest placed 0.87m away from the display monitor. At this distance, one pixel is approximately 1' of visual angle and images were 16.7^0 across. A target was added to half the images at a random location between 6° and 7.5° from the centre. The targets subtended 0.66° of visual angle.

EXPERIMENT 1 – SURFACE ROUGHNESS

This first experiment was designed to investigate how surface roughness influences the difficulty of a simple visual search task (illustrated in Figure 2). The target was a small ellipsoidal indent in a $1/f^{\beta}$-noise surface. Five values of β and three of σ_{RMS} were used to produce a range of surfaces from smooth to rough. For each value of β ten surfaces were created, each one being scaled to three different values of σ_{RMS}, giving a total of 150 surfaces. Each surface was then used twice, once with a target added and once without, resulting in 300 trials. The target was an ellipsoid with $a = b = 10$, $c = 2$, and was subtracted from the surface texture so that it created a hole with volume 10mm^3. Our hypothesis is that as surface roughness increases reaction times and the number of fixations taken to find the target will increase and accuracy in detecting targets will decrease. In addition to this we will also investigate whether a fixation is required to identify the target or not, and how this relationship is influenced by changing the roughness of the background.

RESULTS

Overall, observers' accuracy was high, and in the target absent trials 99.5% of responses were correct. This suggests that the search target was well defined and easily identifiable: observers had no trouble in rejecting background patches. The few false positives that did occur can likely be attributed to observers pressing the wrong response key. There was no indication that increasing surface roughness had any effect on the number of false positives. Table 1 shows overall accuracy for each observer on the target present trials, and Figure 3 the effect of the two surface roughness parameters on accuracy in these trials. A two way repeated measures ANOVA gives significant effects ($p<0.05$) of β, ($F(4,16) = 79$), σ_{RMS}, ($F(2,8) = 58$), and the interaction ($F(8,32) = 13$).

Figure 4 shows the individual and mean reaction time data from target present trials on which the response was correct. The pattern of variation between individuals suggests dif-

ferent speed/accuracy tradeoffs: comparing the graphs below with Table 1 shows that observer 1 (GW) was the slowest but the (joint) most accurate (12.67% of targets missed) while observers 2 and 3 (HW and LM) were the fastest and also missed a greater number of targets (22% and 19.33). Despite these differences, all observers were affected by surface roughness in the same way, with longer reaction times when searching on rougher surfaces. A two-way repeated measures ANOVA gives a significant effect ($p<0.05$) of β ($F(4,16) = 8.8$), σ_{RMS}, ($F(2,8) = 9.0$) and the interaction, ($F(8,32) = 4.5$). The relationship between the number of fixations made on each trial and the two parameters of surface roughness is shown in Figure 5. The effects of both variables and their interaction are significant ($p < 0.05$; β, ($F(4,16) = 8.1$; σ_{RMS}, $F(2,8) = 8.1$; interaction, $F(8,32) = 4.4$). The implication that most variance in reaction time is due to variance in number of fixations, rather than duration, is confirmed by significant correlations between reaction time and number of fixations on each trial (values of r for individual observers range from 0.899 to 0.971, all $p < 0.0001$).

Do observers have to fixate on the target to be able to identify it? In order to investigate this, we looked at the distance on each trial from the target to the centre of the fixation when the response key was pressed. Because it is not possible to define exactly the time at which the decision to press the key is made, we defined 'final fixation to target distance' as the distance from the target to either the fixation during which the response key was pressed, or the fixation before it, whichever was the smaller. This criterion allowed for some variation in the time between the decision to respond and initiation of a saccade away from the target. Specifically, it meant that when an observer made a saccade away from the target after fixating it but before responding with a key press, the shorter distance was counted provided that the response occurred during the next fixation. Two trials in which the response key was pressed several saccades after the target was fixated were removed from the analysis.

Figure 6a shows the distribution of final fixation to target distances over all trials and observers. It appears that the distances fit a bi-modal distribution. The majority of trials, 82%, have a final fixation to target distance of 1° or less. There also appears to be a second, smaller set of trials with a larger final fixation to target distance. These account for 5% of all correct target present trials and appear to fit a Gaussian distribution with a mean of approximately 6°. In these trials, the target was therefore identified without fixation. Figure 6b shows how the mean final fixation to target distance changes with surface roughness. A two-way repeated measures ANOVA gives a significant effect only of β ($F(4,16) = 3.05$, $p = 0.048$). As β increases, and the surface appears less rough, mean distance from final fixation to target increases, as identification without fixation becomes more frequent. The lack of an effect of σ_{RMS} is probably due to a lack of data for the rougher surfaces, where the proportion of correct responses is small).

The large majority, 82%, of correct responses occurred when fixating within 1° of the target. Considering the trials on which the target was present but missed, we can use the figure of 1° as a criterion to determine how often the target was fixated but not identified. This happened 25 times, accounting for 20% of the target missed trials.

DISCUSSION AND CONCLUSIONS

The effects of roughness on visual search performance in Experiment 1, as measured both by accuracy and reaction time, are closely similar to the effects of set size in conventional visual search tasks using arrays of discrete items. When $\sigma_{RMS} = 0.8$, the reaction time vs. β slope is near horizontal implying that search is efficient. As σ_{RMS} increases the magnitude of the gradient increases implying that search is less efficient. At the rough end of the range, the task became very difficult with target hit rates far lower than those commonly encountered in visual search tasks (Figure 3).

We therefore conclude that it is possible to change the parameters of these continuous, synthetic surface textures in ways that have systematic effects on ability to identify a small anomalous region in the surface. The very small number of false positives recorded in the experiment indicates that observers did not have any trouble in identifying the target once they fixated it; rather, the increase in difficulty with rough surfaces came from an inability to identify the target pre-attentively based on the contrast information present. Observers have to switch from using pre-attentive vision to carrying out some sort of serial search strategy, leading to an increase in both the mean number of fixations and the variation (Figure 5).

Analysis of distances between target and fixation when targets are identified demonstrates two patterns; on the majority of trials, fixation is within 1^0 of the target when it is recognised, but on others it falls in a higher range centred around 6^0, indicating recognition of the target in peripheral vision. There is some evidence that the second pattern is more common when surfaces are smoother, which would be expected as the demands placed by the task on acuity of visual processing will be lower on smoother surfaces. Fixation of a target does not ensure that it will be recognised; on 20% of trials when the target was missed, the target was fixated, but not detected, at some point during the search.

EXPERIMENT 2 – TARGET DEPTH/CONTRAST

Experiment 1 investigated the effects of varying properties of the background surface on visual search. In Experiment 2 we go on to consider the effects of changing a property of the target, its depth. As the depth of the target is reduced, the contrast at its edges created by the illumination process decreases, and we would therefore predict that it will become harder to find. To vary target depth, we first created a target in the same way as in the previous experiment, and then reduced its depth by a scaling factor, z_k. Setting z_k equal to 1 gives the same depth (and hence level of contrast) used in Experiment 1.

Pilot studies showed that people had difficulty in identifying the target for $z_k = 0.5$, even when its location was known. Therefore, the following values of z_k were used: 0.6, 0.7, 0.8, 0.9 and 1.0. The target was placed on a subset of images from Experiment 1: β=1.6, 1.65, 1.7 and σ_{RMS} =1.0. These values were chosen as they give a range of roughness over which target detection is neither too hard nor too easy. For each value of β, ten surfaces were created and each surface was used ten times for each of the 5 values of z_k, once with a target and once without. Target locations were determined in the same way as in Experiment 1.

RESULTS AND DISCUSSION

Accuracy of target detection fell as the target was made shallower, to the extent that when z_k was 0.6 or 0.7 the level of accuracy fell considerably below those found in Experiment 1 (see Figure 7a). Both surface roughness β and target depth z_k have significant effects on accuracy (repeated measures ANOVA: $(F(2,8)=89.5, p<0.05$ for β; $F(4,16)=146.6, p<0.05$ for z_k; $F(8,32)=3.5, p<0.05$ for the interaction). Because there are very few correct target present trials for z_k =0.6 and 0.7, reaction times and numbers of fixations are unreliable measures for these cases. The reaction times for z_k =1, 0.9 and 0.8 only are therefore shown in Figure 7b. Over this range, surface roughness and target depth both have significant effects on accuracy (repeated measures ANOVA: $(F(2,8)=7.0, p=0.049$ for β; $F(2,8)=10.3, p = 0.026$ for z_k; $F(2,16)=2.9$, N.S. for the interaction). As in Experiment 1, the results for numbers of fixations follow a similar pattern to reaction times.

EXPERIMENT 3 – TARGET ORIENTATION

In this experiment we modify the target in a different way, making it elliptical and varying its orientation relative to the simulated illumination of the surface used in the rendering process. As an elongated target is rotated, the angle that its long axis makes with the incoming light varies; resulting in variation in the contrast at is edges (see illustration in Figure 8). We would therefore predict that the target will become harder to detect the closer its orientation to that of the illumination.

The target used in this experiment was an ellipse with axes subtending approximately 0.7° by 0.2°. The volume of the indent, its location and the illumination conditions were the same as in Experiment 1. Unlike the previous two experiments, the roughness parameters were kept constant (β and σ_{RMS} were 1.65 and 1.0 respectively). The variable in this experiment was the orientation of the target. 12 orientations were used, (0°, 45°, 60°, 70°, 80°, 85°, 90°, 95°, 100°, 110°, 120°, 135°), with 10 trials for each value. All target orientations are measured from the horizontal. We chose to include more trials with angles close to 90° (the direction of illumination) as these are harder to find (see Figure 8). 120 target-absent trials were included, giving a total of 240 trials.

RESULTS AND DISCUSSION

The relationships between target orientation and both reaction time and accuracy are shown in Figures 9a and 9b respectively. It is clear that there is a sharp drop in accuracy rates as target orientation approaches vertical, and all observers found the search task very difficult for targets orientated at 90°±5°. Again the number of fixations per trial followed a similar pattern to the reaction times.

As we would expect, target detection is hardest when it is oriented parallel to the illumination, but is important to note that the effect is not linear. Instead, there is a narrow band in which orientation has a strong effect on search performance. In the next section, we will seek to establish whether this effect can be modelled in the same way as the results from Experiment 2, simply as an effect of decreasing contrast at target edges.

COMPARING HUMAN AND MODEL DATA

We will now explore the performance of Itti and Koch's (2000) saliency model when presented with the stimuli used in the experiments above. Given that the surface textures have no local semantic information, high level mechanisms should play a very limited role in visual search. If the model gives similar results to humans over the range of parameters explored above, then that would be grounds for classing it as a good model of the low level processes involved in the control of human attention. However, if it fails to give an approximation to human data with these stimuli, then there seems little point in using it to model bottom-up processes in more complex stimuli such as photographs.

METHODS

The version of Itti and Koch's (2000) model that we use is a simplified variant of the iLab Neuromorphic Vision C++ Toolkit developed in Matlab by Walther and Koch (2006), with only minor changes made to the parameters. The parameters that are changed are the ones that specify which resolutions to work at in the Gaussian pyramid. Since the model was originally designed and tested on photographs containing macroscopic objects the resolution settings are quite low, i.e. the image is blurred and reduced in size a lot. While this works well with photographs (where we measure average contrast over fairly large areas) the stimuli used in the following experiments contain a lot of very fine, high frequency information. In order to accommodate this level of detail we have changed the parameters:

params.minLevel = 1; params.maxLevel = 4; params.minDelta = 1; params.maxDelta = 3; params.mapLevel = 2;

We will use the same method of comparison used by Itti and Koch (2000), and consider the number of fixations required to find the target. A maximum limit for the number of fixations was set at 20: this was the mean number of fixations on the target absent trials in the experiments. This provides a measure of how many fixations a human is prepared to make before giving up a search and making a negative response, and allows us to express the model's accuracy as the proportion of trials on which it fixates the target before the maximum number of fixations is reached.

Comparing the number of fixations made by model and human observers is reliable as long as accuracy rates are high, as in Experiment 1. Where they are lower, in Experiments 2 and 3, we also use accuracy rate (the proportion of trials on which an observer finds a target, or the model fixates it within the maximum number of fixations) as a comparison.

A common means of comparing human fixation data with model predictions is to compare fixation locations (e.g. Parkhurst et al., 2002; Peters et al., 2005; Tatler et al., 2005; Tatler, 2007). This method is not appropriate with the stimuli used here, because the statistics of the image vary little across the background and are only distinct at the target. There would therefore be no reason to expect any correlation in the locations of fixations on the background texture made by observers and by the model.

RESULTS

In Figure 10 the mean number of fixations needed to find the target by humans in Experiment 1 is compared with the number needed by the saliency model. Overall the model outperforms human observers, finding the target in fewer fixations than human subjects. This corresponds to the results reported by Itti and Koch (2000). The graph shows that both humans and the model respond to increasing roughness in a similar way: more fixations are required to find the target on a rougher surface than on a smoother one. A similar comparison between the data from Experiment 2 and model performance is shown in Figure 11. Due to low human accuracy with the shallowest surfaces, numbers of fixations to target detection are compared only for z_k = 0.8-1.0 (Figure 11a) while the accuracy data for the whole range is shown in Figure 11b. In this experiment the model does not fit the empirical data as closely as in the first experiment; nevertheless the model and the human observers follow a similar pattern as the depth of the target is reduced. The results from the final experiment, with the elongated target, are shown in figure 12. The model does markedly less well than humans when the target is close to the vertical. Figure 12b shows that the drop in the model's performance is not only larger but also occurs over a wider band of orientations, 90°± 20° (see Figure 12b). In fact, the model performs so badly in this task that, at one point, humans are outperforming the model by up to 60%, a far larger discrepancy than seen in the previous experiments.

DISCUSSION

To our knowledge, existing computational models of search cannot be easily applied to these stimuli. Two, Pomplun's (2003) Area Activation Model and Rutishauser & Koch's (2007) Probabilistic Model, are defined for sets of discrete search items. The fixation targets are defined as centres of gravity or individual items, respectively. As our stimuli only contain one search item, these models are not applicable. In order to generalise them to our task, we would need to use features that are defined for continuous stimuli, such as those used by Itti & Koch (2000), Rao et al (2002) and Pomplun (2006), and devise a way of mapping feature responses to possible fixation locations.

A potentially more relevant model is that of Rao et al (2002). Although it can be applied to continuous stimuli such as ours, it is designed to model a particular type of search behaviour (Zelinsky et al 1997) in which the search items are arranged on the circumference of a semi-circle and the target is easy to identify (typically subjects take only three saccades to locate the target). This creates search paths that are far more consistent between subjects than we found in our experiments. Additionally it is not clear how the model could be extended to simulate more difficult, general searches.

The previous work that is most applicable to the results of this study is probably Najemnik & Geisler's (2005, 2008) construction of an *ideal Bayesian observer* for searches for a target hidden in 1/*f* noise. However, our stimuli differ from theirs in two important ways. Firstly we do not display 1/*f* noise directly; instead we treat it as a height map and render it to give a naturalistic image. Secondly, we investigate the effect of changing β, the magnitude spectra fall-off, and σ_{RMS}, the RMS-roughness, along with variations in target shape and orientation. Although their stimuli share many properties with ours, it is not applicable to our stimuli as

it uses potential target locations as possible fixation points rather than treating the stimuli as a continuous search area. As Najemnik and Geisler's state in their conclusion, they do not offer a heuristic computational model that can be applied to general stimuli.

COMPARISON WITH THE SALIENCY MODEL

Comparison between our experimental results and the performance of Itti and Koch's saliency model suggests that the features used by the model, while capturing some aspects of human behaviour, are not sufficient to give an adequate simulation of visual search for a target on a rough surface. The closest match between human and model search performance occurred with the set of stimuli used in Experiment 1, where the two parameters of surface roughness were varied. Although there were discrepancies in the absolute number of fixations by humans and model, the model reproduced all the effects of background roughness parameters. This is a surprisingly good match, given that the model was not developed to work on such stimuli and has not been assessed in such a way before. When search performance with an elongated target was considered in Experiment 3, however, there was not only a difference in absolute levels of performance but also in the effect of target orientation, with the ability of the model to detect the target falling to low levels over a considerably wider range of orientations than in the case of human observers. We repeated the tests of the model, varying the number of spatial scales and orientations in the filter bank, and found that performance is robust to these changes as long as the spatial scale which best matches the scale of the target is present. Our conclusion is therefore that there is a clear discrepancy in the case of oriented targets, with the model unable to match human performance when they are oriented close to the direction of illumination. What could the cause of this discrepancy be?

The saliency model that we used is likely to diverge from human performance because it does not incorporate eccentricity-dependent processing (Peters et al., 2005; Vincent et al., 2007). However, this gives the model constant spatial resolution at all distances from fixation, while human resolution falls, and so the model would be expected to perform *better* with all targets. Similarly, the model does not incorporate any process of extracting solid shape from shading, which is known to contribute to efficient detection of targets in human visual search (Braun, 1997), but this feature would result in poorer model performance across *all* targets, which is not the case. It is also unclear to what extent the human subjects used shape from shading in order to find the target as the high contrast edge along the lower lip of the target can also be used to identify it.

Another possible reason for poor performance of the model with elongated targets is that, when the target is oriented close to the vertical, the contrast decreases. If the model is generally less sensitive to low contrast than humans, the result would be poorer performance. However, there is no evidence for such a difference in Experiment 2, where contrast at the target is reduced by making it shallower. The two results together cannot be explained by a difference between humans and model in sensitivity to contrast, and we conclude that the results arise specifically because the saliency model is failing to take advantage of the directional nature of the target in Experiment 3, despite having a dedicated orientation channel.

CONCLUSIONS

The results from the above experiments show that synthetic textured surfaces are a promising means of investigating visual search over backgrounds that have a natural appearance and yet have fully controllable statistical properties. In particular, these stimuli lend themselves to assessing the performance of low level features that are used in many models of attention and eye movement patterns. While recent studies by Pomplun (2006) and Rutishauser & Koch (2007) have been investigating how top down processes can influence search, both models still use low level features as key components. We have shown that the saliency model (Itti & Koch, 2000) fails to give an adequate explanation of human performance in visual search tasks using these stimuli, specifically because it lacks sensitivity to elongated stimuli at low contrast.

REFERENCES

Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: a Bayesian system identification analysis. Vision Research, 46, 2824-2833.

Balboa, R. M., & Grzywacz, N. M. (2003). Power spectra and distribution of contrasts of natural images from different habitats. Vision Research , 43, 2527–2537.

Braun, J. (1993). Shape-from-shading is independent of visual attention and may be a 'texton'. Spatial Vision, 7, 311-322.

Chantler, M. J. (1995). Why illuminant direction is fundamental to texture analysis. IEE Proceedings Vision, Image and Signal Processing, 142, 199-206

Chantler, M. J., Petrou, M., Penirschke, A., Schmidt, M., McGunnigle, G. (2005). Classifying surface texture while simultaneously estimating illumination. International Journal of Computer Vision (VISI), 62, 83-96.

Einhäuser, W., & Koing, P. (2003). Does luminance-contrast contribute to a saliency map for overt attention? European Journal of Neuroscience, 17, 1089-1097.

Einhäuser, W., Rutishauser, U., Frady, E.P., Nadler, S., König, P. & Koch, C. (2006). The relation of phase noise and luminance contrast to overt attention in complex visual stimuli. Journal of Vision, 6(11), 1148-58.

Field, D. J. (1987). Relations between the statistics of natural images and the response profiles of cortical cells. Journal of the Optical Society of America , A4, 2379-2394.

Frazor, R. A. & Geisler, W. S. (2006). Local luminance and contrast in natural images. Vision Research, 46, 1585-1598.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. v. Gompel, M. Fischer, W. Murray, & R. Hill, Eye Movement Research: Insights into Mind and Brain (pp. 537-562). Oxford: Elsevier.

Henderson, J. M., Larson, C. L., Zhu, & David C. (in press). Full Scenes produce more activation than close-up scenes and scene-diagnostic objects in parahippocampal and retrosplenial cortex: An fMRI study. Brain and Cognition, in press.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20, 1254-1259.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research , 40, 1489–1506.

Kayser, C., Nielsen, K.J. & Logothetis, N.K. (2006). Fixations in natural scenes: interaction of image structure and image content. Vision Research, 46, 2535-2545.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology, 4, 219-227.

Najemnik, J. and Geisler, W. S. (2005). Optimal eye move ment strategies in visual search. Nature, 434:387–391, 2005.

Najemnik, J. and Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. Journal of Vision, 8(3):1–14, 2008.

Navalpakkam, V. & Itti, L. (2005). Modeling the influence of task on attention. Vision Research, 45, 205-231.

Navalpakkam, V. & Itti L. (2007). Search goal tunes visual features optimally. Neuron, 53, 605-617.

Neider, M. B., & Zelinsky, G. J. (2006a). Searching for camouflaged targets: Effects of target-background similarity on visual search. Vision Research, 46, 2217-2235.

Neider, M. B., & Zelinsky, G. J. (2006b). Scene context guides eye movements during search. Vision Research, 46, 614-621.

Padilla, S., Drbohlav, O., Green, P. R., Spence, A., & Chantler, M. J. (2008). Percieved Roughness of as $1/f^{\beta}$-noise Surfaces . Vision Research (To be published).

Parkhurst, D. J., & Niebur, E. (2004). Texture contrast attracts overt visual attention in natural scenes. European Journal of Neuroscience, 19, 783-789.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. Vision Research , 42, 107–123.

Peters, R., J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. Vision Research, 45, 2397-2416.

Pomplun, M., Shen, J., Reingold, E. M. (2003). Area Activation: A computational model of saccadic selectivity in visual search. Congnitive Science, 27, 299-312.

Pomplun M. (2006). Saccadic selectivity in complex visual search displays. Vision Research, 46, 1886-1900.

Rajashekar, U., Cormack, L. K., & Bovil, A. C. (2002). Visual search: structure from noise. Proceedings of the Eye Tracking Research & Applications Symposium. New Orleans: ACMPress, 119-123.

Rao, R. P. N.; Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. Neural Computation, 9, 721-763.

Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard , D. H. (2002). Eye movements in iconic visual search. Vision Research, 42, 1447-1463.

Rutishauser, U. & Koch, C. (2007). Probabilistic modelling of eye movement data during conjunction search via feature-based attention. Journal of Vision 7, 5, 1-20.

Tatler, B.W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. Journal of Vision, 7(14), 4, 1-17.

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.

van der Schaaf, A. , & van Hateren, J. H. (1996). Modelling the power spectra of natural images: statistics and information. Vision Research , 28, 2759-2770.

Vincent, B.T., Troscianko, T. & Gilchrist, I.D. (2007). Investigating a space-variant weighted salience account of visual selection. Vision Research, 47, 1809-1820.

Voss, R. F. (1988). Fractals in nature: from characterisation to simulation. In H.-O. Peitgen, & D. Saupe, The Science of Fractal Images . 21-70. New York: Springer-Verlag.

Walther, D. & Koch, C. (2006), Modelling attention to salient proto-objects. Neural Networks, 19, 1395-1407.

Wichmann, F.A., Braun, D.I. & Gegenfurtner, K.R. (2006). Phase noise and the classification of natural images. Vision Research, 46, 1520-1529.

Wolfe, J.M. (1994) Guided Search 2.0: A revised model of visual search. Psychonomic Bulletin & Review, 1, 202-238.

Wolfe, J. M. (1998). Visual search. In H Pashler, Attention, London, University College London Press, pp 13-74.

Wolfe, J. M., Oliva, A., Horowitz, T.S., Butcher, S. J., & Bompas, A. (2002). Segmentation of objects from backgrounds in visual search tasks. Vision Research, 42, 2985-3004.

CAPTIONS FOR FIGURES

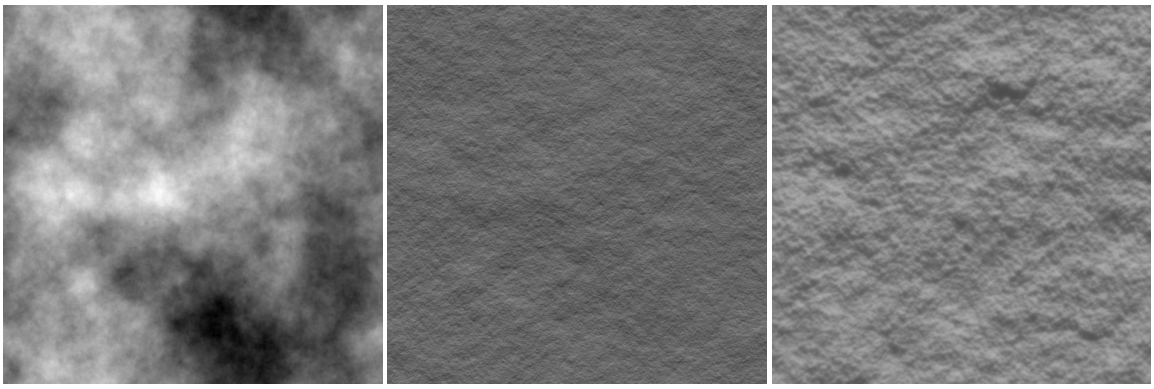

Figure 1: (a) (left) shows an example of a height map with pink noise properties: random phase combined with a $1/f^{\beta}$- magnitude spectrum. (b) (middle) shows the surface obtained by rendering the height map on the left. (c) (right) shows a real life example – rough plaster.

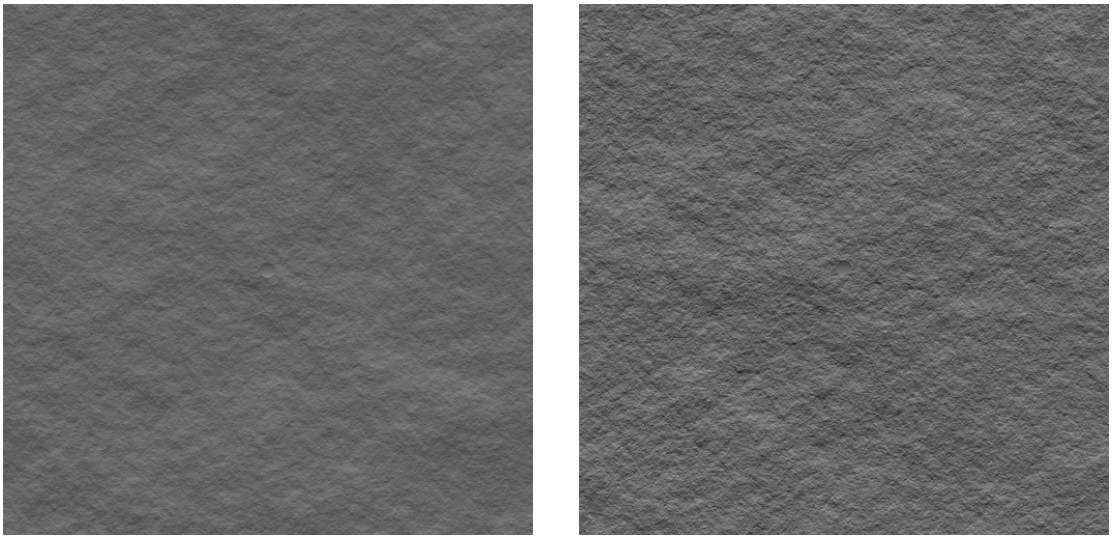

Figure 2: (a) (left) shows a target on a smooth surface, while (b) (right) shows the same target on a rougher surface. In both cases the target is located in the centre of the image.

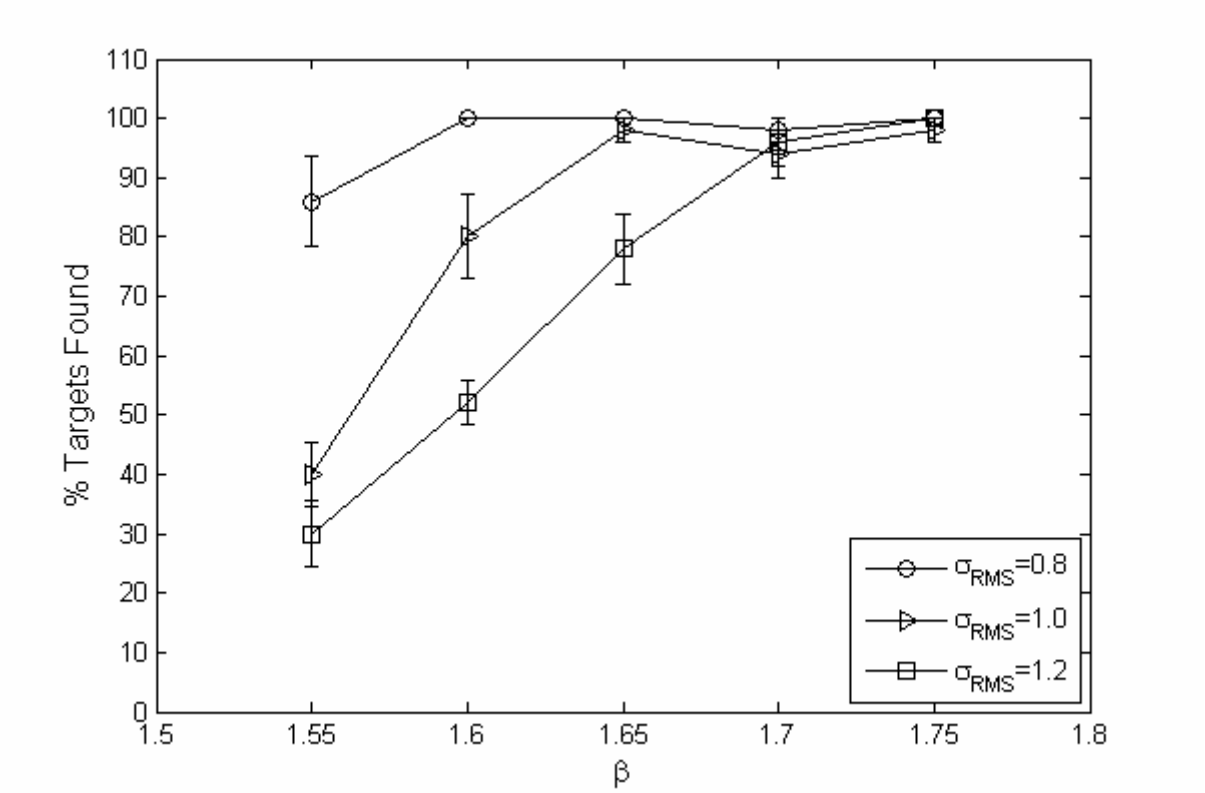

Figure 3: Mean accuracy of target detection, Experiment 1. Error bars: standard errors of means across observers.

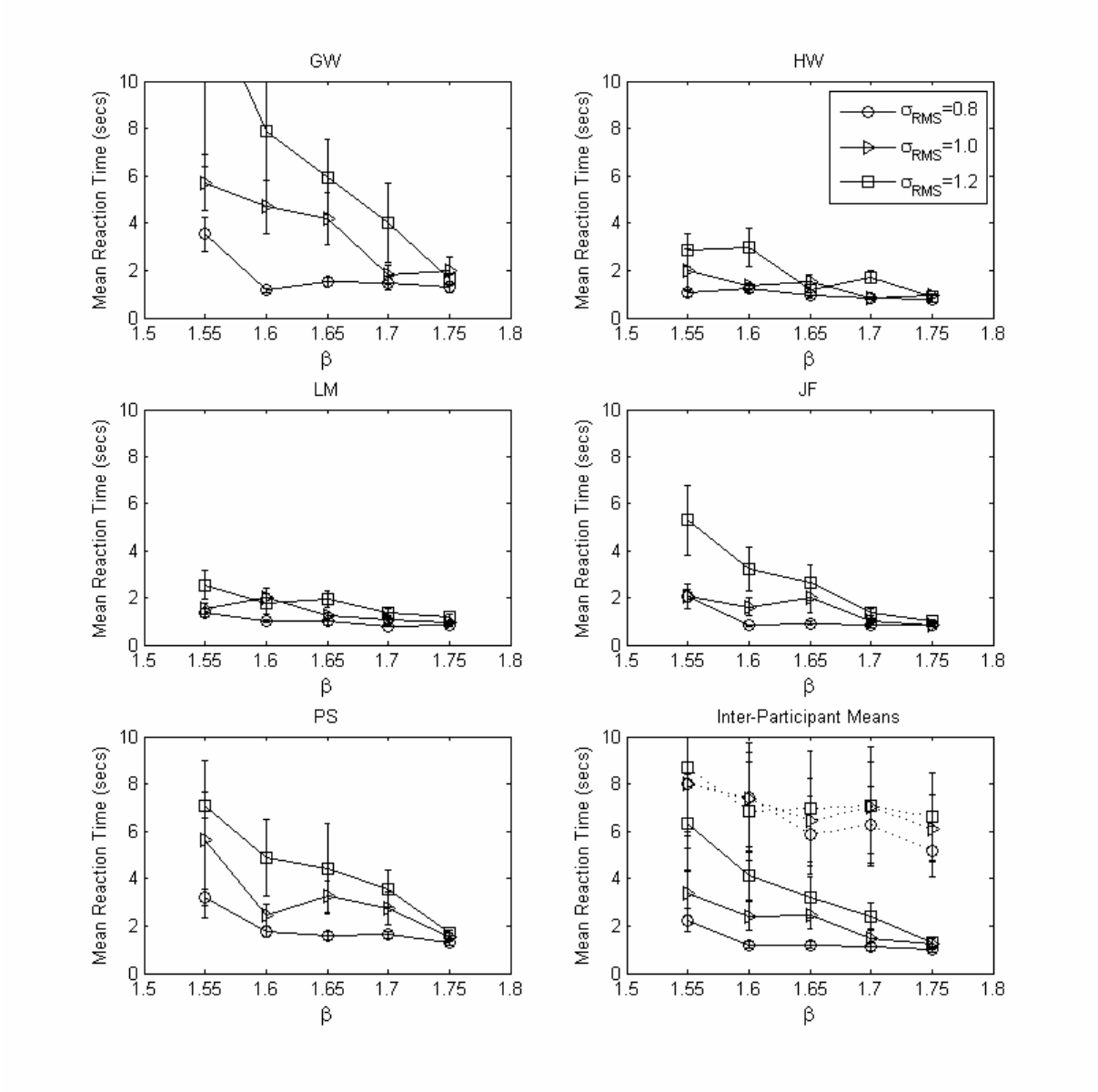

Figure 4: Mean reaction times for each observer in Experiment 1. The bottom right graph shows the mean times across observers. Only trials which were terminated with the correct response were included. Error bars: standard errors of means across trials (individual graphs) or across observers (bottom right).

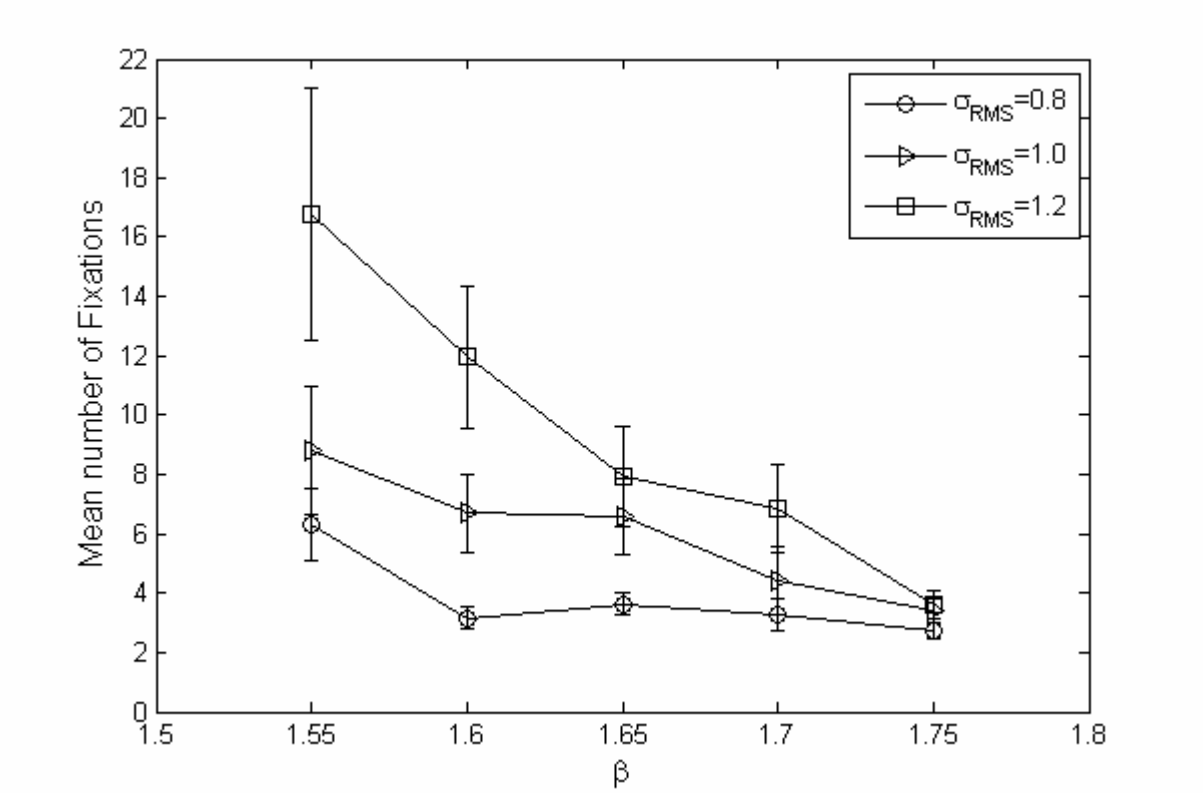

Figure 5: (a) (left) Mean number of fixations on target-present trials in Experiment 1, when the response was correct.

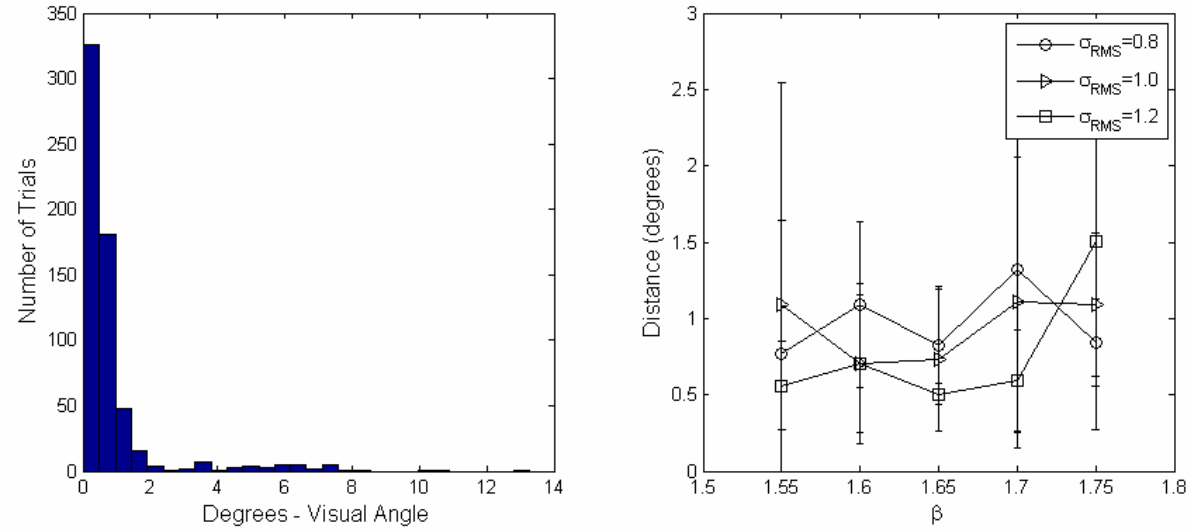

Figure 6: (a) (left) Histogram of final fixation to target distances in Experiment 1. (b) (right) How distances vary with surface roughness. Error bars as in Fig. 3.

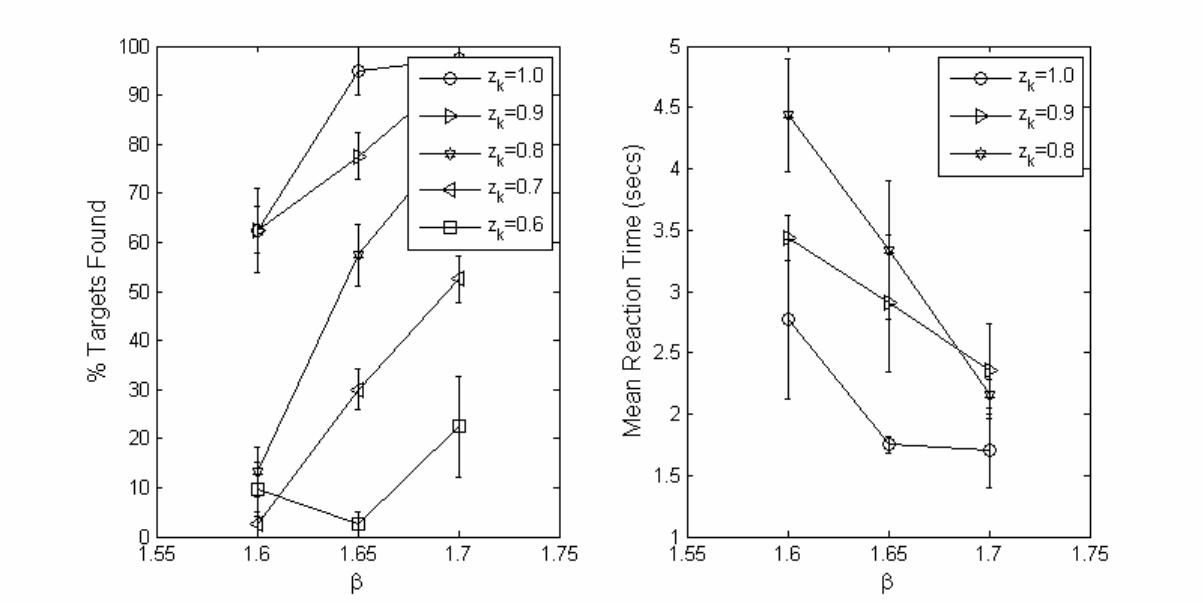

Figure 7: (a) (left) Mean accuracy of target detection, Experiment 2. (b) (right) Mean reaction times for trials where z = 0.8 – 1.0.

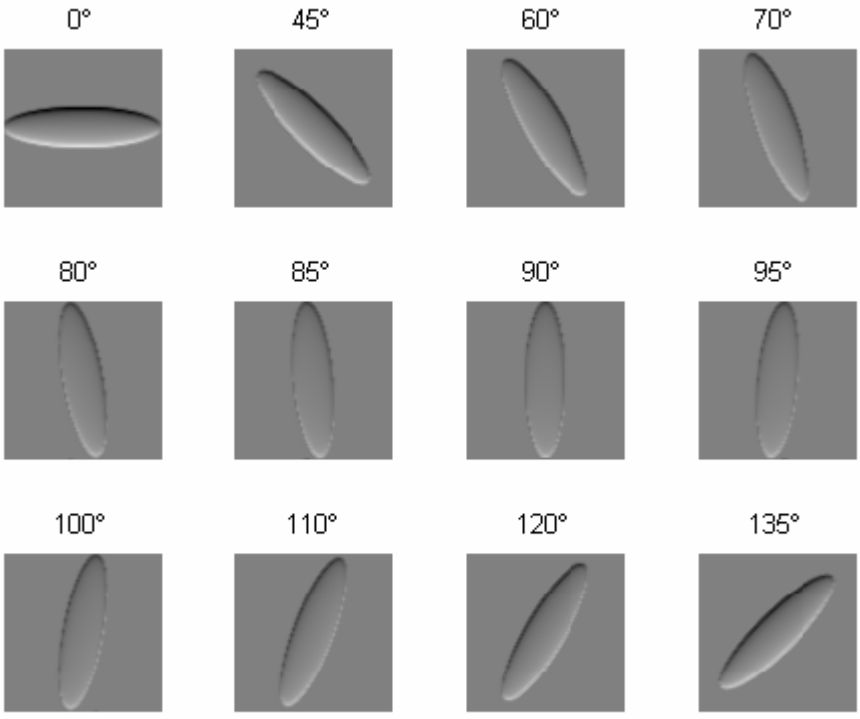

Figure 8: The effect of rotating an elongated target relative to the illumination (here, vertical). Orientations are in degrees relative to the horizontal. Note how contrast at the edges of the target changes with the orientation, reaching a minimum at 90°.

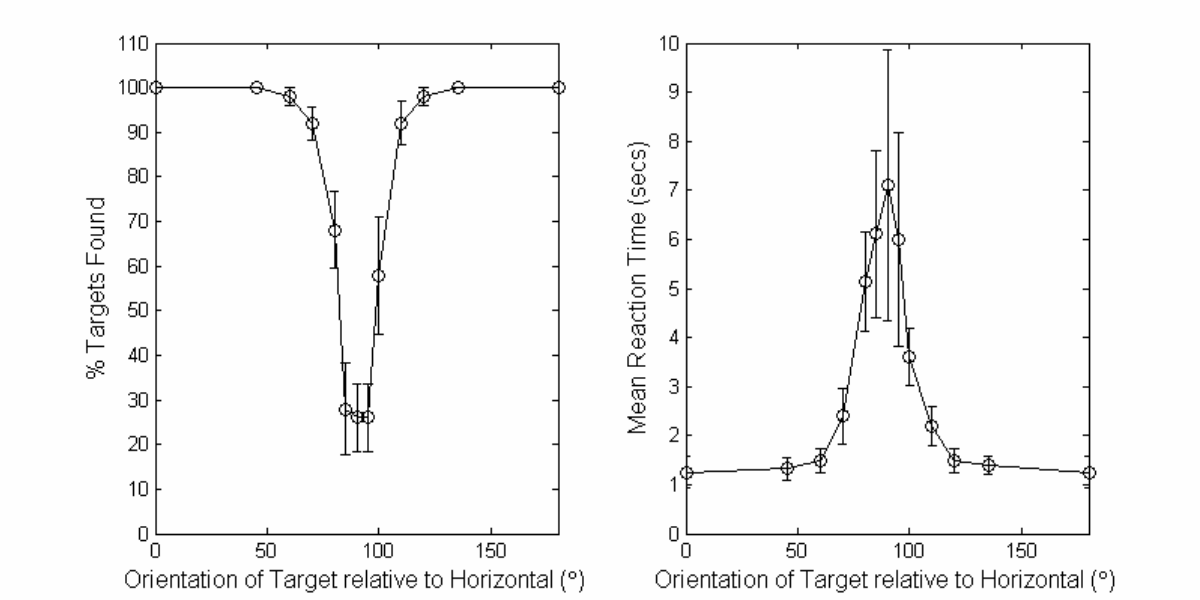

Figure 9: (a) (left) Mean accuracy of target detection, Experiment 3. (b) (right) Mean reaction times.

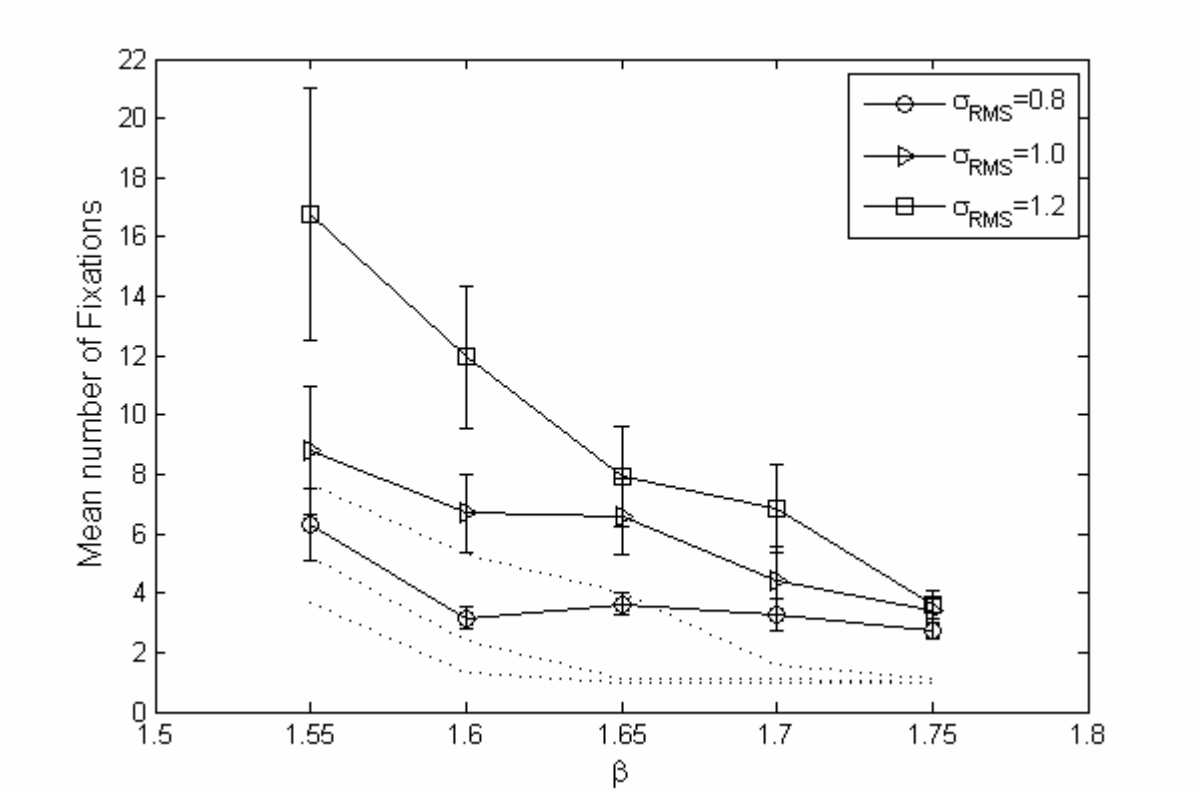

Figure 10: Comparison between the number of fixations to target identification in Experiment 1 (solid lines) and in the output of the model (dotted lines).

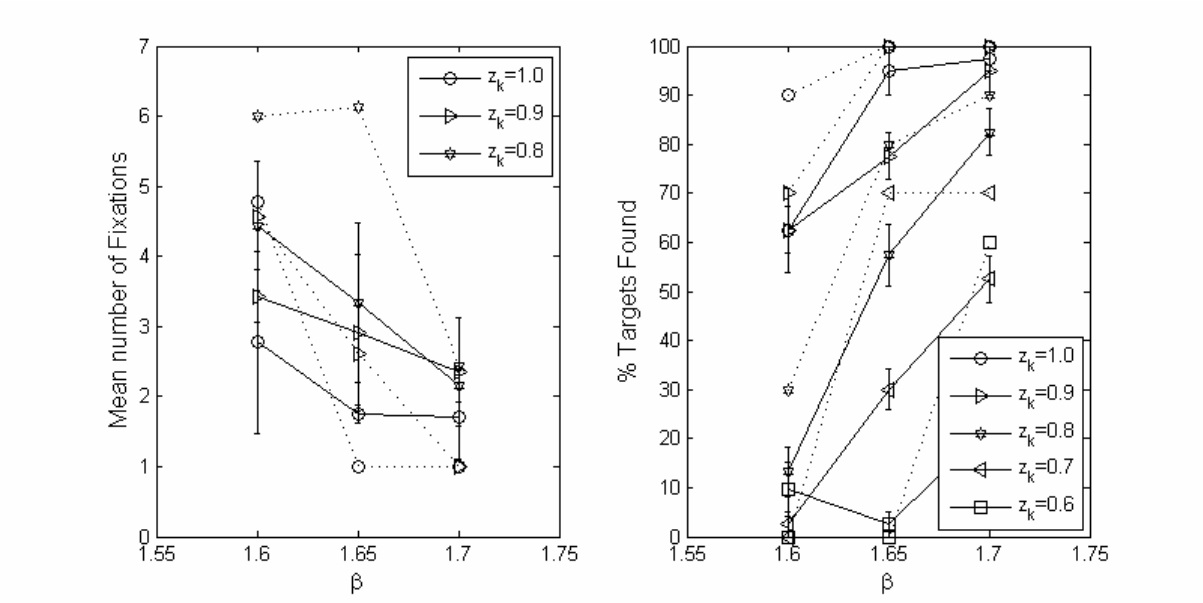

Figure 11: (a) (left) Comparison between the number of fixations to target identification in Experiment 2 (solid lines) and in the output of the model (dotted lines), for values of z from 0.8 to 1.0. (b) (right) Comparison using the accuracy measure, for z = 0.6 to 1.0.

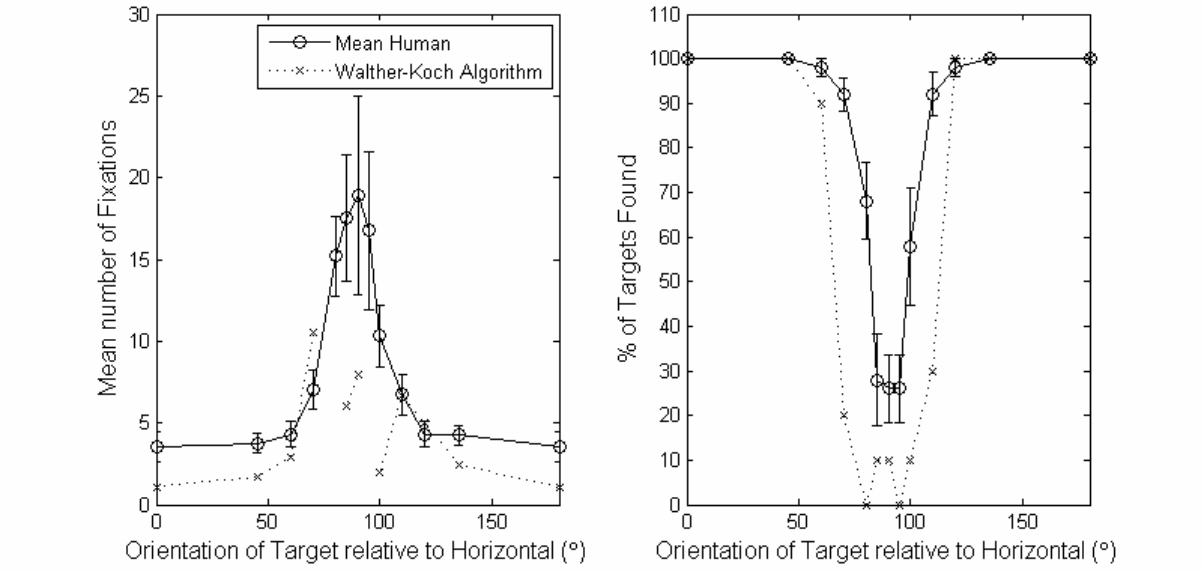

Figure 12: (a) (left) Comparison between the number of fixations to target identification in Experiment 3 (solid lines) and in the output of the model (dotted lines). Missing points for model output indicate that the target was not fixated within the limit. (b) (right) Comparison using the accuracy measure

Participant	GW	HW	LM	JF	PS	Overall
Accuracy for target present trials	87.33%	78%	80.67%	87.33%	83.33%	83.33%

Table 1: Overall accuracies of target detection in Experiment 1 for each observer, and the mean.