Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 46 (2006) 1857–1862
Vision Research
www.elsevier.com/locate/visres

Brief communication

The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task

Benjamin W. Tatler [a,*], Roland J. Baddeley [b], Benjamin T. Vincent [b]

[a] *Department of Psychology, University of Dundee, Dundee DD1 4HN, Scotland, UK*
[b] *Department of Experimental Psychology, University of Bristol 8, Woodland Road, Bristol BS8 1TN, UK*

Received 13 June 2005; received in revised form 15 November 2005

Abstract

We recorded over 90,000 saccades while observers viewed a diverse collection of natural images and measured low level visual features at fixation. The features that discriminated between where observers fixated and where they did not varied considerably with task, and the length of the preceding saccade. Short saccades (<8°) are image feature dependent, long are less so. For free viewing, short saccades target high frequency information, long saccades are scale-invariant. When searching for luminance targets, saccades of all lengths are scale-invariant. We argue that models of saccade behaviour must account not only for task but also for saccade length and that long and short saccades are targeted differently.

Keywords: Eye movement; Salience; Image features; Natural images; Saccade amplitude; Task

1. Introduction

When viewing complex scenes, we are highly selective in the locations that we choose to fixate. While it is clear that the task at hand is important in determining the locations we choose to fixate (Buswell, 1935; Nelson, Cottrell, Movellan, & Sereno, 2004; Yarbus, 1967), low level visual features can also influence eye movements (Findlay, 1981, 1997; Zelinsky, Rao, Hayhoe, & Ballard, 1997), and are likely to play a role in selection even if fixation location choice is dominated by high level factors (Tatler, Baddeley, & Gilchrist, 2005).

A recent framework for investigating the factors involved in saccade targeting is that of the *salience map* (Itti & Koch, 2000; Kadir & Brady, 2001; Koch & Ullman, 1985; Parkhurst & Niebur, 2003; Renninger, Coughlan, & Vergheese, 2005). These authors suggest that a spatial map of the *salience* of potential fixation locations is constructed by combining multiple low level visual features at many spatial scales. One natural question in this framework is whether the visual salience (in terms of a given set of low level features) differs between fixated and non-fixated locations (Parkhurst, Law, & Niebur, 2002; Tatler et al., 2005). Such studies have shown that the visual features at locations selected for foveation differ statistically from those at randomly selected locations. It appears that low level visual features are on average more extreme at fixated than non-fixated locations, and these differences tend to be larger for edges and contrast than luminance and colour. Furthermore, the largest differences between fixated and non-fixated locations is for high frequency information, and this can be interpreted as reflecting a dominance of high frequency information in saccade target selection (Tatler et al., 2005). It is this final point, and its interaction with task, that we will consider in more detail in the present report.

A system in which saccade target selection is dominated by high frequency information in scenes does encounter an obvious problem. Selection of the target to fixate must have occurred prior to the initiation of the saccade that brought the fovea to bear on this location and therefore was selected

* Corresponding author. Fax: +44 0 1382 229993.
E-mail address: b.w.tatler@dundee.ac.uk (B.W. Tatler).

0042-6989/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.12.005

using peripheral vision. Visual acuity declines with increasing eccentricity in the retina (Østerberg, 1935). Thus we are faced with the question of how selection can be driven by high frequency information when it is performed by low resolution peripheral vision.

The decline in visual acuity with retinal eccentricity suggests that it may be important to consider the eccentricity of a location when it is selected for fixation. In their recent study, Tatler et al. (2005) found that the largest difference between fixated and non-fixated locations was for spatial scales of information as high as 10.8 cycles per degree (cpd) when viewing natural images. Given estimates of how the human modulation transfer function changes with eccentricity (Rovamo & Virsu, 1979) and assuming a central visual acuity of 60 cpd (Curcio, Sloan, Kalina, & Hendrickson, 1990) we can calculate that from approximately 14° from the centre of fixation, information at scales of 10.8 cpd or higher should not be resolved. However, a significant proportion of saccades are made to locations in excess of 14° when viewing complex scenes or during real world tasks (see, Land & Hayhoe, 2001). Logically, the high frequency information should not dominate this subset of saccades.

In the present study, we analysed a total of 68,983 eye movements as observers viewed images of complex real world scenes. We then constructed salience maps of the same visual features explored by Tatler et al. (2005) at six spatial scales. The same approach for comparing salience at fixation and at randomly selected locations was employed to speculate upon the relative contribution of different spatial scales of information in target selection. However, in this study we also considered the amplitude of the saccade that had brought the fovea to the fixated location. In this way, we were able to assess the interaction between saccade amplitude and the scale of information present at fixation and thus make inferences about their possible involvement in saccade target selection.

While it has long been understood that task has a strong influence upon the distribution of fixations on complex scenes (e.g., Buswell, 1935; Yarbus, 1967), the influence of task upon any salience map framework that might underlie saccade targeting still remains unclear (see, Tatler et al., 2005). It may be that there are common salience-based targeting mechanisms underlying different tasks (this is possible even in the face of different spatial *locations* being chosen for different tasks), or it may be that the task constraints alter salience-based criteria for saccade target selection. We are by no means the first to consider the limitations of salience-based models of fixation behaviour under varying task conditions. Using an information theoretic approach to model scene statistics at the centre of gaze Krieger, Rentschler, Hauske, Schill, and Zetzsche (2000) highlighted the need for a unitary model of fixation behaviour that integrated high and low level factors; a goal toward which they have since been working (e.g., Schill, Umkehrer, Beinlich, Krieger, & Zetzsche, 2001). Raj, Geisler, Frazor, and Bovik (2005) proposed a model based upon minimising contrast entropy, which they suggested was adequate for certain tasks (in which the observer must gain as much information about the structure of a scene as possible) but would not generalise to all tasks. More explicit models of the role of task in salience-based approaches have been proposed recently. Torralba and colleagues (e.g. Torralba, 2001, 2003; Torralba, Murphy, & Freeman, 2005) have suggested a specific Bayesian framework in which low level salience maps are spatially weighted depending on the most probable location of target objects. Navalpakkam and Itti (2005) have suggested that the high level component is manifest as a bias toward particular features (or feature conjunctions) in the salience map framework, effectively weighting particular feature channels over others.

Given the body of evidence for top down effects in the control of fixation behaviour, we decided to include a manipulation of task within our exploration of saccade length effects on fixation selection. We collected eye movement data under two different task situations. In the first, participants were merely asked to look at the images freely (free viewing). In the second, they were given a search task in which they had to search for a small, localised artificial increase in brightness at a random location in the image: specifically a Gaussian luminance bump that had been added to 50% of the images (search task). For the second task only images in which the target was absent were analysed for this study. Thus, the stimuli viewed under the two task constraints in this report were identical. In this way, we were able to assess whether selection criteria for saccades of various amplitudes varied according to the task.

2. Method

In the free viewing task, 22 participants aged 18 to 29 years (mean = 21.7, $SD = 3.2$) viewed 120 photographic images of real world scenes. In the search task, 30 participants aged 18 to 53 years (mean = 22.9, $SD = 6.6$) viewed the same 120 images. In this task, half of the images had a small ($SD = 0.3°$) Gaussian brightness blip added in a random location. The task was to decide whether there was a brightness blip present and to respond using a button box.

Images were recorded using a Nikon D2 digital SLR using the highest resolution (4 megapixels). Images were displayed in 1600×1200 pixel format on a 21 in. SVGA colour monitor with a refresh rate of 100 Hz and a maximum luminance of 55 cd m^{-2}. The monitor was positioned at a viewing distance of 60 cm; consequently, the images presented subtended 40° horizontally and 30° vertically. Each trial was preceded by a fixation target positioned randomly within 10° of the centre of the screen before displaying the image for 5 s.

Eye movements were recorded during viewing using the SR Research. EyeLink II eye tracker, which samples eye position data at 500 Hz. Eye position data were collected binocularly and analysed for the eye that produced the better spatial accuracy as determined using the calibration. Nine-point target displays were used for calibration and validation of eye position. Saccade detection required a deflection of greater than 0.1°, with a minimum velocity of 35° s^{-1} and a minimum acceleration of 9500° s^{-2}, maintained for at least 4 ms. We used a minimum fixation duration of 50 ms.

Using this procedure, data were collected for 40,011 saccades in the free view task and for 55,170 saccades in the search task. We only analysed eye movements made when viewing images in which the search target was absent in order to ensure that the stimuli viewed in the two tasks were identical. This resulted in 28,972 being available for analysis for the search task.

Image features were made explicit using the same procedures as detailed in Tatler et al. (2005) for luminance, contrast and edge information. The only departure from the image feature extraction methodology is in the spatial scales at which image features were extracted. For these we used filters with standard deviations between 0.625 and 20 cpd (for contrast this refers to the standard deviation of the centre Gaussian, for edge information this refers to the standard deviation of the Gaussian carrier). These filters can alternatively be described in terms of their half widths; in this way, our filters had half widths of between 0.05° and 1.6°.

We extracted salience at fixation for each of the 40,011 fixations in the free view task and the 28,972 fixations in the search task. We also collected image features from the same locations but on different images (corresponding to locations not actually selected for fixation by the observers). Matching of the sampling distribution for selecting non-fixated image statistics in this way is important because it removes artefacts that arise from spatially non-uniform sampling of scenes by the eye such as a central fixation bias (for a discussion of such issues see, Tatler et al., 2005). The relative contribution of each feature to selection was assessed using a signal detection technique; the receiver operator characteristic (ROC; see, Tatler et al., 2005 for details). This metric determines how well fixated and non-fixated locations can be discriminated by their saliencies using a simple threshold. For two distributions that it is not possible to discriminate, the ROC area will be 0.5. For perfect discrimination, the value will be 1.0, and when the system is predicting worse than chance, the area will be less than 0.5. To assess whether the ROC area is significantly different from 0.5, we calculated 99% non-parametric confidence limits of the ROC area by the use of the bootstrap technique (Efron & Tibshirani, 1993).

For each fixation we also extracted the amplitude (in degrees of visual angle) of the preceding saccade. This allowed us to consider whether the selection of visual features at different spatial scales varied according to the distance from fixation to the target location at the time the decision to saccade to that location was made.

3. Results

Fig. 1 shows the distributions of saccade amplitudes for the two tasks. The distributions for the two tasks are similar, but there were significantly more of the longer saccades in the search task than when free viewing the images (Mann–Whitney $z = 55.41$, $p < .001$). As well as showing this small difference in saccade lengths for these two tasks, these distributions are included to give an indication of the number of samples that are used in the following analyses.

Fig. 2 shows the influence of spatial scale of contrast information upon the ability to discriminate fixated and

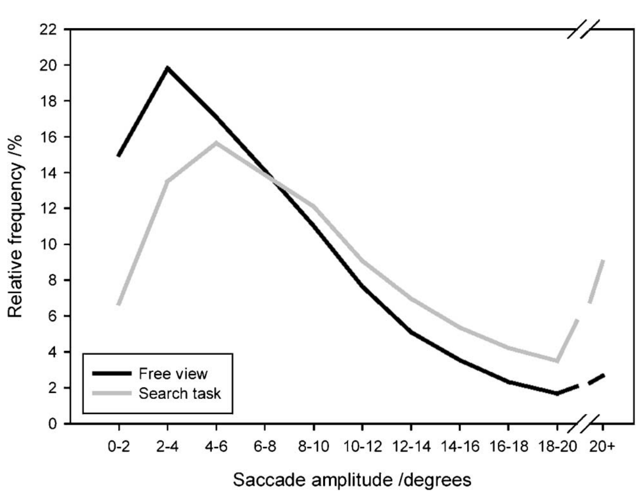

Fig. 1. The distribution of saccade amplitudes when freely viewing a scene (black line) and searching for a luminance target (grey line) in a scene. For the search task, only trials on which the target was absent are included. The same scenes were used in both tasks.

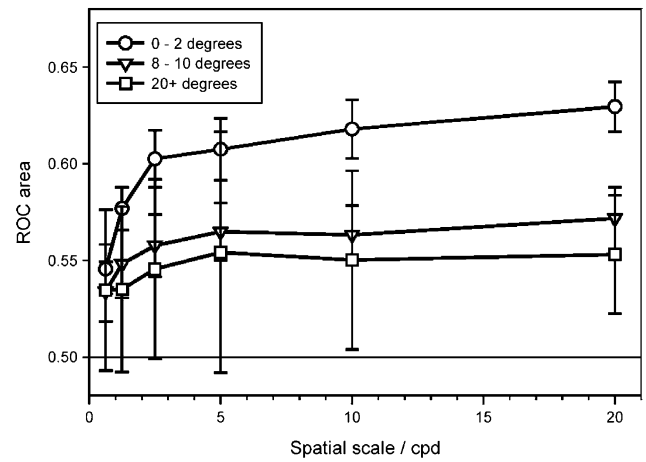

Fig. 2. ROC area values for the selection of contrast information at fixations following short (0°–2°), medium (8°–10°), and long (20+ degrees) range saccades for the six different spatial scales. ROC area values measure the difference between the distributions for fixated and non-fixated locations. An ROC area value of 0.5 indicates no difference. The *y*-scale is much enlarged. Error bars indicate 99% confidence intervals, calculated using a bootstrap technique. For short saccades high frequency information is more discriminatory than low. Conversely, long saccades are scale-invariant.

non-fixated locations in the free viewing condition, for saccades of three different amplitudes. ROC values above 0.5 indicate that the feature was discriminatory between fixated and non-fixated regions. A value below 0.5 suggests that extremes of this feature are avoided. If the 99% confidence intervals do not overlap 0.5, the mean ROC area for that data point is significantly different from chance. Conversely, if they do overlap, there is no significant difference from chance. High frequency information is more discriminatory for short range (0°–2°) saccades than is low frequency information. For long range saccades (over 20°) the predominance of high frequency information diminishes. There is clearly an influence of saccade amplitude on the scale of selection and this is particularly evident when comparing the highest and lowest spatial scales of information.

Fig. 3 explores the influence of saccade amplitude upon the scale of selection in more detail. Data are presented for each of the three features (contrast, edges, and luminance) at the highest (20 cpd; half width of 0.05°) and lowest (0.625 cpd; half width of 1.6°) spatial scales for both of the tasks (free view and search).

There were four main findings for the free viewing task. First, high frequency edge and contrast information are most discriminatory. Second, the difference between fixated and non-fixated edge and contrast information decreases as a function of saccade size to roughly 6°–10°, and is then flat. Thus, for short saccades, high spatial frequency image statistics seem important, but for long saccades, they appear less so. Third, for low spatial frequencies, while image features at fixated and non-fixated locations are significantly different, there is no effect of saccade size. Fourth, high frequency luminance was discriminatory, but low frequency luminance was not and neither of these showed any pronounced effect of saccade amplitude.

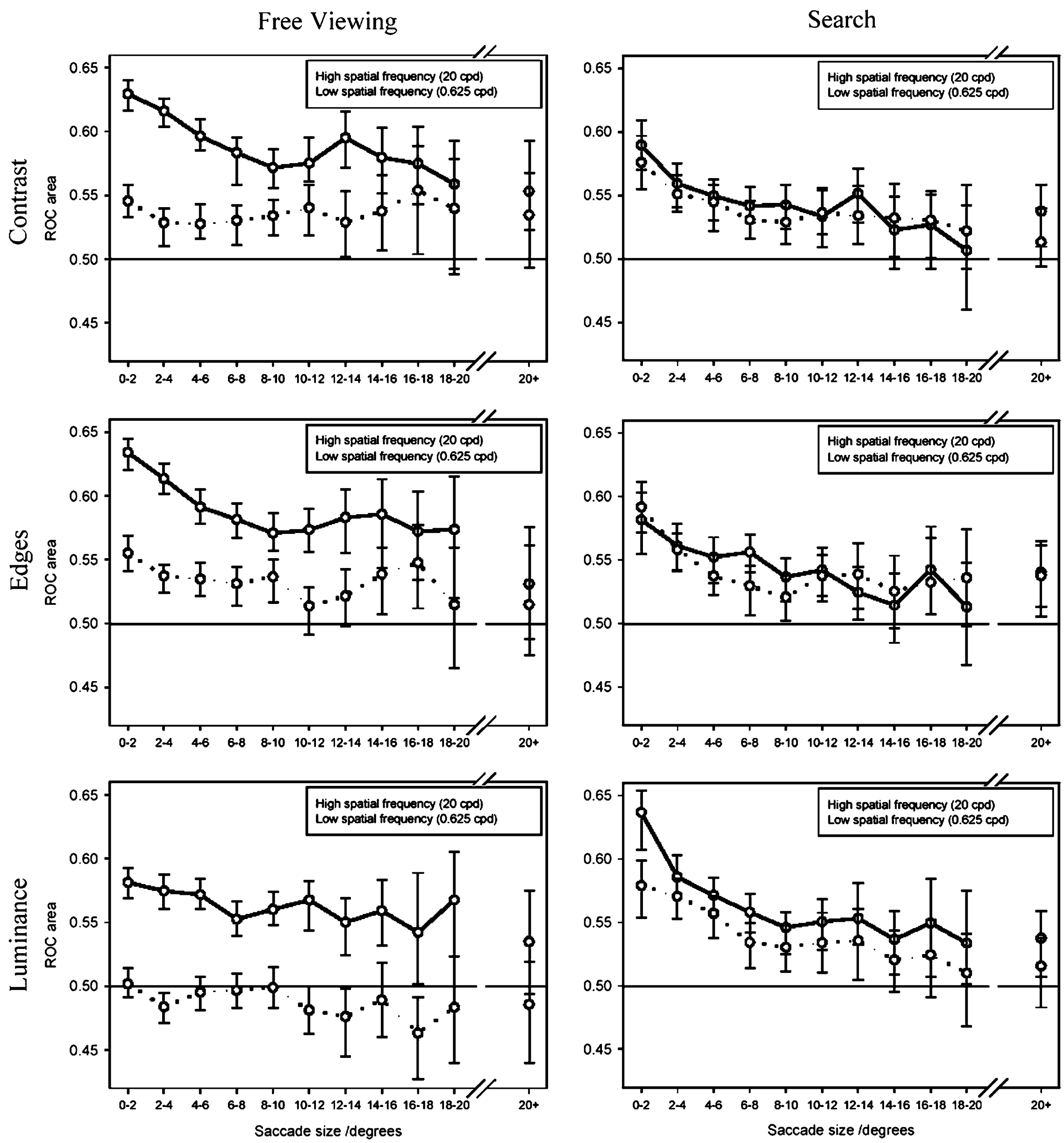

Fig. 3. Visual salience at fixation as a function of saccade amplitude. Data are presented for observers freely viewing scenes (left) and searching for a luminance target in the same scenes (right), for each of the three features investigated: contrast (top), edges (middle), and luminance (bottom). In each plot, the solid line represents data for high spatial frequency information (20 cpd) and the dotted line represents data for low spatial frequency information (0.625 cpd). Thus differences between low and high spatial frequencies can be seen for all three features for small amplitude saccades in the free view task. No differences between scales are seen for the search task. In both tasks there is a tendency for greater differences between fixated and non-fixated regions in the targeting of short saccades rather than long.

A different pattern of results was found for our particular search task. In contrast to the free viewing task where high spatial frequency information was far more discriminatory than low, there was little difference between the two spatial scales in our search task. We again found that contrast and edges were much better at discriminating fixated and non-fixated locations for short saccades than long. However, in contrast to the free viewing task, luminance was also highly discriminatory, especially for the smaller amplitude saccades. All of the effects reported above were highly significant ($p < 10^{-5}$).

Our data therefore show that (i) the image regularities that discriminate between fixated and non-fixated locations are different for long and short saccades, and (ii) the image characteristics that discriminate are different both qualitatively and quantitatively depending on the task, even when the images viewed were identical.

4. Discussion

In this study, we considered whether the features at fixation (at various spatial scales) differed according to the eccentricity of the location when it was selected as a target to saccade to, and the task being performed by the observer. We found that (i) there were clear differences between free viewing and searching images, (ii) in both tasks fixated and non-fixated locations were more discriminable in terms of visual features following short saccades than following long saccades, (iii) for short saccades in the free view task, high frequency information dominated, and (iv) following long saccades in both tasks, there was little difference in the low frequency information present at fixated and non-fixated locations.

Our results have three implications for models of saccade target selection. Most trivially, these results indicate that previous studies will contain a number of systematic biases. For both tasks, the biggest difference in feature statistics was seen for short saccades (<8°). Indeed for longer saccades, the difference between fixated and non-fixated regions, while significant, was close to chance for some of the features. Under normal viewing conditions small amplitude saccades dominate (see, Fig. 1; Land & Hayhoe, 2001). In our free view task 66% of saccades were to locations within 8° of the current centre of gaze, and in our particular search task, 50%. Thus previous reports of the influence of salience upon saccade target selection (Parkhurst et al., 2002; Reinagel & Zador, 1999; Tatler et al., 2005) may not capture the true contribution of visual features in targeting saccades of all lengths, rather reflecting a bias from a (large) subset of small amplitude saccades. The data from the present study suggest that previous models of feature selection may systematically underestimate the involvement of visual features in selecting locations close to the current centre of gaze, and overestimate their involvement in selecting more distant targets.

More importantly, we found that task had a strong effect on which characteristics were discriminatory for saccades of differing amplitudes. In one way this may not be surprising: if, as in our chosen search task, we are looking for a target defined by luminance, then luminance would be expected to be different at the points of fixation. This result is in agreement with Torralba's (2001,2003; Torralba et al., 2005) proposal that incorporates contextual factors into the salience framework, but presents problems for some proposals that model salience in a task independent manner (e.g., Itti & Koch, 2000). Our findings demonstrate that such task independent models are at best an approximation of saccade target selection. Moreover, it is important to note that our findings suggest that it is likely that the nature of the tasks chosen for comparisons such as we make here will greatly influence the results—a search task based upon a different feature, spatial scale or object might well produce very different results. Indeed, we do not even presume to suppose that our search task is generalisable to other search tasks (for models of search behaviour, see e.g., Najemnik & Geisler, 2005; Palmer, Verghese, & Pavel, 2000; Wolfe, 1998). However, our data do reiterate the need to consider the nature of the task in any salience-based approaches. The pattern of results is also not entirely predictable from the task: in our search task, the targets were defined by differences in luminance but as well as luminance, we found highly significant differences in the presence of high and low frequency edges and contrast at fixation following short saccades. The difference we found for short and long saccades does not match that reported by Itti (2006) who found no difference between the visual salience targeted by long and short saccades, for observers viewing movies. The most likely interpretation is again that this is another effect of the difference in task.

For short saccades (<8°), there were clear differences in the relative selection of high and low spatial scales of information between our two tasks. When freely viewing the scenes, high frequency information was highly discriminatory whereas low frequency information was not. Conversely, there was no such difference between high and low frequency information in our search task. This might arise if, when free viewing, observers tend to fixate real objects preferentially. In contrast, when searching for a randomly located luminance target, it is unlikely that viewing will be based upon selecting real objects in the scene and this may account for the difference in scale selectivity between the two tasks.

Lastly, we found large differences in the statistics of visual features at fixation following short and long saccades, with features being highly discriminatory for short saccades, but far less so (and in some cases not at all) for long saccades. One possible explanation for this result is that different selection strategies may dominate for targeting saccades of different amplitudes. If so, our data imply different priorities for exploring nearby locations, perhaps the currently attended object, than for targeting more distant locations, perhaps selecting a new object to be scrutinised. When exploring the current object, a strategy that selects distinctive features of that object for fixation seems highly plausible. In contrast, choosing what to attend to next (I have looked at the cup, now I want to look for a kettle), may be far more dominated by high level constraints such as where kettles are likely to be. The possibility of differing targeting mechanisms for long and short amplitude saccades has been suggested before (Frost & Pöppel, 1976). Frost and Pöppel suggested that saccades to targets further than 10°–15° away are executed using a mechanism involving the superior colliculus, whereas closer saccades are targeted using a mechanism involving the geniculo-cortical pathway. If such mechanistic differences exist, it may be that different priorities underlie target selection. Again these differences are important for models of fixation behaviour. A model that ignores long range, between-object saccades—which appear to be relatively low level feature invariant and are likely to be dominated by higher-level constraints—may offer a relatively good characterisation of short, salience dominated saccades, but may be an oversimplification of saccade target selection. Thus, it may be that producing a unitary salience map of a scene (as most current salience-based models do) is inappropriate; rather models should account for the moment-to-moment location of the centre of gaze, and reflect different targeting priorities at different eccentricities.

References

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology of perception in art*. Chicago: University of Chicago Press.

Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *Journal of Comparative Neurology, 292*(4), 497–523.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall.

Findlay, J. M. (1981). Local and global influences on saccadic eye movements. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 171–179). Hillsdale, NJ: Lawrence Erlbaum Associates.

Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research, 37*(5), 617–631.

Frost, D., & Pöppel, E. (1976). Different programming modes of human saccadic eye-movements as a function of stimulus eccentricity—indications of a functional subdivision of visual-field. *Biological Cybernetics, 23*(1), 39–48.

Itti, L. (2006) Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, in press.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10–12), 1489–1506.

Kadir, T., & Brady, M. (2001). Saliency, scale, and image description. *International Journal of Computer Vision, 45*(2), 83–105.

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

Krieger, G., Rentschler, I., Hauske, G., Schill, K., & Zetzsche, C. (2000). Object and scene analysis by saccadic eye-movements: An investigation with higher-order statistics. *Spatial Vision, 13*(2-3), 201–214.

Land, M. F., & Hayhoe, M. M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*(25-26), 3559–3565.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature, 434*(7031), 387–391.

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45*(2), 205–231.

Nelson, J. D., Cottrell, G. W., Movellan, J. R., & Sereno, M. I. (2004). Yarbus lives: A foveated exploration of saccadic eye movement. *Journal of Vision, 4*(8), 741a.

Østerberg, G (1935). Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica Supplimentum, 6*, 1–103.

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research, 40*(10–12), 1227–1268.

Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision, 16*(2), 125–154.

Parkhurst, D. J., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42*(1), 107–123.

Raj, R., Geisler, W. S., Frazor, R. A., & Bovik, A. C. (2005). Contrast statistics for foveated visual systems: Fixation selection by minimizing contrast entropy. *Journal of the Optical Society of America – Optics Image Science and Vision, 22*(10), 2039–2049.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network-Computation in Neural Systems, 10*(4), 341–350.

Renninger, L. W., Coughlan, J., & Vergheese, P. (2005). An information maximization model of eye movements. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems* (Vol. 17). Cambridge, MA: MIT Press.

Rovamo, J., & Virsu, V. (1979). An estimation and application of the human cortical magnification factor. *Experimental Brain Research, 37*(3), 495–510.

Schill, K., Umkehrer, E., Beinlich, S., Krieger, G., & Zetzsche, C. (2001). Scene analysis with saccadic eye movements: Top-down and bottom-up modeling. *Journal of Electronic Imaging, 10*(1), 152–160.

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research, 45*(5), 643–659.

Torralba, A. (2001). Contextual modulation of target saliency. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems 14* (Vol. 2, pp. 1303–1310). Cambridge, MA: MIT Press.

Torralba, A. (2003). Contextual priming for object detection. *International Journal of Computer Vision, 53*(2), 169–191.

Torralba, A., Murphy, K. P., & Freeman, W. T. (2005). Contextual models for object detection using boosted random fields. In. L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems 17* (NIPS 2004, pp 1401–1409). Cambridge, MA: MIT Press.

Wolfe, J. M. (1998). Visual Search. In H. Pashler (Ed.), *Attention Hove* (pp. 13–74). East Sussex: Psychology Press.

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

Zelinsky, G. J., Rao, R. P. N., Hayhoe, M. M., & Ballard, D. H. (1997). Eye movements reveal the spatiotemporal dynamics of visual search. *Psychological Science, 8*(6), 448–453.