European Journal of Neuroscience, Vol. 19, pp. 783–789, 2004

SHORT COMMUNICATION
Texture contrast attracts overt visual attention in natural scenes

Derrick J. Parkhurst and Ernst Niebur
The Zanvyl Krieger Mind/Brain Institute, The Johns Hopkins University, Baltimore, Maryland, 21218, USA

Keywords: eye movements, image statistics, computational model, salience, second-order

Abstract

In natural vision, the central nervous system actively selects information for detailed processing through mechanisms of visual attention. It is widely held that simple stimulus features such as color, orientation and intensity contribute to the determination of visual salience and thus can act to guide the selection process in a bottom-up fashion. Contrary to this view, Einhäuser, W. & König, P. [(2003) *Eur. J. Neurosci.*, 17, 1089–1097] conclude from their study of human eye movements that luminance contrast does not contribute to the calculation of stimulus salience and that top-down, rather than bottom-up, factors therefore determine attentional allocation in natural scenes. In this article, we dispute their conclusion and argue that the Einhäuser and König study has a number of methodological problems, the most prominent of which is the unintentional introduction of changes in texture contrast. We hypothesize that texture contrast, like luminance contrast, can contribute to the guidance of attention in a bottom-up fashion, and that an appeal to top-down factors is not necessary. To test this hypothesis, we implement a purely bottom-up model of visual selective attention where salience is derived from both luminance and texture contrast. We find that the model can quantitatively account for Einhäuser and König's results and that texture contrast strongly influences attentional guidance in this particular paradigm. The significance of this result for attentional guidance in other paradigms is discussed.

Introduction

Under natural viewing conditions, humans sequentially sample parts of the visual scene by making rapid eye movements, called saccades. Significant progress towards understanding saccades has been made primarily on the mechanisms that control their dynamics, while their guidance is less well understood. Recent studies of visual selective attention have shed light on this question, especially with regard to the guidance of these eye movements when observers view complex, natural scenes.

It is known, from studies utilizing well-controlled but simple experimental stimuli, that two classes of selection mechanisms influence the allocation of attention. 'Bottom-up' selection involves fast, and sometimes compulsory, stimulus-driven mechanisms. That is to say that computational resources are allocated to particular parts of the visual input, based on the properties of that input. For example, attention is automatically attracted by unique features (Treisman & Gelade, 1980), abrupt onsets (Yantis & Jonides, 1984; Yantis & Jonides, 1996) and the appearance of new perceptual objects (Hillstrom & Yantis, 1994) even when irrelevant to the task at hand. When a target in a visual search task is uniquely defined by a stimulus feature (e.g. direction of motion) from the distractors, the target can 'pop-out', automatically attract attention and lead to efficient search (Nakayama & Silverman, 1986). On the other hand, 'top-down' selection is a slower, goal-directed mechanism that influences the allocation of attention based on the observer's expectations, intentions or past experiences. For example, observers can volitionally select objects (Rock & Gutman, 1981) or regions of space (Posner, 1980) for detailed processing largely independent of the stimulus properties at those locations.

The interaction of bottom-up and top-down mechanisms of visual attention in experimental paradigms using simple stimuli has been extensively studied. It is established that salient stimuli will automatically capture attention as long as attention is not spatially focused prior to stimulus onset (Theeuwes, 1990; Theeuwes, 1994; Yantis & Jonides, 1996). Attentional capture occurs when attention is nondeliberately allocated to a stimulus, irrespective of its task relevance (Yantis & Egeth, 1999). In situations where attention is not captured, there exists a strong dependence of stimulus-driven selection on top-down attentional control settings (Folk *et al*., 1992, 1994). However, the influence of top-down mechanisms on bottom-up mechanisms is dependent on the particular task at hand and the strategy that the observer adopts to accomplish that task (Bacon & Egeth, 1994).

Guidance of attention in natural scenes

Although there is good experimental evidence supporting both stimulus-driven and goal-directed mechanisms in well-controlled paradigms with simple stimuli, the relative degree to which these mechanisms determine attentional allocation in complex natural scenes has been examined much less extensively. One technique that can be used to address this question is to record eye movements. When observers view natural scenes they typically make a series of saccades, each followed by a period of fixation. These fixations can be taken as indicators of attentional allocation given that focal attention at the location of a pending eye movement is a necessary precursor for that

Correspondence: Dr Derrick Parkhurst, as above.
E-mail: Derrick.Parkhurst@jhu.edu

Received 11 August 2003, revised 26 November 2003, accepted 1 December 2003

doi:10.1111/j.1460-9568.2003.03183.x

movement (Shepherd *et al*., 1986; Hoffman & Subramaniam, 1995; Kowler *et al*., 1995; Deubel & Schneider, 1996; McPeek *et al*., 1999).

We recently conducted an experiment that took advantage of this relationship to study the extent to which stimulus-driven factors influence the allocation of attention in complex natural scenes (Parkhurst *et al*., 2002). We recorded the eye movements of participants while they free-viewed 300 natural and artificial scenes and examined the relationship between the observed fixation locations and the salience of stimuli at those locations. To quantify the salience of the stimuli in the scenes, we used a purely bottom-up model of visual selective attention (Itti *et al*., 1998). In the model, the processing begins by breaking the visual input up into a series of feature channels, representing color, intensity and orientation information, at a range of spatial scales. These simple feature representations are then converted to center–surround representations, which are optimized to detect local feature differences. Finally, a salience map is calculated by summing the local feature differences across spatial scale and feature type. The salience map is a retinotopic map that indicates the visual importance of stimuli at each location, and is based purely on stimulus features in the scene (Koch & Ullman, 1985).

In our experiment we found that, for each new fixation made after stimulus onset, stimulus salience as calculated by the model was significantly higher than that expected by chance factors alone. The magnitude of this effect was largest for early fixations and tended to decline (but never to or below chance levels) for subsequent fixations. These results are consistent with the evidence from paradigms utilizing simpler artificial stimuli that indicate attention can be guided by bottom-up mechanisms. We concluded from these results that overt attention is significantly dependent on stimulus properties when participants free-view complex, natural and artificial scenes.

Does luminance contrast guide attention in natural scenes?

In a recent report, Einhäuser & König (2003) note that, although the correlation between stimulus salience and fixation locations supports the hypothesis that luminance contrast causally influences attention in a bottom-up way, correlation does not necessarily imply causation. Instead, they suggest that the observed correlation is also consistent with top-down attentional guidance, if one assumes that there exists a correlation between luminance contrast and top-down representations.

To test the causal dependence of attentional allocation on luminance contrast experimentally, they recorded eye movements from five participants viewing unaltered and contrast-altered natural scenes. Contrast was altered by selecting five different locations in an image and increasing or decreasing local contrast by a factor α. Contrast modifications ranged from a decrease in contrast by slightly more than one-half ($\alpha = -0.6$) to a doubling of contrast ($\alpha = +1.0$). The logic of their study is that if bottom-up mechanisms contribute to attentional guidance then a local increase of luminance contrast should increase the probability of attention visiting a location (i.e. attract attention), while a local decrease of contrast should decrease this probability.

To test this prediction, they compared the percentage of fixations at the modified locations in the contrast-altered images to the percentage of fixations at the same locations in the unaltered images. They found an increase in the percentage of fixations for both regions of increased contrast as well as for regions of decreased contrast. These differences were only significant for large manipulations of contrast. To assure that the contrast modifications were detectable, an experiment was conducted where participants were given unlimited time to distinguish which sides of a hybrid image (half altered, half unaltered) contained modifications. The long reaction times in the task (on average ≈ 5 s) and the low accuracy (on average $\approx 70\%$ correct) indicate that the detectability of the contrast modifications was low. Although no tests of significance were reported, it appears that accuracy is well above chance levels only for large increases or decreases in contrast, thus paralleling the fixation results.

Einhäuser & König (2003) argue that these results are inconsistent with a bottom-up mechanism of visual attention where luminance contrast contributes to stimulus salience. They claim that any model in which luminance contrast contributes to stimulus salience must predict an increase in fixation probability for increases in contrast and a decrease in fixation probability for decreases in contrast. Given that they observe an increase in fixation probability for both increases and decreases in luminance contrast, and that these differences are detectable (in the extreme cases), they conclude that luminance contrast does not contribute to stimulus salience through bottom-up mechanisms. They infer that the correlation between stimulus salience and fixation locations observed by Parkhurst *et al*. (2002) reflects instead a non-causal relationship that is a result of the presumed correlation between luminance contrast and top-down representations.

Experimental control in natural scenes

In this section, we discuss a number of methodological problems in the Einhäuser & König (2003) paradigm. First, only eight scenes were used in their entire experiment. Participants saw each scene 30 times, for 8 s each time. This design provided ample opportunity for participants to develop a detailed memory of the scenes. Given that the altered regions were the only parts of the scene that changed on each presentation, participants may well have adopted an explicit top-down strategy to search for, and thus preferentially fixate, the altered regions. This possibility is particularly strong considering that participants were instructed to 'study the images carefully'. These instructions may have suggested to participants that they should memorize the scenes. In fact, this top-down strategy is consistent with the observed pattern of results; more frequent fixations on regions of both increased and decreased contrast. Alternatively, top-down attentional mechanisms based on stimulus familiarity, repeated scene context or scene layout may have implicitly influenced the allocation of attention without the explicit awareness of participants (Noton & Stark, 1971; Wang *et al*., 1994; Henderson & Hollingworth, 1999; Chun, 2000). It is well known that task instructions and experimental design can dramatically influence eye movements (Yarbus, 1967; Andrews & Coppola, 1999; Pelz & Canosa, 2001). Thus it would be difficult to conclude that the eye movements observed in this paradigm are not unduly influenced by top-down factors, which may obscure bottom-up factors that are otherwise effective.

A second problem concerns the generation of the modified images. Einhäuser & König (2003) attempted to alter local contrast without altering local luminance. They manipulated local contrast by adding (or subtracting) some portion of the difference between the luminance at a location and the luminance averaged over the entire image. Because the global luminance is not always a good estimate of the local luminance, this method of contrast manipulation unintentionally induces local luminance artifacts. [The scenes used in the Einhäuser & König (2003) study tend to be shown from the point of view of a human looking at the horizon and thus are lit from above. We found that the intensity at the top of the images used in their experiment was on average 33% higher than that at the bottom and thus, in this case, local luminance was a poor estimate of global luminance.]

These luminance artifacts induce contrast artifacts. Consider a location where the local luminance is lower than the global luminance. If a negative contrast manipulation is introduced using this method, the luminance at this location will increase towards the global luminance.

In this case, a negative contrast manipulation actually increases contrast by creating a patch of average luminance in a region of low luminance. Likewise, a patch of average luminance is introduced into a high-luminance region. In either case, negative contrast manipulations result in increases rather than decreases in contrast. This lack of control is extremely troublesome given that the evidence on which Einhäuser & König (2003) reject bottom-up models of attention rides entirely on negative contrast manipulations. Notably, by simply manipulating contrast using the local luminance instead of the global luminance, these artifacts would have been minimized.

Another serious problem in the Einhäuser & König (2003) study also stems from lack of stimulus control. Even if manipulations of local contrast had been made that kept local luminance constant, uncontrolled changes in texture contrast would still have resulted. Whereas luminance contrast is a first-order stimulus property defined by variation in local luminance, texture contrast is a second-order stimulus property defined by local variation in contrast or other texture elements (Chubb & Sperling, 1988; Cavanagh & Mather, 1989). By purposely altering luminance contrast at one spatial scale, Einhäuser & König (2003) unintentionally altered texture contrast at larger spatial scales.

It is possible that attentional guidance in their paradigm is influenced by texture contrast as well as luminance contrast. In fact, it is known that natural scenes do contain significant variation in texture contrast that is not correlated with luminance contrast and thus serves as an important source of visual information (Schofield, 2000). However, it is not clear whether the manipulations of texture contrast unintentionally induced in their experiment are of a significant magnitude relative to the texture contrast inherently present in natural scenes. To test the possibility that both luminance contrast and texture contrast contribute to attentional guidance, we examine the ability of purely bottom-up models of visual attention that base stimulus salience on luminance contrast and texture contrast to account for the Einhäuser & König (2003) results.

Modelling visual selective attention

In this section, two bottom-up models of attention are implemented. The first model computes stimulus salience from luminance contrast using a center–surround computation. The second model computes stimulus salience from texture contrast by applying a center–surround computation to the normalized luminance-contrast maps. Divisive inhibition is used to normalize the luminance-contrast responses. Such normalization can explain single-cell behaviour in primary visual cortex (Carandini & Heeger, 1994; Carandini *et al*., 1997) and introduces a nonlinearity required for the detection of second-order stimulus features. These models process visual features at a number of spatial scales, which is important for the present study given that natural scenes contain information at many spatial scales (Ruderman & Bialek, 1994). The following subsections provide a description of the two models. A more comprehensive description can be found elsewhere (see Niebur & Koch, 1996; Itti *et al*., 1998; Itti & Koch, 2000; Parkhurst *et al*., 2002; Parkhurst, 2002).

First-order salience map

In order to generate a first-order salience map from a natural scene, the grey-scale input image of the scene is sampled at a range of spatial scales to form a Gaussian image pyramid (Burt & Adelson, 1983). The input image is used as the base level of the pyramid and each subsequent level of the pyramid is given by reducing the linear resolution of the previous level by a factor of two. For the simulations reported in this study, each pyramid had six levels.

To detect luminance contrast, two types of receptive fields were used, one on-center–off-surround ($CS_{\mathrm{on/off}}$) and one off-center–on-surround ($CS_{\mathrm{off/on}}$). A center–surround pyramid was created, from the intensity image pyramid (I), for each receptive field type:

$$CS_{\mathrm{on/off}} = R(L(I,\sigma_c) - L(I,\sigma_s)) \quad (1)$$

$$CS_{\mathrm{off/on}} = R(L(I,\sigma_s) - L(I,\sigma_c)) \quad (2)$$

where $R(x) = 0$ if $x \leq 0$ otherwise $R(x) = x$, and $L(x,\sigma)$ is a low-pass filter implemented as the convolution of x with a Gaussian kernel (SD of σ pixels). For the simulations in this study, $\sigma_c = 0.5$ and $\sigma_s = 2.5$. Divisive inhibition is applied to the responses in each center–surround pyramid. The normalized response is given by:

$$N(r) = r^{\beta}/(r^{\beta} + s^{\beta}) \quad (3)$$

where r is the response at a given location, β is a power term and s is the crossover point where normalization goes from expansive to compressive. For the simulations in this study, $\beta = 2$ and the crossover point is set to a response magnitude corresponding to $\approx 10\%$ contrast ($s = 5$ given the images used in this study). To generate the first-order salience map, the normalized center–surround pyramids are collapsed across spatial scale and receptive field type by resizing each level in the pyramids to the resolution of the salience map (1/32 of the input resolution) and summing.

Second-order salience map

To generate a second-order salience map, capable of signalling texture contrast, a second stage of center–surround receptive fields is implemented. Four center–surround pyramids are created that receive input from the first-stage center–surround pyramids:

$$CS_{\mathrm{on(on/off)}} = R[L(CS_{\mathrm{on/off}},10\,\sigma_c) - L(CS_{\mathrm{on/off}},10\,\sigma_s)] \quad (4)$$

$$CS_{\mathrm{on(off/on)}} = R[L(CS_{\mathrm{on/off}},10\,\sigma_s) - L(CS_{\mathrm{on/off}},10\,\sigma_c)] \quad (5)$$

$$CS_{\mathrm{off(on/off)}} = R[L(CS_{\mathrm{off/on}},10\,\sigma_c) - L(CS_{\mathrm{off/on}},10\,\sigma_s)] \quad (6)$$

$$CS_{\mathrm{off(off/on)}} = R[L(CS_{\mathrm{off/on}},10\,\sigma_s) - L(CS_{\mathrm{off/on}},10\,\sigma_c)] \quad (7)$$

Given that second-order mechanisms operate at a spatial scale in the range of 8–15 times that of first-order mechanisms (Sutter *et al*., 1995; Zhou & Baker, 1996), we introduced an increase in spatial scale by a factor of 10. To generate the second-order salience map, the second-stage center–surround pyramids are summed across spatial scale and receptive field type.

Results

To examine the effects of local contrast modification on salience in the first- and second-order salience maps, Monte Carlo simulations were conducted. Each of the eight natural scenes used in the Einhäuser & König (2003) study were subjected to the contrast modification procedure with α ranging from -0.6 to $+1.0$. Each combination of image and modification level was simulated 25 times with different, randomly selected locations where local contrast manipulations were introduced. A salience map for both the first-order and second-order model was generated for each simulated case.

For each image, we computed the maximum change in salience between the salience map for the contrast-altered image and the salience map for the unaltered image and then divided by the standard deviation of the salience map for the unaltered image. We call this quantity the signal-to-noise ratio and plot it (averaged over simulation and image number) in Fig. 1 as a function of the contrast modification

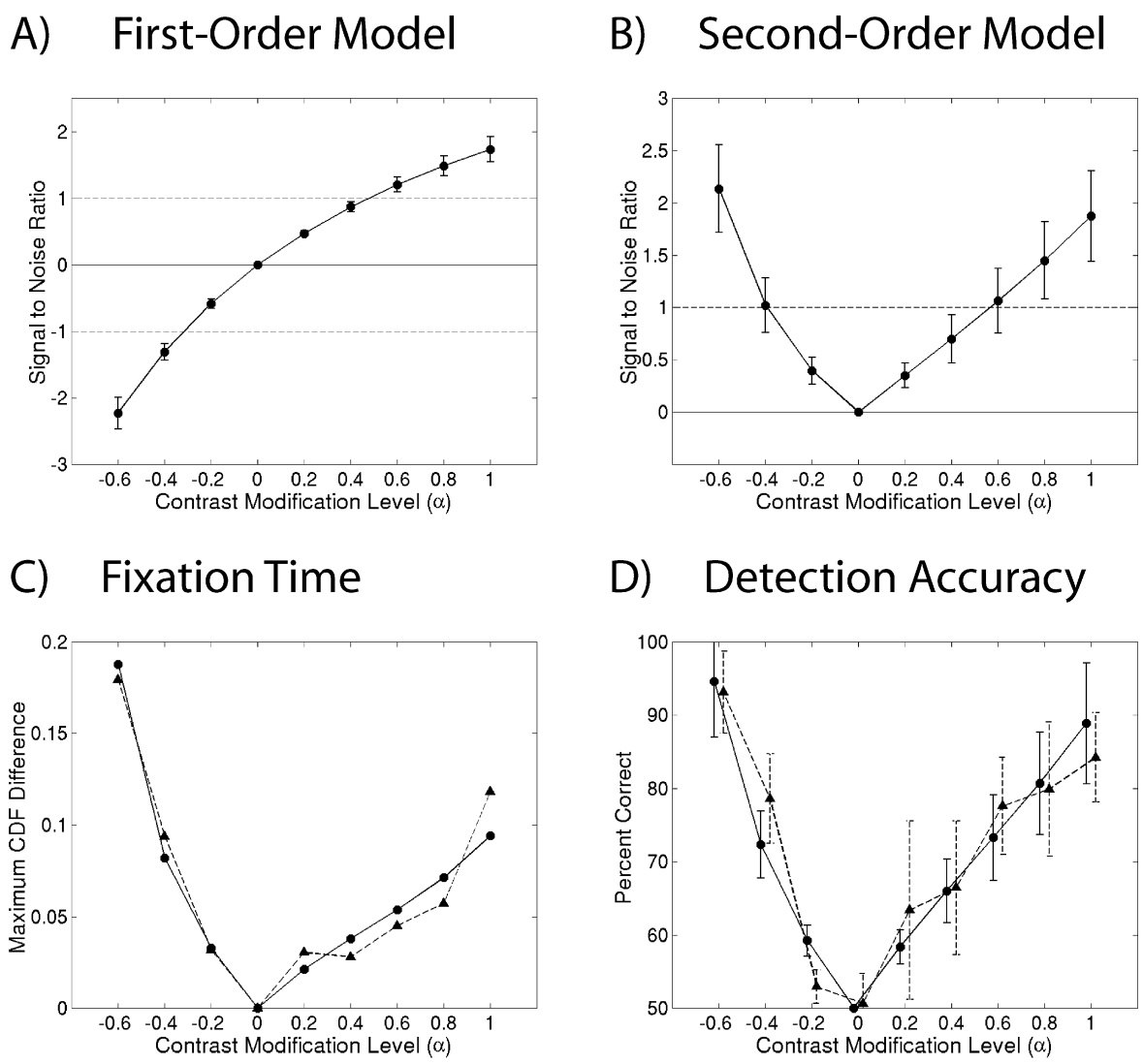

FIG. 1. Simulation results are shown for (A) the first-order model and (B) the second-order model. Signal-to-noise ratios are calculated as the maximum difference between the salience maps generated from the contrast-altered images and the unaltered images divided by the SD of the unaltered salience maps. The error bars represent ± 1 SD across images and thus indicate the variability due to the range of image characteristics inherent in the natural scenes used as stimuli. (C) The maximum differences between actual and control relative fixation time cumulative density functions. Note that for $\alpha = 0$, the analysis method guarantees that the difference is zero. (D) The accuracies in the detection of contrast manipulation experiment. (A–D) Observed data, triangles; modelled data, circles.

level. For signal-to-noise ratios <1 (or >-1 for the negative contrast modifications), the signal is indistinguishable from the noise. Signal-to-noise ratios of 1 and -1 are plotted as dashed lines in the figure. Note that the error bars indicate ± 1 SD of the signal-to-noise ratios taken across the average signal-to-noise ratios obtained for individual images. Thus the error bars are an indication of the variability resulting from the use of different images.

As expected, increasing contrast in the first-order model increases stimulus salience, while decreasing contrast decreases stimulus salience. This result is shown in Fig. 1A as positive and negative signal-to-noise ratios, respectively. A model of overt visual attention based solely on a first-order salience map predicts more fixations for regions of increased contrast and fewer fixations for regions of decreased contrast. This prediction is contrary to the behavioural results of Einhäuser & König (2003), i.e. more fixations for both increases and decreases in contrast, and argues against a bottom-up model based solely on a first-order salience map.

However, the observed pattern of results is predicted by a bottom-up model that bases stimulus salience on texture contrast. The second-order model shows positive signal-to-noise ratios for both positive and negative contrast manipulations. This result is shown in Fig. 1B. For this model, texture differences due to either a decrease or an increase in contrast both increase stimulus salience and thus attract attention.

While the second-order model based on texture contrast alone reproduces the basic pattern of results in the Einhäuser & König (2003) experiment, the question remains as to the relative contributions of luminance and texture contrast to the guidance of attention in this paradigm. To address this question, we fitted the signal-to-noise ratios taken from the combination of the first- and second-order models to the main measure reported by Einhäuser & König (2003), the maximum difference between the observed and control relative fixation time, cumulative density functions. [This measure is obtained by taking the maximum difference between two cumulative density functions, an 'observed' and a 'control' function. Each cumulative density function is calculated by measuring the relative time spent fixating the contrast modification levels present in a single image. Note that a range of contrast modification levels are present in a single image for a single

peak contrast manipulation level because contrast modifications are introduced with a Gaussian falloff around randomly selected locations. The 'observed' function is calculated for each peak contrast manipulation level used in the experiment while the 'control' function is calculated once by taking the same image but using the fixations obtained across all peak contrast manipulation levels. We fitted the data from session 4 (see fig. 3B in Einhäuser & König, 2003) to demonstrate that a bottom-up model can account for attentional guidance even after experience with the images, when any presumed top-down factors would be most influential. However, we note that fits of similar quality are obtained using the data from any session.]

Prior to fitting this combination, the additivity of salience must be considered. It is known that when salience is generated from different stimulus features the resulting salience is significantly less than the sum of the saliences resulting from these features presented in isolation (Nothdurft, 2000). This failure to sum linearly indicates a dependence in the mechanisms that generate salience. In other words, the salience derived from different stimulus features is to some degree redundant. Although the neural mechanism that underlies this effect is unknown, we can model this dependence in a biologically plausible way by assuming that, prior to summing, salience is divisively normalized. This is accomplished by dividing salience in the first-order map by the salience in the second-order map, and vice versa. The effect of this normalization is to scale the contribution of first-order salience to the overall salience relative to the contribution of second-order salience, and vice versa. We also assume in this fit that fixation time is linearly related to salience, as we have not explicitly modelled a saccade generation mechanism.

The optimal fit to the maximum difference between the observed and control relative fixation time density functions is obtained when the second-order map is weighted nine times more than the first-order map. A stronger or weaker weighting results in a reduction in the quality of the fit. The difference results from the Einhäuser & König (2003) study and the best fits of the model are shown in Fig. 1C. The error in this fit averages 0.004 units per data point.

We also addressed the question of the relative importance of luminance and texture contrast by fitting the accuracy results from the detection of contrast manipulations experiment (see fig. 5A in Einhäuser & König, 2003). To accomplish this, we assumed that accuracy is linearly related to salience. However, in detection experiments, observers can combine first-order and second-order cues linearly, and sometimes supra-linearly, to improve performance when visibility is low (Smith & Scott-Samuel, 2001), as it is in this paradigm. Therefore, we fitted the absolute magnitude of the signal-to-noise ratios taken from the linear combination of the first- and second-order models to the observed percentage correct data. Note that we used the absolute values for the model fitted to the accuracy data, but not for the fit to the fixation time density functions (as described above). In the case of contrast detection, we took the absolute magnitude of the signal-to-noise ratios because, in detection paradigms, observers can detect both contrast decrements and contrast increments equally well. In the case of attentional guidance we did not do this because, while stimuli made unique by larger-than-normal contrast values attract attention, stimuli made unique by smaller-than-normal contrast values do not attract attention (Treisman & Gormican, 1988).

The optimal fit to the observed percentage correct means and SDs is obtained when the second-order map is weighted 11 times more than the first-order map. The accuracy results from the Einhäuser & König (2003) study and the best fits of the model are shown in Fig. 1D. The error in this fit averages 2% correct per data point. Note that the optimal weighting, a factor of 11, obtained in this fit is similar to the weighting obtained in the previous fit, a factor of 9.

Discussion

We implemented a bottom-up model of attention and used this model to fit experimental data from the Einhäuser & König (2003) study. A model of attentional guidance based solely on luminance contrast cannot account for these results. However, a model in which both luminance contrast and texture contrast contribute to stimulus salience accounts well for these results. While both luminance and texture contrast contributed to the calculation of stimulus salience in the model, the optimal fits to the fixation results and the contrast detection results both indicate that texture contrast contributed approximately 10 times more to the generation of salience than did luminance contrast. While this result holds for this particular paradigm where the stimulus characteristics of natural scenes were artificially modified, it is not clear that it can be generalized to other paradigms or stimuli. In fact, we have shown that the relative contribution of different stimulus features to salience depends strongly on the properties of the stimulus (Parkhurst *et al*., 2002; Parkhurst & Niebur, 2003). Based on our previous results and the results of this study, we conclude that luminance contrast, texture contrast and a number of other stimulus features including color and orientation contribute to the generation of stimulus salience in a bottom-up fashion to influence the allocation of overt visual attention.

Our model is the first, as far as we are aware, that is capable of predicting the guidance of attention based on second-order features in natural scenes. This model implements second-order feature processing in a bottom-up fashion using a set of parallel feature maps. This model is reasonable given that a number of studies indicate that higher-order representations can guide attention in a bottom-up fashion. For example, it is the onset of new perceptual objects (as opposed to onsets in general) that capture attention (Yantis & Hillstrom, 1994; Yantis & Jonidas, 1996; Rauschenberger & Yantis, 2001). Moreover, observers preferentially fixate locations with two-dimensional image features such as T-junctions and corners when viewing natural scenes (Krieger *et al*., 2000).

While some research indicates that the processing of second-order stimuli is dependent on attention (Yeshurun & Carrasco, 2000) and that second-order stimuli do not lead to rapid visual search as would be expected if second-order stimuli were processed preattentively (Ashida *et al*., 2001), these results do not imply that second-order features influence attention in a top-down fashion. Rather, they imply that attention influences the bottom-up processing of second-order stimuli. This conclusion is consistent with the evidence that early visual areas in the brain, where bottom-up processing occurs, are known to be modulated by attention (for review, see Kastner & Ungerleider, 2000). It is important to note that the characterization of a mechanism of attentional guidance as bottom-up or top-down is independent of its attentional requirements. A mechanism is best characterized as being bottom-up when it is primarily dependent on stimulus properties rather than other factors, which are largely independent of the stimulus properties, such as semantic associations. This distinction is clear when one considers that not only is bottom-up processing influenced by attention, but that top-down processing can occur in the absence of attention. For example, a remembered scene context can implicitly guide attention to a target during visual search when the context is reliably associated with the target's location or identity (for review, see Chun, 2000).

Conclusions

In natural vision, mechanisms of visual attention select information from the visual scene for detailed processing. Many studies using

simple, well-controlled stimuli have indicated that stimulus features such as color, orientation and intensity can act to guide this selection process in a bottom-up fashion, and now a growing body of studies are supporting a similar conclusion for the guidance of attention in natural scenes. However, Einhäuser & König (2003) recently concluded from the pattern of eye movements recorded in a novel experimental paradigm using natural scenes that luminance contrast does not contribute to stimulus salience. We discussed a number of methodological problems with this paradigm and suggested alternative explanations for the observed pattern of eye movements in this paradigm. We then implemented a simple model of visual selective attention to explicitly test the hypothesis that both luminance contrast and texture contrast contribute to stimulus salience. We found that this model is consistent with the observed pattern of behavioural results. We conclude that both luminance contrast and texture contrast contribute to the generation of visual salience and play a role in determining the allocation of overt visual attention.

We emphasize that we do not claim that stimulus salience alone can account for all attentional guidance. Indeed, a significant amount of research with simple, well controlled stimuli indicates that both bottom-up and top-down factors influence the guidance of attention. We believe that the same is true of attentional guidance in natural scenes. The only question that remains is to what degree do bottom-up and top-down factors account for attentional allocation and under what experimental conditions does the balance of control shift. We have shown that bottom-up mechanisms of visual selective attention can account for a significant degree of attentional guidance in a free-viewing paradigm using complex, natural and artificial scenes (Parkhurst *et al*., 2002). Other studies have demonstrated a significant influence of top-down factors when participants are required to perform complex tasks such as driving, making a cup of tea or building a model (Land & Lee, 1994; Land & Hayhoe, 2001; Pelz & Canosa, 2001). More research using a variety of natural stimuli under natural viewing conditions and a variety of tasks is needed to further address this question.

The proposition raised by Einhäuser & König (2003) that stimulus features in natural scenes may in fact be associated with top-down representations is of interest. Unfortunately, we know of no quantitative experimental evidence presently available that directly addresses this question. It is not implausible that, through evolution or visual experience, bottom-up mechanisms have become tuned to those stimulus features which are strongly correlated with top-down representations. Given that top-down mechanisms generally have a slow time-course, it would be advantageous, when possible, to preempt these mechanisms and guide attention to behaviourally relevant stimulus features using a rapid bottom-up mechanism. If this is the case, evidence of an association between stimulus features and top-down representations cannot rule out guidance of attention through bottom-up mechanisms. Furthermore, evidence that attention has been allocated to task-relevant stimuli cannot necessarily be interpreted as evidence of top-down attentional guidance, as appropriately tuned bottom-up mechanisms could be at hand.

Given the inherent difficulty of studying attentional allocation in natural scenes, we feel strongly that computational modelling of visual processing is an important tool. Computational models allow for explicit and quantitative implementations of conceptual hypotheses. These models can be effectively used to make predictions for complex, natural stimuli, where intuition sometimes fails. In our work, we have constrained our model's design using psychophysical as well as neurobiological evidence. However, much is still unknown about the implementation of visual selective attention, for example, whether the salience map is implemented in one anatomically defined area of the brain or whether the salience map is a functional concept and implemented in multiple areas, as suggested by Desimone & Duncan (1995). In either case, it is clear that visual salience plays an important role in the guidance of attention.

Acknowledgements

We thank W. Einhäuser and P. König for providing the natural scenes used in their study. This material is based upon work supported by the National Science Foundation under CAREER Grant no. 9876271 to E.N. D.P. was supported by an NIH-NEI visual neuroscience training fellowship.

References

Andrews, T.J. & Coppola, D.M. (1999) Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Res.*, **39**, 2947–2953.

Ashida, H., Seiffert, A. & Osaka, N. (2001) Ineffcient visual search for second-order motion. *J. Opt. Soc. Am. A*, **18**, 2255–2266.

Bacon, W.F. & Egeth, H.E. (1994) Overriding stimulus-driven attentional capture. *Percept. Psychophys.*, **55**, 485–496.

Burt, P.J. & Adelson, E.H. (1983) The laplacian pyramid as a compact image code. *IEEE Trans. Comms*, **31**, 532–540.

Carandini, M. & Heeger, D. (1994) Summation and division by neurons in primate visual cortex. *Science*, **264**, 1333–1336.

Carandini, M., Heeger, D.J. & Movshon, J.A. (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.*, **17**, 8621–8644.

Cavanagh, P. & Mather, G. (1989) Motion: the long and short of it. *Spatial Vision*, **4**, 103–129.

Chubb, C. & Sperling, G. (1988) Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *J. Opt. Soc. Am. A*, **5**, 1986–2007.

Chun, M.M. (2000) Contextual cueing of visual attention. *Trends Cogn. Sci.*, **4**, 170–177.

Desimone, R. & Duncan, J. (1995) Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.*, **18**, 193–222.

Deubel, H. & Schneider, W.X. (1996) Saccade target selection and object recognition: Evidence for a common attenional mechanism. *Vision Res.*, **36**, 1827–1837.

Einhäuser, W. & König, P. (2003) Does luminance-contrast contribute to a salience map for overt visual attention? *Eur. J. Neurosci.*, **17**, 1089–1097.

Folk, C.L., Remington, R. & Johnston, J.C. (1992) Involuntary covert orienting is contingent on attentional control settings. *J. Exp. Psychol. Hum. Percept. Perf.*, **18**, 1030–1044.

Folk, C.L., Remington, R. & Wright, J.H. (1994) The structure of attentional control: Contingent attentional capture by apparent motion, abrupt onset, and color. *J. Exp. Psychol. Hum. Percept. Perf.*, **20**, 317–329.

Henderson, J.H. & Hollingworth, A. (1999) High-level scene perception. *Annu. Rev. Psychol.*, **50**, 243–271.

Hillstrom, A.P. & Yantis, S. (1994) Visual motion and attentional capture. *Percept. Psychophys.*, **55**, 399–411.

Hoffman, J.E. & Subramaniam, B. (1995) The role of visual attention in saccadic eye movements. *Vision Res.*, **57**, 787–795.

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

Itti, L., Koch, C. & Niebur, E. (1998) A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Machine Intelligence*, **20**, 1254–1259.

Kastner, S. & Ungerleider, L. (2000) Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.*, **23**, 315–341.

Koch, C. & Ullman, S. (1985) Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.*, **4**, 219–227.

Kowler, E., Anderson, E., Dosher, B. & Blaser, E. (1995) The role of attention in the programming of saccades. *Vision Res.*, **35**, 1897–1916.

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**, 201–214.

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

Land, M.F. & Lee, D.N. (1994) Where we look when we steer. *Nature*, **369**, 742–744.

McPeek, R.M., Maljkovic, V. & Nakayama, K. (1999) Saccades require focal attention and are facilitated by a short-term memory. *Vision Res.*, **39**, 1555–1566.

Nakayama, K. & Silverman, G.H. (1986) Serial and parallel processing of visual feature conjunctions. *Nature*, **320**, 264–265.
Niebur, E. & Koch, C. (1996) Control of selective visual attention: Modeling the 'where' pathway. In Touretzky, D.S., Mozer, M.C. & Hasselmo, M.E. (eds), *Advances in Neural Information Processing Systems*, **Vol. 8**. MIT Press, Cambridge, MA, pp. 802–808.
Nothdurft, H.-C. (2000) Salience from feature contrast: additivity across dimensions. *Vision Res.*, **40**, 1183–1201.
Noton, D. & Stark, L. (1971) Scanpaths in eye movements. *Science*, **171**, 308–311.
Parkhurst, D. (2002) Selective attention in natural vision: Using computational models to quantify stimulus-driven attentional allocation. PhD Thesis, The Johns Hopkins University, Baltimore, MD.
Parkhurst, D., Law, K. & Niebur, E. (2002) Modeling the role of salience in the allocation of overt visual selective attention. *Vision Res.*, **42**, 107–123.
Parkhurst, D.J. & Niebur, E. (2003) Scene content selected by active vision. *Spatial Vision*, **6**, 125–154.
Pelz, J.B. & Canosa, R. (2001) Oculomotor behavior and perceptual strategies in complex tasks. *Vision Res.*, **41**, 3587–3596.
Posner, M.I. (1980) Orienting of attention. *Q. J. Exp. Psychol.*, **32**, 3–25.
Rauschenberger, R. & Yantis, S. (2001) Attentional capture by globally defined objects. *Percept. Psychophys.*, **63**, 1250–1261.
Rock, I. & Gutman, D. (1981) The effect of inattention on form perception. *J. Exp. Psychol. Hum. Percept. Perf.*, **7**, 275–285.
Ruderman, D.L. & Bialek, W. (1994) Statistics of natural images: scaling in the woods. *Phys Rev. Lett.*, **73**, 814–817.
Schofield, A. (2000) What does second-order vision see in an image? *Perception*, **29**, 1071–1086.
Shepherd, M., Findlay, J.M. & Hockey, R.J. (1986) The relationship between eye movements and attention. *Q. J. Exp. Psychol.*, **38A**, 475–491.
Smith, A. & Scott-Samuel, N. (2001) First-order and second-order signals combine to improve perceptual accuracy. *J. Opt. Soc. Am. A*, **18**, 2267–2272.
Sutter, A., Sperling, G. & Chubb, C. (1995) Measuring the spatial frequency selectivity of second-order texture mechanisms. *Vision Res.*, **35**, 915–924.
Theeuwes, J. (1990) Perceptual selectivity is task dependent: evidence from selective search. *Acta Psychologica*, **74**, 81–99.
Theeuwes, J. (1994) Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets. *J. Exp. Psychol. Hum. Percept. Perf.*, **20**, 799–806.
Treisman, A. & Gelade, G. (1980) A feature-integration theory of attention. *Cogn. Psychol.*, **12**, 97–136.
Treisman, A. & Gormican, S. (1988) Feature analysis in early vision: Evidence from search asymetries. *Psychol. Rev.*, **95**, 15–48.
Wang, Q., Cavanagh, P. & Green, W. (1994) Familiarity and pop-out in visual search. *Percept. Psychophysics*, **56**, 495–500.
Yantis, S. & Egeth, H.E. (1999) On the distinction between visual salience and stimulus-driven attentional capture. *J. Exp. Psychol. Hum. Percept. Perf.*, **25**, 661–676.
Yantis, S. & Hillstrom, A.P. (1994) Stimulus-driven attentional capture: Evidence from equilumunant visual objects. *J. Exp. Psychol. Hum. Percept. Perf.*, **20**, 95–107.
Yantis, S. & Jonidas, J. (1996) Attentional capture by abrupt onsets: New perceptual objects or visual masking? *J. Exp. Psychol. Hum. Percept. Perf.*, **22**, 1505–1513.
Yantis, S. & Jonides, J. (1984) Abrupt visual onsets and selective attention: Evidence from visual search. *J. Exp. Psychol. Hum. Percept. Perf.*, **10**, 601–621.
Yantis, S. & Jonides, J. (1996) Attentional capture by abrupt visual onsets: New perceptual objects or visual masking? . *J. Exp. Psychol. Hum. Percept. Perf.*, **22**, 1505–1513.
Yarbus, A. (1967) *Eye Movements and Vision*. Plenum Press, New York.
Yeshurun, Y. & Carrasco, M. (2000) The locus of attentional effects in texture segmentation. *Nature Neurosci.*, **3**, 622–627.
Zhou, Y.X. & Baker, C.L.J. (1996) Spatial properties of envelope-responsive cells in area 17 and 18 neurons of the cat. *J. Neurophysiol.*, **75**, 1038–1050.