Spatiotemporal processing of saliency signals in the primate: a behavioral and neurophysiological investigation

by

David J. Berg

A Dissertation Presented to the
FACULTY OF THE PROVOST
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(NEUROSCIENCE)

January 2013

Dedication

I dedicate this work to all the educators, advisers and mentors that have sparked my curiosity to learn, create, and explore the world around me. Most importantly, my father. He taught me at a young age that I could make my imagination come to life with a little ingenuity.

Acknowledgements

I thank everyone at the Doug Munoz Laboratory at Queen's University, Ontario, Canada; particularly Susan Boehnke, Brian White and Robert Marino. The scientific collaborations they provided were paramount to completing the work in this thesis. I truly enjoyed all our interactions both scientific and social. I'd also like to thank Laurent Itti for his academic and financial support, and everyone at iLab for their keen insights, friendly demeanor, and the generally inspiring atmosphere they provided.

Table of Contents

List Of Figures

Summary

To quickly locate and discriminate visual events important for an organisms survival is a complex task. The visual systems of mammals evolved specialized heuristics to quickly aid in detecting visually salient items that may be relevant for survival. Salient items differ statistically from their background, and tend to 'pop-out', or draw an observer's attention automatically (such as red flower among green grass, or the abrupt onset of a light). A standard computational model of saliency processing has been implemented and has served as a quantitative framework to study salient item detection in primates during free viewing of natural scenes. This thesis consists of three experiments that address the spatiotemporal processing of visual saliency signals in the monkey (*Macaca mulatta*), with an emphasis on assessing models of saccadic behavior and neural processing during free viewing of natural scenes.

In the first experiment monkeys freely watched videos of natural scenes while their eye movements were recorded. A computational saliency model received the same video input and made predictions about which screen locations were the most attention grabbing. The study found that the saliency model was predictive of monkey gaze significantly above chance, and that although differences were found, a strong correspondence could be made between results obtained from humans and monkeys. The first experiment established the

free-viewing paradigm in monkeys, and confirmed the general computational approach by validating the predictive power of the computational saliency model against monkey eye movements. This was important because saliency is usually studied in non-human primates under fixation control with simple stimuli, and it was necessary to behaviorally validate the methods and model before probing the brain directly.

Evidence suggests the primate superior colliculus (SC) may play an important role in saliency processing. The SC has an established role in integrating sensory signals for the control of saccades and attention. Its superficial layers receive visual input from most of the early visual brain, and its intermediate layers receive more complex visual and cognitive input from cortex. The SC outputs to the brainstem circuitry in control of saccadic eye-movements and has recently been shown to be important for the selection of items for attentional deployment.

The second experiment was designed to understand the temporal processing of visual signals in the primate SC. A recently proposed theory of temporal saliency computation (surprise) predicted eye movements in humans significantly better than previous saliency models, suggesting adaptation is key to salient item detection. Spiking activity of neurons in the primate SC was monitored while stimuli were repeatedly flashed into the receptive fields of cells. A reduction in the magnitude of the initial transient neural response was observed with stimulus repetition for all visually responsive neurons in the SC. Response decrement was successfully captured by the surprise model which also predicted the effects of presentation rate and rare luminance changes. This experiment was important to understand the temporal aspects of SC neural processing, which had not been fully characterized previously.

In the final experiment, we directly tested the hypothesis that the SC represents visually salient items in natural scenes by using a combination of computational modeling and single-unit monkey electrophysiology. A new computational model of saliency processing, tailored to the SC, was built based on previous findings. Monkeys freely viewed videos of natural scenes presented on a large, high-definition display. Recordings of the monkey's eye position were used to replay to the computational model, the exact, gaze-contingent stimulus that impinged onto the monkey's retina. We found that the spike rates of 35 of 39 cells in the SC were significantly predicted by the saliency model and that during fixations, neural responses could be rank ordered by their saliency responses. To test the necessity of saliency and the importance of each feature, responses were computed for models that only processed individual stimulus features or lacked features. These models performed poorly for individual neurons and across the population of SC cells, suggesting a feature sensitive but non-specific representation. Taken together, the results indicate that during free viewing of natural stimuli SC represents the visual saliency of items in the world.

Chapter 1

Freeviewing of dynamic stimuli by humans and monkeys

1.1 Abstract

Due to extensive homologies, monkeys provide a sophisticated animal model of human visual attention. However, for electrophysiological recording in behaving animals simplified stimuli and controlled eye position are traditionally used. To validate monkeys as a model for human attention during realistic free viewing we contrasted human (n=5) and monkey (n=5) gaze behavior using 115 natural and artificial video clips. Monkeys exhibited broader ranges of saccadic endpoints and amplitudes, and showed differences in fixation and intersaccadic intervals. We compared tendencies of both species to gaze towards scene elements with similar low-level visual attributes using two computational models - luminance contrast and saliency. Saliency was more predictive of both human and monkey gaze, predicting human saccades better than monkey saccades overall. Quantifying interobserver gaze consistency revealed that while humans were highly consistent, monkeys were more heterogeneous and were best predicted by the saliency model. To

address these discrepancies, we further analyzed high-interest gaze targets - those locations simultaneously chosen by at least two monkeys. These were on average very similar to human gaze targets, both in terms of specific locations and saliency values. Although substantial quantitative differences were revealed, strong similarities existed between both species, especially when focusing analysis onto high-interest targets.

1.2 Introduction

Monkeys are widely used as animal models for the study of human cognitive processes, such as visual attention, due to the neural homologies between the species. More and more, there is a shift towards studying vision using natural and dynamic stimuli. When the visual system is examined using such stimuli it responds differently than it does to simple stimuli traditionally used in the laboratory [for reviews see Felsen and Dan, 2005, Kayser et al., 2004, Reinagel, 2001, Simoncelli and Olshausen, 2001]. The system also responds differently when monkeys view such stimuli freely [Dragoi and Sur, 2006, Gallant et al., 1998, Vinje and Gallant, 2000]. What is not yet known is whether humans and monkeys *behave* similarly under such natural viewing conditions. This is important because, although there are similarities in the early stages of visual processing, cortical architecture differences exist in parietal and frontal areas related to attention and cognitive processing [Orban et al., 2004].

Computational models [Itti et al., 1998, Le Meur et al., 2006, Privitera and Stark, 2000] provide a quantitative framework to assess visual behavior and compare species under complex stimulus conditions. For example, model output for the scene can be investigated at actual saccadic target locations. Simple image statistics (such as local contrast, orientation) and deviations from global image statistics exhibit differences between fixated vs non-fixated locations [Parkhurst and Niebur, 2003, Reinagel et al., 1999], and these statistics are factors in guiding attention. Such experiments have been done with monkeys [Dragoi and Sur, 2006] and humans [Itti and Baldi, 2005, Parkhurst et al., 2002, Peters et al., 2005, Tatler et al., 2005] separately. However, viewing behavior has yet to be compared directly using a wide set of complex dynamic natural stimuli (video).

To investigate species correspondence, [Einhäuser et al., 2006] compared 2 monkeys and 7 humans who repeatedly viewed static, grayscale natural images. Computational models equally predicted the species gaze shifts, however, differences in viewing strategies were observed when local image contrast was manipulated. Here, we expand on this significantly by comparing human and monkey free-viewing behavior using video clips ranging in semantic content and species relevance. Additionally, the main computational model of viewing behavior, the saliency model, was adapted to better account for the temporal dynamics of video [Itti, 2006]. We also measured consistency among observers gaze, which provided context specific predictions of saccadic targets that complement the stimulus-driven predictions of the saliency model.

Our results demonstrate correlations between saliency and both human and monkey visual behavior; however, marked differences exist between species in eye movement statistics, model correspondence and interobserver consistency. These differences must be considered when using monkeys as a model of human attention during free viewing. We find that focusing analysis on a subset of high-interest gaze locations - to which two or more monkeys looked simultaneously - can alleviate such differences. We speculate that high-interest locations reveal commonalities between both species, possibly by emphasizing the role of their largely homologous and common low-level visual systems over their likely more different and individualized cognitive systems.

1.3 Methods

1.3.1 Subjects

Eye movements during free viewing were recorded from five human (two male) and five monkey (*Macaca Mulatta*, all male) subjects. Human subjects provided informed consent under a protocol approved by the Institutional Review Board of the University of Southern California. Monkeys were used with approval by the Queens University Animal Care Committee and were in accordance with the Canadian Council on Animal Care policy on the use of laboratory animals and the Policies on the Use of Animals and Humans in Neuroscience Research of the Society for Neuroscience.

1.3.2 Stimulus Presentation

Naïve subjects (both human and monkey) watched 115 video clips (totaling approx. 27 minutes in duration, played in random order) that varied in duration and semantic content. The clips were subjectively categorized into six coarse semantic groups (Building/City, Natural, Sports, Indoor, Non-natural, and Monkey-relevant), as shown in Figure 1.1. Stimuli were collected from television (NTSC source) with a commercial framegrabber (ATI Wonder Pro). Monkey relevant clips were collected at the Queens University animal care facility with a consumer grade digital video camera. Frames were acquired and stored at 30 Hz in raw 640x480 RGB555 format and compressed to MPEG-1 movies (640x480 pixels). Stimuli were presented to human subjects, with head stabilized by a chin rest, on a 101.6 x 57.2 cm LCD TV (Sony Bravia) at a viewing distance of 97.8 cm. This provided a usable field-of-view of 54.9°x32.6°, which was the largest the video-based human eye-tracker could accommodate. Stimulus presentation was orchestrated using a Linux computer running in house programmed presentation software (downloadable at http://iLab.usc.edu/toolkit) under SCHED_FIFO scheduling to ensure proper frame rate presentation [Finney, 2001, Itti and Baldi, 2005]. Subjects were given minimal instructions "watch and enjoy the video clips, try to pay attention, but don't worry about small details." Each video presentation was preceded by a fixation point, and the next video began when the subject pressed the space bar.

The exact same stimuli were also presented via the same Linux system to head-restrained monkeys who were seated 60 cm from a Mitsubishi XC2935C CRT monitor (71.5 x 53.5 cm; 640 x 480 pixels). This provided a usable field-of-view of 61.6°x48.1°. Trial initiation was self-paced. Each video presentation was preceded by a fixation point and the next video was initiated when the monkeys eye position remained within a square electronic window with 5°radius of the central fixation point for 300-500 ms. The monkey subjects were not rewarded systematically for doing this task, but most monkey subjects easily learned to fixate in order to initiate the next clip.

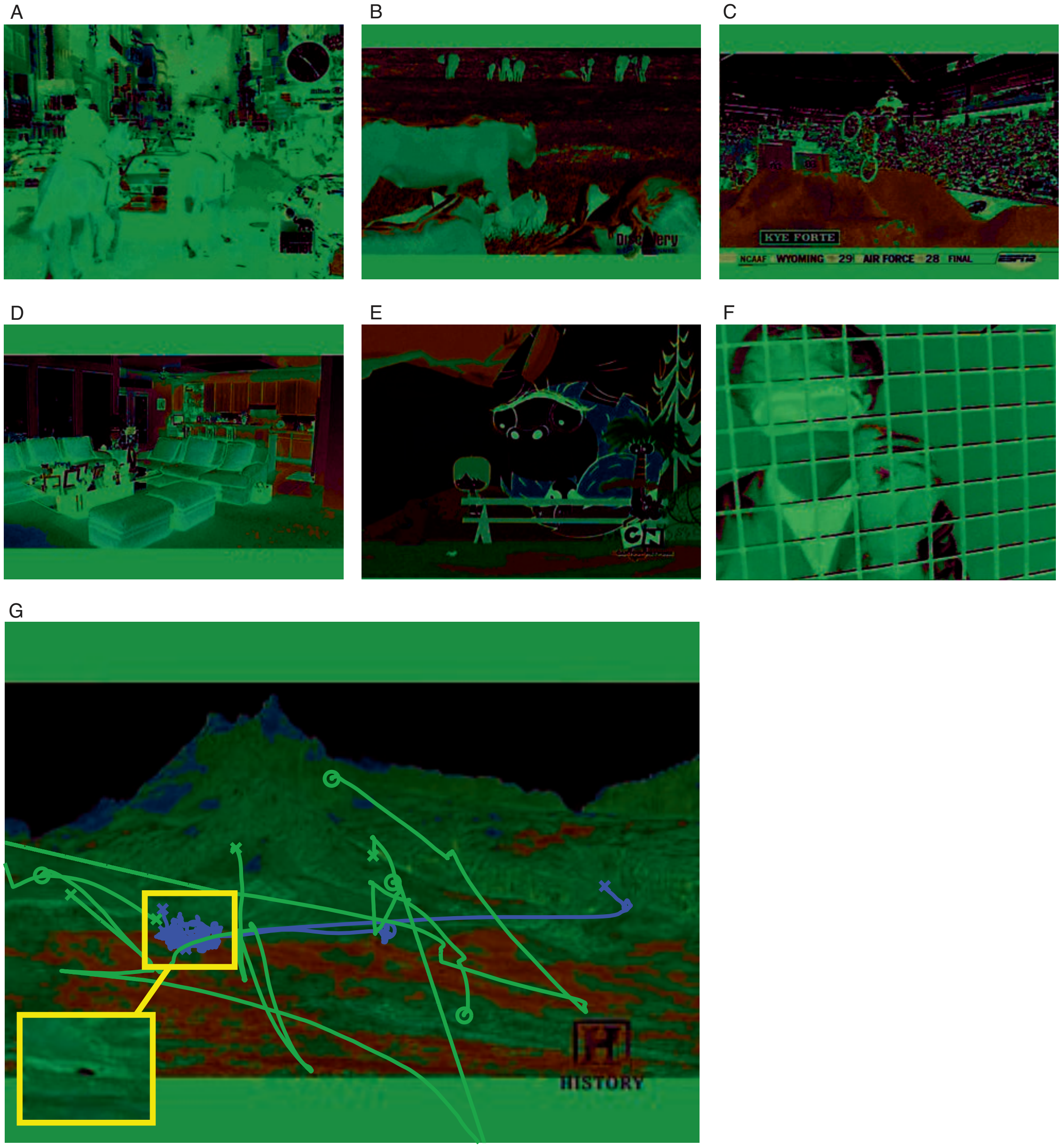


Figure 1.1: The six categories of scene types. Exemplars are shown from the six categories of scene types: (A) building and city, (B) natural, (C) sports, (D) indoor, (E) non-natural (cartoons, random noise, space), and (F) monkey relevant (monkeys, experimenters, facilities). Each group contains scenes with and without main actors (*e.g.* empty room vs talk show). (G) Shows an example of eye movement traces from 4 humans (blue) and 4 monkeys (green) superimposed on a video clip during a relatively stationary three second period. Notice monkeys looked around the screen while humans focused their gaze on the slowly moving car in the background (inset with yellow box).

1.3.3 Human eye-tracking procedure

Human eye movements were recorded using an infrared-video-based eye-tracker (ISCAN RK-464). Pupil and corneal reflection of the right eye were used to calculate gaze position with an accuracy 1°, sampled at 240 Hz. To calibrate the system, subjects were asked to fixate on a central point and then saccade to one of nine target locations distributed across the screen on a 3x3 grid. This procedure was repeated until each location was visited twice. In subsequent offline analysis, the endpoints of saccades to targets were used to perform an affine transform followed by a thin-plate-spline interpolation [Itti and Baldi, 2005] on the eye position data obtained in the free-viewing experiment in order to yield accurate estimate of eye position given the geometry of the eye-tracker and display. Re-calibration was performed every 13 movie clips during the experiment.

1.3.4 Monkey eye-tracking procedure

A stainless steel head post was attached to the skull via an acrylic implant anchored to the skull by stainless steel screws. Eye coils were implanted between the conjunctiva and the sclera of each eye [Judge et al., 1980] allowing for precision recording of eye position using the magnetic search coil technique [Robinson, 1963]. Surgical methods for preparing animals for head-fixed eye movement recordings have been described previously [Marino et al., 2008]. Monkeys were seated in a primate chair with their heads restrained for the duration of an experiment (2-4 hours). Eye position data was digitized at 1000 Hz using data acquisition hardware by Plexon, Inc. Concurrently, timestamps of the time of fixation point onset, acquisition of the fixation target by the monkey, and initiation of the clip were recorded.

To calibrate eye position, monkeys performed a step saccade paradigm in which targets at three eccentricities and eight radial orientations from the fixation point were presented in random order. Monkeys were given a liquid reward if they fixated a target within an square electronic window of 4°radius within 800 ms. During calibration, behavioral paradigms and visual displays were controlled by two Dell 8100 computers running UNIX-based real-time data control and presentation systems [Rex 6.1: Hays Jr et al., 1982]. In order to control for small non-linearities in the field coil, the weighted average of several visits to each target endpoint were later used to perform an affine transform and thin-plate-spline interpolation on the eye position data collected during free viewing of the video clips.

1.3.5 Quantifying eye-movement behavior

In order to quantify viewing behavior, an algorithm was used for both species which parsed the analog eye position data into saccadic, fixational and smooth pursuit eye movements. Traditional techniques to separate these various eye movements did not work well with these data, because many of the eye movement patterns elicited during free viewing of dynamic stimuli were non-traditional (*e.g.* blends of smooth pursuit, optokinetic, and saccadic eye movements). To deal with such idiosyncrasies, standard velocity measurements were combined with a simple windowed Principal Components Analysis (PCA). The eye position data was first smoothed (63 Hz Lowpass Butterworth), and eye positions with velocities greater than 30 deg/sec were marked as possible saccades. Within a sliding window, the PCA was computed and the ratio of explained variances (minimum over maximum) for each of the two dimensions was stored. A ratio near zero indicates a straight line, and hence a likely saccade. The results of several different window sizes were linearly combined to produce a robust and smooth estimate. Eye positions with a ratio near zero, but with insufficient velocity to be marked as a saccade were labeled as smooth pursuit. Remaining data was marked as fixation. Saccades with short ($<$ 80 ms) intervening fixations or smooth pursuits and small differences in saccadic direction ($< 45^\circ$) were assumed to represent re-adjustments of gaze en route to a target, and so were combined into a single saccadic eye movement toward the final target, rather than two or more separate saccades. Additionally, saccades of $< 2^\circ$ in amplitude and $<$ 20 ms in duration were removed in order to decrease the false positive rate of saccade parsing, and to focus analysis on eye movements that more likely reflected a shift of attention to a new target as opposed to minor gaze adjustments on a current target [Itti and Baldi, 2005]. This saccade parsing algorithm is freely available as part of the stimulus presentation software.

For each subject (human or monkey), clips which contained excessive durations ($>$ 30% of clip length) of tracking loss (blinks, loss of signal from search coil or video based tracker) or off-screen eye movements (sleeping, inattentive behavior) were excluded from analysis. The majority of monkey clips were rejected for excessive off-screen eye position (18.6% of the monkey data, .7% for humans). 11.8% of the monkey data (1.4% for humans) was discarded for loss of tracking. In monkeys, the implanted search coil still produces a signal when a subject is in a blink, however, strain on the coil due its implanted position (along with other noise factors) will cause some loss of tracking. Due to technical errors, data was not recorded for 17 clips for 1 monkey and 2 clips for another, accounting for 3.3% of the monkey data. In total, 1.9% of human and 27.3% of monkey eye traces were rejected. Note that the individual rejection percentages do not add to the total percentage rejected due to overlap between clips containing tracking loss and off-screen data. Analysis was consequently performed on different subsets of clips for each observer with the limitation that at least three observers from each species had to have successfully viewed each clip for it to be retained in the analysis.

1.3.6 Implementation of computational models

To assess the visually-guided behavior of humans and monkeys, two validated computational models of visual attention (contrast and saliency) and an interobserver consistency metric were used to predict individual eye movements (Figure 1.2). Models were created and run under Linux using the iLab C++ Neuromorphic Vision Toolkit [Itti, 2004]. First, ait luminance contrast model [Reinagel et al., 1999], defined as the variance of pixel values in 16x16 pixel patches tiling the input image frame (Figure 1.2, left), is a simple, but non-trivial model of attention and serves as a control for the performance of the saliency model. Second, we used the saliency model of visual attention framework [Figure 1.2, center; Itti and Koch, 2000, Itti et al., 1998]. The Itti and Koch model computes salient locations by filtering the movie frames along several feature dimensions (color, intensity, orientation, flicker and motion). Center-surround operations in each feature channel highlight locations which are different from their surroundings. Finally, the channels are normalized and linearly combined to produce a saliency map, which highlights screen locations likely to attract the attention of human or monkey observers. To process our video clips, we used the latest variant of the saliency model which uses Bayesian learners to detect locations that are not only salient in space, but are also salient (or so-called "surprising") over time [Itti, 2006]. This model hence substantially differs from and generalizes other models of stimulus-driven attention [Itti et al., 1998, Le Meur et al., 2006, Privitera and Stark, 2000, Tatler et al., 2005] in that both spatial and temporal events within each feature map that violate locally accumulated beliefs about the input cause high output for that location.

The contrast model contains no temporal dynamics, and consequently would not be expected to outperform the saliency model. Since many simple models would perform significantly above chance, we use the contrast model as a lower bound of performance for any non-trivial model of attention. Additionally, luminance contrast is correlated with many features used in the saliency computation. Comparing the static luminance contrast model with the saliency model gives some insight into the contribution of the dynamic features irrespective of luminance contrast.

To compute a measure of gaze agreement among and between species, an interobserver metric was created separately for each species using a leave-one-out approach (Figure 1.2, right). A master map is created by placing Gaussian blobs (σ =48 pixels) centered at the instantaneous eye positions of a subset of human or monkey observers. For each subject a map is created from the eye positions of the 2-4 other subjects in the same species who viewed the clip. A maximum output for this map is achieved when all subjects look at the same item simultaneously. This map represents a combination of stimulus-driven and goal-directed eye movements and has been used as an upper bound for human gaze prediction [Itti, 2006].

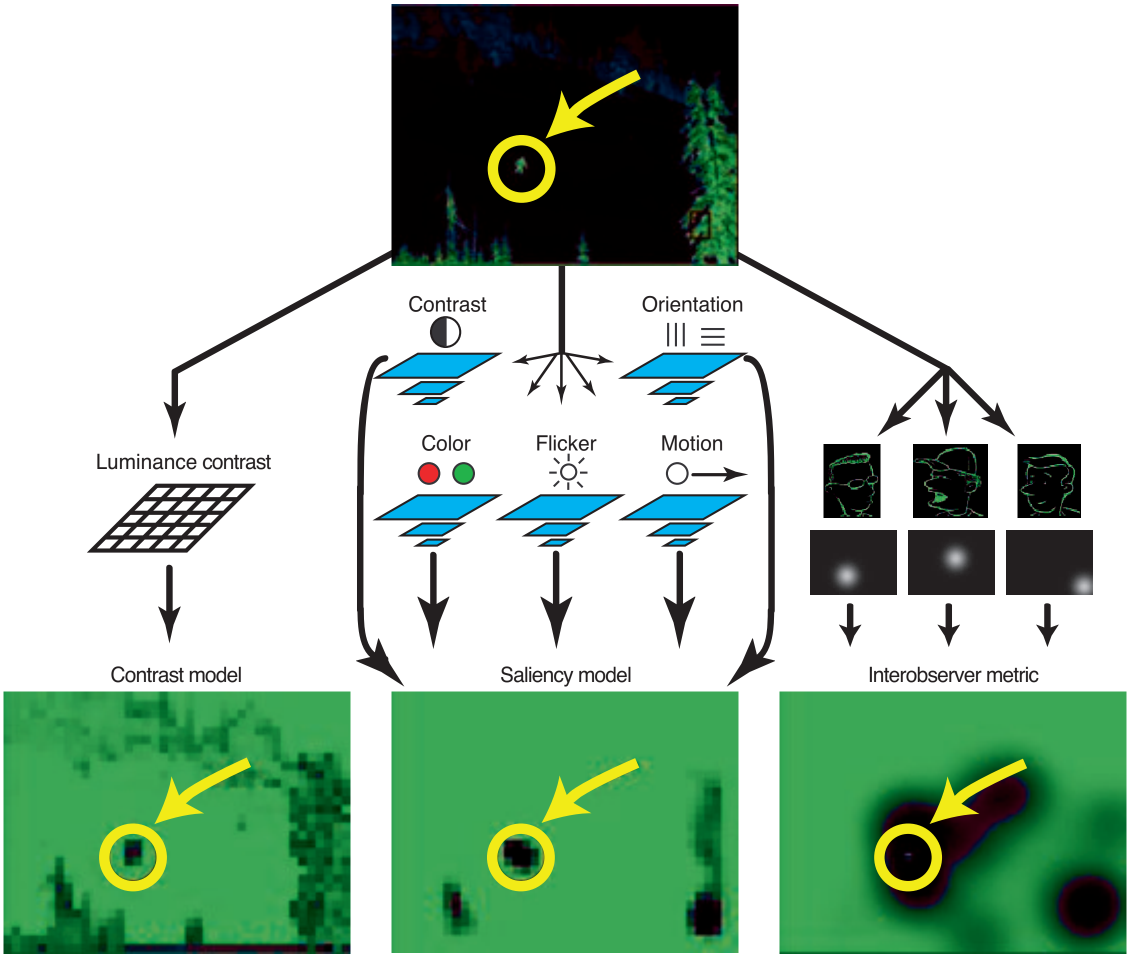


Figure 1.2: Architecture of the contrast and saliency models, and interobserver agreement metric. Left, a simple luminance contrast model computed as the variance of luminance values in 16x16 pixel image patches. Center, the latest implementation of the saliency model [Itti, 2006]. Right, an interobserver agreement metric (see Methods) created by making a heat map from the pooled eye movements of all observers, except the one under test, on a given movie clip (leave-one-out analysis). The yellow circle indicates the endpoint of a saccadic eye movement. At the start of the saccade the maximum value within a 48 pixel radius circular aperture was stored along with 100 values chosen randomly from the saccadic endpoint distribution of all clips and subjects except for the one under test. To test for agreement between or among species the interobserver agreement metric was sampled at the time when the eye landed at its target.

1.3.7 Comparing eye movements to model and metric output

To compute the performance of each model or metric the maximum map values in a circular window (3.6°humans, 4.7°monkeys: A 48 pixels window, but different viewing distances and screen sizes for each species) around human or monkey saccadic endpoints were compared to 100 map values collected from locations randomly chosen from the distribution of saccadic endpoints from all saccades (in the same species) except those generated in the same clip by the same subject as the sample. This approach is similar to the image-shuffled analysis method used by others for static images [Parkhurst and Niebur, 2003, Reinagel et al., 1999, Tatler et al., 2005], and allows for an unbiased measure of model performance despite any accidental correlation between a particular species saccadic endpoint distribution and model output. For a particular subject, at the onset of a saccade we measured the value in each model map at the endpoint of the saccade, *i.e.* the activity in the map just before the saccade. For the interobserver model, the map value was measured at the time of the endpoint of the saccade to assess the congruency of gaze locations, either within or between species.

Differences between saliency at human or monkey gaze targets and at the randomly selected locations were quantified using ordinal dominance analysis [Bamber, 1975]. Model or metric map values at observers saccadic endpoints and random locations were first normalized by the maximum value in the map when the saccade occurred (*i.e.* when the map was sampled). For each model, histograms of values at eye positions and random locations were created. To non-parametrically measure differences between observer and random histograms, a threshold was incremented from 0 to 1, and at each threshold value we tallied the percentage of eye positions and random locations that contained a value greater than the threshold (hits). A rotated ordinal dominance curve (similar to a receiver operating characteristic graph) was created with observer-hits on one axis and random-hits on the other (Figure 1.5, inset). The curve summarizes how well a binary decision rule based on thresholding the map values could discriminate signal (map values at observer eye positions) from noise (random map values). The overall performance can be summarized by the area under this curve. This value is calculated and stored for each of the 100 randomly sampled sets. The mean of the 100 ordinal dominance values is taken as the final ordinal dominance estimate. A model that is no more predictive than chance would have equal random and model hits for each threshold, creating a straight line with an ordinal dominance of 0.5. The interobserver metric is assumed to provide the upper bound of predictability, between 0.5 and 1.0 (see Results), which the best computational models might be expected to approach. Note that an ordinal dominance of 1.0 is not achievable by any model, because there is imperfect agreement among observers, hence it is impossible for a single model to exactly pinpoint the gaze location of each observer.

1.3.8 High-interest gaze targets

For some analyses we defined a subset of saccadic endpoints as high-interest gaze targets. These were locations separated by less than 48 pixels (3.6°humans, 4.7°monkeys) that two or more observers of a given species looked at within 150 ms of one another. For monkeys, filtering the 12,826 saccades used for the overall analysis by these criteria resulted in a subset of 1,812 saccades; for humans, filtering the original 12,148 saccades resulted in a subset of 4,142 saccades.

1.3.9 Statistical analysis

Distributions of model and metric output at gaze targets were statistically compared using the permutation framework (Monte-Carlo simulation, 10,000 repetitions). Confidence intervals for model and metric scores were estimated by repeating the ordinal dominance measurement on a randomly selected half of the data, to form a sampling interval. Tests between species or models were carried out using a permutation test, computed by taking all saccades from both groups under test and randomly assigning each saccade to one of the two groups, irrespective of the actual group membership of the saccades. The difference between mean ordinal dominance values for the two randomly assigned groups was computed and stored. The process was repeated to form a sampling interval. The p value represents the probability of observing a value more extreme than the original group assignment [Good, 2001]. Statistical analysis of the saccadic endpoint distributions was also carried out in the permutation framework, but the symmetric Kullback-Leibler distance function was used in place of ordinal dominance.

1.4 Results

1.4.1 Saccade Metrics

Several differences in the saccade metrics of humans and monkeys were observed. Figure 1.3A, B show the smoothed distribution of saccadic endpoints used for analysis. Hotter colors represent a higher likelihood that a subject made a gaze shift to that location. Human and monkey saccadic endpoint distributions were significantly different (permutation test, $p < 0.0001$), but both species showed the characteristic center bias reported in human experiments using natural photographs [Reinagel et al., 1999, Tatler, 2007]. This may reflect a physiological bias to return the eyes to the center of the orbits [Paré and Munoz, 2001]. Monkeys seemed to explore the spatial extent of the display more thoroughly than humans, who were very center-biased. This difference may be due to a variety of factors including motor differences, cognitive awareness of the main actors and actions which were often near the center of the video, or a general search strategy. The TV channel logo that often appeared in the lower right-hand corner (Figure 1.1G) also attracted a high number of gaze shifts for both species.

Figure 1.3 C and D show the saccadic main sequence for humans and monkeys. The main sequence plots the relationship between saccadic peak velocity and amplitude, and is well know to be an exponential function [Bahill et al., 1975]. The shape of this function is thought to reflect the brainstem circuitry controlling saccades, and is altered when there is damage in the brainstem circuits or muscles controlling saccades [Ramat et al., 2007]. The main sequence data combined across the 5 monkeys was noticeably more variable than the human main sequence. When analyzed on a log-log scale, a linear regression revealed an R^2 of 0.77 (ANOVA test, $F(1, 15168) = 52002$, $p < 0.0001$) for monkeys compared to R^2 of 0.96 (ANOVA test, $F(1, 14835) = 342150$, $p < 0.0001$) for humans. Monkeys were much faster for a given amplitude, and regression lines showed monkeys had significantly higher velocity offset (Figure 1.3). The slope of the line was significantly higher (Figure 1.3, although a small magnitude difference) in humans, indicating a steeper relationship between amplitude and peak velocity. Figure 1.4 compares saccadic amplitude, fixation duration, and intersaccadic interval distributions for monkeys (green bars) and humans (blue bars). The probability distribution of saccadic amplitudes

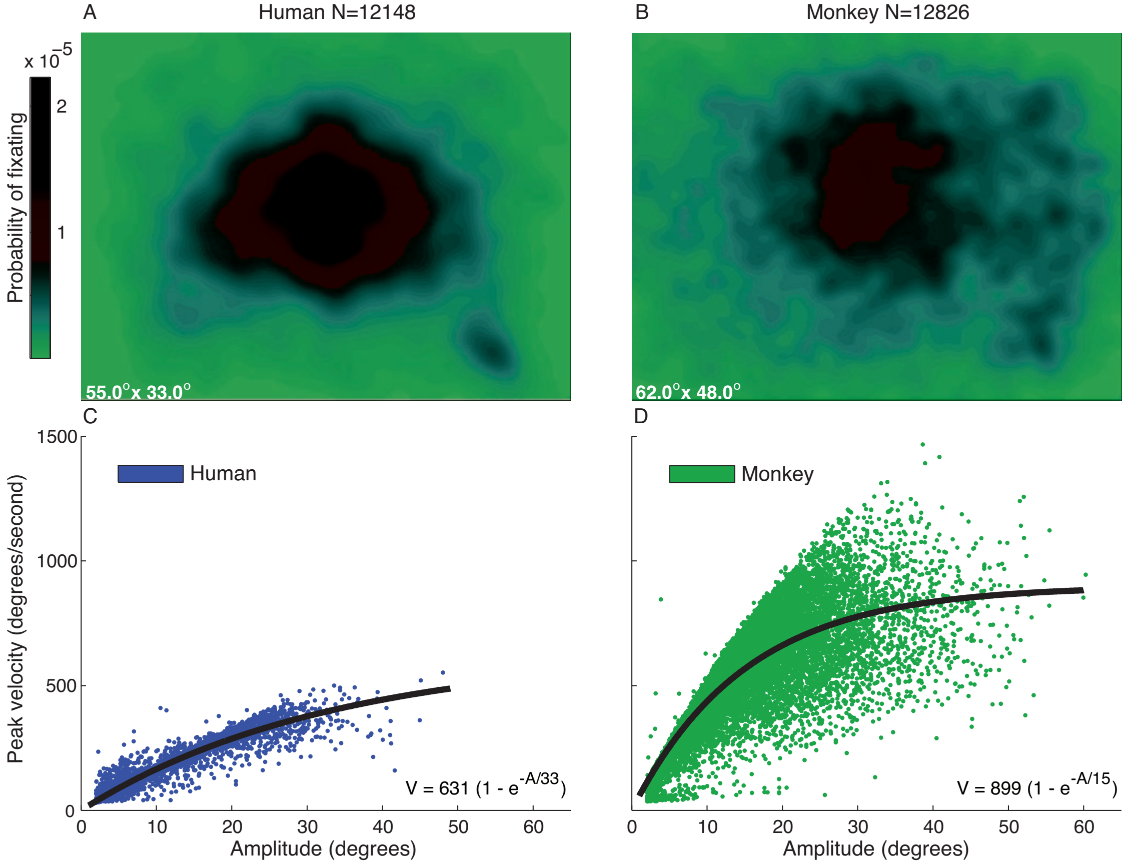
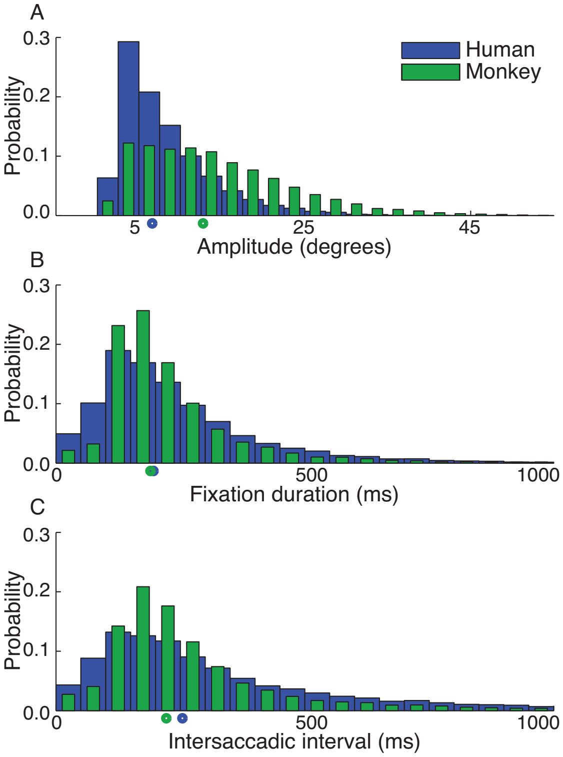


Figure 1.3: Saccade Metrics: Endpoint distributions and main sequences. (A-B) Saccadic endpoint distributions for the 12,138 human and 12,832 monkey saccades (computed after removing noisy data and clips with fewer than three observers, resulting in less data for monkeys) used for comparison with the contrast and saliency models and the interobserver agreement metric. Points were smoothed by convolving each map with a Gaussian kernel ($\sigma = 1.5°$). Hotter colors represent a higher likelihood that a human or monkey gaze shift landed at that screen location. Distributions were significantly different at $p < 0.0001$, using the Kullback-Leibler distance function between distributions in a permutation test (see Methods). (C-D) Main sequence for all saccades (14,837 human and 15,170 monkey, before removing clips with fewer than three observers) recorded from humans (blue) and monkeys (green). The main sequence was computed before combining multi-step saccadic eye movements into a single saccade, yielding separate entries for each component of the multi-step saccade. Main sequences for humans and monkeys were significantly different (ANOVA test, $F(2, 30003) = 58024.55$, $p < 0.0001$), testing for coincident regression lines on a log-log scale. Significant differences were observed for both the slope (ANOVA test, $F(1, 30003) = 1703.29$, $p < 0.0001$) and velocity offset (ANOVA test, $F(1, 30003) = 21805.25$, $p < 0.0001$) components of the main sequence. Black lines fitted to the data were computed by minimizing $V = a(1-e^{-A/s})$ where V and A are saccadic velocities and amplitudes respectively; a and s are the model parameters representing maximum amplitude and slope of the lines.

differed significantly, in that monkeys had a broader distribution and a greater median (Figure 1.4A). This could in part have been because monkey subjects had a slightly wider field of view; however, when amplitudes were re-plotted on a normalized axis, the same qualitative results were obtained (not shown). The probability distributions for fixation durations and intersaccadic intervals also significantly differed between species with humans having slightly longer median durations (Figure 1.4B, C). Monkey fixation and intersaccadic interval distributions were narrower, which possibly indicates a stereotyped fixation pattern (*e.g.* , fixate for 250 ms and then saccade to new place). In contrast, human fixation durations and intersaccadic intervals were spread over a wide range of values.

Figure 1.4: Saccade Metrics: Distributions of saccade amplitude, fixation durations and intersaccadic intervals. Probability histograms for (A) saccadic amplitude, (B) fixation duration after a saccade, and (C) intersaccadic interval (which may include smooth pursuit) for humans (blue), and monkeys (green) calculated before combining multi-step saccades into a single saccade. For display purposes only, the green bars are half the width of the blue bars, which represent the actual interval for both. The time axes are truncated at 1000 ms. Amplitude (Two-tailed Kolmogorov-Smirnov, $D = 0.34$, $n1 = 14837$, $n2 = 15170$, $p < 0.0001$), fixation duration (Two-tailed Kolmogorov-Smirnov, $D = 0.12$, $n1 = 14837$, $n2 = 15170$, $p < 0.0001$), and intersaccadic interval (Two-tailed Kolmogorov-Smirnov, $D = 0.13$, $n1 = 14837$, $n2 = 15170$, $p < 0.0001$) histograms were significantly different. Green and blue circles represent the median scores for each species.



1.4.2 Model predictions of Gaze Shift Endpoints

To further quantify species differences we used a computational model of saliency-based visual attention. In previous human experiments, this model has revealed that observers gaze more frequently towards the salient hot-spots computed by the model in both static images and dynamic scenes [Itti, 2006, Itti and Baldi, 2005, 2009, Parkhurst et al., 2002, Peters et al., 2005]. The model takes as input an image or video clip frame and outputs a salience map that gives a prediction of the screen locations likely to attract attention. The specific implementation details of this model have been described previously [Itti, 2006, Itti et al., 1998].

We measured the mount of computed saliency for each video frame at the endpoints of saccadic eye movements in both species (see Methods), to assess the extent to which humans and monkeys exhibited similar computations of salience [perhaps represented in monkey LIP or the SC Goldberg et al., 2006, Shipp, 2004] and strategies for deploying gaze towards salient locations. To quantify the chance-corrected performance of the saliency model, values at gaze targets were compared to values at gaze targets taken at random from other video clips, giving an ordinal dominance score (see Methods). Measurements from the contrast model and interobserver agreement metric were similarly chance-adjusted. Figure 1.5A shows the comparison of human and monkey ordinal dominance scores for different models and metrics, and Figure 1.5B shows a summary of the statistical analysis. All models and metrics predicted human and monkey gaze targets significantly better than chance (permutation test, $p < 0.0001$), and saliency predicted human and monkey gaze behavior significantly better than the baseline-control contrast model (permutation test, $p < 0.0001$). This finding validated the use of the saliency model as a good predictor of visually guided attentive behavior in both humans and monkeys.

Interestingly, we found that saliency correlated with human behavior significantly better than monkey behavior, over all clips combined (permutation test, $p < 0.0001$). Differences in the likelihood to deploy attention to salient items should be minimized when using monkeys as a model for human attention during free viewing. The saliency differences were, however, small in magnitude compared to the difference in interobserver agreement (Figure 1.5). Comparing saliency scores with interobserver agreement may provide insight into a way to reconcile such differences. Although saliency was a strong predictor of human visually guided behavior, the stimulus-driven nature of the model limited its predictive power. The interobserver agreement metric captured aspects of stimulus-driven (saliency) and top-down (context specific) attentional allocation, the latter of which has also been shown to be a significant factor in guiding human gaze shifts in natural scenes [De Graef et al., 1992, Neider and Zelinsky, 2006, Noton and Stark, 1971, Oliva et al., 2003, Yarbus, 1967]. The interobserver agreement metric was the best predictor of human saccadic targets (permutation test, $p < 0.0001$). Interestingly, this trend did not hold for monkeys and the interobserver agreement metric was significantly less correlated with monkey gaze shifts than the saliency model (permutation test, $p = 0.0027$). That is, the computational saliency model better predicted where one monkey might look than was predicted from the gaze patterns of two to four other monkeys. Any top-down information present in the monkey interobserver agreement metric was insufficient to increase predictability of gaze patterns over a purely stimulus-driven model. Monkey top-down attentional allocation may be completely inconsistent among

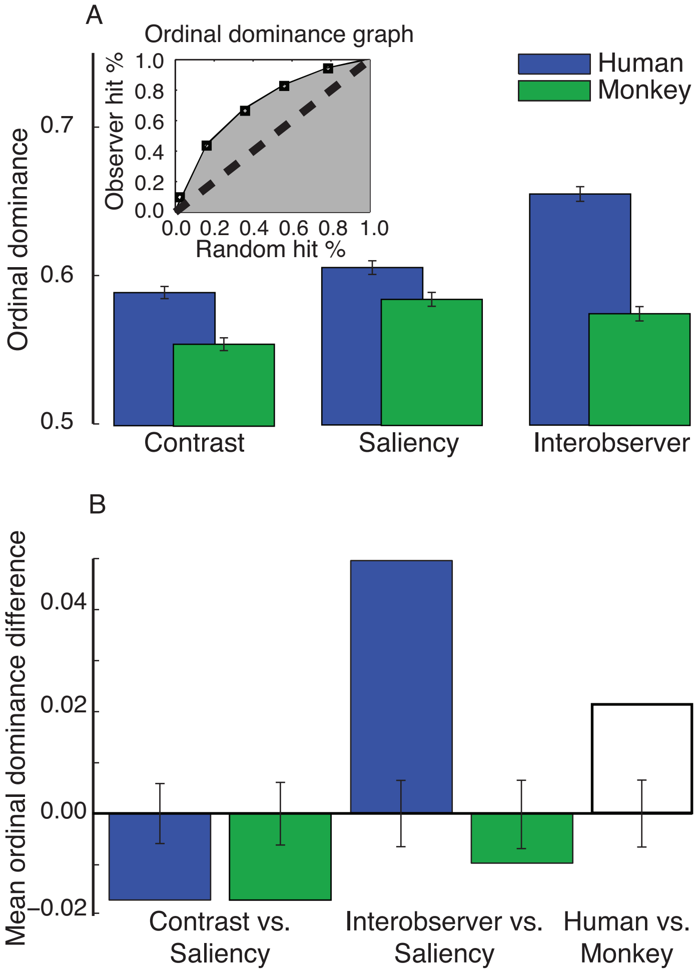


Figure 1.5: Model and metric scores at human and monkey saccadic endpoints. (A) Comparison of the contrast and saliency model, and interobserver agreement metric values at human (blue) and monkey (green) saccadic endpoint locations with values at randomly selected eye positions. Overall, human and monkey gaze shifts were predicted (permutation test, $p < 0.0001$) by all models and metrics greater than chance levels (ordinal dominance of .5). Error bars show the 95% confidence interval on the ordinal dominance estimate (see Methods). (B) Summarizes the statistical differences between species and models as obtained through permutation tests (see Methods). Blue (human), green (monkey) and white (human-monkey) bars show the magnitude of the test statistic (mean ordinal dominance difference) obtained between pairs labeled on the x-axis. Values greater than 0 indicate the first model or species in the pair had a larger ordinal dominance score. Black bars represent the 95% confidence interval of the test statistics sampling distribution. Left, saliency performed better than the baseline-control contrast model for both humans and monkeys (permutation test, $p < 0.0001$). Center, interobserver agreement was more predictive than saliency for humans (permutation test, $p < 0.0001$), however, interobserver agreement was less predictive than saliency for monkeys (permutation test, $p = 0.0027$). Right, the human saliency ordinal dominance score was significantly higher than the monkey score (permutation test, $p < 0.0001$).

observers (*e.g.* Figure 1.1G), leaving saliency to be the best predictor of visually-guided attentive behavior.

Figure 1.6 shows a scatter plot of median normalized (not chance corrected) monkey vs human saliency values at all saccadic endpoints that occurred during each entire clip. This clip-by-clip analysis revealed that saliency values from monkeys and humans were significantly correlated (Figure 1.6). The best fitting line (solid black) had significantly lower slope than the unity line (dashed black), indicating that monkeys saliency scores varied less than humans from clip to clip, and clips that contained higher saliency values for humans contained on average slightly lower saliency values for monkeys. The y-offset, however, was not different from 0 (Figure 1.6), indicating that there was no systematic bias, or baseline shift, in human or monkey raw saliency scores. The majority of the regression line falls below the unity line; hence, on average the saliency scores were lower for monkeys, as was already the case with our aggregate analysis (Figure 1.5). Individual clip content affected deployment of gaze to salient locations for humans and monkeys in a comparable way, however, monkeys may have had a tendency to be less modulated by clip content. This likely reflects differences in semantic understanding of the clips between the two species.

We defined a subset of clips (Figure 1.1F) as monkey relevant. These clips contained scenes from the monkeys daily environment (*e.g.* their housing, familiar monkeys and humans, facilities), and represented a contextual control to ensure monkeys attended to familiar natural scenes similarly to novel ones. The points in the scatter plot for monkey relevant clips (Figure 1.6, green triangles) were in the same distribution as those for other clips. Only considering these monkey relevant clips, a significant linear correlation was

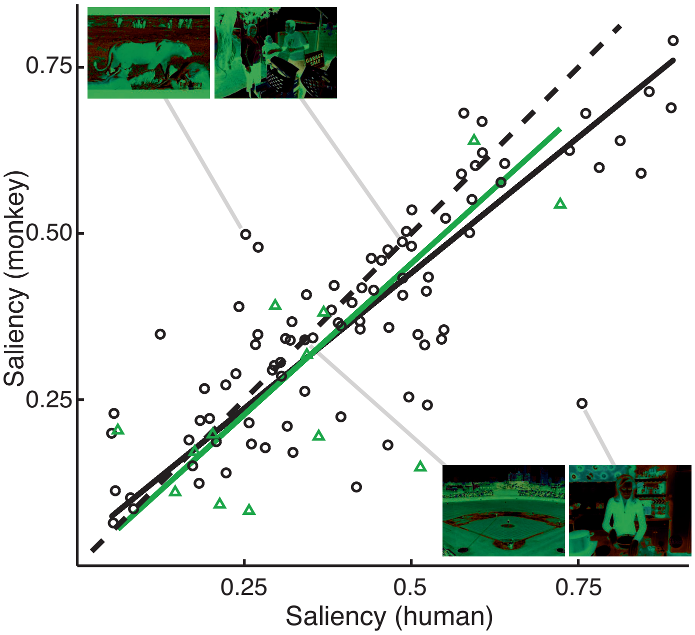


Figure 1.6: Correlation between saliency values at human and monkey eye positions. The scatter plot shows median saliency values considering all saccadic endpoints in a given video clip for monkeys vs humans. Each point represents the median of raw (not chance corrected) saliency values for each video clip, with green triangles indicating clips that would be relevant to a monkey as described in Methods. Human and monkey scores were well correlated (Pearson correlation, $r(98) = 0.80$, $p < 0.0001$). Analysis of coefficients obtained by major axis regression [Sokal and Rohlf, 1995] revealed that the best fitting line ($y = 0.82x + 0.032$, solid black) was significantly different from unity (dotted black) in slope (F-test, $F(1, 98) = 7.26$, $p = 0.0083$) but not y-offset (t-test, $t(98) = 0.089$, $p = 0.38$). The regression line for monkey relevant clips ($y = 0.91x - 0.00021$, solid green) was not significantly different from the regression line for all other clips (chi-square test, $c2(1, N = 100) = 0.2$, $p = 0.65$), computed by testing for coincident lines. Hypothesis testing was performed according to **?**. The example frames in the upper left and lower right corners are from videos where one species had a considerably higher saliency score than the other. The two adjacent frames are from the two videos where human and monkey scores were most similar.

found (Pearson correlation, $r(13) = 0.72$, $p = 0.005$). This line was not significantly different from that calculated for all other clips (Figure 1.6). Taken together this analysis indicates that monkeys were visually attentive to the video clips in a similar fashion to humans, at least as far as saliency is concerned, although from this analysis we can not know if they looked at similar spatial locations at the same time, only that they looked at similarly salient items.

1.4.3 High-interest gaze locations

We wondered if the relatively poor predictability of monkey behavior by the saliency model and interobserver agreement metric might be due to idiosyncratic search strategies and/or cognitive systems by monkeys, which may or may not have been related to the video content. To remove idiosyncratic gaze shifts from the analysis, we determined a subset of high-interest gaze targets those locations that attracted the attention of two or more observers toward the same location at the same time (see Methods). Saliency and interobserver agreement metrics were then reanalyzed based on this subset for each species. Figure 1.7 A shows the effect of filtering by high-interest gaze targets on an interspecies agreement metric. This metric represents the correlation between monkey saccadic target locations, and those target locations selected by humans. This metric was computed by testing monkey saccadic endpoints against the same human-derived interobserver metric that was used for human interobserver agreement analysis. The interspecies agreement metric allowed us to directly measure the extent to which monkey gaze target locations were also looked at by humans. The lowest score the interobserver agreement metric obtained for humans was when all human saccades were analyzed together (Figure

1.5 A, Figure 1.7 A, lower black line). This can serve as a lower bound for our interspecies agreement metric, as to be a good model of human visual behavior monkeys should be as consistent with human gaze targets as humans are with one another. A useful upper bound for this metric is obtained by re-calculating the interobserver agreement metric for saccadic target locations where at least two humans agreed to look (Figure 1.7A, upper black line). We expect the best models of human visual behavior (animal or computational) to approach this level of correlation with humans, as it means the model is often selecting the strong attractors of attention those scene locations that on average attracted the attention of multiple human observers.

When all monkey saccades were considered, the interspecies ordinal dominance score was lower than the score obtained from the human interobserver agreement metric (permutation test, $p < 0.0001$). That is, monkey saccadic target selection was less consistent with human target selection, than humans were with one another. However, the interspecies ordinal dominance score dramatically increased (permutation test, $p < 0.0001$) when analysis was limited to monkey saccades made towards monkey high-interest targets. In fact, the interspecies score for these high-interest monkey saccades fell above our human-derived lower bound (permutation test, $p < 0.0001$), but below our human-derived upper bound (permutation test, $p < 0.0001$). This demonstrates a high correlation between locations where humans and monkeys looked when analysis of monkey saccades was restricted to high-interest locations.

Figure 1.7B compares human and monkey saliency ordinal dominance scores for all gaze targets and high-interest gaze targets. As was shown in Figure 1.5, when all saccades were considered the monkey saliency ordinal dominance score was significantly lower than

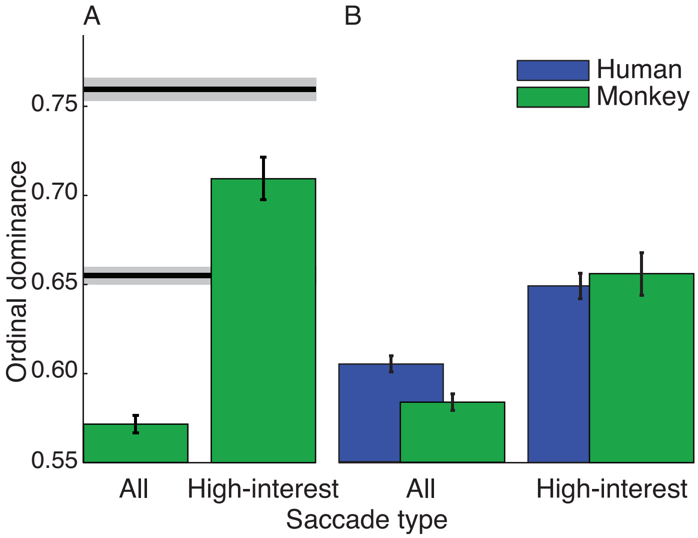


Figure 1.7: Analysis at high-interest gaze locations. To test agreement in saccadic target selection between humans and monkeys, the human interobserver metric was used to predict the gaze locations of monkeys (interspecies agreement metric). (A) Shows ordinal dominance scores for the interspecies agreement metric for all monkey saccadic endpoints, and a subset of "igh-interest" saccadic targets, that multiple monkeys looked at simultaneously. When only high-interest targets were considered, monkey saccadic endpoints were closer to human gaze locations (permutation test, $p < 0.0001$). To serve as a reference, the lower black line is the mean ordinal dominance score of the human interobserver agreement metric. The upper black line is the mean ordinal dominance score of the human interobserver agreement metric when only locations where two or more humans agreed to look were considered. Shaded regions represent the 95% confidence intervals of these estimates. When all monkey gaze targets were considered, the interspecies agreement metric scored lower than the human interobserver agreement metric (permutation test, $p < 0.0001$). However, when only high-interest gaze targets were considered, the interspecies ordinal dominance score fell between the lower and upper bounds derived from our human interobserver metric (permutation test, $p < 0.0001$). (B) Shows saliency ordinal dominance scores for all gaze endpoints and a subset of high-interest gaze locations for humans and monkeys. The ordinal dominance scores for all saccades (Figure 1.5) is re-plotted as a reference. When all monkey gaze targets were considered, the monkey saliency ordinal dominance score was lower than the human score (permutation test, $p < 0.0001$). For the subset of high-interest gaze targets, where two or more monkeys agreed, the ordinal dominance score was increased (permutation test, $p < 0.0001$) and indistinguishable from the human high-interest gaze targets (permutation test, $p = 0.16$), putting the monkeys in the range of human predictability.

the human score, indicating that the saliency model predicted human saccades better than monkey saccades. However, at high-interest gaze targets, the ordinal dominance scores were significantly higher for humans and monkeys (permutation test, $p < 0.0001$), indicating that the saliency model was a better predictor of high-interest gaze targets than of low-interest ones (*e.g.* , when the five observers looked at five different locations) for both species. Note that increasing the number of humans who agreed on a saccadic target to three did not significantly increase the saliency ordinal dominance score (not shown). Thus, in our analysis, gaze locations where two human observers agreed can serve as an upper bound for human gaze predictability. Increasing the number of agreeing monkeys beyond two seemed to increase the ordinal dominance scores linearly (not shown), but more data would be required for hypothesis testing. Interestingly, the saliency ordinal dominance score for monkey high-interest saccadic targets was greater than the human score for all saccades (permutation test, $p < 0.0001$) and was indistinguishable from the score for human high-interest gaze targets (permutation test, $p = 0.16$). That is, scene items that drew the attention of multiple monkeys (high-interest gaze targets) contained similar chance corrected saliency values than those locations that attracted the gaze of multiple humans.

1.5 Discussion

The present study objectively compared, for the first time, human and monkey visually attentive behavior during free viewing of natural dynamic (video) stimuli. In addition to examining saccadic eye movement metrics, several models of visual attention were employed to provide objective metrics by which to compare human and monkey viewing behavior. We found significant differences between human and monkey gaze shifts during free viewing. In summary, monkeys generated faster saccades which spanned a greater range of the screen and were separated by shorter fixation durations. Although both species shifted gaze to locations that were deemed salient by the saliency model, humans were more likely to do so. The gaze locations of other humans were the best predictors of human behavior, but this was not true of monkeys. The saliency model predicted monkey gaze shifts better than the combined gaze behavior of other monkeys. These differences, however, could be minimized if we only examined high-interest gaze locations those that at least two monkeys jointly attended. When the saccades were filtered in this way, monkey behavior became more human like, almost indistinguishable in terms of gaze location and saliency values. This filtering technique focuses analysis on common attractors of attention between species, possibly by emphasizing the role of the shared low-level saccadic selection processes over the more idiosyncratic cognitive processes. High-interest targets minimize differences between the species, providing a method to make the best use of monkeys as a model of human visual behavior under free-viewing conditions.

1.5.1 Monkey-human differences in eye movement metrics

Eye movement metrics under free viewing of video stimuli were found to be quite different between monkeys and humans. Monkeys were less center-biased and made saccades with larger amplitudes on average. This may suggest that monkeys were less interested in the videos actions and actors, which tended to be filmed near the center. Monkeys may have had less cognitive understanding of the scenes, and/or they were more interested in exploring the screen, possibly in search of actions/locations that could have resulted in reward.

At a more mechanical level, monkeys differed from humans in features of their saccadic main sequence (saccadic velocity vs amplitude). Monkeys made much faster saccades for a given amplitude compared to humans, confirming what has been found by Harris et al. [1990]. The main sequences under free-viewing conditions were comparable to those obtained in previous studies using laboratory stimuli with humans [Bahill et al., 1975, 1981, Becker and Fuchs, 1969, Boghen et al., 1974] and monkeys [Quaia et al., 2000, Van Gisbergen et al., 1981] separately. Our data tended to have slower peak velocities, particularly in humans; however, velocities still fell within the normal range defined by Boghen et al. [1974]. Differences in our data may be a feature of free viewing, or idiosyncratic to our subjects and methodology.

Discrepancies between species could be partly accounted for by differences in neural connectivity from the retina through the oculomotor system to the eye muscles, and possibly by differences in the motor plant, *e.g.* smaller viscous reactive forces in monkeys because they have a smaller eyeball. These plant differences probably reflect little on the processes involved in the deployment of visual attention. However, some discrepancies (*e.g.* , intersaccadic intervals, saccadic endpoints distributions) may stem from different scanning strategies employed and should be accounted for when comparing species.

1.5.2 Monkey-human differences in model correstpondence and interobserver agreement

More relevant to understanding visual attention is an examination of image properties at human and monkey gaze positions. To objectively compare species, we examined how computational models predicted saccadic targets of humans and monkeys. We used a model that measures static luminance contrast, which has been shown to be an attractor of gaze in humans and monkeys watching grayscale images [Einhäuser et al., 2006], and a saliency model, which has been shown to capture aspects of stimulus-driven eye movements in humans viewing images [Peters et al., 2005] and videos [Itti, 2006, Itti and Baldi, 2005]. The contrast model, although it does not contain temporal dynamics, serves as a baseline to measure the performance of the saliency model, for even simple models of attention will predict behavior significantly above chance (random sampling). Both models predicted gaze shifts of both species above chance, but the saliency model performed better, as expected. Validation of the saliency model with monkeys suggests the species may possess similar computations of saliency during free viewing, and the model captures aspects of these mechanisms shared among primates. This is encouraging as it validates investigation of the neural substrates of such computations in monkeys.

Interestingly, the computational models predicted human gaze shifts better than monkey gaze shifts. This was surprising, as we had expected monkeys would be more saliency-driven than humans, due to their impoverished knowledge of the clips content (*e.g.* one video clip shows the earth viewed from space, likely a foreign concept to our monkeys). Our finding was also in contrast to results from [Einhäuser et al., 2006] who found monkeys and humans to be equally saliency-driven to grayscale images. However, inconsistency in gaze target selection among monkey observers relative to humans provided some insight into these discrepancies.

Human attention has been described as a combination of stimulus-driven (bottom-up) and contextually-driven or goal-directed (top-down) factors [Itti and Koch, 2001a, Treisman and Gelade, 1980], and monkey attention is likely controlled by similar mechanisms [Fecteau and Munoz, 2006]. The interobserver agreement metric contains elements of both factors while the saliency algorithm captures aspects of bottom-up processing only. As expected, for humans, the interobserver agreement metric provided the best prediction of gaze deployment. It has been suggested that gaze density among observers is increased over scene regions containing semantically inconsistent or highly informative objects [Henderson et al., 1999, Loftus and Mackworth, 1978]. Hence, the gaze consistency among our humans likely reflects their shared notion of semantically informative regions in the clips. Monkey gaze, however, was best predicted by the saliency model. This suggests that monkeys made many idiosyncratic eye movements, possibly related to each monkeys unique interpretation of the scene, the goal of the experiment, or inattentiveness to the stimuli. Monkeys may have been engaged by the clips, but shared less top-down knowledge of how to follow the main actions compared with humans. Alternatively, it may be that as a result of their training, monkeys were in part examining the screen looking to unlock the task or find a screen location or action that would lead to a reward. Such a search strategy is supported by the stereotyped fixational pattern (more narrow distribution of intersaccadic intervals). In either case, since their top-down interpretation seems inconsistent, saliency based computations may serve as the lowest common denominator in deploying gaze in natural scenes for monkey observers.

Perhaps the most relevant question to consider, given the observed differences, is to what degree monkeys looked at the same places that humans looked. To address this, we focused analysis on high-interest targets, those locations that were gazed at by two or more monkeys simultaneously. This effectively forced consistency on our monkey data by filtering out some idiosyncratic eye movements that may have been due to differences in top-down scene interpretation or general attentiveness to the stimuli. An interspecies agreement metric revealed that when all saccade data was used, monkey saccadic targets were not as consistent with humans, as humans were with each other. In other words, monkeys didnt often look where humans looked. This is not unexpected, as monkeys were inconsistent with each other. However, when the analysis was repeated using only the subset of monkey high-interest saccadic targets, those targets were dramatically closer to locations where, on average, humans looked (Figure 1.7A). High-interest gaze targets for monkeys became consistent with human visual behavior and were within the expected range of human interobserver agreement scores. These saccadic targets may focus our monkey analysis on scene locations that were of common interest to both species, narrowing the gap between human and monkey visual behavior during free viewing of dynamic scenes.

Interestingly, those same high-interest targets that correlated well with human behavior, were also highly salient; in fact, indistinguishable from human high-interest gaze targets in terms their chance corrected saliency scores. Highly salient items, as predicted by our model, may have simultaneously attracted the attention of multiple monkey observers. Since the monkey high-interest targets are also close to human gaze targets, this may indicate that saliency was the common factor in driving human and monkey attention to those locations. Analysis of monkey high-interest saccades minimized species differences both in terms of specific saccadic targets and saliency model agreement. This analysis may emphasize the shared bottom-up attentional processes among humans and monkeys, filtering out the more individualized cognitive processes.

This result may be particularly relevant when using monkeys in experiments requiring neural recording or imaging during free viewing of dynamic or natural scenes. Restricting analysis of neural responses to stimuli that attracted the gaze of at least two monkeys would ensure that the monkeys behavior would be as consistent as possible with human behavior under such conditions. While doing so eliminates a significant portion of the data, more data can be collected more easily under free viewing compared with traditional single-trial methods. This technique may emphasize common attentional mechanisms between species, thus making the best use of our animal model to generate results meaningful to human behavior and cognition.

Chapter 2

Visual adaptation and novelty responses in the superior colliculus

2.1 Abstract

The brain's ability to ignore repeating, often redundant, information while enhancing novel information processing is paramount to survival. When stimuli are repeatedly presented, the response of visually-sensitive neurons decreases in magnitude, *i.e.* neurons adapt or habituate, although the mechanism is not yet known. We monitored activity of visual neurons in the superior colliculus (SC) of rhesus monkeys who actively fixated while repeated visual events were presented. We dissociated adaptation from habituation as mechanisms of the response decrement by using a Bayesian model of adaptation, and by employing a paradigm including rare trials that included an oddball stimulus that was either brighter or dimmer. If the mechanism is adaptation, response recovery should be seen only for the brighter stimulus; if habituation, response recovery ('dishabituation') should be seen for both the brighter and dimmer stimulus. We observed a reduction in the magnitude of the initial transient response and an increase in response onset latency with

stimulus repetition for all visually responsive neurons in the SC. Response decrement was successfully captured by the adaptation model which also predicted the effects of presentation rate and rare luminance changes. However, in a subset of neurons with sustained activity to visual stimuli, a novelty signal akin to dishabituation was observed late in the visual response profile to both brighter and dimmer stimuli and was not captured by the model. This suggests that SC neurons integrate both rapidly discounted information about repeating stimuli and novelty information about oddball events, to support efficient selection in a cluttered dynamic world.

2.2 Introduction

Efficient selection of important events among temporal clutter requires ignoring repeating stimuli, thereby emphasizing novel and potentially important ones. This simple form of non-associative learning has been referred to as adaptation, habituation and repetition suppression, depending on the era and field of study [Clifford et al., 2007, Grill-Spector et al., 2006, Kohn, 2007, Krekelberg et al., 2006]. From an information processing perspective, adaptation serves to adjust the operating point of a sensory system, to maximize the efficiency of sensory coding and increase differential sensitivity to novel events [David et al., 2004, Dean et al., 2005, Dragoi, 2002, Dragoi et al., 2002, Müller et al., 1999]. This can be achieved through incremental updating over time of a Bayesian prior, which can then bias the processing of incoming sensory data [Itti and Baldi, 2005, Stocker and Simoncelli, 2006].

Electrophysiological evidence of response reduction with stimulus repetition has been observed throughout the visual system, from retina [Brown et al., 2001, Hosoya et al., 2005, Smirnakis et al., 1997] and thalamus [Solomon et al., 2004], to visual cortex [Maffei et al., 1973, Motter, 2006, Movshon and Lennie, 1979, Müller et al., 1999] and frontal eye fields [Mayo and Sommer, 2008]. These studies usually focus on perception; however, stimulus repetition effects also have profound, though less studied, consequences on visual orienting: the latency and magnitude of the visual response influences the timing of eye and head movements to foveate the stimulus [Corneil et al., 2008, Dorris et al., 2002, Fecteau et al., 2004].

The ideal place to study visual repetition effects is the superior colliculus (SC) - the phylogenetically conserved hub of the visual orienting system [Dean et al., 1989, Huerta and Harting, 1983, Ingle, 1975, May, 2006, Munoz et al., 2000] - which is integrated with all other visual areas in the brain. Many visually responsive neurons in the SC also have movement responses time-locked to saccades [Mohler and Wurtz, 1976]. The superficial layers of the SC (SC_s) receive visual input directly from the retina and from early visual cortex, while the intermediate layers (SC_i) receive more complex visual and cognitive input from various cortical areas, the basal ganglia and cerebellum [May, 2006]. Therefore, early (*e.g.* retinal) and late (*e.g.* cortical) sources of visual adaptation can be compared directly by examining repetition effects across all SC visually-responsive neurons located in different layers.

We explored how the magnitude and onset latency of SC visual responses changed with repetition. We modeled these changes using a Bayesian approach to provide a quantitative definition of adaptation, which was then used to predict the consequences of changes to stimulus timing and intensity. To dissociate simple adaptation from higher-level learning processes (*e.g.* habituation), we compared responses to rarely-presented brighter or dimmer oddball stimuli. Response decrement due to adaptation should follow the adaptation model and recover with a brighter but not a dimmer stimulus; however, response decrement due to habituation should recover (dishabituation) after any novel stimulus change [Bernard, 1988, Sokolov, 1963].

2.3 Methods

All procedures were approved by the Queen's University Animal Care Committee and were in full compliance with the Canadian Council on Animal Care guidelines on the care and use of laboratory animals. Experiments were performed using two male rhesus monkeys (*Macaca mulatta)* weighing between 9-12 kg. The surgical techniques to prepare the animal for behavioral and physiological recordings have been described previously in detail [Marino et al., 2008]. Briefly, monkeys were implanted with a head post for head fixation, a recording chamber over the SC and eye coils to measure eye position with the search coil technique. The evening prior to surgery, the animal was placed under Nil per Os (NPO, water *ad lib*), and a prophylactic treatment of antibiotics was initiated [5.0 mg/kg enrofloxacin (Baytril)]. On the day of the surgery, anesthesia was induced by ketamine (6.7 mg/kg im). A catheter was placed intravenously to deliver fluids (lactated Ringer) at a rate of 10 mL/kg/h to a maximum of 60 mL/kg throughout the duration of the surgical procedure. Glycopyrolate (0.013 mg/kg im) was administered to control salivation, bronchial secretions, and to optimize heart rate (HR). An initial dose was delivered at the start of surgery followed by a second dose 4 hours into the surgery. General anesthesia was maintained with gaseous isofluorene (2-2.5%) after an endotracheal tube was inserted (under sedation induced by an intravenous bolus of propofol, 2.5 mg/kg). HR, pulse, pulse oximetry saturation (SpO_2), respiration rate, fluid levels, circulation, and temperature were monitored throughout the surgical procedure. The analgesic buprenorphine (0.01-0.02 mg/kg i.m.) was administered throughout the surgery and during recovery (8-12 hours). The antiinflammatory agent ketoprofen (2.0 mg/kg 1st dose, 1.0 mg/kg additional doses) was administered at the end of the surgery (prior to arousal), the day after the surgery and every day thereafter (as required). Monkeys were given 2 weeks to recover prior to onset of behavioral training.

Monkeys were trained to perform a variety of oculomotor tasks for liquid reward. Real-time control of the experimental task and visual display was achieved using REX version 6.0. Monkeys were seated in a primate chair 60 cm away from a CRT monitor (Mitsubishi XC2935C, 75 Hz refresh rate, 71.5 x 53.5 cm; usable field of view of 62° x 48°). Visual stimuli were presented within a darkened environment. Dark adaptation was prevented by dimly illuminating the monitor screen for 800 ms during the inter-trial interval. Physiological activity was monitored from 109 single neurons using tungsten electrodes (Frederick Haer, 0.5-5.0 mΩ with stimulus events and spike times collected, and waveforms digitized, through the Plexon MAP system. Further analysis was performed off-line with custom Matlab-based software.

2.3.1 Cell classification

When a neuron was first isolated, its visual receptive field was established using a simple fixation task in which white light stimuli (42.5 cd/m^2, 100 ms duration, 0.25° diameter spot) were presented in pseudorandom order to 182 possible locations distributed across 60° (horizontal) x 50° (vertical) of visual angle, the order of which was designed so that no two subsequent stimuli appeared within the typical response field of a SC neuron in order to prevent adaptation effects. The centroid of the receptive field was then determined using a cubic spline function and this location was used for all subsequent study of the neuron. Because we were interested in adaptation of visual responses, we limited data collection to encountered cells that had a visual response.

We then collected further information to characterize the neuron relative to known SC cell types. First, we made careful measures of microelectrode depth referred to the dorsal surface of the SC (as determined by the electrode depth that first elicited multiunit visual-only activity). Second, neural recordings taken during four interleaved saccade tasks [step, gap, memory-delay, and visual-delay; described in detail elsewhere; Munoz and Wurtz, 1993] were used to classify visual and motor responses; critically, the visual delay task dissociated visual and motor activity (Figure 2.1A). In this task, the animal starts each trial by fixating a central fixation point (FP). A target was then presented randomly in the center of the response field or at a location opposite the vertical and horizontal meridian. To receive a reward the animal had to maintain fixation until the FP disappeared (the delay period: 500-800 ms randomized) and then make a saccade to the target.

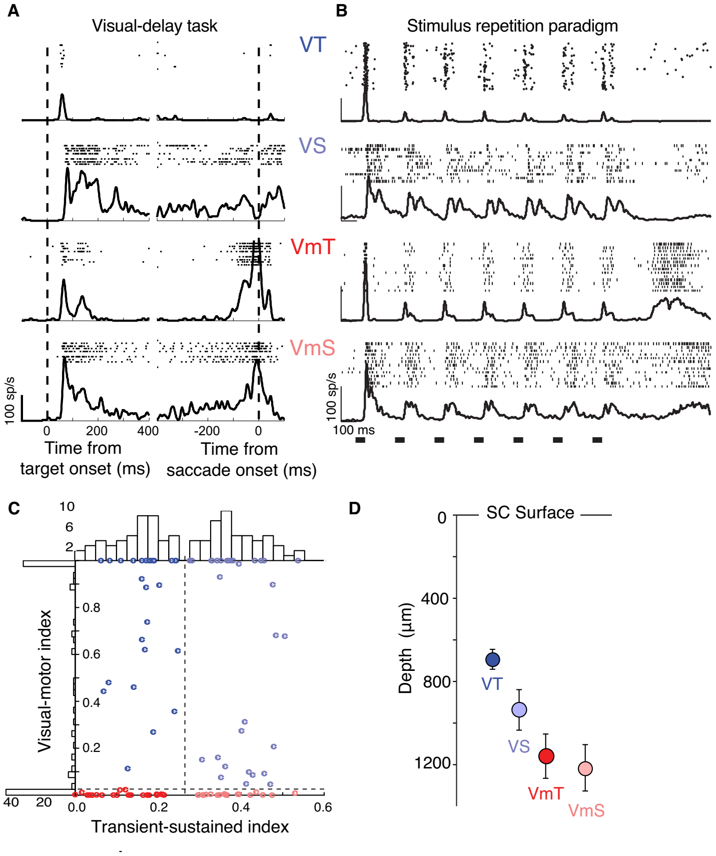


Figure 2.1: (A) Raster plots and spike density waveforms (σ=5 ms) recorded from a representative visual transient (VT), visual sustained (VS), visuomotor transient (VmT), and visuomotor sustained (VmS) neuron to a delayed saccade task, which was used to facilitate neural classification. Data are aligned on target appearance (left column) and saccade onset (right column) in the delayed saccade task when the target appeared in the neuron's response field. (B) The response of the same single neurons to the 7 stimuli in the standard repetition paradigm. The black bars across bottom of abscissa represent the stimulus timing. Spikes for individual trials are presented in raster format (only a subset of trials shown for display purposes) and overlaid with a mean spike density function (σ=5 ms). (C) Scatter plots and histograms of the metrics used to classify cells. The transient-sustained index is plotted against the visual-motor index for each cell (color indicates cell class), with smaller numbers indicating a more motor and more transient responses, respectively, as measured from responses in the visual delay task shown above. The histograms show the number of cells with each parameter value using a bin width of 0.025 units. The dashed lines show the cutoff values that were used to separate classes of neurons into the 4 categories.(D) The mean depth for each cell class. The cell classes had significantly unequal variances (Bartlett's test, T(3)=15.50, $p = 0.0014$), and consequently a Kruskal-Wallis test was conducted to evaluate differences in depth among the four cell classes. Cell classes significantly differed in depth (C2(3,108)=20.10, $p = 0.0002$), and pair-wise Wilcoxon rank-sum tests (Bonferroni corrected) indicated that visual transient cells significantly differed from both visual motor transient and visual motor sustained (z=3.78, $p < 0.001$ and z=3.79, $p < 0.001$, respectively).

We refer to visual responses as ‘transient’ (a short visual burst) or ‘sustained’ (visual burst followed by an extended period of low frequency activity) as was described previously [White et al., 2009]. This terminology is in line with descriptions of visual neurons in the geniculostriate pathway. We classified neurons as visual transient (VT), visual sustained (VS), visuomotor transient (VmT) and visuomotor sustained (VmS, see Figure 2.1B for single-unit examples) using two indices: a visual-motor index and a transient-sustained index. The visual-motor index was constructed with information from the saccade-aligned spike density function (Gaussian, σ=5 ms) from the visual-delay task (Figure 2.1A). The spike density function was first low-pass filtered by iterative convolution with a 5-tap binomial kernel: 100 iterations in the forward direction and 100 in the backward direction. The result had no phase shift and approximated convolution with a Gaussian of σ=14.12 ms. The timing and magnitude of the peaks and troughs of the waveform were then estimated by finding the zero crossings of the numerical gradient. A strong peak in activity from 25 ms pre-saccadic to 5 ms post-saccadic initiation was taken as evidence for a visuomotor neuron, and we quantified this feature with a simple probability measure:

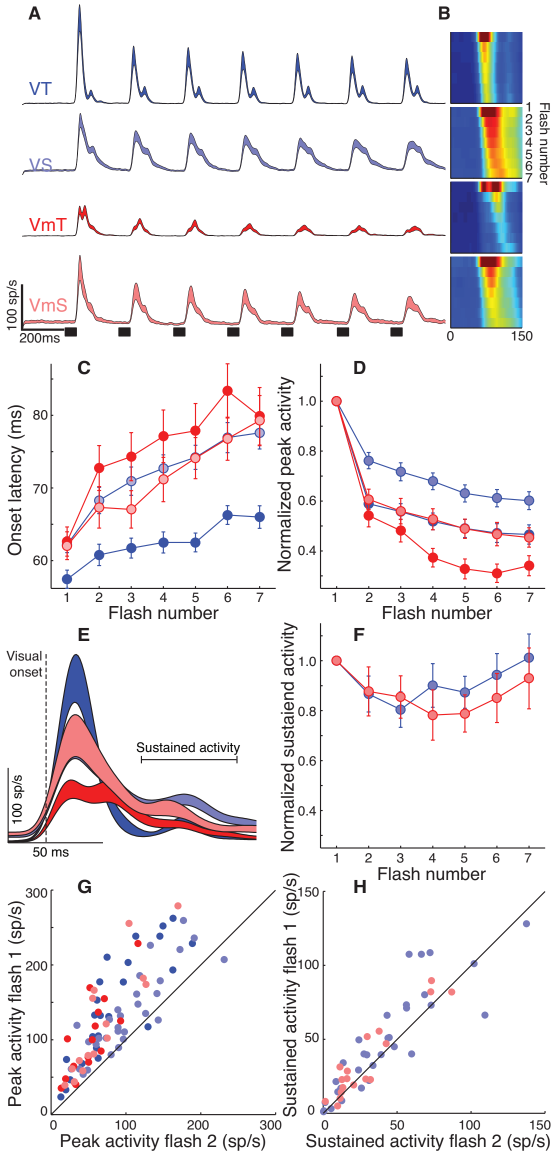
$$P = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\theta - T}{\sigma\sqrt{2}}\right)\right] \tag{2.1}$$

Where T and σ were the mean and standard deviation of non-perisaccadic peaks, θ was the value of the perisaccadic peak, and $\mathrm{erf}(x)$ is the Gauss error function. Large probabilities indicated the presence of motor activity, and if no peak was found, a probability of 0.0 was assigned. To confirm that the peak activity was related to a robust motor response and not to residual sustained visual activity or noise, the smallest trough was measured in a small window (±25 ms) around saccade onset. We computed the probability that activity at the trough (or at saccade initiation if no trough existed) was higher than the average pre-saccadic baseline activity (-900 ms-50 ms pre-saccade) using Equation (2.1), but where T and σ were the mean and standard deviation of the baseline, and θ was the value at the trough. Finally, the visual-motor index was computed as $1 - P_p P_t$, where P_p was the probability from the peak measurement, and P_t was the probability from the trough measurement. We considered cells with a visual-motor index < 0.025 to be visuomotor cells.

To compute the transient-sustained index, we aligned spike density functions to target appearance in the visual-delay task (Figure 2.1A) and divided the post-stimulus visual response into early (transient) and later (sustained) components. Each time point in the first 400 ms of post-stimulus activity was compared to a baseline (700 ms pre-target) using (2.1), where T and σ were the mean and standard deviation of the baseline, and θ was the value of the time point. Intervals of post-stimulus activity where each point in the interval had a high probability ($p \geq 0.99$) were identified and if the first region had raised activity for greater than 10 ms, its start and end (maximum of 400 ms) identified the early component; otherwise, the whole post-stimulus interval from 0-400 ms was taken as the early component. The later component was then identified as the remaining interval until 550 ms after stimulus appearance (minimal delay interval). The visual-transient index was then calculated as: $\frac{S}{T+S}$, where S was the mean activity in the later (sustained) component, and T was the mean activity in the early (transient) component. The distribution of index values for each metric is shown in Figure 2.1C, To divide the cells into transient and sustained classes we chose a value of 0.2625, which was a natural division in the distribution of index values.

SC neurons have well-characterized responses ranging from purely visual to purely motor [Mays and Sparks, 1980, McPeek and Keller, 2002, Mohler and Wurtz, 1976, Munoz and Wurtz, 1995]. Visual only cells with transient visual responses and no saccade-related activity (VT) tended to be located more superficially than the other classes of visually-responsive cells (see Figure 2.1D). Thus VT cells were typically found in the upper superficial gray layer (*e.g.* SC_S), where retinal Y-type cells terminate directly and indirectly through magnocellular lateral geniculate nucleus and V1 [May, 2006]. Visual-only cells that had sustained visual responses (VS) typically paused during saccadic eye movements (Figure 2.1A). Previously we have shown VS neurons to be sensitive to color signals whereas VT cells were not [White et al., 2009], suggesting parvocellular input. These features, along with a mean depth of about 900 μ (Figure 2.1D), suggest they were located in the lower superficial layers. This area receives visual input from higher occipital and parietal areas [Graham et al., 1979, Tigges and Tigges, 1981]. The VS neurons we identified were likely the same as "visual-tonic" neurons described previously [Li and Basso, 2008, McPeek and Keller, 2002]. Finally, visuomotor cells with transient or sustained activity - VmT or VmS - were easily characterized because of their bursts of activity time-locked to saccades and their bursts of visual activity time locked to stimulus onset (Figure 2.1A). Our sample of visuomotor neurons was by necessity biased to those with robust visual responses and were always found more than 1mm below the dorsal surface (Figure 2.1D).

Figure 2.2: (A) Spike density functions for visual transient (VT, n=32), visual sustained (VS, n=32), visuomotor transient (VmT, n=16), and visuomotor sustained (VmS, n=18) neurons in response to 7 repeated stimuli (shown as small dark bars at bottom of trace) in the center of each neuron's response field. (B) Color coding of intensity of neural activity in response to the 7 stimuli (time of the response to a given stimulus on the horizontal, response to each stimulus descending vertically, color coded for normalized spike rate). Note the shift in onset latency with each stimulus repetition. (C) Changes in mean response onset latency across stimulus number for each neural type. (D) Changes in peak response magnitude across stimulus number for each neural type, normalized to the response on the first stimulus. (E) Population spike density waveforms in response to the first target stimulus, aligned on response onset to show the early (transient) and later (sustained) components of the visual response. (F) Normalized mean sustained activity (50 ms to 100 ms after onset of visual response) is plotted for the 7 stimuli for VS and VmS neuron populations. (G,H) Scatter plots showing the relationship between the response to the first and second stimulus for the transient peak (G) and sustained portion (H) of the neural response. Standardized major axis regression analysis revealed that this relationship had a slopegreater than unity for the peak activity (F test, F(1,96)=72.32, $p < 0.01$), but not for the sustained activity (F test, F(1,48)=0.99, $p = 0.32$).



2.3.2 Behavioral Task

Monkeys actively fixated a central fixation spot (grayscale circle of 0.25°diameter presented at 1.1 cd/m^2) while a series of seven light flashes (*i.e.* stimuli, 0.25°diameter, 55 ms duration, Figure 2.1B, 2.2A) were presented in the receptive field of the monitored neuron. In the main paradigm, these 7 stimuli were separated by intervals of 200 ms duration (*i.e.* 255 ms interstimulus interval (ISI)). Monkeys received a small liquid reward for maintaining fixation within a small computer-controlled window (1-3°square window) for the duration of each trial. If fixation was broken prior to the end of the trial, the trial was aborted, eliminated from further analysis, and recycled back into the trial sequence. In the main paradigm, 70% of the trials ("Control" trials) consisted of 7 equiluminant stimuli (1.1 cd/m^2) and 30% of the trials ("Oddball" trials) were identical except that the fourth stimulus could be brighter (10%, 5 cd/m^2), dimmer (10%, 0.1cd/m^2), or absent (10% of trials). These trial types were randomly interleaved. Trains of 7 stimuli were chosen because they allowed for examination of responses before and after presentation of the oddball, and it was a comfortable trial duration for the monkey to maintain steady fixation. The ISI of 255 ms was chosen because maximal inhibition of return was observed in monkeys at a cue-target onset asynchrony of that interval [Fecteau et al., 2004].

For a subset of neurons (N = 19 tested and 17 analyzed, 2 removed for having no response to some stimuli), the ISI was varied systematically between 155, 255, or 455 ms in the control condition only, to investigate the effects of ISI on the repetition effect. Typically the 255 ms ISI block was collected first because that was part of the main paradigm used with the oddball trials. If neuronal isolation remained strong after that paradigm, additional files were obtained at other ISIs collected in pseudo random order.

2.3.3 Neural Analysis

Single neurons or pairs of single neurons were recorded from a single electrode, isolated online using the window discriminator in Plexon, and verified and optimized offline using Plexon's Offline Sorter. The timing of events in the trial sequence was then calculated automatically using custom Matlab (Matlab 6.1 Mathworks Inc) software during offline analysis. We recorded from a total of 109 neurons in the control task with oddball trials. Of these, recordings from 98 neurons (60 from monkey Q, 38 from monkey Y) had sufficiently good spike isolation throughout recording, a mean visual response greater than 40 spikes/sec, responses to all 7 stimuli during the control trials, and at least 6 trials of each oddball condition. Typically, there were 10-20 repetitions of each oddball condition and 70-140 repetitions for the control condition. Two spike density functions were created for each trial of each condition by convolving the trains of action potentials with a Gaussian kernal (σ=5) or by convolving with a combination of growth and decay functions that resemble a postsynaptic potential given by:

$$R(t) = \left(1 - e^{\frac{-t}{\tau_g}}\right)\left(e^{\frac{-t}{\tau_d}}\right) \tag{2.2}$$

where R is the firing rate as a function of time, τ_g is a time constant for the growth phase, and τ_d is a time constant for the decay phase. Time constants of 1 and 20 for the growth and decay phases, respectively, were chosen following others [Thompson et al., 1996]. Spike density functions were aligned on the first stimulus onset and activity from repeated trials were averaged to generate a mean spike density function (for both functions separately) for each neuron for each condition. The magnitude of the first peak response to each visual stimulus was calculated off the Gaussian spike density function and the onset of that response was calculated from the spike density function created by rate function R, which provides a more accurate measurement of onset time. This was done for the responses to each of the 7 stimuli in every condition, for each neuron. To find the first peak in the visual response and its onset, a custom computer algorithm written in Matlab looked for all peaks in the spike density function in the epoch from 50 ms after stimulus onset until end of the ISI. It then marked the highest peak (usually the first one) on a visual display. Each first peak calculation was manually examined and changed if it was incorrect (*e.g.* if a late noisy peak was incorrectly chosen as the first main peak by the algorithm). Once the peak was determined, the onset latency of that visual response was calculated by an algorithm which looked backward in time (maximum of 40 ms back) along the descending slope from the first peak in order to find the point at which the response became significantly greater then the mean neural activity in an epoch spanning 25 ms before to 25 ms after the onset of the visual stimulus that generated that response. Again, these were each examined on the visual display, and manually adjusted if necessary (*e.g.* if unusual noise levels, or sustained activity from the preceding response unduly lengthened the ROL calculation of the algorithm)

2.3.4 ROC Analysis

The receiver operating characteristic (ROC) was used to quantify the time course of differences between control and oddball conditions after the presentation of the 4th stimulus in the sequence. For each cell we computed the area under the ROC curve between the control condition and each oddball condition in a 50 ms sliding window centered on the time point of interest (gray box bottom left of 2.6A). The window's left edge started at the onset of the visual response and advanced 1 ms in time (depicted by the solid line) until the right edge reached the onset of the next (5th) stimulus. This resulted in one ROC area measurement for each cell and each time point on the interval shown (which started at $\frac{1}{2}$ the window size, 25 ms after the onset of the visual response). To control for variation in timing of each cell's response to the 4th stimulus, it was necessary to first align each spike density function to the onset of visual activity to the 4th stimulus. As a result of this realignment the waveforms from different neurons and conditions had slightly different lengths since the time of the visual onset varied depending on cell type and condition, and consequently, the ROC analysis was performed over a different length of time for each cell and condition. As a result, fewer cells entered the ROC area calculation near the end of the analysis interval. The length of the analysis interval shown in Figure 6 (below VT spike density function) was the maximum interval that could be chosen that still contained greater than 50% of the cells for all classes and conditions (all but visuomotor transient cells had more than 80% at the end of this interval). All cells in all conditions had at least 122 ms from the onset of the visual response to the onset of the next stimulus, and the median was 152 ms.

2.3.5 Implementation of the computational model

The Bayesian model of adaptation is summarized in Figure 2.3A. The model is based on Surprise theory using a Poisson-Gamma model which is described in detail elsewhere [Baldi and Itti, 2010, Itti and Baldi, 2005] but summarized here noting differences in our implementation. The model consists of two stages of Bayesian learning which are identical except for their input sources (Figure 2.3A), so for clarity the equation subscripts are omitted from the following discussion. We consider that each Bayesian learner receives 1-dimensional Poisson-distributed spike trains (from the retina and visual cortex or from the previous stage of learning) represented internally as a family of models, $M(\lambda)$, which are all the possible 1-dimensional Poisson distributions of firing rates ($\lambda > 0$). Each learner builds probability distributions (hypotheses or beliefs), $P(M(\lambda))$, about which of these models best represents the current state of the stimulus. As is typical in iterative Bayesian learning, the prior and posterior are chosen from the same functional form (conjugate priors) so that the posterior at one time step is used as the prior for the next. When the data is Poisson-distributed ($D = \lambda$) the Gamma probability density function is the conjugate prior, $P(M(\lambda))$:

$$P(M(\lambda)) = \gamma(\lambda; \alpha, \beta) = \frac{\beta^{\alpha} \lambda^{\alpha-1} e^{-\beta\alpha}}{\Gamma(\alpha)} \tag{2.3}$$

with shape $\alpha > 0$, inverse scale $\beta > 0$, and where $\Gamma(\alpha)$ is the Euler Gamma function of α. Given an input sample $D = \lambda$, the posterior distribution of beliefs over the possible input firing rates is also a Gamma distribution characterized by:

$$\acute{\alpha} = \zeta\alpha + \lambda + \varepsilon \text{ and } \acute{\beta} = \zeta\beta + 1 \tag{2.4}$$

where $\acute{\alpha}$ and $\acute{\beta}$ are the shape and inverse scale of the posterior distribution, ζ is a temporal parameter (forgetting factor, $0 < \zeta < 1$) which determines the rate of learning, and ε is a constant representing noise. The second stage of Bayesian learning takes as input the expected value of the first stage's posterior distribution:

$$E[P(M(\lambda)|D)] = E[\gamma(\lambda; \alpha, \beta)] = \frac{\alpha}{\beta} \tag{2.5}$$

The output of the system is then calculated from the final Bayesian learner as the Kullback-Leibler divergence [Kullback and Leibler, 1951] between prior and posterior distributions over all possible firing rates, which summarizes the amount of learning or adaptation which just resulted from observing the data D:

$$\begin{aligned} KL(P(M(\lambda)), P(M(\lambda)|D)) = KL(\gamma(\lambda; \alpha, \beta), \gamma(\lambda; \acute{\alpha}, \acute{\beta})) = \\ \acute{\alpha} \log \frac{\beta}{\acute{\beta}} + \log \frac{\Gamma(\acute{\alpha})}{\Gamma(\alpha)} + \acute{\beta}\frac{\alpha}{\beta} + (\alpha - \acute{\alpha})\Psi(\alpha) \end{aligned} \tag{2.6}$$

where $\Psi(\alpha)$ is the digamma function of α. This differs from the Itti & Baldi implementation [Baldi and Itti, 2010, Itti and Baldi, 2005] where the KL divergence is computed at each learning stage, and the system output is the product of the outputs at each stage. Figure 2.3B shows the time dynamics of the system for each stage in response to a control trial.

The Itti & Baldi implementation uses five learning stages each having the same temporal parameter. In this experiment we found that only two stages, but allowing each stages temporal parameter to be different, adequately predicted the peak firing rates of the neurons. Additionally, in their implementation of Equation (2.4) the temporal parameter is applied to the prior distribution's α and β parameters before computing the Bayesian update. As a result, there is always a baseline output. We computed the update so that if the posterior and prior are the same, the output of the system is 0.

2.3.6 Model Fitting

Model parameters were estimated for each neuron individually by fitting the peaks in the model's output to the seven peak magnitudes in each neurons response profile from the ISI 255 ms condition. Each model neuron consisted of three parameters: two time constants (ζ, eq.(2.4)) that controlled the speed of learning in the two Bayesian learners and the baseline noise parameter (ϵ, eq. (2.4)) which was globally set for all model neurons. The data were fit initially with all three parameters free for each cell. After fitting, a probability density function of the baseline parameter was estimated using kernel density estimation with automatic bandwidth selection [Jones et al., 1996], which was implemented in the R statistical package. A quadratic function was fit (least squares method) to 3 points centered on the maximum of this curve, and the analytic maximum of the quadratic function was used as an estimate of the most likely value of the baseline parameter. Fitting was then performed again with the baseline parameter fixed for all cells to the most likely value, reducing the model to the two temporal parameters. The best parameters were determined by using the Nelder-Mead simplex method [Lagarias

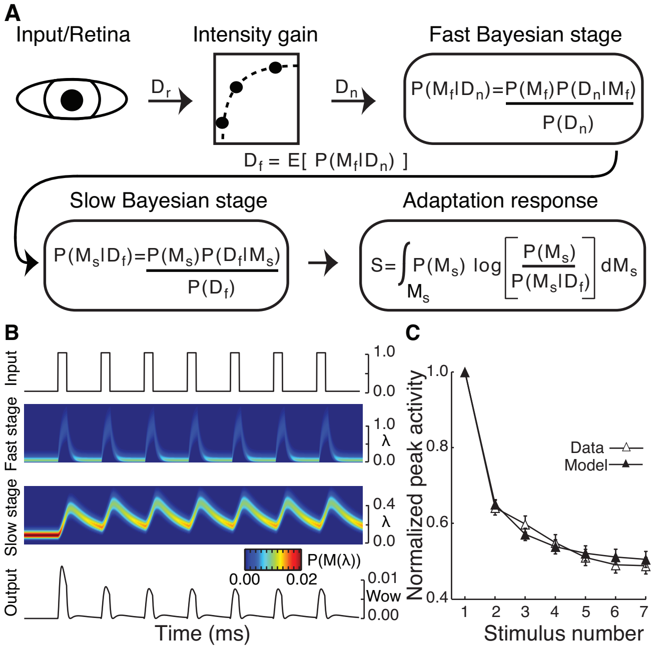


Figure 2.3: (A) Schematic of the Bayesian adaptation model. Light stimulating the retina was modeled as a square wave of unity amplitude (D_r; 1.1 and 0.9 for the brighter oddball conditions) and passed through a static gain function that was constant for all model neurons (see Methods). Two stages of Bayesian learning supply the adaptation dynamics. In each stage (subscripts omitted), the learning process builds hypotheses or beliefs (probability distribution) over a class of internal models M that represent all possible values of its input. As new sensory data D_r is collected, Bayes theorem provides the mechanics to turn a prior set of hypotheses $P(M)$ about which model best characterizes the input data into a posterior set of hypotheses $P(M|D)$, given the likelihood of the data $P(D|M)$ under the assumptions of model M. The fast Bayesian stage quickly adapts to the input and passes the expectation of its posterior beliefs D_f as input to the second Bayesian stage. A posterior set of beliefs is computed in the same fashion as the fast learner, but with a slower learning dynamic. The adaptation response is then calculated for every data observation as the Kullback-Leibler (KL) divergence [Kullback and Leibler, 1951] between the slow learner's prior and posterior hypotheses, signaling the amount of shift in the model's beliefs caused by each new observation. (B) Detailed view of the model dynamics across each stage during a control trial (see methods for a detailed description of the model). Top trace represents the input stimulus from control trials. The two central images show, for each Bayesian learner, the distribution of beliefs about which of the possible Poisson firing rates (y-axis) best characterizes the input over the course of a single trial (x-axis). Hotter colors indicate that, at a given point in time, there is a higher belief (probability) in a particular firing rate. The bottom panel shows the final output of the system. (C) Population mean and standard error of the model (filled symbols) and neural (open symbols) normalized peak responses to the 7 stimuli in the control condition.

et al., 1998] built into Matlab to minimize the error of the following process: First the input signal (7 stimuli) was simulated as a square wave with unit amplitude and the adaptation model's response was computed for a given parameter set. Parameters were encoded such that the second stage's learning rate was guaranteed to be slower than the first stage's. Model and cell responses were normalized by the response to the first stimulus. Normalization eliminated the need for scaling parameters in the model without affecting the morphology of the adaptation. The error was then computed as the median absolute difference between the model's peak response to each stimulus and the cell's peak responses (disregarding the first stimulus which always had zero error). Several alternative error functions were explored, and median absolute error was chosen because it gave highly significant, and qualitatively the best, overall fits and predictions. Several other error functions also gave significant results. Additionally, a single-parameter model significantly fit the data; however, the two-parameter model produced less total error for all conditions combined, and less median error for all but the 155 ms ISI condition. Qualitatively, the mean population responses of the neural data were in agreement with the mean population responses of the two parameter model.

To account for the relationship between stimulus brightness and a cell's peak firing rate, the adaptation model used a gain function (Figure 2.3A). Because only three intensity levels were considered, this amounted to finding two gain factors to represent the 10% brighter and 10% dimmer stimulus. After finding the best temporal parameters during control trials, a single set of gain parameters was chosen that minimized the error between all model neurons and all real neurons simultaneously, only considering the brighter and dimmer conditions. For this data set, the gain factors were 1.1 and 0.9, respectively.

Because the gain factors were very close to the 10% brighter and 10% dimmer input, the gain function could have been omitted with little loss of model fit quality.

2.3.7 Model Evaluation

To evaluate the model fits without assumptions about the distribution of data or errors, several statistics were computed in the permutation (randomization) framework. Goodness-of-fit was assessed by using the median of the absolute error between all neuron and model responses, for all conditions, as a test statistic in a repeated measures permutation design (stimulus 2 to 7 for each condition). To assess whether cell and model responses came from the same underlying distribution, the permutation equivalent of a two-factor repeated measures ANOVA was performed. The final test (paired-error or reliability test) indicated whether, overall, the model was able to predict neuronal responses better than other cells from the same class (which might be thought of as an upper-bound). First, for each condition separately, all pairwise combinations of neurons (restricted to within neuron class) were evaluated with the error function. This distribution of values represents the errors that occurred when each cell was used to predict other cell responses, and served as a summary of the variability (reliability) of the repetition effect within a class of neurons. Higher values indicated that neurons responded very differently from one another. Using the permutation equivalent of a two-factor repeated measures ANOVA, this distribution was compared to the distribution of errors generated by model predictions. Figures 2.4D, and 2.5C show the distribution of model errors for the ISI, and oddball manipulations, respectively. This test compared directly the quality of our model fits to the variability of the repetition effect. We reason that a well performing model should be on average as, or more, consistent with the neurons' response than neurons of the same class are with each other. All permutation tests were carried out using the Monte Carlo method with 30,000 iterations.

2.4 Results

2.4.1 Effects of stimulus repetition

Figure 2.2 illustrates the main effect of this study - the large response decrement that occurred with repeated stimulation (7 stimuli) of the receptive field of visually-responsive neurons in the SC. Response decrement was observed for all four types of visual neurons classified: visual transient (VT, n=32), visual sustained (VS, n=32), visuomotor transient (VmT, n=16), and visuomotor sustained (VmS, n=18) neurons (see Figure1B for examples of individual neuron responses). Following the appearance of the first stimulus, neurons of each cell type discharged a robust phasic response (Figure 2.2A, E). The early transient part of this response was dramatically affected by repeated stimulation: the peak response magnitude decreased (Figure 2.2A, B, D) and response onset latency (ROL) increased (Figure 2.2B, C). A mixed analysis of variance (ANOVA) with a between-subjects factor (4 cell classes) and a within-subjects factor (7 stimuli) was conducted which revealed significant main effects of cell class [Peak: F(3,94)=9.15,$p < 0.01$; ROL: F(3,94)=8.8, $p < 0.01$], stimulus repetition [Peak: F(6,94)=378.1, $p < 0.01$; ROL: F(6,94)=89.9, $p < 0.01$] and an interaction [Peak: F(18,564)=5.8, $p < 0.01$; ROL: F(18,564)=3.3, $p < 0.01$].

All cell types decreased their peak response magnitude with repetition (Peak: VT [F(6,186)=143.06,$p < 0.001$]; VS[F(6,186)=71.69, $p < 0.001$)]; VmT[F(6,90)=114, $p < 0.001$)]; VmS[F(6,102)=71.5, $p < 0.001$)] and the majority of the decrease occurred on the second stimulation. This was verified statistically: the ratio of peak magnitude between the 1^{st} and 2^{nd} stimuli (mean= 0.36) was greater than the ratio of peak magnitude between the 2^{nd} and 7^{th} stimuli (mean=0.15) [t(97)=7.9, $p < 0.001$; paired t-test]. This relationship was confirmed for all cell types independently (p=0.04 or less). ROL increased in a mostly linear fashion with repetition for all cell types (VT [F(6,186)=18.1, $p < 0.001$]; VS[F(6,186)=42.5, $p < 0.001$)]; VmT[F(6,90)=15.9, $p < 0.001$)]; VmS[F(6,102)=17.5, $p < 0.001$)].

In summary, VT neurons, likely found in the most superficial retino-recipient SC layers [May, 2006], had strong adaptation (∼50%) but the smallest ROL increase (∼10 ms) of all cell types. VS neurons, likely found in lower superficial layers [Tigges and Tigges, 1981] showed the least adaptation (∼35%) but a large increase in ROL with repetition (∼15 ms). Finally, Vm neurons of the intermediate SC layers showed both strong adaptation (>50%) and a large increase in ROL (>15 ms), particularly the VmT cells. Indeed some VmT cells (not described here) completely lost their visual response after only a few trials [Goldberg and Wurtz, 1972a] and thus could not be studied in our paradigms.

To determine whether the later components of the visual response in cells with a significant sustained component (VS and VmS as defined by our cell classes) were also affected by repetition, we calculated the average activity from 50-100 ms after the response onset (Figure 2.2E). The sustained activity was less affected by repetition (Figure 2.2F, H) than the early transient component (Figure 2.2D, G). An ANOVA on the sustained activity of VS and VmS neurons showed a far smaller main effect of repetition [F(6, 48)=6.75, $p < 0.01$] compared with that seen in the transient component, and no main effect of cell class, nor an interaction (F's<1).

2.4.2 Modeling the response decrement using a Bayesian framework

The effect of stimulus repetition on response magnitude was modeled using a simple Bayesian model of stimulus adaptation which monitored the temporal dynamics of streams of stimuli (see Figure 2.3A, B, and Methods). The model relies on a recently developed Bayes-optimal theory of novelty, shown to provide a quantitative account of adaptation in early visual neurons [Baldi and Itti, 2010, Itti and Baldi, 2005] This model provided a principled theoretical foundation for quantifying the effects of adaptation in terms of a hypothetical optimal Bayesian learner: stimuli that over time gave rise to no significant learning caused a rapid decrease in response (adaptation); in contrast, stimuli that caused a shift in the model's current estimates gave rise to significant learning and to vigorous model responses. Each neuron was modeled individually as three stages consisting of a static gain function and two Bayesian learners (Figure 2.3A). Parameters were estimated by fitting each model neuron's peak responses to a real neuron's peak responses to all seven stimuli (see Methods). The model was able to significantly fit the repetition effect (goodness-of-fit test, $p < 0.01$), and the population responses for model and real neurons overlapped (Figure 2.3C).

2.4.3 The effect of stimulus presentation rate

We generated predictions for the repetition effect's dependence on the rate of stimulus presentation by altering the inter stimulus interval (ISI) of inputs to the population of model neurons. If the response decrement followed the adaptation model predictions then decreasing the ISI to 100 ms would cause a stronger repetition effect, while increasing the ISI to 400 ms would allow recovery of the effect of previous stimulation. We tested these predictions in a subset of 17 neurons (3 VT, 7 VS, 3 VmT, 4 VmS) by repeating control trials (7 identical stimuli of 55 ms duration) with these different ISIs (onset to onset time = 155, 255 and 455 ms). Figure 2.4A shows the combined spike density functions from this subpopulation. There was a clear effect of ISI on both peak response magnitude [F(2,32)=29.14, $p < 0.01$)] and ROL [F(2,32)=28.5, $p < 0.01$]. That is, the shorter the ISI, the more dramatic the repetition effect. This was confirmed by an interaction between ISI and repetition [Peak: F(12,192)=8.44, $p < 0.01$; ROL: F(12,192)=6.9, $p < 0.01$] . Reducing ISI led to an increase in ROL (Figure 4B), and a reduction in response magnitude (Figure 2.4C). The main effect of repetition, as expected, was significant [Peak: F(6,96)=46.29, $p < 0.01$; ROL: F(6, 96)=32.27, $p < 0.01$]. Remarkably, we found that our simple model was able to predict the pattern of response magnitudes observed in these other ISI conditions (Figure 4C) without a change in parameters (goodness-of-fit test, $p < 0.01$) and was a better predictor of neural activity than other neurons (see Figure 2.4D).

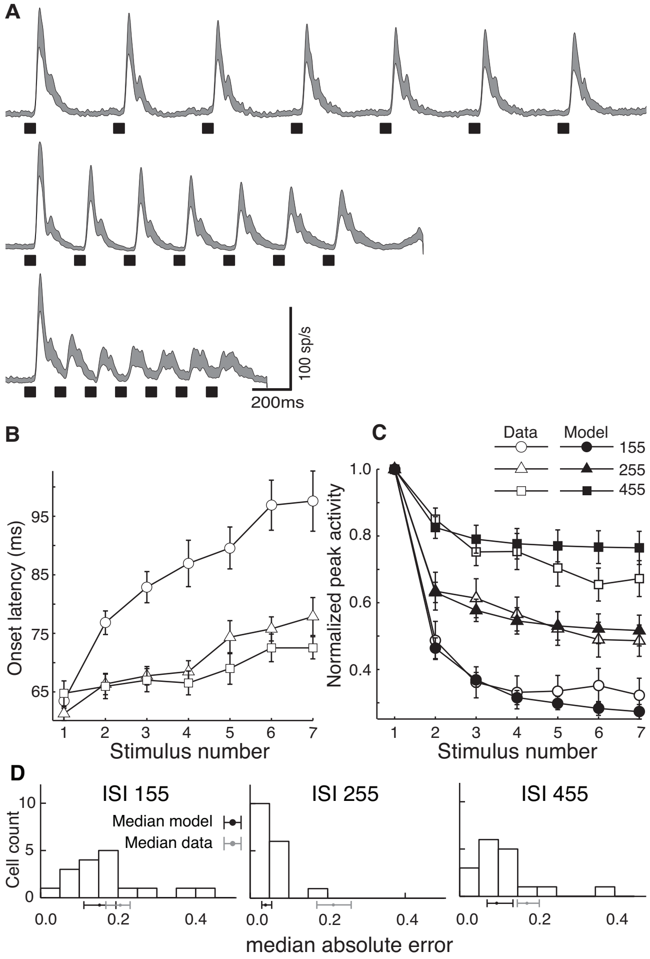


Figure 2.4: Effect of changing interstimulus interval (ISI) on the repetition effect. (A) Population spike density waveforms recorded from 17 visually responsive neurons in the SC in response to 7 stimuli (55 ms) presented with ISIs of 155, 255, 455 ms. As ISI increased, the repetition effect was reduced. At short ISIs the response onset latency (B) was increased and peak response magnitude (C) was decreased with stimulus repetition. Without changing the parameters used to generate the model fit to control trials (2.3C), the model (closed symbols) predicted the neurons peak response magnitude (open symbols) across the different rates of stimulus presentation (two-factor repeated measures permutation-test, $p = 0.89$). (D) The paired-error test (see Methods) indicated that, on average, each model was a better predictor of the peak activity of its corresponding real neuron, than other neurons of the same class ($p < 0.01$). Histograms of median absolute errors between each neuron and its corresponding model for the three ISI conditions are shown in (D). The black dot and line below each axis show the median error and 95% bootstrapped confidence interval (30,000 iterations) of model errors. The gray dot and line show the median error and 95% bootstrapped confidence interval from the distribution of pairwise errors between actual neurons (see Methods). Notice that in all cases the models' median error is less than the median error between neurons.

2.4.4 The effect of rare changes in stimulus luminance

We also modeled the effect of inserting rarely presented luminance oddball stimuli (brighter, dimmer, absent) into the stimulus sequences. If the pattern of changes were due purely to adaptation, the neural response should follow the predictions of the adaptation model and recover somewhat with a brighter stimulus, but not a dimmer stimulus, and have an opposite trend on the stimulus following the oddball. Alternatively, neurons could show response recovery to all the rare stimuli, akin to 'dishabituation'. To test these predictions, we examined the visual response of SC neurons to sequences of 7 stimuli where the 4^{th} was of higher intensity (10% of trials), lower intensity (10% of trials), was absent (10% of trials), or had no change (70%). Figure 2.5A illustrates the population responses recorded from each type of neuron for the 3^{rd}, 4^{th} and 5^{th} stimuli. We found that the peak magnitude of the neural data conformed to the predictions of the adaptation model using the same parameters as in the control condition (see Figure 2.5B to contrast physiological data with model fits). The model fit the data significantly (goodness-of-fit test, $p < 0.01$), and was a better predictor of neural activity than other neurons in the same class (Figure 2.5C).

Figure 2.5 D and E show the normalized difference [(oddball - control) / (oddball + control)] between the oddball and control trials for ROL and the peak magnitude, respectively (all cell types were collapsed because the changes in the early part of the visual response were qualitatively the same in all cell types). Significance was tested using a Bonferroni corrected t-test (critical t = 2.49, $p < 0.05$). As expected, there was no difference between control and oddball trials on the 3^{rd} stimulus for any condition

(all t's < 2.49). Presentation of the brighter stimulus in the 4^{th} position led to a larger magnitude response [t(97)=5.67] at a shorter latency [t(97)=-6.36], while presentation of the dimmer stimulus showed the opposite effect - smaller peak response [t(97)=-3.7] with a longer latency [t(97)=8.79]. Furthermore, the changes in the latency and magnitude of the 4^{th} response had predictable consequences on the earliest part of the response to the 5^{th} stimulus. If the 4^{th} stimulus was brighter, the response to the 5^{th} was reduced [t(97)=-6.22] and arrived later in time [t(97)=6.72] compared to the control condition. In contrast, when the 4^{th} stimulus was dimmer, the response to the 5^{th} stimulus was larger in magnitude [t(97)=7.6] but not significantly earlier in time [t(97)=-2.36]. In the absent 4^{th} stimulus condition, the response to the 5^{th} stimulus was much larger in magnitude [t(97)=11.24] and occurred earlier in time [t(97)=-6.02]. These analyses indicate that the pattern of changes observed in the timing and magnitude of the early transient component of the response to oddball stimuli are consistent with predictions of Bayesian adaptation to stimulus intensity.

2.4.5 Sustained responses to novel events

The best fitting adaptation models often did not produce any output during the inter trial interval (sustained response), but a sustained response that showed some modulation to repeated stimuli was observed in sustained cell classes (Figure 2.2E,F). To investigate whether the sustained activity could possibly reflect something other than simple adaptation, we analyzed the later part of the visual response to oddball trials (Figure 2.6). First, we realigned the visual responses in control and oddball conditions to the onset of the response to the 4^{th} stimulus (see traces on left side of each panel in Figure 2.6),

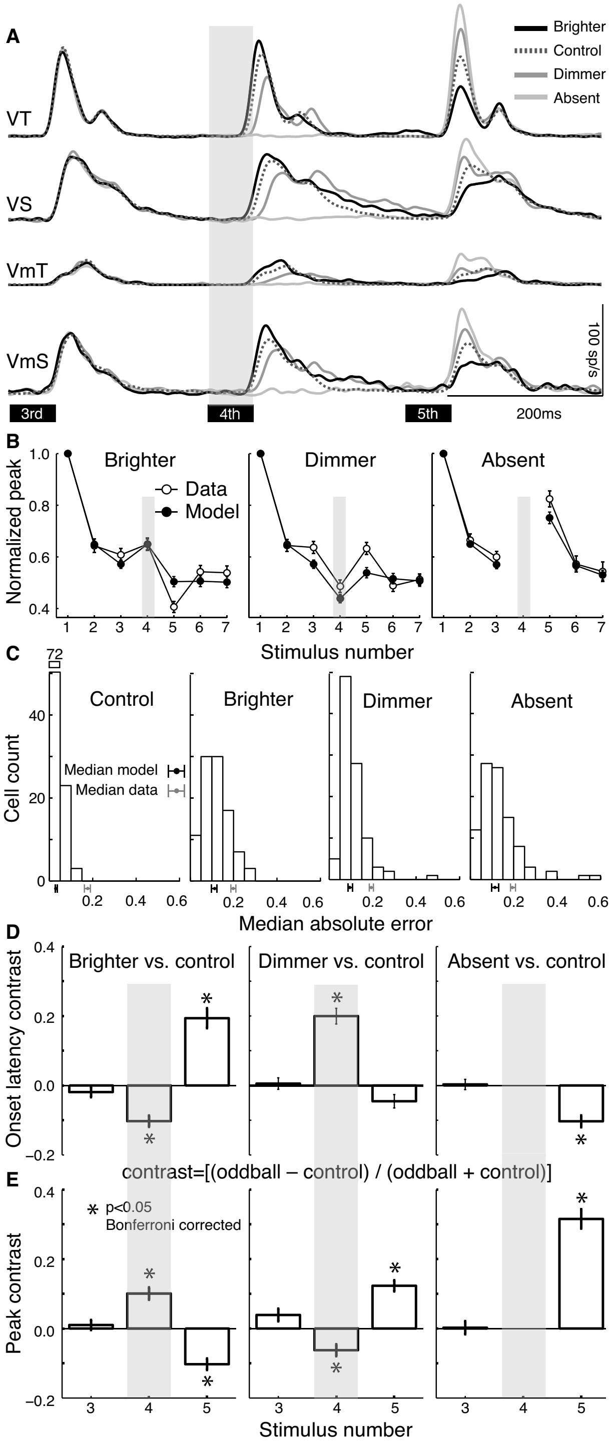


Figure 2.5: Changes in the early transient part of visual responses to oddball stimuli presented in the 4^{th} stimulus position in the sequence as depicted by the gray shaded bar across the different panels. (A) Population spike density waveforms showing the response to the 3^{rd}, 4^{th} and 5^{th} stimuli are plotted for the 4 classes of visually responsive neurons studied (same neurons as 2.2). (B) Changes in the normalized peak response magnitude across stimulus number for model (closed symbols) and neural responses (open symbols) were not significantly different (two-factor repeated measures permutation-test, $p = 0.61$). The paired-error test (see Methods) indicated that, on average, each model was a better predictor of the peak activity of its corresponding real neuron, than other neurons of the same class ($p < 0.01$). Histograms of median absolute errors between each neuron and its corresponding model for the control and oddball conditions are shown in (C), conventions are the same as in 2.4D. Mean normalized difference (*i.e.* contrast) in ROL (D) and mean normalized difference in peak response magnitude (E) between the control condition and each oddball condition calculated as [(control-oddball)/(control+oddball)]. Negative and positive values mean that oddball conditions had lower and higher values respectively compared with control conditions.

correcting for the ROL difference between brighter and dimmer stimuli and between cells (Figure 2.5D). We then performed a receiver operating characteristic (ROC) analysis on the response spanning from 25-120 ms after response onset (right side of each panel in Figure 2.6) to determine when the response to the control and oddball trials became significantly different (see Methods). This analysis interval started earlier than that used in Figure2 to show the effect of the first visual volley in the ROC plots. For all cell types except VmT, immediately after the onset of the visual response there was a significant difference in the transient response in the brighter condition that became insignificant approximately 30 ms after the ROL. That is, the peak transient activity faithfully reflected stimulus intensity - the brighter stimulus elicited the strongest response, the dimmer stimulus the weakest. However, later in the response of the VS and VmS cells (bottom panels), there was a significant increase in activity for both the brighter and dimmer oddball conditions, possibly representing a dishabituation signal reflecting the novelty of the oddball stimuli. The time point after ROL when the brighter and dimmer stimulus responses diverged from control responses was at 73 ms and 68 ms, respectively, for VS neurons, and 95 ms and 81 ms, respectively, for VmS neurons (see vertical dotted lines, and p-values plotted below the ROC area curves). To demonstrate how consistent this was across individual neurons, in Figure 2.6, E-F the mean sustained firing rate after the 4^{th} stimulus for control trials is plotted against that for oddball trials for VS and VmS neurons respectively. Points falling above the unity line show neurons whose rate was higher after the oddball stimulus vs. control stimulus (sustained epoch 80-110 and 90-120 for VS and VmS neurons respectively). In the inset graphs we show the grand mean firing rate with standard error bars for the control and oddball stimuli. The rate for oddball stimuli was signficantly greater than control for each comparison (paired t-test, 1-tailed; all $p < 0.002$) There was no change in the later portion of the visual response for VT and VmT neurons (Figure 2.6, A-B).

In sum, the oddball manipulation shows that the pattern of effects seen in the peak of the transient response was consistent with the adaptation to light intensity computed by our model; however, a significant dishabituation signal (enhanced response to both brighter and dimmer stimuli) not seen in the model response was present in the later part of the visual response only in neurons with sustained activity.

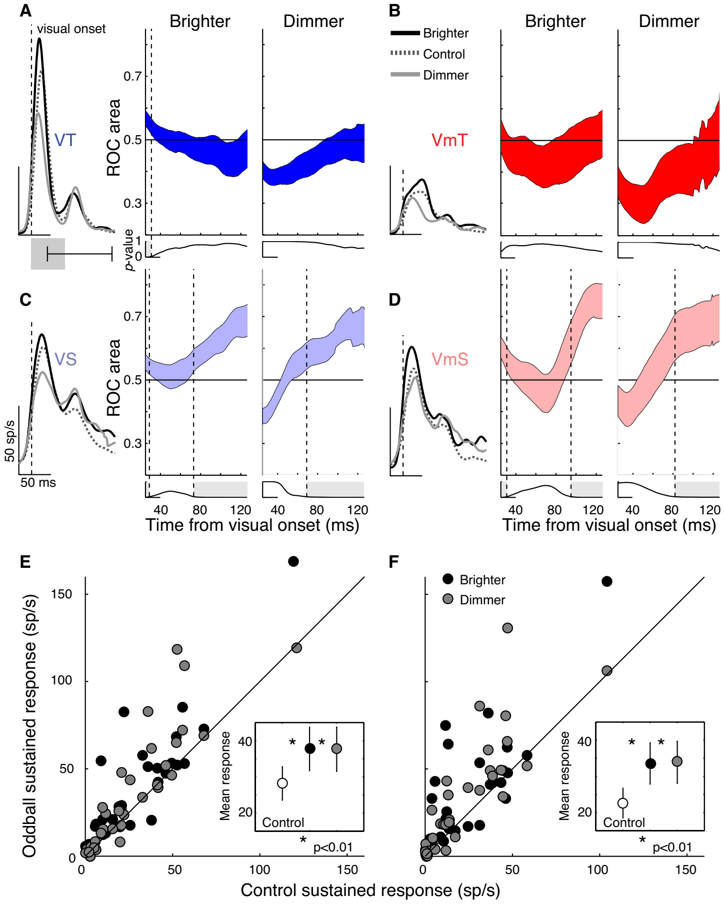


Figure 2.6: Changes in the later sustained part of visual response to oddball stimuli presented in the 4^{th} stimulus position are shown for each neuron type (A) VT; (B) VmT; (C) VS; (D) VmS. An ROC analysis was performed to determine at what points in time the later (sustained) part of the visual response became significantly different between control trials and either brighter or dimmer trials. The overlaid spike density functions show the average activity for the control, brighter, and dimmer conditions aligned to the onset of the transient visual response (see Methods) to the 4^{th} stimulus. The filled colored regions represent the standard error of the ROC area across all cells of the same class, for the brighter and dimmer conditions separately. A ROC area of 0.5 or less indicates no difference between the control and oddball conditions at that particular time point, while values greater than 0.5 indicate the oddball condition had more activity on average. Each point was tested with a 1-tailed t-test to determine if the ROC area was significantly greater than 0.5. The p-value of this test is plotted below each ROC area plot and the vertical dotted lines and light gray shaded regions indicate when the p-value crossed the significance threshold ($p < 0.05$). All ROC area's for all time points were normally distributed (Kolmogorov-Smirnov, $p < 0.05$). (E) Scatterplot of the mean sustained activity for the 4^{th} stimulus in control trials vs. mean sustained activity in the brighter or dimmer oddball stimuli for VS neurons. Inset graph shows the population mean sustained activity with standard error bars for the control, brighter and dimmer stimuli. Asterisk shows significant difference between oddball and control activity rates (paired t-test, 1-tailed). (F) As described in E, but for VmS neurons.

2.5 Discussion

The timing and magnitude of the visual response of SC neurons underwent significant modification following stimulus repetition: the earliest part of the visual response decreased in magnitude and increased in latency with repetition (Figure2). The modulation of this early response with repetition was successfully modeled using our Bayesian adaptation model (Figure3) and predictions made about the effect of changing the rate of stimulus presentation (Figure4) and the intensity of rare stimuli (Figure5) were confirmed with neural data. The repetition effect was strongly dependent on the rate of stimulus presentation (Figure 2.4), with the repetition effect increasing in magnitude as the interval between stimuli was reduced. For brighter or dimmer oddball stimuli, the main features of the repetition effect followed simple adaptation to light - larger, earlier responses to the brighter, and smaller, later responses to the dimmer oddball stimuli and the opposite pattern in response to the next (non-oddball) stimulus. In contrast, the later, sustained component of the visual response was modulated much less by repetition, as observed previously in V4 [Motter, 2006], and was inconsistent with our Bayesian adaptation model. Finally, in response to either brighter or dimmer oddball stimuli we observed an increase in response (*e.g.* a dishabituation) in this later sustained firing, suggestive of a “novelty response”.

2.5.1 Comparison to other studies

Reductions in response magnitude with repetition have been previously observed in cortical areas including V1 [Müller et al., 1999], V4 [Motter, 2006] and frontal eye fields [Mayo and Sommer, 2008], and also in single neurons of the SC [Goldberg and Wurtz, 1972a, Woods and Frost, 1977] and multi-unit activity of the SC$_s$ [Mayo and Sommer, 2008]. In the context of an attentional cueing task, a repetition effect has been described in the SC [Bell et al., 2004, Dorris et al., 2002, Fecteau et al., 2004, Robinson and Kertzman, 1995] and LIP [Robinson and Kertzman, 1995]. The present report is the first to systematically explore the repetition effect using long stimulus sequences studied across different cell types and layers in the SC, the first to report the increase in response onset latency with repetition in the SC, and the first to explore the mechanism for this response decrement through modeling and experimental manipulations (oddball stimuli).

We observed a significant increase in ROL with repetition and with changes in stimulus intensity (oddball experiment). Modulation of ROL with intensity is consistent with previous reports on intensity modulations in the SC [Bell et al., 2006, Li and Basso, 2008]. Increases of ROL with stimulus repetition are evident in some results from V4 [Hudson et al., 2009, Motter, 2006], although have not been explicitly described in the SC. There is one study in the SC which did not show an ROL increase with repetition [Mayo and Sommer, 2008] and there were interesting stimulus differences between their study and the present one that may explain why: the two stimuli in their sequence were shifted spatially in order to activate different retinal receptive fields but the same relatively large receptive fields in the frontal eye fields or SC$_s$. Thus, their failure to see the ROL increases with repetition, while observing the decrease in response magnitude, may suggest that the ROL effect occurs very early in visual processing (*e.g.* retina, lateral geniculate nucleus of the thalamus or input to V1) while the magnitude decrease occurs more centrally (*e.g.* V1 or beyond), although the anatomical locus of these effects remain to be explicitly tested. There are a few possible explanations for the ROL increase: there could be a complete elimination of the earliest spikes of the response due to adaptation which artificially shifts the ROL, or there could be reduced numbers of cells converging to provide the response, thus delaying when EPSPs can generate the first spikes. The increase in ROL is reminiscent of the increase in ROL as stimulus contrast is reduced [Bell et al., 2006, Li and Basso, 2008], almost as if repetition was reducing the contrast of subsequent stimuli.

2.5.2 Mechanisms of Adaptation

Grill-Spector and colleagues [Grill-Spector et al., 2006] recently proposed 3 models for the mechanisms underlying adaptation. Adaptation may reflect a proportional reduction in firing rate to repetition (*i.e.* Fatigue), a change in the tuning of neural responses for the repeated stimulus (*i.e.* Sharpening), or a reduction in processing time for repeated stimuli (*i.e.* Facilitation). The Facilitation model can be discarded based upon our data, because it predicts that the latency of the response (ROL) would be earlier with repetition, and we uniformly found the opposite. The sharpening model is possible, although it would predict some neurons would have no response with repetition, and some (the best tuned for that stimulus) would show little response decrement. We found that, generally, SC neurons showed a graded reduction in response. Some form of the Fatigue model is therefore most likely to account for the repetition effect observed on the early transient part of the visual response, and it is also the most closely related to our Bayesian model of adaptation. An important addition, however, is that we found a two-stage model with fast and slow dynamics necessary to best explain our neural data, a refinement which indicates that more than one mechanism (but possibly still within a single neuron) with different temporal sensitivities may be contributing to the adaptation effect. Note however, that none of these models can yet account for the increase in ROL with repetition.

Alternatively, some portion of the response reduction may be affected locally in the SC by increasing global inhibition from the basal ganglia. The SC_i projects the transient visual response monosynaptically to the Substantia Nigra compacta [Redgrave and Gurney, 2006], which is then processed through the basal ganglia, and the Substantia Nigra pars reticulata (SNr) projects back to the SC_i [Hikosaka et al., 1983, Jiang et al., 2003] to modulate neuronal firing via GABAergic synapses [Isa et al., 1998, Kaneda et al., 2008]. A visual transient that is not accompanied by a response or reward (as in our simple fixation task) could result in increased SNr inhibition with each repetition (or less disinhibition), and thus reduced subsequent responses. The same mechanism could also account for our dishabituation effect following an intensity 'oddball' stimulus. VS and VmS neurons responded to oddball stimuli that were either brighter or dimmer with an increase in late sustained activity with a latency around 140-160 ms after stimulus onset. A transient reduction of the inhibition from SNr (disinhibition) after an oddball stimulus is recognized by the basal ganglia as novel, could account for the later increase in the sustained component of VS and VmS neuronal activity (*i.e.* reduced SNr inhibition). This 'novelty signal' might then in turn be broadcast to entire visual system from the SC [Boehnke and Munoz, 2008].

2.5.3 Implications for learning theory

In this paper we designed a paradigm to study simple learning phenomena in a behaving primate, which have been studied previously in equisite detail in Aplysia [Carew et al., 1971, Castellucci et al., 1970]. Given the differences in the complexity of the organisms, it is not clear that terminology and concepts are easily transferable, but some discussion is at least warranted. The response decrement with repetition we observed on the initial transient part of the visual response has been called 'habituation' in V4 [Motter, 2006] and 'adaptation' in FEF [Mayo and Sommer, 2008]. Given how that transient response changed with our stimulus intensity oddballs, we believe this decrement in the transient component is best described as adaptation. We have described the increased sustained activity after the brighter or dimmer oddball stimuli as a 'dishabituation-like' or 'novelty' signal. It is also possible that that increase represents a phenomenon called sensitization [Hawkins et al., 2006, Marcus et al., 1988], which amplifies responses like the dishabituation process. Sensitization has been shown to be an independent process from dishabituation because, at least in Aplysia, it develops at a different time [Rankin and Carew, 1988]. Our experiment was not designed to differentiate these two processes, though sensitization usually requires a noxious stimulus, which we did not employ. We also did not objectively determine the discriminability of our 3 stimuli, although the neurons clearly differentiated them. The use of brighter and dimmer stimuli as oddballs had the advantage of simplicity and allowed for the dissociation of habituation from adaptation. However, since the stimuli were identical in shape, size and color, there may have been a counteracting generalization process, which prevented a larger recovery of response (dishabituation/sensitization) than might have been possible with a more distinctly different stimulus. These are questions for future studies. Importantly, this paper represents an initial step in extending to primates the detailed understanding of these simple learning phenomenon achieved in simpler animals like Aplysia, and the oculomotor system is a great candidate system to ask these questions.

2.5.4 Implications for psychophysical studies

Our results are consistent with psychophysical findings on stimulus duration perception [Eagleman, 2008], where repeating stimuli are perceived as shorter in duration compared with an initial stimulus [Pariyadath and Eagleman, 2008, Rose et al., 1995] and any novel stimulus presented [Pariyadath and Eagleman, 2007, Tse et al., 2004]. In our sustained cell types, repetition reduced the size of visual responses and novelty (oddball brighter or dimmer stimuli) caused an increased firing in the later sustained epoch. Thus, the first stimuli and any novel stimuli had a larger sustained response compared with repeated stimuli, and may represent a neural correlate of the aforementioned perceptual findings. The timing of the novelty response also matches that of the N2 component of the human event-related potential to visual oddball stimuli [Folstein and Van Petten, 2008], a component thought to reflect detection of novelty or mismatch. We did not observe any response, early or late, when the fourth stimulus was absent (see Figure 2.5a). A “stimulus omission” mismatch response in audition only occurs when the onset to onset time of the sequence of stimuli is less than 150 ms [Yabe et al., 1997] so perhaps it is not surprising that it was not observed. Late ERP responses such as the p300 are observed to omitted visual stimuli [Tarkka and Stokic, 1998], however, the timing of such a respose would coincide with the time our neurons were responding to the 5^{th} stimulus. The enhancement of the 5^{th} stimulus response after a missing stimuli might in part reflect a P300, though it is difficult to know.

A previous visual event (attentional cue) also has implications for processing of a subsequent visual target for a manual or saccadic response: at separation intervals similar to those used here the response to a subsequent target stimulus is slowed [Fecteau and Munoz, 2006, Klein, 2000]. We show that continued repetition of a visual stimulus (akin to having multiple cues) while fixating further reduces and delays the visual response. This presumably would lead to even slower reaction times and greater IOR. Indeed, recently it was shown that IOR for manual responses increased as the number of repeating cues increased [Dukewich and Boehnke, 2008].

2.5.5 Information processing in the Superior Colliculus

Our results demonstrate that SC neurons' peak transient responses are consistent with a model of adaption which outputs an information quantity related to the amount of learning caused by a new stimulus based on recent stimulation history. This is quite different from the most widely used quantitative definition of information [Shannon et al., 1949], where the information content of a piece of data, or a stimulus, is related to its probability (*i.e.* rare events are very informative). Although useful for the hi-fidelity transmission of data, Shannon information doesn't quantify the subjective impact of stimuli on an observer - an important quantity when processing temporally changing signals.

Adaptation in the SC serves to rapidly decrease the early neural representation of repeating visual events at a particular spatial location (reducing the chance of reflexive orienting to that location), and to increase the representation of temporal outliers. Visual events which are not oriented upon first presentation, and subsequently repeat, are not likely to contain immediately relevant information and there is little to be learned. In this sense, adaptation acts as a simple and fast heuristic to bias selection away from behaviorally irrelevant events in the absence of goal directed orienting signals. Behaviorally relevant events may also manifest as more subtle changes in a stream of stimuli, and orienting to these novel events may require reinstatement of a previously adapted response. The slower dishabituation signal observed later in the response profile may serve as an additional heuristic to support orienting, albeit delayed, to temporally adapted yet novel stimuli. Our data suggests that by combining these heuristics the primate orienting system achieves an efficient trade-off between fast selection of temporal outliers and slower detection of novel events.

Chapter 3

A computational model of visual saliency processing in the primate superior colliculus

3.1 Abstract

To quickly locate and discriminate visual events important for an organisms survival, the visual and ocular motor systems of mammals evolved specialized heuristics to aid in detecting and orienting to visually salient items. The superior colliculus (SC) is a visual-orienting structure that serves an important role in transforming sensory signals encoding locations of stimuli in space, to commands for the control of gaze and attentional shifts. The SC of primates has been shown to be sensitive to a wide range of visual stimuli; yet, no studies have characterized processing of complex natural visual stimuli in a task-free condition. We directly test the hypothesis that the SC represents visually salient items in natural scenes by using a combination of computational modeling and single-unit monkey electrophysiology. We recorded extracellular spike trains from visually responsive neurons in the SC (N=39) of two monkeys (*Macaca mulatta*) while the monkeys freely viewed videos of natural scenes presented on a large, high-definition display. Offline, recordings of the monkey's eye position were used to replay to a computational saliency model of processing in the SC, the exact, gaze-contingent stimulus that impinged onto the monkey's retina. We found that the spike rates of 35 of 39 cells in the SC were significantly predicted by the saliency model (permutation test, $p < .05$), and that during fixations, neural responses could be rank ordered by their saliency responses. To test the necessity of saliency and the importance of each feature, responses were computed for models that only processed individual stimulus features or lacked features. These models performed poorly for individual neurons and across the population of SC cells, suggesting a feature sensitive but non-specific representation. Taken together, the results indicate that during free viewing of natural stimuli SC activity may represent a saliency map of visually conspicuous locations.

3.2 Introduction

Quickly locating and discriminating visual events important for an organisms survival is in principle a task with high computational cost [Tsotsos, 1990]. As a result, the visual system of mammals evolved specialized pre-attentive (bottom-up) mechanisms [Neisser, 1967] that operate in parallel across the visual field to quickly locate regions which may contain behaviorally relevant items. These are purely stimulus driven qualities - salient items 'pop-out' from the scene in that attention is automatically drawn to them [Neisser, 1967, Theeuwes et al., 1994, Treisman and Gelade, 1980, Wolfe and Horowitz, 2004]. Koch and Ullman [1985] proposed a theoretical model of selective visual attention in which basic stimulus attributes [see Wolfe and Horowitz, 2004, for an exhaustive list of stimulus factors guiding attention] are computed in parallel to create separate feature maps. Feature maps are then combined without losing topography into a single two-dimensional master map (saliency map), that represents the most visually conspicuous locations in a scene without reference to any particular features type [see Itti and Koch, 2001a, Shipp, 2004, for a review of saliency map theory].

The standard computational saliency model was implemented by Itti et al. [1998] and extended to video [Itti, 2006, Itti and Baldi, 2009]. The model has provided a quantitative framework to study stimulus driven visual target selection under complex stimulus conditions. The model operates by processing an input image or video with different digital filters at multiple spatial scales. Across scale center-surround difference and non-linear spatial competition promote features which are different from their surroundings. The different features are then linearly combined to produce a saliency map that indicates the locations a human or monkey observer would find bottom-up attention grabbing. Using natural scenes, evidence for this model comes from eye-tracking experiments that have found significant difference in saliency between fixated vs non-fixated locations in humans and monkeys [Berg et al., 2009, Itti and Baldi, 2005, Parkhurst and Niebur, 2003, Peters et al., 2005, Reinagel et al., 1999].

The neurobiological substrates of saliency computations are not specified in the model, and an ideal place to study saliency is the primate superior colliculus (SC). The SC is a phylogenetically old and largely conserved visual orienting structure [Dean et al., 1989, Ingle, 1975, May, 2006] that integrates visual signals in the superficial layers (SC_s), with activity related to saccadic and attentional selection in the intermediate layers [SC_i Goldberg and Wurtz, 1972b, Krauzlis et al., 2004, Lovejoy and Krauzlis, 2009, Mohler and

Wurtz, 1976, Muller et al., 2005]. The SC contains a topographic map of visual space in retinal coordinates, in which the SC$_s$ and SC$_i$ are in correspondence [Marino et al., 2008]. Neurons in the SC$_s$ are visually responsive with receptive fields described as having a center-surround structure [Cynader and Berman, 1972, Schiller and Koerner, 1971]. SC$_s$ neurons receive predominately sensory input from the retina [Hubel et al., 1975], striate and prestriate cortex [Fries, 1984, Lui et al., 1995], and medial temporal [Lui et al., 1995, Maunsell and van Essen, 1983], and respond to a variety of stimuli including visual onsets, offsets, bars, movement [Cynader and Berman, 1972, Prévost et al., 2007, Schiller and Koerner, 1971], and relative motion [Davidson and Bender, 1991], but are not tuned to any particular stimulus property (*e.g.* orientation of a bar). Neurons in the SC$_i$ inherit many of the response properties of SC$_s$ neurons due to the direct columnar projection [Isa and Saito, 2001, Isa et al., 1998, Katsuta and Isa, 2003, Phongphanphanee et al., 2008], but also receive input from higher-level extrastriate visual cortex [Lui et al., 1995] and are sensitive to color features [White et al., 2009]. Evidence from the Isa lab indicates a rich local connectivity in the SC$_s$ and SC$_i$ which could support intrinsic spatial competition [Isa and Hall, 2009].

Using the computational saliency model framework [Itti et al., 1998] we created a new model of sensory inputs and spatial processing in the SC, and extend the previous behavioral work by directly testing a biologically plausible model of saliency processing against SC neural responses collected under natural stimulus conditions. We found significant agreement between the saliency model and neural responses overall, and found that during fixation neural responses could be ranked by their saliency values. Finally, models which were impoverished by removing one or more channels did not perform as well as the full model. Taken together, the data supplies evidence for the saliency map theory of visual processing in the SC.

3.3 Methods

3.3.1 Subjects

Eye movements and neural responses during free viewing were recorded from two monkey (*Macaca Mulatta*, both male) subjects. Monkeys were used with approval by the Queens University Animal Care Committee and were in accordance with the Canadian Council on Animal Care policy on the use of laboratory animals, and the Policies on the Use of Animals and Humans in Neuroscience Research of the Society for Neuroscience. A stainless steel head post was attached to the skull via an acrylic implant anchored to the skull by stainless steel screws. Eye coils were implanted between the conjunctiva and the sclera of each eye [Judge et al., 1980] allowing for precision recording of eye position using the magnetic search coil technique [Robinson, 1963]. The surgical techniques to prepare the animal for behavioral and physiological recordings have been described elsewhere [Marino et al., 2008].

3.3.2 Stimulus presentation

Monkeys were seated in a primate chair with their heads restrained for the duration of an experiment (2-5 hours). Subjects were positioned 70 cm from a Sony Bravia LCD TV (120 cm x 60 cm), giving a usable field of view of 80.3°horizontal and 51.9°vertical.

To calibrate eye position, monkeys performed a step saccade paradigm in which targets at nine eccentricities and eight radial orientations from the fixation point were presented in random order. Monkeys were given a liquid reward if they fixated a target within a square electronic window of 3°radius within 800 ms.

When a neuron was first isolated, its visual receptive field was mapped using a task in which the monkey fixated while white stimuli (42.5 cd/m^2, 100 ms duration, 0.25° diameter circle) were presented in a darkened environment. Stimuli were presented in pseudorandom order (no subsequent stimuli were presented at the same location) at 182 locations distributed across 60° (horizontal) x 50° (vertical) of visual angle. The monkey subjects were given liquid reward after each trial. Real-time control and display of the calibration task and mapping task was achieved using REX version 6.0.

Free-viewing stimulus presentation proceeded similarly to Berg et al. [2009]. Stimuli were taken from several high-quality, hi-definition commercial and in-house created sources. Commercial sources included clips from the BBC Planet Earth collection, BBC Wild Pacific, BBC Wild India, and several in-house collections filmed in locations in Los Angeles, CA and Kingston, ON on a high-definition camcorder (Canon Vixia HF S20). In total, 516 clips were used in the stimulus set. Files were converted from their raw M2TS format into 40 Mbits/s MPEG-4 (deinterlacing when required) and displayed full screen at 1920x1080 pixels. Due to the poor per frame timing of commercial LCD TV's a 2 in^2 area of the lower left corner of the screen was sacrificed to mount a photodiode. On each frame an alternating black and white square 1 in^2 in size was added to the lower left corner of the video, and the flashing area was hidden with non-reflective tape. The photodiode was recorded concurrently with the neural response and the eye position (1000

Hz), and was used offline to recover the exact onset and duration of each frame. Stimuli were presented using a Linux computer running in-house programmed presentation software (downloadable at http://iLab.usc.edu/toolkit) under SCHED_FIFO scheduling to ensure proper frame rate presentation [Finney, 2001]. Initiation of a video presentation was controlled by the REX system in communication with the presentation computer (RJ-45, UDP communication). Monkeys were presented stimuli in a darkened room with the reflective frame of the TV covered with black cloth. Each video was randomly selected from the set and preceded by a fixation point. The next video was initiated when the monkeys eye position remained within a square electronic window with 5°radius of the central fixation point for 150 ms. After all videos were randomly selected, the set was allowed to repeat. The monkey subjects were given liquid rewarded after each movie.

3.3.3 Data acquisition

Eye position data was digitized at 1000 Hz using data acquisition hardware by Plexon, Inc. Concurrently, timestamps of the time of fixation point onset, acquisition of the fixation target by the monkey, and initiation and ending of the clip were recorded. Physiological activity was monitored from single neurons or pairs of single neurons using tungsten electrodes (Frederick Haer, 0.5-5.0 mΩ) with stimulus events and spike times collected, and waveforms digitized, through the Plexon MAP system. Isolation was performed online using the window discriminator in Plexon, and verified and optimized offline using Plexon's Offline Sorter. In total, 45 neurons were recorded for analysis. Individual video clips for each neuron were discarded if a minimum spike rate of 40 sp/s was not reached, or if there was time-stamp mismatch between stimulus presentation and control computers, or if there was an error with the photodiode signal. A cell was discarded from analysis if a minimum of 25 clips were not recorded or did not remain after preprocessing. 28 cells from monkey Q and 11 cells from monkey Y were used in the final analysis.

3.3.4 Saccade detection

Saccades were detected by first filtering the eye position with a 3-tap Butterworth filter at 50 Hz cut-toff. Velocity was computed from a finite difference and smoothed by a 3-tap Butterworth filter with 40 Hz cut-toff. Time points with greater than 15 deg/s velocity were marked as potential saccades. The start and end of the saccade was then refined by applying a hysteresis filter that started from a high-point on the velocity curve and traversed forward and backward until velocity stopped decreasing. Finally, saccades less than $\frac{1}{2}^{\circ}$in amplitude were rejected.

3.3.5 Eye position calibration

In order to control for nonlinearities when using the magnetic search coil a calibration procedure was performed using the data from the step saccade task. For each rewarded trial, the saccade to the target and the surrounding fixations were identified. A principle components analysis was computed on the fixation intervals and trials were rejected if either Eigenvalue was greater than 50, which indicated a highly variable fixation. Trials were also rejected if the fixation before the saccade was not at least 150 ms, and the fixation after the saccade at least 200 ms. The median of the fixation points were taken as the fixation locations. Multiple visits to a calibration location were collapsed by finding the mode of the distribution of points using the mean-shift algorithm [Fukunaga and

Hostetler, 1975] with a bandwidth of 10 pixels. If multiple clusters emerged, the cluster with maximum count was used. If two clusters had the same count, the bandwidth was increased and the procedure repeated until a cluster emerged or the maximum bandwidth of 15 pixels was reached, at which point the location was rejected. An affine transformation with outlier rejection recovered the linear component of the transformation and a thin-plate-spline the nonlinear component [Berg et al., 2009, as in]. In each recording session, the initial fixations (if greater than 150 ms) to the fixation point before the presentation of the video were extracted. The same mean-shift procedure used to determine calibration locations was used to find the offset for a linear drift correction.

3.3.6 Implementation of computational model

The general architecture of the SC saliency model follows that of Koch and Ullman [1985] and Itti et al. [1998], and the model was created and run under Linux using the iLab C++ Neuromorphic Vision Toolkit [Itti, 2004] on a cluster utilizing over 300 CPU's. The model is feed-forward in nature consisting of 6 main components: retinal input, raw feature map computation, space-variant mapping, receptive-field computation.

3.3.6.1 Retinal input

Each high-definition video frame (80.3°x 51.9°of the monkeys field of view) was first shifted to retinal coordinates (i.e so that the monkey's point of gaze was always at the center of the input to the model) replacing any empty values with black to match the monkeys environment. Next the frame was embedded in a larger black background image to simulate 100°x 100°of the the monkey's visual world (the size of our SC saliency map).

The eye-movement data was sub sampled and the video was processed at 200 Hz to accurately capture the visual dynamics caused by the monkey's eye movements (each eye-tracker sample gives rise to a new retinal image). Since the exact frame onset and length were collected from a photodiode, each frame could be accurately linked to the eye position samples that occurred during the frames presentation. The retinal image was rescaled by decimating and smoothing with a 3-tap binomial filter. The RGB image was then converted to the Derrington-Krauskopf-Lennie (DKL) colorspace which corresponds to the type of segregation that exists along the geniculostriate pathway in early vision [Derrington et al., 1984]. The computation was performed by gamma correcting the RGB values based on measurements from a photo-spectrometer (PR-650, Photoresearch, CA, USA). The RGB values could then be converted to the CIE XYZ colorspace and then to the Stockman and Sharpe cone fundamentals [Stockman et al., 2000]. The DKL space combines long, medium and short wavelength cone responses to produce a colorspace with axes that approximately represent luminance, red-green opponency, and blue-yellow opponency.

3.3.6.2 Raw feature computation

Schiller and Koerner [1971] and Cynader and Berman [1972] demonstrated that neurons in the SC_s and SC_i respond to a variety of stimuli including visual onsets, offsets, bars and movement, and are insensitive to the shape, orientation, wavelength and direction of motion of the stimulus. Many SC neurons were also described as 'event' detectors which responded transiently to both onset and offset of stimuli, and 'jerk' detectors which preferred short rapid movements [Schiller and Koerner, 1971]. More recently, SC neurons have been shown to respond well to relative motion [Davidson and Bender, 1991, when the motion at the cell's receptive field center differs from that of the surround] and color [White et al., 2009]. Based on these findings the model consisted of 6 high-level feature maps: luminance, red-green opponency, blue-yellow opponency, flicker (abrupt onsets and offsets), edges and bars, and motion.

High-level feature maps were created by linearly combining the responses to different filters (feature maps) of the same type (*e.g.* a 45°and 90°edge detector). In total, there were 60 feature maps. The first three high-level features (luminance and two chromatic) contained only a single feature maps taken from the DKL conversion. The other 57 maps were computed from the luminance feature map.

The luminance output was buffered for 9 frames and processed by three-dimensional 2^{nd} derivative of Gaussian separable steerable filters [Derpanis and Gryn, 2005]. This formulation provided a single efficient computational framework to compute static edges, flicker (abrupt onsets), and spatiotemporal motion energy. The steerable set allows the creation of filter responses at any space-time orientation from a small basis set. The filters were computed in quadrature pair and the magnitude was taken as the filter response. The result is a filter sensitive to both step edges and bars, as in models of V1 complex cells [Adelson and Bergen, 1985, Pollen and Ronner, 1983]. The choice of order of the Gaussian derivative was largely due to reduce computation, as higher order derivatives require a larger basis set. Static edges were computed at 4 orientations and two spatial scales (retinal and $\frac{1}{2}$ retinal) for a total of 8 maps. Flicker was computed at only the retinal scale yielding a single map. Local motion signals were computed at 3 speeds and 4 orientations at both the retinal and $\frac{1}{2}$ retinal scales. Local motion signals were converted to opponent motion by subtracting maps at the same speed but opposite motion directions, producing 48 maps.

3.3.6.3 Space-variant transformation

The primate retinostriate [Schwartz, 1977, 1980] and retinotectal projections [McIlwain, 1975] create a nonhomogenius mapping of visual space such that more neural surface is dedicated to processing signals at the fovea, the center of gaze, and less in the periphery. In the colliculus, Ottes et al. [1986] used data from Robinson [1972] to recover a logarithmic mapping of the retinal surface to the SC surface. In the model's isotropic case, it is equivalent to the complex logarithmic mapping of Schwartz [1980]. One disadvantage of these mappings is that they are discontinuous at the vertical meridian, which makes standard image processing techniques unusable. Instead of a conformational mapping to the SC surface, we performed a simpler log-polar resampling of the input image which captures key features of the retinostriate and retinotectal mapping [Wiebe and Basu, 1997]. Further processing complications were illuminated because the neighborhood relations between hemifields are maintained [Harting, 2004, possibly neurally implemented through cross collicular connections] and the result of the transform is a square image on which standard image processing can be applied.

Model SC units were simulated on a two dimensional (200 x 200) grid representing approximately 100°x 100°of visual angle. To compute the point in visual space (or image space) corresponding to each SC unit's receptive field center, we used the inverse of the basic variable resolution transform [Wiebe and Basu, 1997]. The scaling parameters of the transform were set by simulating a square grid of SC surface (4.5 mm rostral-caudal and

3.5 mm medial-temporal for each hemifield) and using the inverse conformational mapping Ottes et al. [1986] to find the corresponding points in visual space. The SC model map points were also projected to visual coordinates using the inverse basic variable resolution transform, and a least squares fit [minimized with a simplex-algorithm Lagarias et al., 1998, using 1000 random restarts] found the best scaling parameters.

3.3.6.4 Receptive-field computation

SC Neurons are reported to have a center-surround receptive field structure such that a disk stimulus larger than a preferred size begins to inhibit responses. Preferred stimulus size ranges from .75°near the fovea to approximately 5°at 40°eccentricity [Cynader and Berman, 1972, Schiller and Koerner, 1971]; however, SC neurons have large activation fields of 10°-40°peripherally [Cynader and Berman, 1972] that are invariant to the exact position of the preferred stimulus [Cynader and Berman, 1972, Schiller and Koerner, 1971]. Center-surround structure is also observed in the SC$_s$ sensitivity to relative motion [Davidson and Bender, 1991], and in afferents to the SC: retinal [Croner and Kaplan, 1995], V1 [Cavanaugh et al., 2002], and MT [Allman et al., 1985]. The size specificity with large invariant activation fields suggests a simple model where the SC pools responses of Difference-of-Gaussian (DoG) detectors at nearby spatial locations [McIlwain, 1975].

Receptive fields are modeled by first computing a Gaussian scale space [Crowley and Riff, 2003]. A Gaussian scale space is an image representation where each level in the space represents increasing lowpass filtering of the input. A DoG detector can be built by subtracting a sample at a lower level from one at the same spatial location in a higher level. The spatial locations of the samples are given by the space-variant transform. The level (receptive field size) was computed by estimations from the Cynader and Berman [1972], Schiller and Koerner [1971] reports. To create a DoG detector sensitive for an optimal stimulus size S in degrees, and DoG ratio K, the excitatory Gaussian size in degrees is $\sigma_E = \frac{S\sqrt[2]{2}\sqrt[2]{K-1}\sqrt[2]{K+1}}{4K\sqrt[2]{\log K}}$ and the inhibitory is $\sigma_I = K\sigma_E$. DoG receptive fields were constructed linearly with eccentricity with an optimal stimulus size of .75°in the fovea, and 5°at 40°eccentricity.

The space-variant remapping and center-surround receptive fields were computed for each feature map. For the luminance and chromatic features, we used $K = 6.7$ based on experiments in retinal ganglion cells [Croner and Kaplan, 1995]. For these features, the absolute value of the DoG response was taken giving rise to double opponency responses (*e.g.* responds to red surrounded by green and green surrounded by red). For the spatiotemporal feature maps, $K = 3.2$ was used based on experiments in V1 [Cavanaugh et al., 2002] and the DoG responses were half-wave rectified.

To model any long-range competition in cortex [Grinvald et al., 1994] and SC [Isa and Hall, 2009, Munoz and Istvan, 1998, Olivier et al., 1999] a single iteration of the operator developed by Itti and Koch [2001b] was applied to each feature map. Briefly, the map was first filtered with a large DoG (3°excitatory, 9°inhibitory). The result was added back to the feature map and a constant subtracted to represent global inhibition, followed by a final half-wave rectification.

Each point in each 200x200 map was then replaced by the sum in a 3 pixel circular neighborhood to simulate the large (but size selective) activation fields observed in the SC. This resulted in model activation fields which were approximately 2°at 10°eccentricity and 10°at 40°eccentricity [Cynader and Berman, 1972, similar to those in the SC$_s$]. Due to pooling and symmetrically filtering in the SC model space, the activation fields also have an asymmetry such that they slightly narrow toward the fovea, as discussed elsewhere [Marino et al., 2008, McIlwain, 1975].

3.3.7 Data fitting and saliency response

For every neuron and video in the analysis, 60 feature maps were computed for every retinal frame and values were collected over the duration of each video at the location corresponding to the cell's receptive field center (obtained from the receptive field mapping paradigm). Spike trains during each clip were binned (5 ms intervals) and Gaussian filtered (σ=50 ms) to create spike density functions. The data fitting took place for each cell separately using a leave-one-out training such that each clip was excluded from the set once for testing, and the model was trained on all the rest.

The collected feature values in the test clip were normalized by the max in the training set (across all clips for each feature separately) and linearly combined into the 6 high-level features. Before combining the high-level features into the saliency response, each feature was optimally aligned to the spike density function. The optimal delay for each feature was computed by calculating the mutual information Antos and Kontoyiannis [2001, computed using a plug-in method] at different time delays (up to 150 ms) between test-set features and test-set neural responses, considering all clips. After optimal alignment, the square root of each high-level feature was taken and the responses were linearly combined to create the saliency response.

3.4 Results

Figure 3.1A,B show the main experimental paradigm (see Methods for a detailed description) and the computational model. Briefly, eye movements and single-unit neural responses in the superior colliculus were recorded from 2 male monkeys (*Macaca mulatta*) while the monkeys freely watched high-definition video clips of natural scenes (516 videos, 3-30 seconds, ∼80°x∼50°field of view). Each clip was initiated by the monkey fixating a central point, and the monkeys quickly learned the task without training.

Schiller and Koerner [1971] and Cynader and Berman [1972] demonstrated that neurons in the SC respond to a variety of stimuli including visual onsets, offsets, bars and movement, and are insensitive to the shape, orientation, wavelength and direction of motion of the stimulus. More recently, sensitivity to relative motion [Davidson and Bender, 1991] and color [White et al., 2009] has been demonstrated. Based on these findings we constructed a computational model of saliency processing in the SC that consisted of 6 high-level feature maps: luminance, 2 chromatic contrasts (derived from the Derrington-Krauskopf-Lennie (DKL) colorspace derived from retinal ganglion cells [Derrington et al., 1984]), flicker (abrupt onsets and offsets), static edges and bars, and motion (see methods for a detailed description and discussion of the SC saliency model).

The eye movements recordings were used to replay to the computational saliency model the exact gaze-contingent stimulus the monkey viewed. For every video clip in the analysis, saliency feature maps were computed in retinal coordinates and feature values were collected at the location in the maps that corresponded to the cell's receptive field center (obtained through a receptive field mapping paradigm, see Methods). A training procedure found the best alignment between high-level saliency features and the neural response. The optimally aligned high-level features were then linearly combined into the saliency model response and compared to an estimate of the neuron's firing rate (see Methods). Figure 3.1C shows an example of the saliency model output and a SC neurons activity during a video clip.

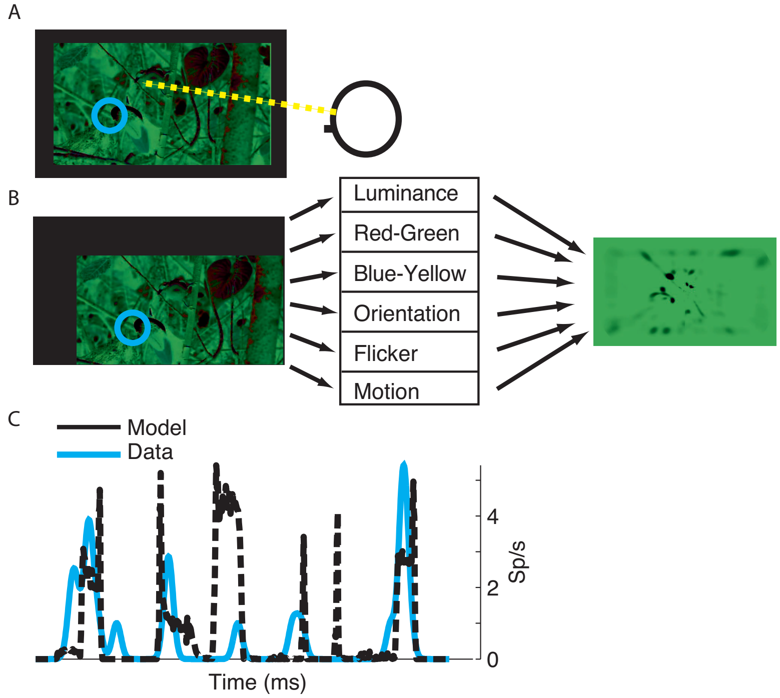


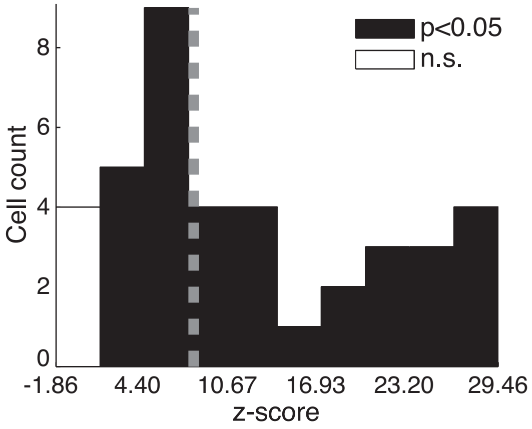
Figure 3.1: Experimental paradigm. (A) Monkey subjects watched high-definition natural scene videos while eye position and neural activity of single SC neurons was recorded. (B) The computational saliency model of SC processing was presented a gaze contingent version of the stimulus (shifted to the monkey's current eye position). The model computed feature maps sensitive to luminance, red-green opponency, blue-yellow opponency, flicker, edges and motion, that were combined to produced a saliency map simulation of activity in the SC. (C) Spike trains were converted to rate functions and the time series of neural activity was compared to saliency model output at the location in the saliency map that corresponded to each cell's receptive field center.

3.4.1 Comparing neural and model responses

Mutual information (MI) [computed using the plug-in method Antos and Kontoyiannis, 2001] was used to assess the overall degree of correspondence between the saliency model's output and neural activity. Mutual information measures the statistical dependence between two variables and is sensitive to nonlinear relationships. For each cell separately, MI was computed using the time-series (including fixations and saccades) of model and neural responses from all video clips together, resulting in a single MI score for each cell. Statistical significance of each cell's MI score was obtained by randomly shuffling model and neural responses so that the neural response from one clip could be assigned to the model response from another. Computing MI for the random shuffling and repeating the process forms a sampling distribution of shuffled responses (1000 samples) and the p value is determined by the ratio of samples greater than or equal to the cell's MI score.

35 of the 39 cells used in the analysis had MI scores significantly above chance (permutation test, $p < 0.05$). Raw MI scores were converted to z-scores to examine the population. Figure 3.2 shows the distribution of z-scores for the population of cells. The population median (7.89) was significantly greater than zero (sign-test, $p < .01$). This analysis reveals that the majority of cells recorded had a significant statistical relationship with the saliency model, and that, across the population, the saliency model was a significant predictor of neural activity. Subsequent analysis was performed only on the 35 cells with significant MI scores.

Figure 3.2: The distribution of mutual information z-scores (see Results) for the population of recorded cells. Black bars indicate cells that had significantly high mutual information with the model (permutation-test, $p < .05$). The gray dashed line shows the median z-score (7.89).



3.4.2 Saliency during fixations

When a saccade was made and the eye landed at its new target, a salient item may have been brought into the receptive field of the cell, or a salient item may have abruptly appeared during the fixation. We tested the degree to which saliency values during fixations (200 ms or greater in length) ranked with neural responses. Spike rates for each cell were normalized by their max response across all video clips. Saccades were detected (see Methods) for each clip, and fixations were determined as periods between saccades. Median saliency and spike rate values for each fixation interval were computed and ranked (all clips and cells together) by saliency level: high saliency (N=7580), medium saliency (N=7580) and low saliency (N=7582).

Figure 3.3A shows the mean (time locked to the onset of the fixation) neural response for the three saliency levels. Figure 3.3B shows the median neural activity during the fixation intervals, organized in the same way. A repeated measures 2-factor (cell x saliency level) permutation test [Edgington and Onghena, 2007] was performed on the data in Figure 3.3B and confirmed a significant difference between neural responses when organized by saliency level (permutation test, $p < .01$). The results reveled, on average, neural activity during fixations could be ranked by saliency activity during fixations.

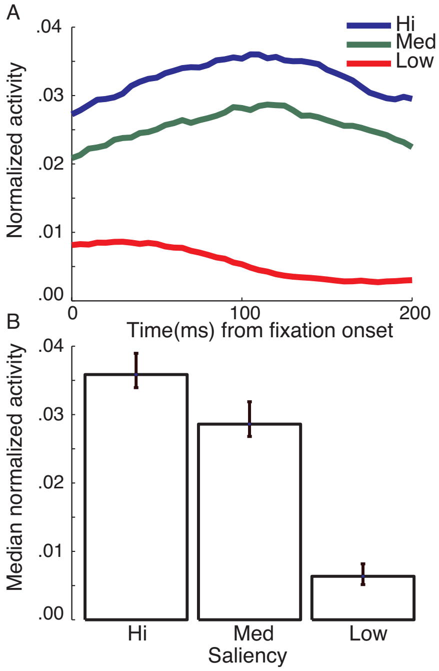


Figure 3.3: Neural responses during fixation of salient items. (A) Neural responses during fixations were ranked by the median of saliency responses during the fixation, and split into three equally sized groups (high saliency, medium saliency, low saliency). Plotted is the mean of the normalized and fixation aligned neural response for the three saliency groups. (B) The bar graph shows the median of the normalized neural response during each fixation interval, sorted by the three saliency groups. Error bars show the 95% confidence intervals of the median estimated by boot strapping (1000 iterations).

3.4.3 Single-feature and leave-one-out analysis

The single-feature analysis was designed to determine the necessity of the saliency model and if for some cells, and across the population, a simpler model could perform better than the saliency model. Models were constructed such that they only contained a single high-level feature, and mutual information between the neural response and each single-feature model was computed (as was done for the full model). The single-feature and full model MI scores were compared by computing a simple information contrast score: $IC = \frac{(I_{single} - I_{full})}{(I_{single} + I_{full})}$, where I_{full} is the MI score of the full model and I_{single} is the MI score of the single-feature model. Values below 0 indicated a cell with a higher degree of correspondence to the full model. Figure 3.4A shows the information contrast for each single-feature model (rows) and each cell (columns) and the median information contrast across the population (right table). All medians were negative and a sign test confirmed that for each single-feature model, the full model performed significantly better across the population of cells (sign test, $p < .01$).

Models were also constructed that contained all features except a particular high-level feature, and were quantified in the same way as the single-feature models. Due to the correlations between single high-level features, this analysis provides a better estimate of each features importance than does the single-feature analysis, and gives a measure of the non-redundant contribution of each feature. The more negative the values, the more the full model's score suffered from the loss of the feature. Figure 3.4B shows the information contrast for each leave-one-out model (rows) and for each cell (columns) and the median information contrast across the population (right table). All medians were negative and a sign test confirmed that for each leave-one-out model, except for the model lacking edges, the full model performed significantly better (sign test, $p < .01$).

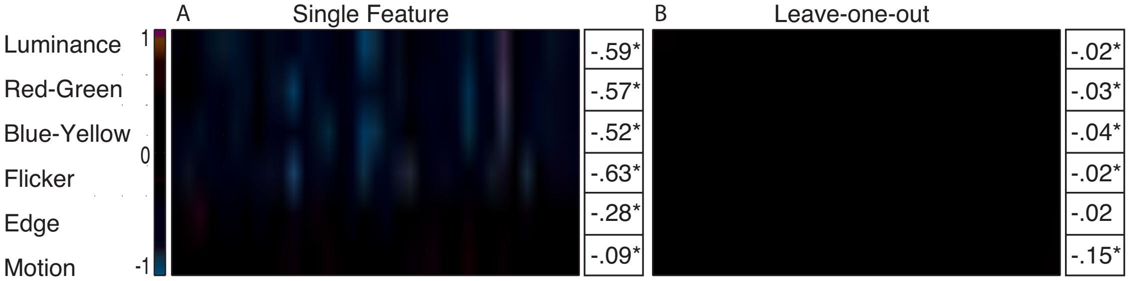


Figure 3.4: Single-channel and leave-one-out analysis. (A) Each row shows the information contrast (see Results) between models with only single features compared to the full saliency model. Each column shows a different cell, and cells are sorted such that the cell with the highest mutual information with the model (see figure 3.2A) are in the furthest right columns. Negative values (more blue hues) indicate the full saliency model performed better than the single-channel model. The adjacent table shows the median contrast value across all cells. Population values that significantly differ from 0 are marked with an asterisk (sign-test, p¡.01). (B) The analysis is the same as (A) but information contrast is computed with leave-one-out models where a single feature was removed from the full model. Each row indicates the channel that was removed.

3.5 Discussion

The present study objectively compared, for the first time, a computational model of saliency to spiking activity of neurons in the primate superior colliculus (SC) to test the hypothesis that the SC represents a saliency map during free viewing of natural dynamic stimuli. In summary, a new computational model of saliency processing in the SC was built. Eye movements and single-unit responses in the SC were recorded from monkeys while they freely watched videos of natural scenes presented on a large high-definition display. The computational model was replayed the gaze contingent stimulus and saliency values were collected at the location in the saliency map that corresponded to the cell's receptive field. Approximately 90% of the cells recorded had mutual information with the saliency model that was significantly above chance (on average, approximately 8 standard deviations above chance). Furthermore, saliency during fixations was predictive of the magnitude of the neural activity during fixations. Models which only contained single features or had features removed, demonstrated that, across the population, no single feature was sufficient to explain the result, and that most saliency features were needed. Taken together, this data suggests that during free viewing of natural stimuli, the SC contains activity consistent with saliency map theory.

3.5.1 Computational SC saliency model

The general architecture of the SC saliency model was inspired by classic saliency map theory Itti et al. [1998], Koch and Ullman [1985], with modifications to account for properties of the SC (see Methods for a detailed discussion of the model). The primate retinostriate [Schwartz, 1977, 1980] and retinotectal projections [McIlwain, 1975] create a space-variant mapping [described by **?**] of visual space such that a larger portion of the topographic map in the SC is dedicated to the center of gaze (fovea), and a smaller portion to the periphery. SC neurons are reported to have a preferred stimulus size [Cynader and Berman, 1972, Schiller and Koerner, 1971]; however, have large activation fields [Cynader and Berman, 1972] that are invariant to the exact position of the preferred stimulus [Cynader and Berman, 1972, Schiller and Koerner, 1971]. Both the preferred size and activation field size increase with increasing eccentricity. SC receptive fields are also described as having a slight asymmetry in visual space [Marino et al., 2008], predominantly due to the afferent mapping [McIlwain, 1975].

The size specificity, large invariant activation fields, and receptive field asymmetry, suggests a simple model in which the SC symmetrically pools responses (in space-variant coordinates) of Difference-of-Gaussian (DoG) detectors at nearby spatial locations [McIlwain, 1975]. This model has implications for how center-surround computations in the SC model should be constructed, and for how visual stimuli interact on the SC map. A new type of saliency map based on these findings was constructed that approximated the space-variant SC transform and eccentricity dependent feature detection, and as a consequence, required a gaze-contingent input (see Methods).

3.5.2 Feature analysis

The single-feature analysis revealed that across the population, and for the majority of cells, neural responses were best predicted by the full saliency model; however, several neurons had positive information contrast values indicating that they were better predicted by single-feature models. One cell preferred only the luminance feature, another only the static edge features, and 7 cells preferred only motion.

The leave-one-out analysis confirmed that, across the population of neurons, the full model's predictive power was significantly diminished by the removal of any high-level features. However, this was not the case for static edges (median information contrast=-.02, sign test, $p = .18$), even though the edge feature was one of the more predictive features in the single-feature analysis. This is likely because the high-level motion feature is made of edges moving at different orientations and velocities, and there is correlation between these features and static edges. Even though static edges may be a predictive feature, when combined with moving edges in moving video sequences there is little non-redundant contribution.

The construction of the high-level motion feature (see Methods) makes it highly sensitive to relative motion - when the motion in the center of the receptive field is different from that of the receptive field surround. The single-feature analysis revealed that motion was the most predictive single-feature, and the leave-one-out analysis confirmed that motion was the largest contributor to the saliency model's predictive power. Furthermore, only 3 cells preferring a model without motion; conversely, many cells preferred models without static features. Relative motion signals have been reported in the SC [Bender and Davidson, 1986, Davidson and Bender, 1991] and inactivation studies indicate that cortex plays a strong role in shaping this response [Davidson et al., 1992]. Given the amount of patterned motion that occurs in natural scenes, it is not surprising that this feature contributed significantly to the saliency response.

Bibliography

E.H. Adelson and J.R. Bergen. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A*, 2(2):284–299, 1985.

J. Allman, F. Miezin, E. McGuinness, et al. Direction-and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (mt). *Perception*, 14(2):105–126, 1985.

A. Antos and I. Kontoyiannis. Estimating the entropy of discrete distributions. In *IEEE International Symposium on Imformation Theory*, pages 45–45, 2001.

A.T. Bahill, M.R. Clark, and L. Stark. The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, 24(3-4):191–204, 1975.

AT Bahill, A. Brockenbrough, and BT Troost. Variability and development of a normative data base for saccadic eye movements. *Investigative ophthalmology & visual science*, 21(1):116–125, 1981.

P. Baldi and L. Itti. Of bits and wows: A bayesian theory of surprise with applications to attention. *Neural Networks*, 23(5):649–666, 2010.

D. Bamber. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of mathematical psychology*, 12(4):387–415, 1975.

W. Becker and AF Fuchs. Further properties of the human saccadic system: eye movements and correction saccades with and without visual fixation points. *Vision Research*, 9(10):1247–1258, 1969.

A.H. Bell, J.H. Fecteau, and D.P. Munoz. Using auditory and visual stimuli to investigate the behavioral and neuronal consequences of reflexive covert orienting. *Journal of neurophysiology*, 91(5):2172–2184, 2004.

AH Bell, MA Meredith, AJ Van Opstal, and DP Munoz. Stimulus intensity modifies saccadic reaction time and visual response latency in the superior colliculus. *Experimental Brain Research*, 174(1):53–59, 2006.

DB Bender and RM Davidson. Global visual processing in the monkey superior colliculus. *Brain research*, 381(2):372–375, 1986.

D.J. Berg, S.E. Boehnke, R.A. Marino, D.P. Munoz, and L. Itti. Free viewing of dynamic stimuli by humans and monkeys. *Journal of Vision*, 9(5), 2009.

B. Bernard. A cognitive theory of consciousness, 1988.

S.E. Boehnke and D.P. Munoz. On the importance of the transient visual response in the superior colliculus. *Current opinion in neurobiology*, 18(6):544–551, 2008. ISSN 0959-4388.

D. Boghen, BT Troost, RB Daroff, LF Dell'Osso, and JE Birkett. Velocity characteristics of normal human saccades. *Investigative Ophthalmology & Visual Science*, 13(8):619–623, 1974.

S.P. Brown, R.H. Masland, et al. Spatial scale and cellular substrate of contrast adaptation by retinal ganglion cells. *Nature neuroscience*, 4(1):44–51, 2001.

T.J. Carew, V.F. Castellucci, and E.R. Kandel. An analysis of dishabituation and sensitization of the gill-withdrawal reflex in aplysia. *International Journal of Neuroscience*, 2(2):79–98, 1971.

V. Castellucci, H. Pinsker, I. Kupfermann, and E.R. Kandel. Neuronal mechanisms of habituation and dishabituation of the gill-withdrawal reflex in aplysia. *Science*, 167 (3926):1745–1748, 1970.

J.R. Cavanaugh, W. Bair, and J.A. Movshon. Nature and interaction of signals from the receptive field center and surround in macaque v1 neurons. *Journal of Neurophysiology*, 88(5):2530–2546, 2002.

C.W.G. Clifford, M.A. Webster, G.B. Stanley, A.A. Stocker, A. Kohn, T.O. Sharpee, O. Schwartz, et al. Visual adaptation: Neural, psychological and computational aspects. *Vision research*, 47(25):3125–3131, 2007.

B.D. Corneil, D.P. Munoz, B.B. Chapman, T. Admans, and S.L. Cushing. Neuromuscular consequences of reflexive covert orienting. *Nature neuroscience*, 11(1):13–15, 2008.

L.J. Croner and E. Kaplan. Receptive fields of p and m ganglion cells across the primate retina. *Vision research*, 35(1):7–24, 1995.

J. Crowley and O. Riff. Fast computation of scale normalised gaussian receptive fields. In *Scale space methods in computer vision*, pages 584–598. Springer, 2003.

M. Cynader and N. Berman. Receptive-field organization of monkey superior colliculus. *Journal of Neurophysiology*, 35(2):187, 1972. ISSN 0022-3077.

S.V. David, W.E. Vinje, and J.L. Gallant. Natural stimulus statistics alter the receptive field structure of v1 neurons. *The Journal of Neuroscience*, 24(31):6991–7006, 2004.

RM Davidson and DB Bender. Selectivity for relative motion in the monkey superior colliculus. *Journal of neurophysiology*, 65(5):1115–1133, 1991.

RM Davidson, TJ Joly, and DB Bender. Effect of corticotectal tract lesions on relative motion selectivity in the monkey superior colliculus. *Experimental brain research*, 92 (2):246–258, 1992.

P. De Graef, A. De Troy, and G. d'Ydewalle. Local and global contextual constraints on the identification of objects in scenes. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 46(3):489, 1992.

I. Dean, N.S. Harper, and D. McAlpine. Neural population coding of sound level adapts to stimulus statistics. *Nature neuroscience*, 8(12):1684–1689, 2005.

P. Dean, P. Redgrave, and GWM Westby. Event or emergency? Two response systems in the mammalian superior colliculus. *Trends in neurosciences*, 12(4):137–147, 1989. ISSN 0166-2236.

K.G. Derpanis and J.M. Gryn. Three-dimensional nth derivative of gaussian separable steerable filters. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 3, pages III–553. IEEE, 2005.

A.M. Derrington, J. Krauskopf, and P. Lennie. Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of Physiology*, 357(1):241–265, 1984.

M.C. Dorris, R.M. Klein, S. Everling, and D.P. Munoz. Contribution of the primate superior colliculus to inhibition of return. *Journal of Cognitive Neuroscience*, 14(8): 1256–1263, 2002.

V. Dragoi. A feedforward model of suppressive and facilitatory habituation effects. *Biological cybernetics*, 86(6):419–426, 2002.

V. Dragoi and M. Sur. Image structure at the center of gaze during free viewing. *Journal of cognitive neuroscience*, 18(5):737–748, 2006.

V. Dragoi, J. Sharma, E.K. Miller, M. Sur, et al. Dynamics of neuronal sensitivity in visual cortex and local feature discrimination. *Nature neuroscience*, 5(9):883–891, 2002.

K.R. Dukewich and S.E. Boehnke. Cue repetition increases inhibition of return. *Neuroscience letters*, 448(3):231–235, 2008.

D.M. Eagleman. Human time perception and its illusions. *Current opinion in neurobiology*, 18(2):131–136, 2008.

E. Edgington and P. Onghena. *Randomization tests*, volume 191. Chapman & Hall/CRC, 2007.

W. Einhäuser, W. Kruse, K.P. Hoffmann, and P. König. Differences of monkey and human overt attention under natural conditions. *Vision research*, 46(8-9):1194–1209, 2006.

J.H. Fecteau and D.P. Munoz. Salience, relevance, and firing: a priority map for target selection. *Trends in cognitive sciences*, 10(8):382–390, 2006. ISSN 1364-6613.

J.H. Fecteau, A.H. Bell, and D.P. Munoz. Neural correlates of the automatic and goal-driven biases in orienting spatial attention. *Journal of Neurophysiology*, 92(3):1728–1737, 2004.

G. Felsen and Y. Dan. A natural approach to studying vision. *Nature neuroscience*, 8 (12):1643–1646, 2005.

S.A. Finney. Real-time data collection in linux: A case study. *Behavior Research Methods*, 33(2):167–173, 2001.

J.R. Folstein and C. Van Petten. Influence of cognitive control and mismatch on the n2 component of the erp: a review. *Psychophysiology*, 45(1):152–170, 2008.

W. Fries. Cortical projections to the superior colliculus in the macaque monkey: a retrograde study using horseradish peroxidase. *The Journal of Comparative Neurology*, 230(1):55–76, 1984. ISSN 1096-9861.

K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *Information Theory, IEEE Transactions on*, 21(1): 32–40, 1975.

J.L. Gallant, C.E. Connor, and D.C. Van Essen. Neural activity in areas v1, v2 and v4 during free viewing of natural scenes compared to controlled viewing. *Neuroreport*, 9 (9):2153, 1998.

M.E. Goldberg and R.H. Wurtz. Activity of superior colliculus in behaving monkey. II. Effect of attention on neuronal responses. *Journal of Neurophysiology*, 35(4):560, 1972a. ISSN 0022-3077.

M.E. Goldberg and R.H. Wurtz. Activity of superior colliculus in behaving monkey. i. visual receptive fields of single neurons. *J Neurophysiol*, 35(4):542–559, 1972b.

M.E. Goldberg, J.W. Bisley, K.D. Powell, and J. Gottlieb. Saccades, salience and attention: the role of the lateral intraparietal area in visual behavior. *Progress in brain research*, 155:157–175, 2006.

P.I. Good. *Resampling methods: a practical guide to data analysis.* Birkhauser, 2001.

J. Graham, C.S. Lin, and JH Kaas. Subcortical projections of six visual cortical areas in the owl monkey, aotus trivirgatus. *The Journal of comparative neurology*, 187(3): 557–580, 1979.

K. Grill-Spector, R. Henson, and A. Martin. Repetition and the brain: neural models of stimulus-specific effects. *Trends in cognitive sciences*, 10(1):14–23, 2006.

A. Grinvald, E.E. Lieke, R.D. Frostig, and R. Hildesheim. Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *The Journal of neuroscience*, 14(5):2545–2568, 1994.

C.M. Harris, J. Wallman, and C.A. Scudder. Fourier analysis of saccades in monkeys and humans. *Journal of neurophysiology*, 63(4):877–886, 1990.

J.K. Harting. Descending pathways from the superior colliculus: an autoradiographic analysis in the rhesus monkey (macaca mulatta). *The Journal of comparative neurology*, 173(3):583–612, 2004.

R.D. Hawkins, T.E. Cohen, and E.R. Kandel. Dishabituation in aplysia can involve either reversal of habituation or superimposed sensitization. *Learning & Memory*, 13 (3):397–403, 2006.

AV Hays Jr, BJ Richmond, and LM Optican. Unix-based multiple-process system, for real-time data acquisition and control. *Journal of Neuroscience*, 28(17), 1982.

J.M. Henderson, P.A. Weeks Jr, and A. Hollingworth. The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance; Journal of Experimental Psychology: Human Perception and Performance*, 25(1):210, 1999.

O. Hikosaka, R.H. Wurtz, et al. Visual and oculomotor functions of monkey substantia nigra pars reticulata. iii. memory-contingent visual and saccade responses. *J Neurophysiol*, 49(5):1268–1284, 1983.

T. Hosoya, S.A. Baccus, and M. Meister. Dynamic predictive coding by the retina. *Nature*, 436(7047):71–77, 2005.

D.H. Hubel, S. LeVay, and T.N. Wiesel. Mode of termination of retinotectal fibers in macaque monkey: an autoradiographic study. *Brain Research*, 96(1):25–40, 1975. ISSN 0006-8993.

A.E. Hudson, N.D. Schiff, J.D. Victor, and K.P. Purpura. Attentional modulation of adaptation in v4. *European Journal of Neuroscience*, 30(1):151–171, 2009.

M.F. Huerta and J.K. Harting. Sublamination within the superficial gray layer of the squirrel monkey: an analysis of the tectopulvinar projection using anterograde and retrograde transport methods. *Brain research*, 261(1):119–126, 1983.

D. Ingle. Sensorimotor function of the midbrain tectum. II. Classes of visually guided behavior. *Neurosciences Research Program Bulletin*, 13(2):180, 1975. ISSN 0028-3967.

T. Isa and W.C. Hall. Exploring the superior colliculus in vitro. *Journal of neurophysiology*, 102(5):2581, 2009. ISSN 0022-3077.

T. Isa and Y. Saito. The direct visuo-motor pathway in mammalian superior colliculus; novel perspective on the interlaminar connection. *Neuroscience Research*, 41(2):107–113, 2001. ISSN 0168-0102.

T. Isa, T. Endo, and Y. Saito. The visuo-motor pathway in the local circuit of the rat superior colliculus. *Journal of Neuroscience*, 18(20):8496, 1998.

L. Itti. The ilab neuromorphic vision c++ toolkit: Free tools for the next generation of vision algorithms. *The Neuromorphic Engineer*, 1(1), 2004.

L. Itti. Quantitative modelling of perceptual salience at human eye position. *Visual cognition*, 14(4-8):959–984, 2006.

L. Itti and P. Baldi. A principled approach to detecting surprising events in video. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 631–637. IEEE, 2005.

L. Itti and P. Baldi. Bayesian surprise attracts human attention. *Vision research*, 49(10): 1295–1306, 2009.

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

L. Itti and C. Koch. Computational modeling of visual attention. *Nature reviews neuroscience*, 2(3):194–203, 2001a.

L. Itti and C. Koch. Feature combination strategies for saliency-based visual attention systems. *Journal of Electronic Imaging*, 10(1):161–169, 2001b.

L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20 (11):1254–1259, 1998. ISSN 0162-8828.

H. Jiang, B.E. Stein, and J.G. McHaffie. Opposing basal ganglia processes shape midbrain visuomotor activity bilaterally. *Nature*, 423(6943):982–986, 2003. ISSN 0028-0836.

M.C. Jones, J.S. Marron, and S.J. Sheather. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, pages 401–407, 1996.

S.J. Judge, B.J. Richmond, and C. Chu. Implantation of magnetic search coils for measurement of eye position: An improved method. *Vision Research*, 20:535—538, 1980.

K. Kaneda, K. Isa, Y. Yanagawa, and T. Isa. Nigral inhibition of gabaergic neurons in mouse superior colliculus. *The Journal of Neuroscience*, 28(43):11071–11078, 2008.

H. Katsuta and T. Isa. Release from GABAA receptor-mediated inhibition unmasks interlaminar connection within superior colliculus in anesthetized adult rats. *Neuroscience research*, 46(1):73–83, 2003. ISSN 0168-0102.

C. Kayser, K.P. Körding, and P. König. Processing of complex stimuli and natural scenes in the visual cortex. *Current opinion in neurobiology*, 14(4):468–473, 2004.

R.M. Klein. Inhibition of return. *Trends in Cognitive Sciences*, 4(4):138–147, 2000. ISSN 1364-6613.

C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiol*, 4(4):219–27, 1985.

A. Kohn. Visual adaptation: physiology, mechanisms, and functional benefits. *Journal of neurophysiology*, 97(5):3155–3164, 2007.

R.J. Krauzlis, D. Liston, and C.D. Carello. Target selection and the superior colliculus: goals, choices and hypotheses. *Vision research*, 44(12):1445–1451, 2004. ISSN 0042-6989.

B. Krekelberg, G.M. Boynton, and R.J.A. Van Wezel. Adaptation: from single cells to bold signals. *Trends in neurosciences*, 29(5):250–256, 2006.

S. Kullback and R.A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

J.C. Lagarias, J.A. Reeds, M.H. Wright, and P.E. Wright. Convergence properties of the nelder-mead simplex method in low dimensions. *Siam journal of optimization*, 9: 112–147, 1998.

O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau. A coherent computational approach to model bottom-up visual attention. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(5):802–817, 2006.

X. Li and M.A. Basso. Preparing to move increases the sensitivity of superior colliculus neurons. *The Journal of Neuroscience*, 28(17):4561–4577, 2008.

G.R. Loftus and N.H. Mackworth. Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 4(4):565, 1978.

L.P. Lovejoy and R.J. Krauzlis. Inactivation of primate superior colliculus impairs covert selection of signals for perceptual judgments. *Nature neuroscience*, 13(2):261–266, 2009. ISSN 1097-6256.

F. Lui, K.M. Gregory, R.H.I. Blanks, and R.A. Giolli. Projections from visual areas of the cerebral cortex to pretectal nuclear complex, terminal accessory optic nuclei, and superior colliculus in macaque monkey. *The Journal of Comparative Neurology*, 363 (3):439–460, 1995. ISSN 1096-9861.

L. Maffei, A. Fiorentini, and S. Bisti. Neural correlate of perceptual adaptation to gratings. *Science*, 182(4116):1036–1038, 1973.

E.A. Marcus, T.G. Nolen, C.H. Rankin, and T.J. Carew. Behavioral dissociation of dishabituation, sensitization, and inhibition in aplysia. *Science*, 241(4862):210–213, 1988.

R.A. Marino, C.K. Rodgers, R. Levy, and D.P. Munoz. Spatial relationships of visuomotor transformations in the superior colliculus map. *Journal of neurophysiology*, 100(5):2564, 2008. ISSN 0022-3077.

JH Maunsell and D.C. van Essen. The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *Journal of Neuroscience*, 3(12):2563, 1983.

P.J. May. The mammalian superior colliculus: laminar structure and connections. *Progress in Brain Research*, 151:321–378, 2006. ISSN 0079-6123.

J.P. Mayo and M.A. Sommer. Neuronal adaptation caused by sequential visual stimulation in the frontal eye field. *Journal of neurophysiology*, 100(4):1923–1935, 2008.

L.E. Mays and D.L. Sparks. Dissociation of visual and saccade-related responses in superior colliculus neurons. *Journal of Neurophysiology*, 43(1):207–232, 1980.

J.T. McIlwain. Visual receptive fields and their images in superior colliculus of the cat. *Journal of Neurophysiology*, 38(2):219–230, 1975.

R.M. McPeek and E.L. Keller. Saccade target selection in the superior colliculus during a visual search task. *Journal of Neurophysiology*, 88(4):2019, 2002. ISSN 0022-3077.

C.W. Mohler and R.H. Wurtz. Organization of monkey superior colliculus: intermediate layer cells discharging before eye movements. *Journal of Neurophysiology*, 39(4):722, 1976. ISSN 0022-3077.

B.C. Motter. Modulation of transient and sustained response components of v4 neurons by temporal crowding in flashed stimulus sequences. *The Journal of neuroscience*, 26 (38):9683–9694, 2006.

J.A. Movshon and P. Lennie. Pattern-selective adaptation in visual cortical neurones. *Nature*, 278(5707):850–852, 1979.

J.R. Müller, A.B. Metha, J. Krauskopf, and P. Lennie. Rapid adaptation in visual cortex to the structure of images. *Science*, 285(5432):1405, 1999.

J.R. Muller, M.G. Philiastides, and W.T. Newsome. Microstimulation of the superior colliculus focuses attention without moving the eyes. *Proceedings of the National Academy of Sciences of the United States of America*, 102(3):524, 2005.

D.P. Munoz and P.J. Istvan. Lateral inhibitory interactions in the intermediate layers of the monkey superior colliculus. *Journal of Neurophysiology*, 79(3):1193, 1998. ISSN 0022-3077.

D.P. Munoz and R.H. Wurtz. Fixation cells in monkey superior colliculus. I. Characteristics of cell discharge. *Journal of Neurophysiology*, 70(2):559, 1993. ISSN 0022-3077.

D.P. Munoz and R.H. Wurtz. Saccade-related activity in monkey superior colliculus. I. Characteristics of burst and buildup cells. *Journal of Neurophysiology*, 73(6):2313, 1995. ISSN 0022-3077.

DP Munoz, MC Dorris, M. Pare, S. Everling, et al. On your mark, get set: brainstem circuitry underlying saccadic initiation. *Canadian journal of physiology and pharmacology*, 78(11):934–957, 2000.

M.B. Neider and G.J. Zelinsky. Scene context guides eye movements during visual search. *Vision research*, 46(5):614–621, 2006.

U. Neisser. *Cognitive psychology.* Appleton-Century-Crofts New York, 1967. ISBN 0390665096.

D. Noton and L. Stark. Scanpaths in eye movements during pattern perception. *Science*, 171(3968):308–311, 1971.

A. Oliva, A. Torralba, M.S. Castelhano, and J.M. Henderson. Top-down control of visual attention in object detection. In *International Conference on Image Processing*, volume 1, pages I–253. IEEE, 2003.

E. Olivier, MC Dorris, and DP Munoz. Lateral interactions in the superior colliculus, not an extended fixation zone, can account for the remote distractor effect. *Behavioral and Brain Sciences*, 22(04):694–695, 1999. ISSN 0140-525X.

G.A. Orban, D. Van Essen, and W. Vanduffel. Comparative mapping of higher visual areas in monkeys and humans. *Trends in cognitive sciences*, 8(7):315–324, 2004.

F.P. Ottes, J.A.M. Van Gisbergen, and J.J. Eggermont. Visuomotor fields of the superior colliculus: a quantitative model. *Vision Research*, 26(6):857–873, 1986.

M. Paré and D.P. Munoz. Expression of a re-centering bias in saccade regulation by superior colliculus neurons. *Experimental Brain Research*, 137(3):354–368, 2001.

V. Pariyadath and D. Eagleman. The effect of predictability on subjective duration. *PLoS One*, 2(11):e1264, 2007.

V. Pariyadath and D.M. Eagleman. Brief subjective durations contract with repetition. *Journal of vision*, 8(16), 2008.

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

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

R.J. Peters, A. Iyer, L. Itti, and C. Koch. Components of bottom-up gaze allocation in natural images. *Vision research*, 45(18):2397–2416, 2005.

P. Phongphanphanee, K. Kaneda, and T. Isa. Spatiotemporal profiles of field potentials in mouse superior colliculus analyzed by multichannel recording. *Journal of Neuroscience*, 28(37):9309, 2008.

D.A. Pollen and S.F. Ronner. Visual cortical neurons as localized spatial frequency filters. *Systems, Man and Cybernetics, IEEE Transactions on*, (5):907–916, 1983.

F. Prévost, F. Lepore, and J.P. Guillemot. Spatio-temporal receptive field properties of cells in the rat superior colliculus. *Brain research*, 1142:80–91, 2007. ISSN 0006-8993.

C.M. Privitera and L.W. Stark. Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(9):970–982, 2000.

C. Quaia, M. Paré, R.H. Wurtz, and L.M. Optican. Extent of compensation for variations in monkey saccadic eye movements. *Experimental brain research*, 132(1):39–51, 2000.

S. Ramat, R.J. Leigh, D.S. Zee, and L.M. Optican. What clinical disorders tell us about the neural control of saccadic eye movements. *Brain*, 130(1):10–35, 2007.

C.H. Rankin and T.J. Carew. Dishabituation and sensitization emerge as separate processes during development in aplysia. *The Journal of neuroscience*, 8(1):197–211, 1988.

P. Redgrave and K. Gurney. The short-latency dopamine signal: a role in discovering novel actions? *Nature Reviews Neuroscience*, 7(12):967–975, 2006.

P. Reinagel. How do visual neurons respond in the real world? *Current opinion in Neurobiology*, 11(4):437–442, 2001.

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

D.A. Robinson. A method of measuring eye movemnent using a scieral search coil in a magnetic field. *Bio-medical Electronics, IEEE Transactions on*, 10(4):137–145, 1963.

DA Robinson. Eye movements evoked by collicular stimulation in the alert monkey. *Vision Research*, 12(11):1795–1808, 1972. ISSN 0042-6989.

D.L. Robinson and C. Kertzman. Covert orienting of attention in macaques. iii. contributions of the superior colliculus. *Journal of neurophysiology*, 74(2):713–721, 1995.

D. Rose, J. Summers, et al. Duration illusions in a train of visual stimuli. *PERCEPTION-LONDON-*, 24:1177–1177, 1995.

P.H. Schiller and F. Koerner. Discharge characteristics of single units in superior colliculus of the alert rhesus monkey. *Journal of Neurophysiology*, 34(5):920, 1971. ISSN 0022-3077.

E.L. Schwartz. Spatial mapping in the primate sensory projection: analytic structure and relevance to perception. *Biological cybernetics*, 25(4):181–194, 1977.

E.L. Schwartz. Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. *Vision research*, 20(8):645–669, 1980.

C.E. Shannon, W. Weaver, R.E. Blahut, and B. Hajek. *The mathematical theory of communication*, volume 117. University of Illinois press Urbana, 1949.

S. Shipp. The brain circuitry of attention. *Trends in cognitive sciences*, 8(5):223–230, 2004. ISSN 1364-6613.

E.P. Simoncelli and B.A. Olshausen. Natural image statistics and neural representation. *Annual review of neuroscience*, 24(1):1193–1216, 2001.

S.M. Smirnakis, M.J. Berry, D.K. Warland, W. Bialek, and M. Meister. Adaptation of retinal processing to image contrast and spatial scale. 1997.

R.R. Sokal and F.J. Rohlf. *Biometry: the principles and practice of statistics in biological research*. WH Freeman, 1995.

E.N. Sokolov. Higher nervous functions: The orienting reflex. *Annual review of physiology*, 25(1):545–580, 1963.

S.G. Solomon, J.W. Peirce, N.T. Dhruv, and P. Lennie. Profound contrast adaptation early in the visual pathway. *Neuron*, 42(1):155–162, 2004.

A.A. Stocker and E.P. Simoncelli. Noise characteristics and prior expectations in human visual speed perception. *Nature neuroscience*, 9(4):578–585, 2006.

A. Stockman, L.T. Sharpe, et al. The spectral sensitivities of the middle-and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision research*, 40(13):1711, 2000.

I.M. Tarkka and D.S. Stokic. Source localization of p300 from oddball, single stimulus, and omitted-stimulus paradigms. *Brain topography*, 11(2):141–151, 1998.

B.W. Tatler. 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), 2007.

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

J. Theeuwes et al. Endogenous and exogenous control of visual selection. *PERCEPTION-LONDON-*, 23:429–429, 1994.

K.G. Thompson, D.P. Hanes, N.P. Bichot, and J.D. Schall. Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. *Journal of Neurophysiology*, 76(6):4040–4055, 1996.

J. Tigges and M. Tigges. Distribution of retinofugal and corticofugal axon terminals in the superior colliculus of squirrel monkey. *Investigative ophthalmology & visual science*, 20(2):149–158, 1981.

A.M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136, 1980. ISSN 0010-0285.

P.U. Tse, J. Intriligator, J. Rivest, and P. Cavanagh. Attention and the subjective expansion of time. *Attention, Perception, & Psychophysics*, 66(7):1171–1189, 2004.

J.K. Tsotsos. Analyzing vision at the complexity level. *Behavioral and Brain Sciences*, 13(3):423–469, 1990.

JA Van Gisbergen, DA Robinson, and S. Gielen. A quantitative analysis of generation of saccadic eye movements by burst neurons. *Journal of Neurophysiology*, 45(3):417–442, 1981.

W.E. Vinje and J.L. Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287(5456):1273–1276, 2000.

B.J. White, S.E. Boehnke, R.A. Marino, L. Itti, and D.P. Munoz. Color-Related Signals in the Primate Superior Colliculus. *Journal of Neuroscience*, 29(39):12159, 2009.

K.J. Wiebe and A. Basu. Modelling ecologically specialized biological visual systems. *Pattern Recognition*, 30(10):1687–1703, 1997.

J.M. Wolfe and T.S. Horowitz. What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5(6):495–501, 2004. ISSN 1471-003X.

EJ Woods and BJ Frost. Adaptation and habituation characteristics of tectal neurons in the pigeon. *Experimental Brain Research*, 27(3):347–354, 1977.

H. Yabe, M. Tervaniemi, K. Reinikainen, et al. Temporal window of integration revealed by mmn to sound omission. *NeuroReport*, 8(8):1971, 1997.

A.L. Yarbus. *Eye movements and vision.* Plenum press, 1967.