Title —

Biomimetic Oculomotor Control

Authors —

[1]Tomohiro Shibata, [12]Sethu Vijayakumar, [3]Jörg Conradt, and [12]Stefan Schaal

Affiliation —

[1]Kawato Dynamic Brain Project, ERATO, Japan Science and Technology Corporation

[2]Computer Science and Neuroscience, Univ. of Sounthern California nbn

[3]Institute of Neuroinformatics, University/ETH Zurich

Postal Addresses of Authors —

[1]Japan Science and Technology Corporation

2-2-2 Hikari-dai, Seika, Soraku, Kyoto 619-0288, JAPAN

[2]Computer Science and Neuroscience, Univ. of Sounthern California

USC HEDCO Neuroscience Building 103, Los Angeles, CA90089-2520, USA

Email Addresses of Authors —

tom@atr.co.jp, sethu@usc.edu, conradt@usc.edu, sschaal@usc.edu

Corresponding Author —

Tomohiro Shibata,

(Direct) +81 774 95 1232 , (Fax) +81 774 95 1259

(E-mail)tom@atr.co.jp

Running title —

Biomimetic Oculomotor Control

Title —

Biomimetic Oculomotor Control

Abstract—

Oculomotor control in a humanoid robot faces similar problems as biological oculomotor systems, i.e., capturing targets accurately on a very narrow fovea, dealing with large delays in the control system, the stabilization of gaze in face of unknown perturbations of the body, selective attention, and the complexity of stereo vision. In this paper, we suggest control circuits to realize three of the most basic oculomotor behaviors and their integration - the vestibulo-ocular and optokinetic reflex (VOR-OKR) for gaze stabilization, smooth pursuit for tracking moving objects, and saccades for overt visual attention. Each of these behaviors and the mechanism for their integration was derived with inspiration from computational theories as well as behavioral and physiological data in neuroscience. Our implementations on a humanoid robot demonstrate good performance of the oculomotor behaviors, which proves to be a viable strategy to explore novel control mechanisms for humanoid robotics. Conversely, insights gained from our models have been able to directly influence views and provide new directions for computational neuroscience research.

Key words—

oculomotor control, computational neuroscience, feedback-error-learning, predictive control, visual attention, online statistical learning

1 Introduction

1.1 Research Objectives

The goal of our research is to understand the principles of information processing in the human brain, with a focus on basic sensorimotor control and the hope to expand this scope increasingly more towards more cognitive topics. As a research strategy, we chose an approach that emphasizes the interplay between computational neuroscience and humanoid robotics. In this approach, research topics are initially investigated from the present stage of knowledge of neurobiology and, subsequently, abstract computational models are created that can be implemented on a humanoid robot to accomplish interesting behavioral goals. Control theory and learning theory are employed to examine the validity of the models. The success of the models in actual robotic implementation is investigated. Theoretical and experimental insights are then used to re-evaluate biological data and the present stage of modeling, which usually leads to suggestions for improvement in both neuroscientific research and computational modeling. For the purpose of this research strategy, we developed a humanoid robot system with 30 degree-of-freedoms (DOFs), each of which is hydraulically operated and mimics the compliance of humans by means of impedance control in each joint. Kinematics and dynamics of the robot are as close as possible to the human counterpart.

In this paper, we present results of our research in the field of oculomotor control. Oculomotor control is one of the best investigated areas in computational neuroscience due to three reasons. Firstly the primate oculomotor systems are relatively simple. For example, the monkey's oculomotor system can be approximated by a second order linear system. It has only 3 DOFs per eye, and often only 1 DOF is used in neurobiological experiments. Secondly, it has a rich set of oculomotor behaviors including reflexes and adaptation. In the case of reflex behaviors, experimenters can eliminate the effects of attention, and easily elicit the reflex behavior by simple stimuli such as rotation of the head, spot lights, random dot displays, etc. Among the most well-known oculomotor reflexes are the vestibulo-ocular (VOR) and optokinetic response (OKR). The case of voluntary eye movement such as saccades and smooth pursuit have also been investigated extensively. It is also known that each of these behaviors are highly adaptive and the mechanisms for their plasticity have been a topic of much interest.

Moreover, the oculomotor behaviors cooperate to accomplish oculomotor tasks in an efficient manner. For example, the VOR has shorter latency and can thus, stabilize the image on the retina more efficiently than the OKR, whose retinal slip-based negative feedback system operates with around 100ms latency. But the OKR is also essential to eliminate the residual error by the VOR. Thus, the framework of oculomotor control is ideally suited to draw comparisons between biological knowledge, computational models, and empirical evaluations in robotic experiments.

Many artificial vision systems Ballard and Brown (1993)Aloimonos *et al.* (1987) have been developed to include oculomotor control techniques. Although most of artificial oculomotor systems are inspired by biology's active vision Panerai *et al.* (2000)Ferrell (1996)Capurro *et al.* (1996)Berthouze and Kuniyoshi (1998)Murray *et al.* (1995), only few implementations of oculomotor systems can be found that try to emphasize biological plausibility - a feature which we feel will contribute towards understanding and designing new paradigms in the field of computational neuroscience.

1.2 Robotic Head Setup and Control

We will present computational models for the three oculomotor behaviors we examined and the corresponding experimental results. In all experiments, the same platform - the vision head (see Fig. 1) of our Humanoid RobotShibata and Schaal (2001) - was used. The robot head has 7 DOFs in total, a neck with 3 DOFs and two camera eyes, each equipped with 2 independent DOFs, arranged as pan and tilt. In order to provide high-resolution vision simultaneously with large-field peripheral vision, the robot employs two cameras per eye, a foveal camera (24 deg view-angle horizontally) and a wide-angle camera(100 deg view-angle horizontally). This setup mimics the foveated retinal structure of primates, and it is also essential for an artificial vision system in order to obtain high resolution vision of objects of interest while still being able to perceive events in the peripheral environment (See Fig 2). In order to mimic the semicircular canal of biological systems, we attached a three-axis gyro-sensor circuit to the head. From the sensors of this circuit, the head angular velocity signal is acquired.

The learning controller is implemented with the real-time operating system VxWorks using several parallel Motorola PowerPC processors in a VME rack. Visual processing is performed out of specialized hardware, a Fujitsu tracking vision board and QuickMag color vision tracking system. The Fujitsu tracking vision board calculates retinal error (position error) and retinal slip (velocity error) information of each eye at 30Hz. The QuickMag system returns the centroid of blobs of pixels of pre-specified colors in the environment. Up to six different colors can be tracked simultaneously at 60Hz sampling rate.

The oculomotor control loop runs at 420 Hz, while the vision control loop runs at 60 Hz due to restrictions the QuickMag video processing rate. The oculomotor control loop implements a strong spring and damping term such that the nonlinearities of the oculomotor system due to hydraulics and attached cables become negligible.

1.3 Organization of the paper

In the following sections, we elucidate our model for the synthesis and control of VOR-OKR, smooth pursuit, and saccadic behaviors. In Section 5, we discuss the issue of their integration. Section 6 presents experimental results on real robotic hardware involving all the three oculomotor behaviors, followed by a discussion and conclusion section.

2 Vestibulo-Ocular-Reflex (VOR) and Opto-kinetic-Response (OKR)

The VOR reflex serves to keep the eyes fixed on a target in the case that there is a mechanical perturbation of the head, e.g., as caused by locomotion. The OKR has a similar functionality, just that it is triggered by a movement of the entire visual field, which it tries to compensate for - a typical movement that would be elicited in a movie theater when the entire scene on the screen moves.

2.1 Model

This section outlines the computational model of VOR-OKR we developed Shibata and Schaal (2001), shown schematically in Fig. 3. The inputs to the VOR-OKR system are i) the visual target in camera coordinates and ii) an angular velocity signal generated from a gyroscopic sensor due to perturbations of the robot's body; since the sensor is attached to the head, the signal is referred to as "head angular velocity". From the target position and eye position, retinal error and retinal slip can be computed. In the simplest case, the ideal compensatory desired movement of the eyes would be the negative of the retinal slip, but, in general, a nonlinear transformation from retinal slip error velocity to eye movement is needed due to off-axis effect, i.e. the fact that head axis and eye axis are not collinear. The retinal error signals are also used as input to a PD controller in the bottom part of Fig. 3. The gains of this PD controller have to be kept rather small due to the delays incurred in visual information processing. The output of the PD controller serves primarily as a teacher signal to the feedback error learning system. However, it is also needed to stabilize the crude feedforward controller in the shaded block of Figure 3. Without the feedback input, the feedforward controller would only be marginally stable due to the floating integrator. As described in a later section, eligibility traces, realized by a second order filter, are also used as inputs to the learning system.

The entire control systems is quite similar to what has been discovered in the primate oculomotor system. The a priori existing feedforward controller provides some crude functionality of the VOR while the feedback controller provides acceptable OKR performance for slowly changing visual targets and acts as a compensatory negative feedback for the VOR module. These systems form what is called the "direct pathway" of oculomotor control in biology.

By adding a learning controller trained with the feedback-error-learning Kawato (1990) strategy in the indirect pathway (see Fig. 3), excellent VOR performance can be accomplished even if the feedback pathway has large delays (see Section 2.2). Furthermore, the OKR performance is improved to some extent due to this indirect pathway. This learning network, known to be located in the primate cerebellum, acquires an inverse dynamics model of the oculomotor plant during the course of learning that compensates for the missing performance of the crude feedforward controller in the shaded box (c.f. Fig. 3). The coordination of a direct and indirect pathway is analogous to how the cerebellar pathway acts in parallel to the brainstem pathways Gomi and Kawato (1992). As discussed in Shibata and Schaal (2001), this control system is equally suitable for both biological and robotic oculomotor control. However, for the robotic control, we augment the biologically plausible model with a fast nonlinear learning scheme based on nonparametric regression networks Vijayakumar and Schaal (2000a).

2.2 Learning with delayed-error signal

For successful feedback-error-learning, the time-alignment between input signals and the feedback-error signal is theoretically crucial, and thus, additional techniques are required in the case of delayed sensory feedback. For instance, if a perturbation of the head or body has frequency components that are much faster than the delay in the feedback pathway during VOR learning, the phase delay in the feedback pathway becomes large resulting in very slow learning or in the worst case, learning can even become unstable.

To solve this "temporal credit assignment problem", the concept of eligibility traces has been suggested in both biological modeling and machine learning Barto *et al.* (1983). For neurons in the brain, it is assumed that a second messenger would tag a synapse as eligible for modification. This "tag" would decay with an appropriate time constant, thus forming a temporal eligibility window. Schweighofer et al. proposed a biologically plausible learning model for saccadic eye movement, and modeled the second messenger as a second order linear filter of the input signals to the learning systemSchweighofer *et al.* (1996). For this purpose, note that second order filter is better than first order filter since the impulse response of second order filter has a unimodal peak at a delay time determined by the time constant of the filter. For successful learning, the delay time only has to roughly coincide with the actual delay of the sensory feedback. Applying this technique to feedback-error learning, we complete our final learning control system (see Fig. 3) where the impulse response of a second order linear filter is added just before the "learning" box.

We investigated the efficacy of this technique and confirmed that this is more robust than just using an inaccurate fixed delay time element under the following conditions: (1) the actual delay is roughly less than 150ms, (2) the actual delay is somehow fluctuating on the order of tens of milliseconds, and (3) motion frequency is high, i.e. around 3 Hz or more. These properties are derived from the second order linear filter which changes its phase shift depending on the frequency of the input signal. Condition (1) is satisfied since our humanoid visoin system has a delay of around 70 ms. Condition (2) can be caused by more complicated visual processing where processing time is stochastic or input signal dependent. During the initial transients of learning, slightly incorrect predictions of the learning module can cause rather fast movement of the eyes, resulting in conditions mentioned in (3). The second order filter of the eligibility trace avoids that feedback error learning is destabilized during these transients, which strongly improves robustness and accuracy of learning.

3 Smooth Pursuit

Smooth pursuit refers to the oculomotor behavior of smoothly tracking a moving target on the fovea - a task which requires high accuracy; for instance such behaviors are needed to visually inspect moving object. For example, in constant velocity or in a sinusoidal target motion, the ratio of tracking velocity to target velocity(a.k.a. the smooth pursuit gain) is almost 1.0 Stark *et al.* (1962). From the control theoretical view, this performance cannot be achieved simply by type 0 servo due to the long delays inherent in visual information processing (e.g. around 70 ms in our humanoid vision system, and around 100 ms in the human brain). There is strong evidence that biological smooth pursuit seems to implement some predictive controller, for e.g., the report by Whittaker and Eaholtz Whittaker and Eaholtz (1982). In their experiment, human subjects tracked a sinusoidal target motion. After the target disappeared, sinusoidal post-pursuit eye motion continued to follow the expected trajectory of target.

In the field of robot vision, many projects investigated visual servoing, but, to our knowledge, without examining a smooth pursuit controller that has similar features and performance as that in primates. One of the most related pieces of research is in Bradshaw et al. Bradshaw *et al.* (1997), which employed a Kalman filter for prediction. However, these authors assumed prior knowledge of the target dynamics and, thus, avoided to address how unknown target motion can be tracked accurately. In contrast, in this paper, we present a biologically motivated smooth pursuit controller that learns to predict the visual target velocity in head coordinates based on fast on-line statistical learning of the target dynamics. In the following sections, we will first explain the setup of our smooth pursuit model, then, explain the learning component, and, in Section 3.2, describe the mechanism of how it learns to predict inspite of delayed input signals.

3.1 Model

Fig. 4 presents one of the simplest examples of our smooth pursuit model. It consists of three subsystems, i.e., a feedback controller, a target velocity predictor, and an inverse model controller of the oculomotor system. The feedback controller is not enough to accomplish smooth pursuit, since its pathway includes long delays. The predictor computes the present target velocity based on estimation of the past target state and fast learning of the target dynamics. The predicted target velocity information is input to the inverse model controller as a desired velocity command. In this control diagram, s and $1/s$ are Laplace transform operators denoting differentiation and integration, respectively. Δ stands for a constant delay element. e, $\dot{\mathrm{e}}$, E, and $\dot{\mathrm{E}}$ are the retinal error, the retinal slip, the eye angular position, and the eye angular velocity, respectively. As depicted in this diagram and without loss of generality, we assume the oculomotor plant as well as the visual target has a second order linear dynamics. As we mentioned before, the assumption of second order linear dynamics for the oculomotor plant is very common both in biology and in robot vision.

The predictor outputs an estimate of the current target velocity $\hat{\dot{\mathrm{x}}}(\mathrm{t})$ out of a history of past estimated target angular positions $\mathrm{x}(t-\Delta)$ and velocities $\dot{\mathrm{x}}(t-\Delta)$. In linear systems, the state predictor of a n-th order linear system can be defined as:

$$\boldsymbol{x}_{\mathrm{t}+1} = A\boldsymbol{x}_{\mathrm{t}} \tag{1}$$

where $\boldsymbol{x}$ is the $n \times 1$ state vector and A is the $n \times n$ state transition matrix. As we are only interested in velocity prediction in this paper, we reduce equation 1 to focus only on the the states that are identified with target velocities, not positions:

$$\hat{\dot{\mathrm{x}}}_{\mathrm{t}+1} = \boldsymbol{A}_2\boldsymbol{x}_{\mathrm{t}} \tag{2}$$

where $\boldsymbol{A}_2$ is the the appropriate submatrix of A corresponding to the target velocity component.

The inverse model controller receives sum of the predictor output and the PD feedback command as desired velocity. It should be noted that using only the desired velocity rather than both the position and velocity signals is the prudent thing to do. The inverse model control follows the specified desired trajectory such as position and velocity. If learning predictor would output both position and velocity, they might result in a very crude and inconsistent desired trajectory, which can make entire system unstable. Here, the positional feedback term can be regarded as an integrated error term for the inverse model control block.

It should be emphasized here that our smooth pursuit model has similar performance and features as that in primates. First, our model can achieve smooth pursuit with velocity gain one due to the predictor. Second, by multi-step prediction, the desired trajectory can be maintained even if there is no retinal signal after the target disappeared, or the target is occluded. Third, our model can cope with complex target motion rather than just constant velocity motion as long as the predictor knows the discretized target dynamics. Next, we describe the mechanism of how such dynamics can be acquired by on-line learning.

3.2 Learning the discrete predictor from the delayed signals

The learning scheme in Fig. 4 may appear difficult to implement as it has to learn the target dynamics out of the history of past estimated target states and the delayed retinal error. As will become apparent in the next paragraphs, however, a straightforward development allows us to solve this learning problem.

At time t, the predictor can only see the delayed estimated target state $\boldsymbol{x}_{\mathrm{t}-\Delta}$. The corresponding discrete target velocity prediction is represented as

$$\hat{\dot{\mathrm{x}}}_{\mathrm{t}} = \mathrm{f}(\boldsymbol{x}_{\mathrm{t}-\Delta}, \boldsymbol{w}_{\mathrm{t}}) \tag{3}$$

where $\boldsymbol{w}$ is a parameter vector. Let $\dot{\xi}$, the velocity prediction error, equal $\dot{\mathrm{x}} - \hat{\dot{\mathrm{x}}}$, and let the loss function J be the simple squared error:

$$J = \frac{1}{2}\dot{\xi}_{\mathrm{t}}^2 \tag{4}$$

Thus, a gradient descent learning rule for $\boldsymbol{w}$ can be written as:

$$(\frac{dw_i}{dt})_{\mathrm{t}} = -\epsilon(\frac{\partial J}{\partial w_i})_{\mathrm{t}} = \epsilon(\frac{\partial f}{\partial w_i})_{\mathrm{t}-\Delta}\dot{\xi}_{\mathrm{t}} \tag{5}$$

with ϵ denoting the learning rate. If we can make the assumption that the predicted target velocity $\hat{\dot{\mathrm{x}}}$ will be tracked accurately by the robot without delay, we can regard the retinal slip as the prediction error given by $\dot{\xi} = \dot{\mathrm{x}} - \hat{\dot{\mathrm{x}}} \simeq \dot{\mathrm{x}} - \dot{\mathrm{E}} = \dot{\mathrm{e}}$. The learning rule can thus be rewritten as:

$$(\frac{dw_i}{dt})_{\mathrm{t}} = \epsilon(\frac{\partial f}{\partial w_i})_{\mathrm{t}-\Delta}\dot{\mathrm{e}}_{\mathrm{t}} \tag{6}$$

Note that the time alignment of the predictor output f and the error $\dot{\xi}$ ($\simeq \dot{\mathrm{e}}$) needs to be correct for successful minimization of the loss function J. Since the predictor has no access to $\dot{\mathrm{e}}_{\mathrm{t}}$ at time t, a modified learning rule is required. We achieve this by introducing a delayed form of Eq. 6:

$$(\frac{dw_i}{dt})_{\mathrm{t}} = \epsilon(\frac{\partial f}{\partial w_i})_{\mathrm{t}-2\Delta}\dot{\mathrm{e}}_{\mathrm{t}-\Delta} \tag{7}$$

Thus, the predictor is required to keep the information $\partial f/\partial w_i$ in memory for the duration of Δ. In summary, it is important to use the most recent information for prediction, but to use one delayed by Δ for learning in order to achieve successful learning and control. Note that the delay Δ can be implemented as described in 2.2.

The assumption we made that the predicted target velocity can be tracked accurately by the robot without delay assumes the existence of an accurate inverse model controller of the oculomotor plant. What happens if this assumption is not true In this case, the RNN has to learn a composite task including the target dynamics and the dynamics of the extended plant, i.e., a plant that cannot canceled out and remains due to the inaccurate inverse model controller, in order to minimize the retinal slip. Indeed, in a simulation with the circuit depicted in Fig. 4, we confirmed that learning is successful when the inverse dynamics model was imperfect. This is theoretically expected since the extended plant here is, at most, second order linear which could be canceled out by somehow modulating the predictor, which is also a second order linear system in the RNN.

4 Saccade and Overt Visual Attention

Visual attention involves directing a "spotlight" of attention Koch and Ullman (1984) to interesting areas, extracted from a multitude of sensory inputs. Most commonly, attention will require to move the body, head, eyes, or a combination of these in order to acquire the target of interest with high-resolution foveal vision, referred to as 'overt' attention, as opposed to covert attention which does not involve movement.

There has been extensive work in modeling attention and understanding the neurobiological mechanisms of generating the visual "spotlight" of attention Neibur and Koch (1998), both from a top-downParasuraman (1998) and a bottom-up perspective Itti and Koch (1999, 2000) - albeit mainly for static images. From the perspective of overt shift of foci, there has been some work on saccadic eye motion generation using spatial filters Rao and Ballard (1995), saccadic motor planning by integrating visual information Kopecz and Schoner (1995), social robotics Breazeal and Scassellati (1999), and humanoid robotics Driscoll *et al.* (1998). In contrast to this previous work, our research focus lies on creating a biologically inspired approach to visual attention and oculomotor control by employing theoretically sound computational elements that were derived from models of cortical neural networks, and that can serve for comparisons with biological behavior. We also emphasize real-time performance and the integration of the attention system on a full-body humanoid robot that is not stationary in world coordinates. As will be shown below, these features require additional computational consideration such as the remapping of a saliency map for attention after body movement. In the following sections, we will first give an overview of the attentional system's modules, then explain the computational principles of each module, before we provide some experimental evaluations on our humanoid robot.

4.1 Sensor pre-processing and integration

The key element of our Sensory Processing block (Fig. 5) is a competitive dynamical neural network, derived in Amari and Arbib's Amari and Arbib (1977) *neural fields* approach for modeling cortical information processing. The goal of this network is to take as input spatially localized stimuli, have them compete to become the next saccade target, and finally output the winning target. For this purpose, the sensory input pre-processing stage takes the raw visual flow $V_F(\mathbf{x}, \mathbf{t})$ as inputs to the *stimulus dynamics*, a first order dynamical system. Using $\mathbf{x}$ to denote the position of a stimulus in camera coordinates, the stimulus dynamics is:

$$\dot{S}(\mathbf{x}) = -\alpha S(\mathbf{x}) + VisInp(\mathbf{x}, \mathbf{t}) \tag{8}$$

where

$$VisInp(\mathbf{x}, t) = \int_R G(\mathbf{x}, t) * exp(-\mathbf{x}^2/2\sigma^2) d\mathbf{x} \tag{9}$$

$$G(\mathbf{x}, t) = V_F(\mathbf{x}, t) + \gamma * \lfloor \dot{V}_F(\mathbf{x}, t) \rfloor_+ \tag{10}$$

Eq.(10) enhances the raw visual flow vector when it is increasing to emphasize new stimuli in the scene, while Eq.(9) implements a Gaussian spatial smoother of the stimuli to reduce the effects of noise. The variable α was set to a value of 100 in our experiments while the values of γ and σ were adapted based on the noise content of the environment and sensing equipment. The top of Fig. 6a shows an example of a typical stimulus pattern in the two dimensional neural network due to a moving object at the top-left of the camera image. In general, we could have multimodal sensory inputs, e.g. from color detectors, edge detectors, audio input, etc., feeding into Eq.(10) as a sensory signal. As suggested by Itti and Koch Itti and Koch (1999, 2000), it would be useful to weight these inputs according to their importance in the scene, usually based on some top-down feedback or task-specific biasing (e.g., if we know that color is more important than motion).

This stimulus dynamics feeds into a *saliency map* Koch and Ullman (1984), essentially a winner-take-all (WTA) network which decides on a winning stimulus from many simultaneous stimuli in the camera field. The winning stimulus will become the next saccade target or focus of overt attention. The WTA network is realized based on the theory of *neural fields*, a spatial neural network inspired by the dynamics of short range excitatory and long range inhibitory interactions in the neo-cortex Amari (1977); Amari and Arbib (1977). The activation dynamics $u(\mathbf{x}, t)$ of the saliency map is expressed as:

$$\begin{aligned} \tau \dot{u}(\mathbf{x}) &= -u(\mathbf{x}) + S(\mathbf{x}) + h \\ &\quad + \sum_{\mathbf{x}'} \mathbf{w}(\mathbf{x}, \mathbf{x}') \sigma(u(\mathbf{x}')) \end{aligned} \tag{11}$$

Here, h is the base line activation level within the field, $S(\mathbf{x}, \mathbf{t})$ is the external stimulus input (Eq.8), $\mathbf{w}(\mathbf{x}, \mathbf{x}')$ describes the coupling strength between all the units of the network, and $\sigma(u)$ controls the local threshold of activation. Depending on the choice of parameter h and the form of σ and $\mathbf{w}$, the activation dynamics of Eq.(11) can have various stable equilibrium points Amari (1977). We are interested in a solution which has uniform activation at base line level in the absence of external stimuli, and which forms a unimodal activation pattern at the most significant stimulus in the presence of stimuli that are possibly dispersed throughout the spatial network. This is achieved by choosing a transfer function:

$$\sigma(u) = 1/(e^{(-cu)} + 1) \tag{12}$$

with constant $c >> 1$ and an interaction kernel with short range excitation and long-range inhibition term H_0:

$$\mathbf{w}(\mathbf{x}, \mathbf{x}') = ke^{-(\mathbf{x}-\mathbf{x}')^2/\sigma_w^2} - H_0 \tag{13}$$

The constants were fixed at $\tau = 0.01$, $h = -0.5$, $H_0 = 0.75$, $k = 4$, $\sigma_w^2 = 1.4$, and $c = 5000$, the values of which were decided based the magnitude of the stimulus dynamics $S(\mathbf{x}, \mathbf{t})$, as outlined in Amari (1977).

In addition to the stimulus driven dynamics, we also suppress the activation of the most recently attended location by adding a large negative activation in Eq.(10) at the location of the last saccade target. This strategy implements an *inhibition of return* Itti and Koch (2000) and ensures that the robot does not keep attending to the same location in the continuous presence of an interesting stimuli. While the negative stimulus added is instantaneous, the time constant of the activation dynamics essentially controls the decay of this *inhibition*, ensuring the attended locations are cycled back in due time if there is persistent activation. The plots at the bottom of Fig. 6(a)(b) illustrate the behavior of the activation dynamics just before and after an attention shift, including the effect of the negative activation after the saccade.

4.2 Planning and generation of motor commands

Given a new saccade target, extracted from the saliency map, the direction of gaze needs to be shifted to the center of this target. Since fifth order splines are a good approximation of biological movement trajectories Kawato (1999); Barnes (1993), we use this model to compute a desired trajectory from the current position $\mathbf{x_0}$ to the target $\mathbf{x_f}$, all expressed in camera coordinates. We do not claim that trajectory planning occurs in biology by using the techniques described; in fact this topic is a matter of active research in the motor control community. Here, our aim is to generate trajectories that closely resemble natural motion.

The camera-space trajectory is converted to joint space by inverse kinematics computations based on Resolved Motion Rate Control (RMRC) Liegeois (1977). We assume that only head and eye motion is needed to shift the gaze to the visual target, an assumption that is justified given that the target was already visible in the peripheral field of view. For the time being, the inverse kinematics computation is performed for the right eye only, while the left eye performs exactly the same motion as the right eye. Thus, we need to map from a 2D camera space of the right eye to a 5D joint space, comprised of pan and tilt of the camera, and 3 DOFs of the robot's neck. To obtain a unique inverse, we employ Liegeois Liegeois (1977) pseudo-inverse with optimization:

$$\dot{\boldsymbol{\theta}} = \mathbf{J}^{\#}\dot{\mathbf{x}} + (\mathbf{I} - \mathbf{J}\mathbf{J}^{\#})\mathbf{k_{null}} \tag{14}$$

$$\text{where } \mathbf{J}^{\#} = \mathbf{J^T}(\mathbf{J}\mathbf{J^T})^{-\mathbf{1}}$$

$\mathbf{k_{null}}$ is the gradient of an optimization criterion w.r.t. the joint angles $\boldsymbol{\theta}$. The second term of the Eq.(14) is the part that controls the movement in the null space of the head-eye system. Any contribution to $\dot{\boldsymbol{\theta}}$ from this term will not change the direction of gaze but will only change how much we use the head or eye DOFs to realize that gaze. As optimization criterion we chose:

$$L = \frac{1}{2}\sum_i w_i(\theta_i - \theta_{def,i})^2 \tag{15}$$

resulting in

$$k_{null,i} = \frac{\partial L}{\partial \theta_i} = w_i(\theta_i - \theta_{def,i}) \tag{16}$$

$$\tag{17}$$

This criterion keeps the redundant DOFs as close as possible to a default posture $\boldsymbol{\theta}_{def}$. Adding the weights w_i allows giving more or less importance to enforcing the optimization criterion for certain DOF–this feature is useful to create natural looking head-eye coordination.

Once the desired trajectory is converted to joint space, it is tracked by an inverse dynamics controller using a learned inverse dynamics model Vijayakumar and Schaal (2000b).

5 Integration of Oculomotor Behaviors

In this section, we will attempt to integrate the three independent oculomotor behaviors that we have described so far with the aim of improving their overall performance capability. While there has been recent research on integrating the behaviors focusing on it's emergent properties Kuniyoshi and Berthouze (1998), we take a rather traditional approach (e.g, Brown (1990b)Brown (1990a)Murray *et al.* (1995)Takanishi *et al.* (1997)) of investigating how they can be integrated from the computational viewpoint.

To begin with, we will consider the saccadic behavior as a separate subsytem which supports the VOR-OKR and smooth pursuit modules in the collaborative integration of these oculomotor behaviors. This is primarily due to the reason that saccadic movements have an objective of implementing point attention in space, a goal which runs counter to the objective of smooth pursuit. Moreover, saccadic movements stimulate the entire retinal field and more sophisticated integration schemes have to be implemented to avoid interference with the other behaviors - a topic of future research.

5.1 Coordinates for integration

One of the most important issues in integrating the oculomotor behaviours is the question of which coordinates to perform the integration in. We have two candidates: one is velocity command space, and the other is motor command space.

In neuroscience, researchers have suggested an integration mechanism in which the final common path (FCP) received all the different velocity commands and output the final motor command to the oculomotor plant. In other words, the FCP is regarded as an inverse model controller. This idea is consistent with what we mentioned in section 3 that receiving only the desired velocity is required.

However, one possible problem with this formulation is the fact that the velocity commands should be normalized in order to be simply summed up. For example, a desired velocity generated by a smooth pursuit circuit is not realized if the velocity command has a different scale compared to the desired velocity input of the inverse model learned through the VOR. In this case, however, there is a possibility for the predictor to learn to output necessary velocity commands suitable for the inverse model controller.

If one combines the behavior using the motor command space, such normalization problem does not occur - although we needs one inverse model for each oculomotor behavior, which is computationally inefficient. This issue is still a matter of ongoing debate in the computational neuroscience field. In the current model, we adopt the motor command space due to ease of implementation.

5.2 Integration of VOR-OKR with smooth pursuit

The VOR-OKR and smooth pursuit behaviors should cooperate, especially since moving the head during smooth pursuit is often useful to widen the tracking range and also, VOR-OKR could help sustain good tracking performance in the presence of unforseen perturbations.

Integrating VOR-OKR behaviors with smooth pursuit can be extremely simple in many cases. In the case when only the eye is used to perform the smooth pursuit, all we need to do is to sum up the motor commands from the two modules. If there is a perturbation of the body (and hence, the head) while the smooth pursuit is in progress, the VOR-OKR reflex will kick in to compensate the eye movement for this perturbation. Another possibility is that the head is moved in order to achieve a better range for the target tracking - a voluntary motion rather than a perturbation. However, even under this circumstance, the VOR-OKR behavior will help to sustain efficient target following by cancelling out the effect of the head/body movement. We can think of the resulting movement as a planning in combined eye and head coordinates, where the VOR-OKR system is helping to generate the compensatory corrective movements in the eyes as a result of the head motion. However, more sophisticated algorithms are necessary for implementing finely coordinated head-eye motion (planned with certain optimization criterion) without experiencing the negative interference effects.

5.3 Velocity following through corrective saccades

Based on studies and data collected in neuroscience, it is believed that in primates, all oculomotor behaviors other that saccades follow the velocity of stimulus or target and not their position. This is consistent with our smooth pursuit model which predicts current target velocity without dependence on target position. Also, learning uses only the retinal slip (velocity) and not the retinal error. This is also along the lines of biological observations that motion information obtained from the retina is the retinal slip and not retinal error.

From a computational perspective, retinal slip - although expensive to compute - is extremely suited for parallel (fast) implementations due to it's simplicity. Moreover, working with retinal slip is much more robust to noise compared to the retinal error computation because it does not suffer from drifting or changing image patterns over frames. Since important reflexes such as the VOR-OKR should be fast and robust, it is natural that they follow velocity signals. There are commerically available optical flow computation hardware (e.g. Fujitsu Tracking Vision) which can perform these operations efficiently and robustly in real time.

The method of using only velocity signals has an inherent drawback - the positional errors accumulate over time to give an increasing steady state lag. It is here that the saccades contribute in the overall integration - periodically, the system makes corrective saccades to correct for this positional error in order to ensure accurate pursuit and to keep the target visible in the narrow foveal vision.

Incorporating the above considerations, we modify the diagrams in Fig. 3 and Fig. 4 to follow velocity as shown in Fig. 7. The position gain in the feedback loop is removed. Instead, the saccadic behavior will take over periodically to correct for the errors. During the process of saccadic correction (which is a very fast movement of less than 100ms), the retinal slip information is shutdown or supressed from reaching the smooth pursuit module to prevent spurious effects (c.f. Fig. 7) . This shutdown triggers the operation of the multi-step dynamics prediction in the smooth pursuit module, ensuring that even though we do not get any retinal slip information, the smooth pursuit continues unhindered based on the prediction from the learned dynamics of the target. The retinal error required in the VOR-OKR module is calculated by the integration of the retinal slip.

6 Experimental results

In this section, we will present experimental results of the VOR-OKR, smooth pursuit and saccades implemented on our humanoid robot

6.1 VOR-OKR

We performed a pilot perturbation experiment to demonstrate the basic stabilization capability of our VOR-OKR implementation. Fig. 8(a) shows the perturbation of the head in one DOF. The system had no knowledge of the oculomotor dynamics. Learning was started for all the four DOFs simultaneously. Fig. 8 (b)-(e) show the time course of the retinal error of all four DOFs, which demonstrate that fast learning in all four DOFs were achieved simultaneously.

Since the visual processing in our system introduces a delay of around 70 ms in the retinal signals, one of the important points to demonstrate about the capability of our system is to show how the eligibility traces can improve the efficiency of VOR learning. For this purpose, head movement was generated by three superimposed sinusoidal signals with frequencies of 0.6, 2.0, and 3.0 Hz and amplitude of 0.1 rad, respectively. A frequency of 3.0 Hz is high enough to result in blurred visual images. Fig. 9 shows the time course of the rectified (i.e., one second ensemble mean of squared value) retinal error during learning obtained from a moving average using a one second time window. While the dashed line represents data obtained from learning without eligibility traces, the solid line shows data acquired with eligibility traces. This figure shows that eligibility traces are necessary for successful learning as the the retinal error does not decrease without using the traces. It should also be noted that learning proceeds quite rapidly such that in less than a half a minute, the initial errors are reduced by a factor of two. Longer learning results in a further reduction of the error Shibata and Schaal (2001).

6.2 Smooth Pursuit

We present some results highlighting the basic tracking capability of our smooth pursuit controller. In this experiment, the the system did not know anything about the visual target dynamics in advance, but already obtained the inverse dynamics model of each eye. The motion of the visual target, a red ball, was given by the industrial manipulator. The motion was two dimensional - a simple sinusoidal of frequency 1 Hz signal shown in Fig. 10(a). Fig. 10(b) shows the time course of the rectified retinal error while Fig. 10(c) shows the time course of the rectified retinal slip. We see a very rapid convergence of learning of the target dynamics, while the convergence of the retinal error was relatively slow. Note that, in our learning scheme, the goal of the learning module is prediction of the target velocity, and its prediction error, the retinal slip, conveys only velocity information. Therefore, we see the rapid convergence in Fig. 10(c). As shown in Fig. 4, the positional error can be decreased by a feedback pathway including an integrator and P gain. It should be emphasized that the rectified retinal error is fairly small from the beginning of learning.

Note that our model does not rely on any specific or particular learning method. For tracking a target having a simple linear dynamics such as a swinging pendulum, we employed an adapted version of recursive least squares (RLS) Ljung and Soderstrom (1986), and obtained excellent results. For more complex target motion with nonlinear dynamics, we replaced RLS in our smooth pursuit controller with nonparametric regression networks we developed Vijayakumar and Schaal (2000b) Vijayakumar *et al.* (2001). This system was used to learn a periodic motion generated by van der Pol equations implemented on a separate industrial robot in our laboratory. Fig. 11 shows the excellent learning results of this experiment.

6.3 Saccades

We implemented the visual attention system on our humanoid robot. The stimulus dynamics and saliency map had 44x44 nodes, i.e., twice the length and width of the 22x22 nodes of the visual flow grid of the peripheral vision. This extended size assured that after a saccade, the remapping of the saliency map and stimulus dynamics could maintain stimuli outside of the peripheral vision for some time. The Jacobian needed for the inverse kinematics computation was estimated with linear regression from data collected from moving the head-eye system on randomized periodic trajectories for a few minutes. Due to the limited range of motion of the eye and head DOFs, the Jacobian could be assumed to be constant throughout the entire range of motion of head-eye system, which was confirmed by the excellent coefficient of determination of the regression of the Jacobian. The saliency map was able to determine winning targets at about 10Hz, which is comparable to the capabilities of the human attentional system.

An illustration of the working of the attentional system is provided in Fig. 12. The top image shows the robot's right eye peripheral view of the lab, focusing on the person in the middle of the image. At the bottom left part of the image, another person was waving a racket to attract the robot's attention. This motion elicited a saccade, recognizable from the middle image of Fig. 12 which shows the visual blur that the robot experienced during the movement. The bottom image of Fig. 12 demonstrates that after the saccade, the robot was correctly focusing on the new target. Note that the three images were sampled at 30Hz, indicating that the robot performed a very fast head-eye saccade of about 100ms duration, which is again comparable to human performance.

6.4 Integration of Oculomotor Behaviors

We conducted an experiment to analyze the combined effects of the three oculomotor behaviors. A visual target was moved horizontally using a driving signal that followed a sinusoid with its amplitude 0.25 [rad] and frequency 0.7 [Hz].

Fig. 13 demonstrates that our model integrating the three oculomotor behaviors has the capability to keep capturing the target given unknown and significant perturbation during tracking. The top graph (A) shows the time course of the perturbation. As shown, initially no perturbation was given. After 20 seconds, suddenly perturbation generated by two superimposed sinusoids of amplitude 0.1 [rad] and frequencies 1.0 [Hz] and 1.2 [Hz] respectively was injected into the system. The middle graph shows the rectified mean retinal errors; the solid line corresponds to the case of three behaviors cooperating, and the dotted line corresponds the case of no oculomotor control. Even after the introduction of the perturbation at a timeline of 20 seconds, it shows the rectified mean retinal error was much less than 0.1 [rad], which means the target was always on the foveal image stably. It should be noted that the rectified mean retinal error is decreasing through this experiment, which point to the benefits of continuously adaptive cooperation of the VOR-OKR and smooth pursuit. The bottom graph (C) presents the same analysis for rectified mean retinal slips, again showing the significant difference as seen in (B).

7 Conclusion

In this paper, we have presented our research on humanoid oculomotor control, focusing on models of the VOR-OKR reflex system, smooth pursuit, saccades, and their integration based on concepts of computational neuroscience. We have demonstrated the excellent performance of each oculomotor behavior and their coordination on the vision head that is an integral part of our humanoid robot - a specialized robotic platform developed with a strong emphasis on computational brain science research. In all given examples, the robot control mechanisms were derived based on principles of computational neuroscience and was proved to be able to generate viable solutions for robotic control with good performance.

Our efforts not only present novel control design paradigms for the humanoid robot, but also aim to contribute to brain science by proposing new biological control models and circuits and posing interesting problems towards exploratory neuroscience. Under the experiments with the VOR-OKR behavior, we have rendered a new role for the OKR pathway in that it may be used to stabilize the floating integrator in the direct pathway. We have also described our novel analysis on how biomimetic eligibility traces can be advantageous in comparison to engineering dead-time element under the biologically natural condition. Our smooth pursuit model has to potential to contribute to computational neuroscience modelling in a direct way. It is a quite simple but novel model that can explain many behavioral and physiological data. In particular, it is the first model claiming that primates' brain might learn target dynamics solely based on the retinal slip, and perform predictive control based on the dynamics. The saccade generation model integrates the saliency detection, motor control and coordinate maintainance functions in a close knit loop and successfully implements covert visual attention in real time. Finally, we have discussed the issues involved in integrating these oculomotor behaviors and ensuring that they cooperate without negative effects on the combined performance. As one of the preliminary results, we have been successful in implementing the VOR-OKR and smooth pursuit behavior by accessing only the retinal velocity information. This is more robust than trying to compute with retinal error information and any positional errors that accumulate are periodically corrected through the saccade behavior. We will work on these issue more elaborately for the more general cases, and hope to move on to more cognitive topics in the future.

References

Aloimonos, J., Weiss, I., and Bandyopadhyay, A. (1987). Active vision. *International Journal of Computer Vision*, **1**(4), 333–356.

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, **27**, 77–87.

Amari, S. and Arbib, M. (1977). Competition and cooperation in neural nets. In Metzler, editor, *Systems neuroscience*, pages 119–165, New York. Academic Press.

Ballard, D. and Brown, C. (1993). Principles of animate vision. In Y. Aloimonos, editor, *Active Perception*, pages 245–282. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Barnes, G. (1993). Visual-vestibular interaction in the control of head and eye movement: The role of visual feedback and predictive mechanisms. *Prrogress in Neurobiology*, **41**, 435–472.

Barto, A., Sutton, R., and Anderson, C. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Trans Syst, Man, and Cybern*, **SMC-13**, 834–846.

Berthouze, L. and Kuniyoshi, Y. (1998). Emergence and categorization of coordinated visual behavior through embodied interaction. *Autonomous Robot*, **15**(3/4), 369–379.

Bradshaw, K., Reid, I., and Murray, D. (1997). The active recovery of 3D motion trajectories and their use in prediction. *IEEE Trans on Pat Anal and Mach Intell*, **19**(3).

Breazeal, C. and Scassellati, B. (1999). A context dependent attention system for a humanoid robot. In *Proc Int Joint Conf Artif Intell (IJCAI-99)*, pages 1146–1151.

Brown, C. (1990a). Gaze controls with interactions and delays. *IEEE Trans on SMC*, **20**(1), 518–527.

Brown, C. (1990b). Prediction and cooperation in gaze control. *Biol Cybern*, **63**, 61–70.

Capurro, C., Panerai, F., and Sandini, G. (1996). Vergence and tracking fusing log-polar images. In *Proc Int Conf Pat Recog (ICPR'96)*, pages 740–744.

Driscoll, J., Peters II, R., and Cave, K. (1998). A visual attention network for a humanoid robot. In *Proc Int Conf Intell Robot Syst (IROS-98)*, pages 1968–1974.

Ferrell, C. (1996). Orientation behavior using registered topographic maps. In *Proc the 4th Int Conf on Simulation of Adaptive Behavior*, pages 124–131.

Gomi, H. and Kawato, M. (1992). Adaptive feedback control models of the vestibulocerebellum and spinocerebellum. *Biological Cybernetics*, **68**, 105–114.

Itti, L. and Koch, C. (1999). A comparison of feature combination strategies for saliency-based visual attention systems. In *Proc SPIE Human Vision and Electronic Imaging IV (HVEI'99)*, volume 3644, pages 473–82.

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

Kawato, M. (1990). Feedback-error-learning neural network for supervised motor learning. In R. Eckmiller, editor, *Adv Neural Comput*, pages 365–372. Elsevier, North-Holland.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Curr Opin Neurobiol*, **9**, 718–727.

Koch, C. and Ullman, S. (1984). Selecting one among the many: A simple network implementing shifts in selective visual attention. A.I. Memo 770, Massachusetts Institute of Technology.

Kopecz, K. and Schoner, G. (1995). Saccadic motor planning by integrating visual information and pre-information on neural dynamic fields. *Biological Cybernetics*, **73**, 49–60.

Kuniyoshi, Y. and Berthouze, L. (1998). Neural learning of embodied interaction dynamics. *Neural Networks*, **11**, 1259–1276.

Liegeois, A. (1977). Automatic supervisory control of the configuration and behavior of multibody mechanisms. *IEEE Trans Syst, Man, and Cybern*, **7**, 868–871.

Ljung, L. and Soderstrom, T. (1986). *Theory and Practice of Recursive Identification*. MIT Press.

Murray, D., Reid, I., Bradshaw, K., McLauchlan, P., Sharkey, P., and Fairley, S. (1995). Active exploration of dynamic and static scenes. In C. Brown and D. Terzopoulos, editors, *Real-time Computer Vision*, pages 117–140.

Neibur, E. and Koch, C. (1998). Computational architectures for attention. In R. Parasuraman, editor, *The Attentive Brain*, pages 163–186, Cambridge, MA. MIT Press.

Panerai, F., Metta, G., and Sandini, G. (2000). Adaptive image stabilization: a need for vision-based active robotic agents. In *Proc SAB 2000*, Paris, France.

Parasuraman, R. (1998). *The Attentive Brain*. MIT Press, Cambridge, MA.

Rao, R. and Ballard, D. (1995). Learning saccadic eye movements using multiscale spatial filters. In G. Tesauro, D. Touretzky, and T. Leen, editors, *Advances in Neural Information Processing Systems*, volume 7, pages 893–900, Cambridge, MA. MIT Press.

Schweighofer, N., Arbib, M., and Dominey, P. (1996). A model of the cerebellum in adaptive control of saccadic gain. *Biol Cybern*, **75**, 19–28.

Shibata, T. and Schaal, S. (2001). Biomimetic gaze stabilization based on feedback-error-learning with nonparametric regression networks. *Neural Networks*, **14**(2), 201–216.

Stark, L., Vossius, G., and Young, L. (1962). Predictive control of eye tracking movements. *Institute of Radio Engineers Trans in Human Factors and Electronics*, **3**, 52–57.

Takanish, A., Matsuno, T., and Kato, I. (1997). Development of a anthropomorphic head-eye robot with two eyes-coordinated head-eye motion and pursuing motion in the depth direction. In *Proc IEEE Int Conf Intell Robot Syst (IROS97)*, pages 799–804.

Vijayakumar, S. and Schaal, S. (2000a). Fast and efficient incremental learning for high dimensional movement systems. In *IEEE Int Conf Robot and Automat*, pages 1894–1899.

Vijayakumar, S. and Schaal, S. (2000b). LWPR : An O(n) algorithm for incremental real time learning in high dimensional space. In *Proc Int Conf Mach Learn (ICML 2000)*, pages 1079–1086.

Vijayakumar, S., D'Souza, A., Shibata, T., Contadt, J., and Schaal, S. (2001). Statistical learning for humanoid robots. *Autonomous Robots*. (in press).

Whittaker, S. and Eaholtz, G. (1982). Learning patterns of eye motion for foveal pursuit. *Assoc Res Vis Ophthal*, **23**(3), 393–397.

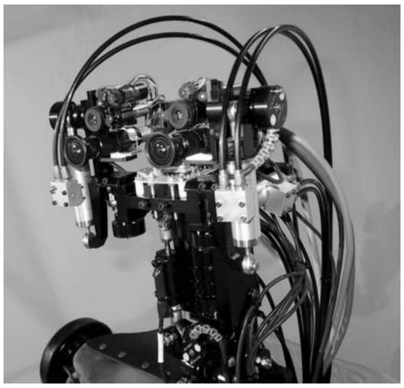


Figure 1: Humanoid vision head

Figure 2: (left) The hunanoid robot performing the task of pole-balancing; (right) monitor output of all four cameras (upper: foveal vision, lower: peripheral vision)

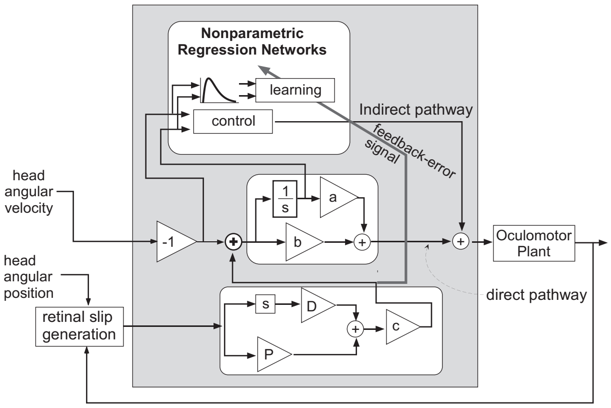


Figure 3: Our VOR-OKR model

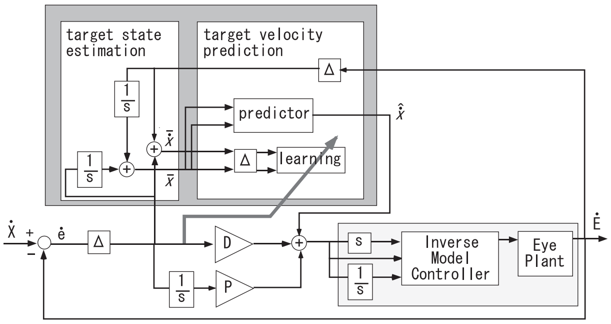


Figure 4: Simple example of our smooth pursuit model

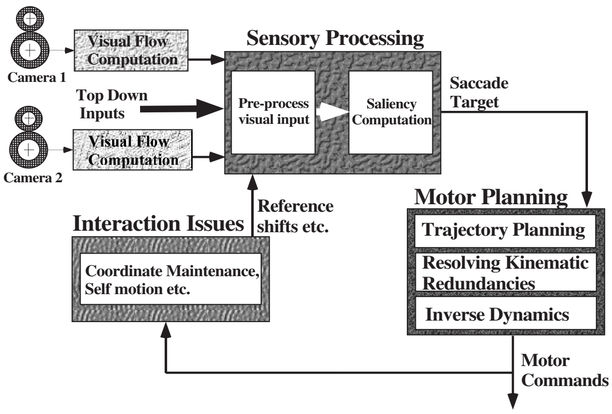
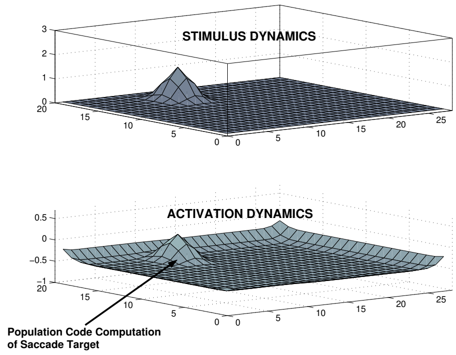
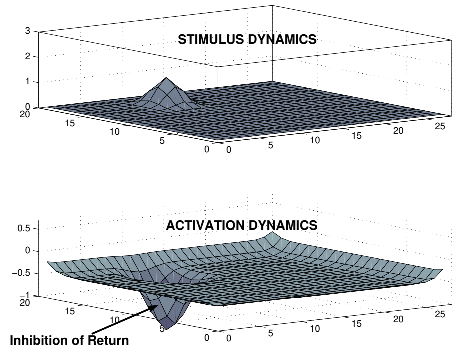


Figure 5: A schematic block diagram of the various modules involved in the system for implementing overt visual attention



(a)



(b)

Figure 6: A snap shot of the stimulus and activation dynamics just (a) before and (b) after the saccade

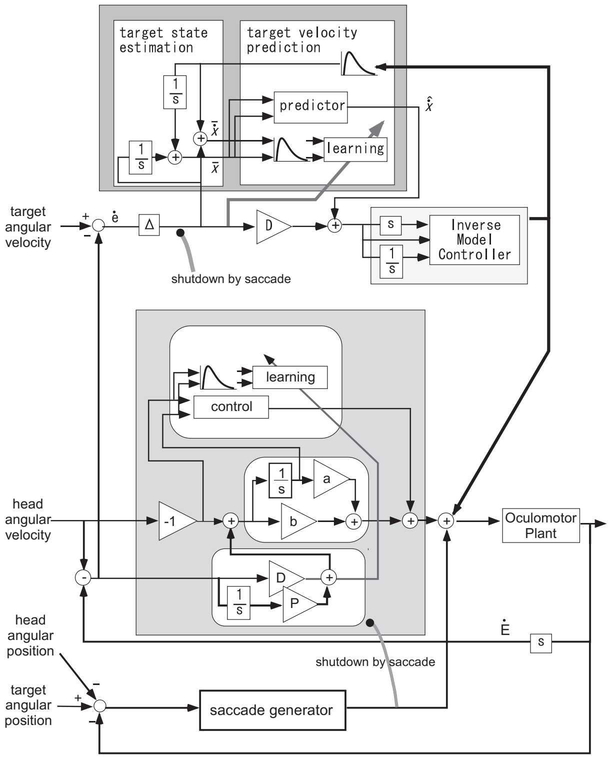


Figure 7: Integrated model of the VOR-OKR, smooth pursuit and saccade. Motor commands out of the VOR-OKR and smooth pursuit are summed. The saccade module corrects the positional errors periodically.

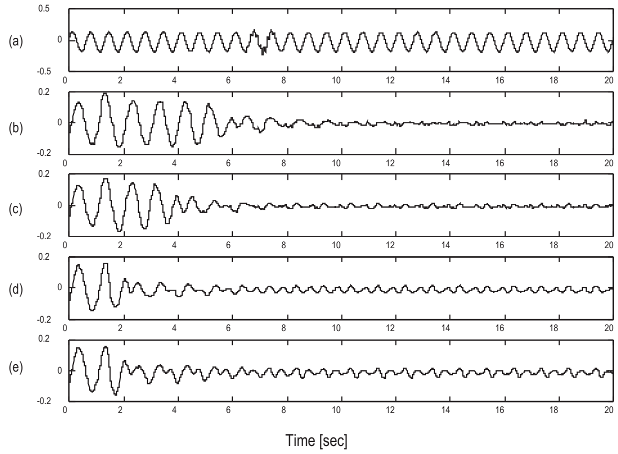


Figure 8: (a) Perturbation signal to the head; (b)-(e) Time course of the retinal error of all four DOFs of the eyes, (b) left pan, (c) right pan, (d) left tilt, (e) right tilt

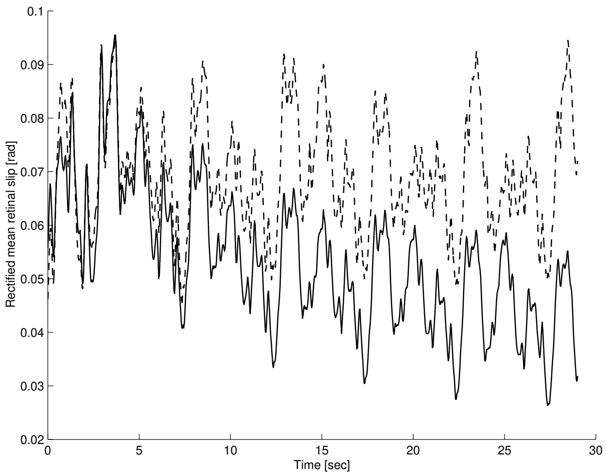


Figure 9: Time course of the rectified mean retinal error; with (solid line) and without (dashed line) eligibility trace

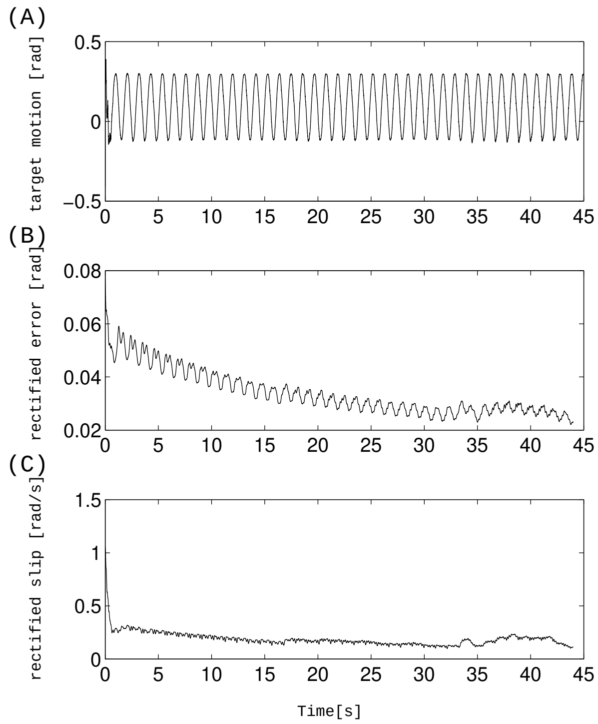


Figure 10: Left eye tracking a 2D sinusoidal motion. (a) Time course of the visual target motion, (b) Time course of the rectified mean retinal error; (c), Time course of the rectified mean retinal slip

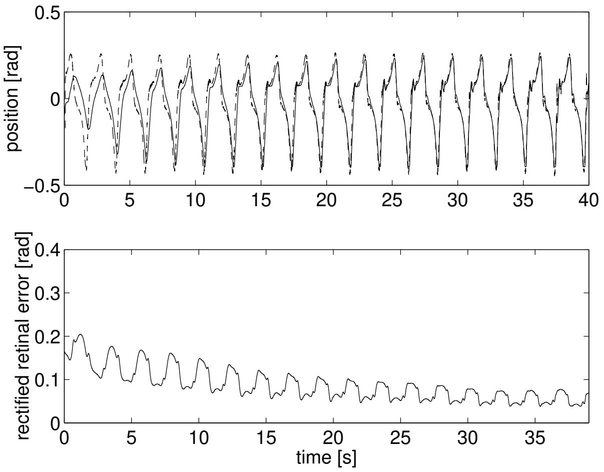


Figure 11: Smooth pursuit of a target following a trajectory that is generated by van der Pol equation. The upper figure shows the time course of the angular position of the visual target (dotted) and the eye (solid). The lower figure presents the time course of the rectified mean retinal error (smoothed with moving average of time window 1 s)

Figure 12: Snap shots of the robot's peripheral view before, during, and after an attentional head-eye saccade, taken at 30 Hz sampling rate. Superimposed on the images is the visual flow field.

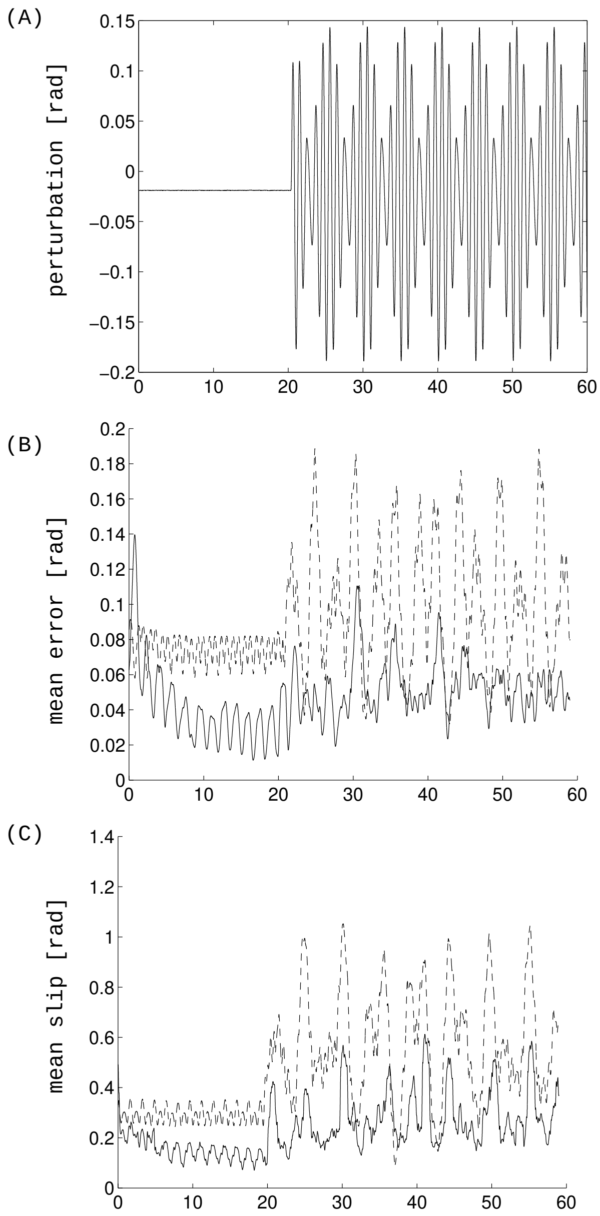


Figure 13: Experimental results demonstrating the effect of coordinating the three oculomotor behaviors: (a) time course of the perturbation. (b) the rectified mean retinal errors; the solid line corresponds to the case of three behaviors cooperating, and the dotted line corresponds the case of no oculomotor control. (c) the rectified mean retinal slips; the solid line corresponds to the case of three behaviors cooperating, and the dotted line corresponds to the case of no oculomotor control.