Noname manuscript No.
(will be inserted by the editor)

A Neural Model of Coordinated Head and Eye Movement Control

Wasif Muhammad · Michael W. Spratling

Received: date / Accepted: date

Abstract Gaze shifts require the coordinated movement of both the eyes and the head in both animals and humanoid robots. To achieve this the brain and the robot control system needs to be able to perform complex non-linear sensory-motor transformations between many degrees of freedom and resolve the redundancy in such a system. In this article we propose a hierarchical neural network model for performing 3-D coordinated gaze shifts. The network is based on the PC/BC-DIM (Predictive Coding/Biased Competition with Divisive Input Modulation) basis function model. The proposed model consists of independent eyes and head controlled circuits with mutual interactions for the appropriate adjustment of coordination behaviour. Based on the initial eyes and head positions the network resolves redundancies involved in 3-D gaze shifts and produces accurate gaze control without any kinematic analysis or imposing any constraints. Furthermore the behaviour of the proposed model is consistent with coordinated eye and head movements observed in primates.

Keywords basis function network · sensory-sensory transformation · sensory-motor transformation · function approximation · eye-head gaze shift · redundancy

Wasif Muhammad
Department of Informatics
King's College London
E-mail: wasif.muhammad@kcl.ac.uk

Michael W. Spratling
Department of Informatics
King's College London

1 Introduction

Coordinated eye-head gaze[1] shifts to targets of interest are very common in humans and other animals. Such coordinated movements may be required when the target of interest appears in the peripheral visual field [15] or outside of the oculomotor range [56]. In such cases visual sensory information brings forth well organized and coordinated actions in 3-D eye and head motor spaces. Two important questions are: how does this sensory information drive eye and head movements in different motor spaces and how much do both contribute to shift gaze when the head is unrestrained. These questions have been intensively investigated with restrained and unrestrained head in three species *i.e.*, human [27,40,60,59,32,29,10,13,38,1,14,18,23,65,64,12], monkey [3,57,58,56,9,28,8,39] and cat [16,17,34,35,37,2,54].

The visual sensory information about a target in the 3-D world goes through a complex sensory-motor transformation in order to shift gaze. The initial eyes and head position coupled with visual sensory information has an important role in coordinated eye-head gaze shifts [9,8,28], hence the sensory-motor transformation has to incorporate an efferent copy of the eyes and head positions. Therefore transformation of 4-D binocular retinal information while integrating 9-D efferent copy of eyes and head position (*i.e.*, vertical, horizontal and torsional (about line of sight) components) produces required action in 9-D eyes and head motor space for each gaze shift. Furthermore, for both eyes and the head this sensory-motor transformation is inherently non-linear in nature [21] because of entailing non-linear eyes and head motor plants [63, 62,61].

The eye and head can contribute infinite many possible ways to shift gaze to a target of interest. For example, if the target of interest is at 50° to the right of visual axis a coordinated movement of both head and eye to foveate this target can be achieved with an eye+head contribution of 20°+30° or 9.91°+40.09° or 55.5°-5.5° and so on. Furthermore 3-D gaze shifts to visual targets are highly redundant because of human head torsional redundancy [4,22,5]. The results obtained from primates always showed a lawful relationship between eye and head gaze contributions [12,8,4,9] while resolving redundancies involved in each gaze shift.

1.1 Previous Work

There are several studies in the literature for endowing humanoid robots with the ability to perform coordinated eye-head gaze shifts. These works have used a diversity of approaches to tackle the problem.

Takanishi and colleagues suggested an eye-head gaze control system employing trigonometric transformation of the target visual information from the eye to head coordinates based on the target depth information perceived from

[1] Gaze is defined as the position of visual axis in space calculated by adding eye position relative to head (E) and head position relative to space (H) [18].

both eyes [53]. In [43], Shibata and colleagues developed a biomimetic gaze shift model based on fifth-order splines to approximate the generation of human similar gaze shift trajectories. The visual information was transformed from eye-centred coordinates to eyes and head joint angles using Liegeois' pseudo-inverse with optimization. In another work a coordinated eye-head control system was developed by Maini and colleagues, where the head movement was controlled by a PID position controller having a trapezoidal velocity profile whereas the eyes position was controlled by a velocity control algorithm [26]. Srinivasa and Grossberg described a self-organizing network for coordinated eyes-head movements to visual targets. A linear kinematic transformation was used to transform the eye-centred information to neck coordinates. The proposed network had the ability to exploit the inherent eyes and head redundancies to exhibit robust performance and to overcome disturbances and constraints which had not been encountered during training [51]. Lopes and colleagues formulated a state-space control framework using proprioceptive feedback for coordinating eye-head movements during target tracking [25]. Kardamakis and Moschovakis employed an optimal control method to simulate independent controlled eye and head circuits for coordinated eye-head gaze shifts. A minimum effort rule, employing short duration for the eye and head movements to optimally select the eye and head control signals, was used for the organization of eye-head gaze shifts. The eye and head control signals kept the movement duration of both as short as possible while minimizing the squared sum magnitude of these motor commands [20]. Saeb and colleagues proposed an open-loop feedforward neural model that combined an optimality principle and incremental local learning mechanism into a unified control scheme for coordinated eye and head movements [42]. In another work a gaze control system was developed based on the adaptive Kalman filter for target tracking by Milighetti and colleagues. The robot head redundancy was used with the weighted pseudo-inverse of the task Jacobian while involving local optimization criteria for human-like motion to shape the inverse kinematic solution [31]. Law and colleagues described a biologically constrained developmental learning model for eye-head gaze control [24].

Earlier work in robotics for coordinated eye-head gaze shift is either using hard-coded kinematic transformation [51,25,26,43,31,53,20] or transformation with a neural model [42] but using linear eye and head plants and a reduced number of degrees of freedom (DOFs) in motor space or constrained transformation [24] to simplify the complexity. However, all previous work either did not consider at all eye-head contribution and head torsional redundancies [51,24–26,43,31,53] or resolved eye-head contribution redundancy using an empirical gaze optimization procedure [20] or with an action cost optimization procedure [42] but without head torsional redundancy.

1.2 Our Solution

In [33] we built a three stage PC/BC-DIM basis function network to control eye movements. In this article we extend that network using another PC/BC-DIM stage to also control head movements. The resulting model can perform non-linear transformation of visual sensory information to redundant DOFs motor space while resolving eyes-head coordination redundancy. The proposed model is an independent eyes and head controlled forward neural network with interacting eyes and head control circuits similar to recent biological models [39, 8, 7]. To demonstrate this model, 3-D coordinated eye-head gaze shift tasks were performed with the iCub humanoid robot simulator having 7 DOFs for binocular eyes and head motor spaces. Specifically, we showed that this new method can be used to learn a hierarchy of basis function-like networks for transforming retinotopic sensory information into a head-centred and finally to a body-centred representation of visual space. We further showed that this transformed body-centred representation can be used for control of coordinated eye-head movements to shift gaze and to bring salient visual information onto the most sensitive part of the binocular retina called the fovea. The foremost advantage of the proposed model is to provide a biologically plausible 3-D eyes-head coordination model for humanoid robots. In this work eyes-head coordination and head torsional redundancies were resolved in a biological similar way without involving any optimization procedure, constraint or kinematic analysis. To our knowledge this is the pioneer work for 3-D coordinated eye-head gaze shift in robotics and which showed biological similar results without involving any pre or post kinematic analysis or imposing any constraints or using any optimization technique.

2 Methods

2.1 The PC/BC-DIM Algorithm

All experiments reported here were performed using the PC/BC-DIM algorithm. PC/BC-DIM is a version of Predictive Coding (PC) [41] reformulated to make it compatible with Biased Competition (BC) theories of cortical function [44, 45] and that is implemented using Divisive Input Modulation (DIM) [50] as the method for updating error and prediction neuron activations. DIM calculates reconstruction errors using division, which is in contrast to other implementations of PC that calculate reconstruction errors using subtraction [19]. PC/BC-DIM is a hierarchical neural network. Each level, or processing stage, in the hierarchy is implemented using the neural circuitry illustrated in Fig. 1a. A single PC/BC-DIM processing stage thus consists of three separate neural populations. The behaviour of the neurons in these three populations is determined by the following equations:

$$\mathbf{r} = \mathbf{V}\mathbf{y} \tag{1}$$

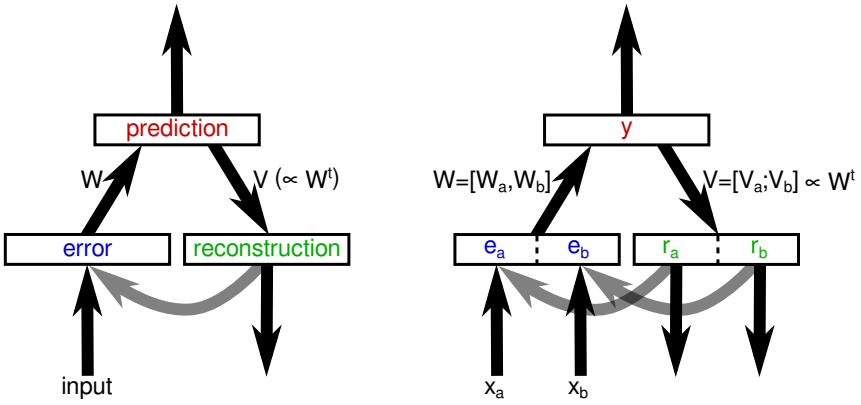


Fig. 1 (a) A single processing stage in the PC/BC-DIM neural network architecture. Rectangles represent populations of neurons and arrows represent connections between those populations. The population of prediction neurons constitute a model of the input environment. Individual neurons represent distinct causes that can underlie the input (*i.e.*, latent variables). The belief that each cause explains the current input is encoded in the activation level, $\mathbf{y}$, and is used to reconstruct the expected input given the predicted causes. This reconstruction, $\mathbf{r}$, is calculated using a linear generative model (see equation 1). Each column of the feedback weight matrix $\mathbf{V}$ represents an "elementary component", "basis vector", or "dictionary element", and the reconstruction is thus a linear combination of those components. Each element of the reconstruction is compared to the corresponding element of the actual input, $\mathbf{x}$, in order to calculate the residual error, $\mathbf{e}$, between the predicted input and the actual input (see equation 2). The errors are subsequently used to update the predictions (via the feedforward weights $\mathbf{W}$, see equation 3) in order to make them better able to account for the input, and hence, to reduce the error at subsequent iterations. The responses of the neurons in all three populations are updated iteratively to recursively calculate the values of $\mathbf{y}$, $\mathbf{r}$, and $\mathbf{e}$. The weights $\mathbf{V}$ are the transpose of the weights $\mathbf{W}$, but are normalised with the maximum value in each column. The activations of the prediction neurons or the reconstruction neurons may be used as inputs to other PC/BC-DIM processing stages. The inputs to this processing stage may come from the prediction neurons of this or another processing stage, or the reconstruction neurons of another processing stage, or may be external, sensory-driven, signals. The inputs can also be a combination of any of the above. (b) When inputs come from multiple sources, it is convenient to consider the population of error neurons to be partitioned into sub-populations which receive these separate sources of input. As there is a one-to-one correspondence between error neurons and reconstruction neurons, this means that the reconstruction neuron population can be partitioned similarly.

$$\mathbf{e} = \mathbf{x} \oslash (\epsilon_2 + \mathbf{r}) \tag{2}$$

$$\mathbf{y} \leftarrow (\epsilon_1 + \mathbf{y}) \otimes \mathbf{We} \tag{3}$$

Where $\mathbf{x}$ is a (m by 1) vector of input activations, $\mathbf{e}$ is a (m by 1) vector of error neuron activations; $\mathbf{r}$ is a (m by 1) vector of reconstruction neuron activations; $\mathbf{y}$ is a (n by 1) vector of prediction neuron activations; $\mathbf{W}$ is a (n by m) matrix of feedforward synaptic weight values; $\mathbf{V}$ is a (m by n) matrix of feedback synaptic weight values; ϵ_1 and ϵ_2 are parameters; and $\oslash$ and $\otimes$ indicate element-wise division and multiplication respectively. For all the experiments described in this paper ϵ_1 and ϵ_2 were both given the value 1×10^{-9}. The value of both parameters has little influence on the results but

the same values were also used in previous work [33]. Parameter ϵ_1 prevents prediction neurons becoming permanently non-responsive. It also sets each prediction neuron's baseline activity rate and controls the rate at which its activity increases when an input stimulus is presented within its receptive field (RF). Parameter ϵ_2 prevents division-by-zero errors and determines the minimum strength that an input is required to have in order to effect prediction neuron response. As in all previous work with PC/BC-DIM, these parameters have been given small values compared to typical values of $\mathbf{y}$ and $\mathbf{x}$, and hence, have negligible effects on the steady-state activity of the network. The matrix $\mathbf{V}$ is equal to the transpose of the $\mathbf{W}$, but each column is normalised to have a maximum value of one. Hence, the feedforward and feedback weights are simply rescaled versions of each other. Given that the $\mathbf{V}$ weights are fixed to the $\mathbf{W}$ weights there is only one set of free parameters, $\mathbf{W}$, and references to the "synaptic weights" refer to the elements of $\mathbf{W}$. Here, as in previous work with PC/BC-DIM only non-negative weights, inputs, and activations are used. Initially the values of $\mathbf{y}$ are all set to zero, although random initialisation of the prediction node activations can also be used with little influence on the results. Equations 1, 2 and 3 are then iteratively updated with the new values of $\mathbf{y}$ calculated by equation 3 substituted into equation 1 and 3 to recursively calculate the neural activations. This iterative process was terminated after 150 iterations in all the experiments reported here.

The values of $\mathbf{y}$ represent predictions of the causes underlying the inputs to the network. The values of $\mathbf{r}$ represent the expected inputs given the predicted causes. The values of $\mathbf{e}$ represent the residual error between the reconstruction, $\mathbf{r}$, and the actual input, $\mathbf{x}$. The full range of possible causes that the network can represent are defined by the weights, $\mathbf{W}$ (and $\mathbf{V}$). Each row of $\mathbf{W}$ (which correspond to the weights targeting an individual prediction neuron) can be thought of as a "basis vector" or "elementary component" or "preferred stimulus", and $\mathbf{W}$ as a whole can be thought of as a "dictionary" or "codebook" of possible representations, or a model of the external environment. The activation dynamics described above result in the PC/BC-DIM algorithm selecting a (typically sparse) subset of active prediction neurons whose RFs (which correspond to basis functions) best explain the underlying causes of the sensory input. The strength of activation reflects the strength with which each basis function is required to be present in order to accurately reconstruct the input. This strength of response also reflects the probability with which that basis function (the preferred stimulus of the active prediction neuron) is believed to be present, taking into account the evidence provided by the input signal and the full range of alternative explanations encoded in the RFs of the whole population of prediction neurons.

When inputs come from multiple sources it is convenient to consider the vector of input signals, $\mathbf{x}$, the vector of error neuron activations, $\mathbf{e}$, and the vector of reconstruction neuron responses, $\mathbf{r}$, to be partitioned into multiple parts corresponding to these separate sources of input (see Fig. 1b). Each partition of the input will correspond to certain columns of $\mathbf{W}$ (and rows of $\mathbf{V}$). While it is conceptually convenient to think about separate partitions of the

inputs, neural populations and synaptic weights, it does not in any way alter the mathematics of the model. In equations 1, 2 and 3, $\mathbf{x}$ is a concatenation of all partitions of the input, $\mathbf{e}$ and $\mathbf{r}$ represent the activations of all the error and reconstruction neurons; and $\mathbf{W}$ and $\mathbf{V}$ represent the synaptic weight values for all partitions.

2.2 Performing Transformations with a PC/BC-DIM Network

As described above, the prediction neurons in a PC/BC-DIM network behave like basis function neurons. Fig. 1b illustrates how this can be exploited to perform a simple mapping from two input variables to an output variable by the basis function network. If a sub-set of the prediction neurons represent combinations of inputs that correspond to the same value of the output, then it is necessary to "pool" the responses from this sub-set of prediction neurons to produce this output whenever one of these combinations is presented to the inputs. Figure 2 shows two ways in which this can be implemented. The first method (Fig. 2a) involves using a separate population of pooling neurons that are activated by the responses of the prediction neurons. This method has been used in previous work [46,47] and is directly equivalent to a standard basis function network. The second method (Fig. 2b) used in [33,49,48] involves defining additional neurons within the reconstruction neuron population that perform the same role as the pooling neurons in the first method. In this article the second method will be used.

2.3 The Proposed PC/BC-DIM Network for Eyes-Head Coordination

To shift gaze the proposed network model utilizes a sequence of eye-head sensory-sensory and sensory-motor transformations as the eye-head coordination strategy. To demonstrate how this strategy is implemented in a PC/BC-DIM network, a 1-D eye-head coordination network is used for simplicity. This network is shown in Fig. 3. The function of the PC/BC-DIM network shown in Fig. 3 is demonstrated with mappings between four variables in Fig. 4. Its function in performing the proposed eye-head coordination strategy is illustrated in Fig. 5.

To implement the eye-head coordination strategy, sensory-sensory and sensory-motor mappings were performed in five steps. In the first step, the retina-centred information about the visual target coupled with the current eye position was provided as input to the first processing stage to perform a sensory-sensory transformation in order to produce a head-centred representation. This head-centred information was provided as input to the second processing stage along with the current head position to perform another sensory-sensory transformation in order to produce a body-centred representation. In the second step, retinal foveal activity and the body-centred representation were used as input to perform a sensory-motor transformation to determine the value of the

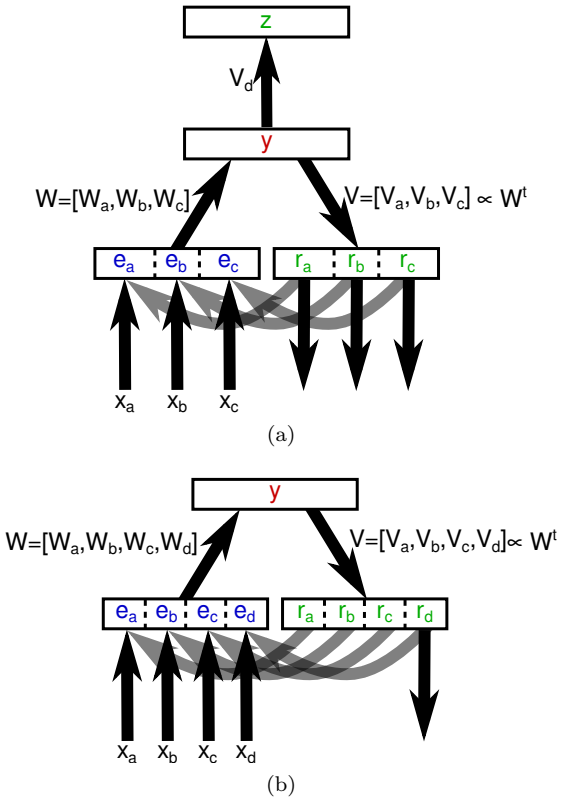


Fig. 2 Methods of using PC/BC-DIM as a basis function network. For the simple task of mapping from three input variables ($\mathbf{x}_a$, $\mathbf{x}_b$ and $\mathbf{x}_c$) to an output variable ($\mathbf{x}_d$). (a) The prediction neurons have RFs in the three input spaces (defined by the weights $\mathbf{W}_a$, $\mathbf{W}_b$ and $\mathbf{W}_c$) that make them selective to specific combinations of input stimuli. A population of pooling neurons receives input, via weights $\mathbf{V}_d$, from the prediction neurons in order to generate the output. The responses of the pooling neurons, $\mathbf{z}$, are calculated as a linear weighted sum of their input, *i.e.*, $\mathbf{z} = \mathbf{V}_d\mathbf{y}$. (b) The PC/BC-DIM network receives an additional source of input. Dealing with this extra partition of the input requires the definition of additional columns of feedforward synaptic weights, $\mathbf{W}$, and additional rows of the feedback weights, $\mathbf{V}$. If the additional feedback weights, $\mathbf{V}_d$, are identical to the pooling weights used in the architecture shown in (a), then (given equation 1), the responses of the fourth partition of the reconstruction neurons, $\mathbf{r}_d$, will be identical to the responses of the pooling neurons in (a), *i.e.*, $\mathbf{r}_d = \mathbf{V}_d\mathbf{y}$. If the feedforward weights associated with the fourth partition, $\mathbf{W}_d$, are rescaled versions of the corresponding additional feedback weights, $\mathbf{V}_d$, then the network can perform mappings not only from $\mathbf{x}_a$, $\mathbf{x}_b$ and $\mathbf{x}_c$ to $\mathbf{x}_d$, but also from $\mathbf{x}_a$, $\mathbf{x}_b$ and $\mathbf{x}_d$ to $\mathbf{x}_c$, and from $\mathbf{x}_a$, $\mathbf{x}_c$ and $\mathbf{x}_d$ to $\mathbf{x}_b$, and from $\mathbf{x}_b$, $\mathbf{x}_c$ and $\mathbf{x}_d$ to $\mathbf{x}_a$ (see Fig. 4).

eye position required to shift gaze towards the target. Using this eye position value, the eye performed a saccade. In the third step, retinal foveal activity, the eye position determined in the previous step and the body-centred representation were used as inputs to perform another sensory-motor transformation to determine the head contribution to the gaze shift. Using this motor command

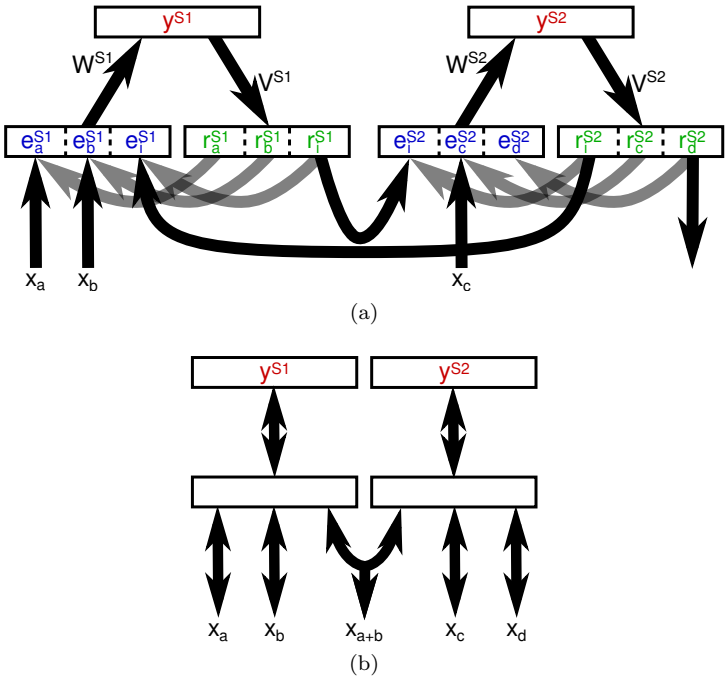
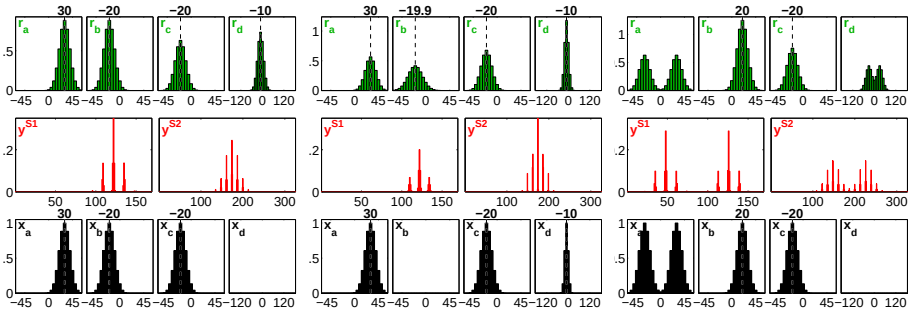


Fig. 3 A hierarchical architecture, consisting of two interconnected PC/BC-DIM networks. (a) A hierarchical architecture, consisting of two interconnected PC/BC-DIM networks, for calculating the same function as shown in Fig. 2. The first network calculates an intermediate result ($\mathbf{x}_{a+b}$) in the third partition of it's reconstruction neurons. This intermediate result provides an input to the second PC/BC-DIM network. The second network's reconstruction of this intermediate representation is fed-back as input to the first PC/BC-DIM network. (b) By superimposing error and reconstruction neurons the network in Fig. 3a can be shown in a simplified format. The network can be used for 1-D coordinated eye-head gaze shift. For 1-D eye-head coordination, 1-D retina-centred information is transformed to 1-D head-centred and then to 1-D body-centred information. The network calculates $\mathbf{x}_d$ (*i.e.*, body-centred representation) given $\mathbf{x}_a$ (*i.e.*, 1-D retina-centred representation), $\mathbf{x}_b$ (*i.e.*, 1-D eye position), and $\mathbf{x}_c$ (*i.e.*, 1-D head position). The first PC/BC-DIM network calculates an intermediate result ($\mathbf{x}_{a+b}$) in the third partition of it's reconstruction neurons as head-centred representation. This intermediate result *i.e.*, head-centred provides an input to the second PC/BC-DIM network. The second network's reconstruction of this intermediate representation is fed-back as input to the first PC/BC-DIM network. This hierarchical mapping of the network is shown in Fig. 5.

the head was moved. At the end of these movements the eye position relative to target in space could be incorrect. The fourth and fifth steps were used to correct the eye position. To approximate the correct eye position in the head, in the fourth step a sensory-sensory transformation was performed to update the head-centred representation using the updated retinal activity after the previous gaze shift. In the fifth step, a sensory-motor transformation was performed with retinal foveal activity, updated head-centred representation and the body-centred representation as input to determine the correct eye position. The input-output mapping of the 1-D eye-head coordination strategy is



(a) (b) (c)

Fig. 4 Mapping between four variables using the two-stage (hierarchical) PC/BC-DIM network illustrated in Fig. 3. The PC/BC-DIM network has been wired-up to approximate the function $\mathbf{x}_d = \mathbf{x}_a + \mathbf{x}_b + \mathbf{x}_c$. In each sub-figure the lower histograms show the inputs, the middle histograms show the prediction neuron activations, and the upper histograms show the reconstruction neuron responses. The x-axis of each histogram is labelled with the variable value, except for the histogram representing the prediction neuron responses which is labelled by neuron number. The y-axes of each histogram are in arbitrary units representing firing rate. (a) When the three inputs representing $\mathbf{x}_a$, $\mathbf{x}_b$, and $\mathbf{x}_c$ are presented (lower histograms), the reconstruction neurons generate an output (upper histograms) that represents the correct value of $\mathbf{x}_d$ (as well as outputs representing the given values of $\mathbf{x}_a$, $\mathbf{x}_b$, and $\mathbf{x}_c$). (b) When the three inputs representing $\mathbf{x}_a$, $\mathbf{x}_c$ and $\mathbf{x}_d$ are presented (lower histograms), the reconstruction neurons generate an output (upper histograms) that estimates the correct value of $\mathbf{x}_b$ (as well as outputs representing the given values of $\mathbf{x}_a$, $\mathbf{x}_c$ and $\mathbf{x}_d$). (c) As (a) but with two values of $\mathbf{x}_a$ represented by a bi-modal input to the first partition. The network correctly calculates two values for $\mathbf{x}_d$ represented by the peaks of the bi-modal distribution produced by the reconstruction neurons in the last partition.

demonstrated in Fig. 5. The determined head position in the step three and the computed eye position relative to head in the step five resolved the kinematic redundancy involved in the eye-head system and chose one gaze plan to move the eye and head towards the visual target.

The PC/BC-DIM 3-D eyes-head coordination network model shown in Fig. 6 uses four processing stages of the PC/BC-DIM neural model to learn body-centred representation of visual space. The proposed network is shown in the simplified format used in Fig. 3b. The mathematical model remains unchanged. The proposed eyes-head coordination network contains a PC/BC--DIM processing stage (shown on the left of Fig. 6) that performs mappings between the position of a visual target on the left retina, the position of the left eye in the skull (the left eye pan and tilt), and the head-centred location of the left-eye visual target. An identical PC/BC-DIM processing stage, shown at the second position of Fig. 6, performs the same transformations for the right eye. A third PC/BC-DIM processing stage translates between the individual head-centred representation centred on the left and right eyes, and a global head-centred representation of visual space, that can be driven by targets viewed by either or both eyes. The fourth and the last processing stage in Fig. 6 uses the global head-centred representation and an efferent copy of the head position (*i.e.*, the head pan, tilt and swing) as inputs to produce the

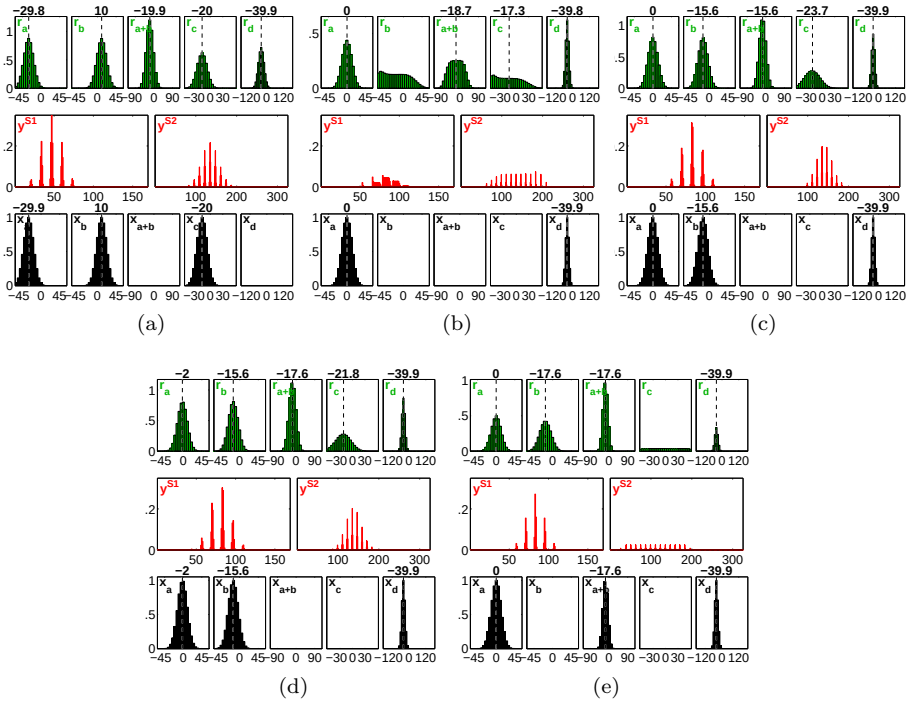


Fig. 5 Example mappings performed by the 1-D hierarchical PC/BC-DIM network shown in Fig. 3 to implement the eye-head coordination strategy. The black histograms in each sub-plot show the input provided to the network whereas the red histograms show response of prediction neurons activations and the green histograms show response of reconstruction neurons. (a) The population coded input was provided at $\mathbf{x}_a$ (*i.e.*, 1-D retina-centred input), $\mathbf{x}_b$ (*i.e.*, 1-D eye position) and $\mathbf{x}_c$ (*i.e.*, 1-D head position) to approximate $\mathbf{x}_{a+b}$ (*i.e.*, 1-D head-centred representation) in first stage and $\mathbf{x}_d$ (*i.e.*, 1-D body-centred representation) in second stage as shown in upper histogram. The intermediate result propagated between two network stages, shown with curved arrow in Fig. 3b, represents the 1-D head-centred representation. (b) Using retina foveal activity $\mathbf{x}_a$ (*i.e.*, peak centred at zero) and learnt body-centred representation $\mathbf{x}_d$, the eye position $\mathbf{x}_b$ relative to target in space was computed. (c) The retina foveal activity $\mathbf{x}_a$, eye position $\mathbf{x}_b$ computed in previous step and body-centred representation $\mathbf{x}_d$ were provided as input to compute head position $\mathbf{x}_c$ relative to target in space. Using the eye position $\mathbf{x}_b$ and head position $\mathbf{x}_c$ gaze was shifted. (d) The eye-head gaze shift in (c) changed the position of eye relative to target in head so peak of retinal activity will not be centred at fovea, therefore correction was required to correct position of eyes relative to target in head. Using current updated retinal activity $\mathbf{x}_a$ and current eye position $\mathbf{x}_b$ new head-centred representation $\mathbf{x}_{a+b}$ was computed as shown in mapping. (e) Then using this head-centred representation $\mathbf{x}_{a+b}$ and retina foveal activity $\mathbf{x}_a$ as input, correct eye position in head $\mathbf{x}_b$ was produced by the network.

body-centred representation of visual targets. The same eye-head coordination strategy was used in the 3-D PC/BC-DIM eyes-head coordination network as described for the 1-D case however now sensory-sensory and sensory-motor mappings were performed using 2-D retinal activities and a 2-D efferent copy of both eyes positions and 3-D head orientation. Moreover, a corrective saccade was initialized when retinal activation at each fovea was less than 0.8 times the

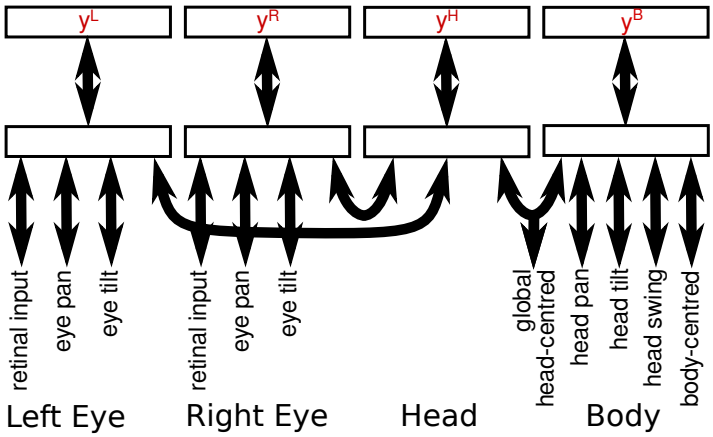


Fig. 6 The hierarchical PC/BC-DIM network for 3-D eye-head coordination drawn using the simplified format.

maximum retinal activity. Since the body of the robot was stationary the body-centred representation was used as a measure of visual target location in space. The determined head position relative to the target in space in step three (as describe above) resolved the torsional redundancy in the head system. It chose one gaze plan to move the head towards the visual target which also resolved the redundancy in terms of head position in the gaze shift plan. The fifth step resolved the remaining redundancy of eyes-head system in terms of eyes position in head. The detailed results of redundancy resolution are shown in the result section. The eyes position in space and approximated head position in space were both predicted based on binocular foveal activity as input for the same body-centred representation. Therefore, binocular retina foveal activity was used as a key input to resolve redundancy by bringing the visual target near the horizontal axis of both eyes as is the case in primates and felines [54].

The retinal input (*i.e.*, $\mathbf{x}_a$) to both the first and second processing stages was encoded using a 2-dimensional uniform array of neurons with Gaussian RFs as used in [33]. For a given visual target, the responses of each retinal neuron was proportional to the overlap of the visual target with its receptive field. These responses were concatenated into a vector to provide the input to the PC/BC-DIM network.

For the purpose of the simulations reported in section 3 the retinotopic input to the model, the input encoded by the retinal neurons described above, are images captured from the iCub cameras. However, the environment in which the iCub is placed is very impoverished consisting of one highly salient object in front of a blank background. In more realistic environments, it would be necessary to process the raw images to derive a retinotopically organised representation to act as the input to the model. This retinotopic input would encode the locations of targets for possible saccades. It is assumed that this could be achieved by processing the images to form a saliency map [36].

The eye position signals *i.e.*, the eye pan and the eye tilt for both eyes and the head position signals *i.e.*, head pan, head tilt and head torsion/swing were each encoded using a 1-dimensional array of neurons with Gaussian RFs that were uniformly distributed between the maximum and minimum values. Decoding these values was performed using standard population vector decoding to find the mean of the distribution of responses [11].

2.4 Training the Eyes-Head Coordination Control PC/BC-DIM Network

The 1-D eye-head coordination network used above to illustrate how PC/BC-DIM can perform simple mappings (*i.e.*, the results shown in Fig. 5) was hard-wired to perform the eye-head coordination strategy. To shift coordinated gaze with the 3-D PC/BC-DIM eyes-head coordination network and to perform more complex or unknown mappings requires some method of learning the appropriate connectivity. Previous work has shown that this can be achieved using unsupervised learning [46,6]. However, this learning procedure is slow and rather impractical. A faster, but biologically implausible, procedure is used in this work for training the weights. The same method was used in [33].

The first three processing stages in Fig. 6 were trained to learn head-centred representations of visual targets as described in [33]. The fourth processing stage was trained to learn body-centred representation of visual space. The head-centred representation of a visual target was determined using the eye-centred representation of the visual target and the efferent copy of eyes position by transformation with the eye control network. In the next stage, the head-centred representation was combined with the efferent copy of head position to determine the body-centred representation of the visual target. Therefore for training of the fourth processing stage, each training set was framed with the head-centred representation of the visual target and the efferent copy of head orientation. A single, stationary, visual target was presented to the static body iCub robot. With both eyes at their rest positions (*i.e.*, eyes pan and tilt 0°), the head was moved systematically to generate distinct combinations of head pan, tilt and swing and retinal inputs that corresponded to the same body-centred target position. Using retinal and eyes position information the global head-centred representation was obtained. The global head-centred information and head pan, tilt and swing values were represented by different prediction neurons in the fourth PC/BC-DIM processing stage. Each of these prediction neurons was also connected to a single reconstruction neuron in the fifth partition of the fourth processing stage which represents that body-centred location. Having trained the network to represent one body-centred location, the visual target was moved to another location and this training procedure was repeated. Repeating this process systematically for a range of different target positions enabled the fourth processing stage of the PC/BC-DIM network to learn body-centred representations of visual space. The PC/BC-DIM eyes-head coordination network was trained with redundancy in eyes-head gaze shift plans and redundancy in head torsional values

since for one body-centred location all head poses were used to learn the body-centred representation.

One issue with the above method is to decide on how many positions to place the visual target during training. Clearly the target needs to appear over the full range of positions that the robot needs to learn. However, how finely does this grid of possible locations need to be sampled? Too fine a sampling will lead to a network with an excess of prediction neurons and fifth partition reconstruction neurons. A second issue is to decide how many head movements the robot needs to make to learn about one body-centred location. Again, it is clearly necessary for the head movements to cover the full range of possible head positions, but how finely does this range need to be sampled? Too fine a sampling will lead to a network with an excess of prediction neurons. To address these issues the following procedure was used. As a body-centred visual target appears in visual field (*i.e.*, in monocular or binocular view) with certain head orientation (*i.e.*, with certain value of head pan, tilt and swing) the network initially does not learn this location but in fact performs a sensory-sensory mapping in order to estimate of the body-centred location of the visual target (as described in section 2.3). The PC/BC-DIM network was then used to perform a sensory-motor mapping in order to calculate the eye and head motor commands (as described in section 2.3) required to bring the visual target into the retina of both eyes. These movements were performed. If successful, the target would now be in the centre of both eyes, and no learning was performed. If unsuccessful and the target was not in view of both eyes, then the network was trained so that it would be able to perform these sensory-sensory and sensory-motor transformations in the future. If the visual target was at a new body-centred location, then a new reconstruction neuron was added to the fifth partition of forth processing stage, otherwise the body-centred location was already associated with a fifth partition reconstruction neuron. The vector providing input to the fifth partition (*i.e.*, $\mathbf{x}_e$) was set to all zeros, except for the single element corresponding the fifth partition reconstruction neuron representing the current body-centred location, which was given a value of one. A new prediction neuron was added to the network. This prediction neuron was given weights corresponding to the inputs received by the first four partitions prior to the movement and the newly calculated input to the fifth partition. Specifically, a new row of $\mathbf{W}$ was created and set equal to $[\tilde{\mathbf{x}}_a; \tilde{\mathbf{x}}_b; \tilde{\mathbf{x}}_c; \tilde{\mathbf{x}}_d; \tilde{\mathbf{x}}_e]^T$ and a new column of $\mathbf{V}$ was created and set equal to $[\hat{\mathbf{x}}_a; \hat{\mathbf{x}}_b; \hat{\mathbf{x}}_c; \hat{\mathbf{x}}_d; \hat{\mathbf{x}}_e]$ (where $\tilde{\mathbf{x}}$ is equal to $\mathbf{x}$ after it has been normalised to sum to one; and $\hat{\mathbf{x}}$ is equal to $\mathbf{x}$ after it has been normalised to have a maximum value of one).

3 Results

The proposed 3-D eyes-head coordination network was trained and tested in the simulated iCub humanoid robot platform [55,30] with static body, visual targets were boxes (with width, height and length of 0.038) simulated with no

gravity and with a depth range from 0° to 20°. In all experiments each retinal image was encoded using a population of Gaussian RFs of standard deviation 7 pixels uniformly distributed on a rectangular grid such that the spacing between RF centres was 14 pixels, and in total 81 RFs were used to uniformly tile the input image as in [33]. The size of each iCub retinal image was 128x128 pixels, which corresponds to 25.6x26.4 degrees of visual angle. Head pan had a range of -40° to +40°, tilt ranged from -30° to +30° and head swing had a range of -20° to +20° and were varied in steps of 1° during learning. Whereas eye pan had a range of -20° to +20° and tilt ranged from -12° to +12°. The eye and head position signals were encoded with 1-dimensional Gaussian RFs evenly spaced every 4° and with spread 2° as in our previous work [33].

All experiments reported below were performed by following the eyes-head coordination strategy described in section 2.3. The gaze amplitude was calculated from the change in position of the visual axis during gaze shift *i.e.*, from the gaze starting position/gaze onset to the gaze end position/gaze offset. Whereas the eyes and head contribution was calculated using the change in eyes and head position during gaze shift. The proposed eyes-head coordination network is not only capable of shifting saccadic gaze to targets of interest but also performs convergent eyes movement to focus on the target as we have shown for saccade and vergence control in previous work [33].

3.1 Accuracy

To quantitatively measure the gaze accuracy with the iCub simulator, the robot's eyes and head were placed at a random pose, and then a visual target was generated at a random location and depth but so that it was visible to at least one eye. The visual input corresponding to the target, together with the efferent copy of the eye pan/tilt and head pan/tilt/swing positions were used to determine the body-centred representation of the target (see section 2.3). This body-centred information with binocular retina foveal activities was used to compute eye positions needed to foveate the target. Using retina foveal activities, the calculated eyes positions and the computed body-centred representation, the desired head position was also computed. This sequence of calculating eyes and head movements enables the eyes to start moving earlier than the head as in primates [7,60]. After this initial gaze shift, if the binocular retina activities centred at the foveae were less than 0.8 of the maximum then a corrective saccade was performed (see section 2.3). Fig. 7 shows an example simulation with the iCub robot. The post-gaze distance was measured between the foveal locations and position of target in the retinal images for 100 trials. The amplitude of gaze shifts were sorted and grouped in a range of 5° along with respective values of post-gaze error. The mean value of gaze amplitude and the mean and standard deviation of the post-gaze distance in each group was calculated as shown in Fig. 8. The mean value of post-gaze distance was 2.09° and SD was 0.49° which compares to an accuracy for large gaze shifts in primates of $< 3°$ [56].

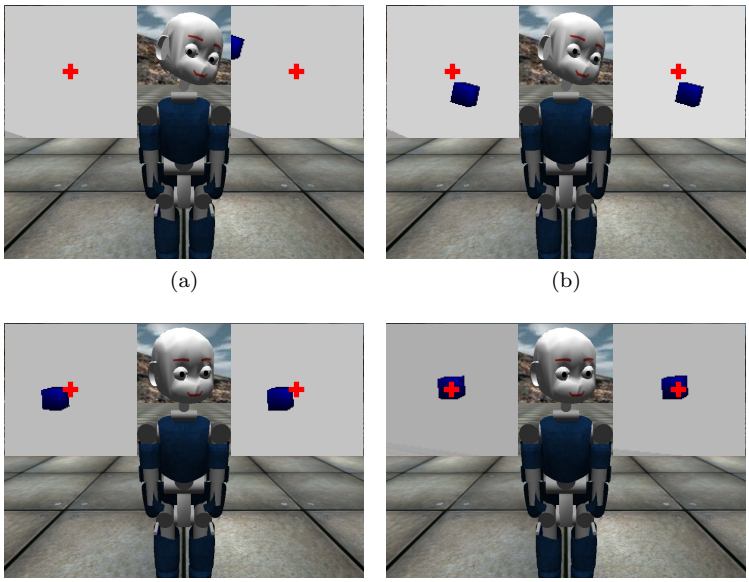


Fig. 7 Example simulation of eyes-head gaze shift. The two windows to the left and right of the iCub show the views of both eyes. The box within these windows is the visual target and the cross hairs mark the location of fovea in middle of retina (the cross hairs were not visible to the robot). (a) Before gaze shift initial pose of eyes and head. (b) After binocular eyes gaze shift. (c) After the head movement. In this example, the head movement causes the target to overshoot the foveal area of binocular vision. (d) After a corrective saccade.

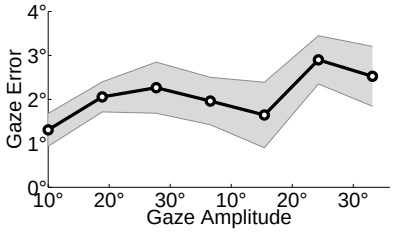


Fig. 8 Gaze accuracy in terms of post-gaze shift error for the trained PC/BC-DIM eyes-head coordination network.

3.2 Effects of gaze direction on large eyes-head gaze shifts

In humans during large gaze shifts head movements contribute more along the horizontal meridian compared to the vertical meridian whereas eye movements contribute more vertically [12]. To measure and quantify the eyes-head gaze contribution during large gaze shifts the following experiment was performed using four visual targets placed at the corners of a square at 40° oblique eccentricity and a fifth target was placed at the center of the square. The experimental procedure as described in [12] was adopted where a random sequence

of gaze shifts between targets were controlled through verbal commands *i.e.*, top-left, bottom-right *etc.*. The robot also performed a random sequence of gaze shifts for 100 trails between visual targets in the square pattern. To imitate verbal directions, at first a sensory-sensory transformation was performed for all targets in the square pattern and the corresponding body-centred representations were recorded. Then random selection was made between these remembered body-centred representations to shift gaze. In this experiment after each gaze offset, right eye in head and head in space motor commands were recorded. The combination of right eye in head vector and head in space vector was defined as right eye in space vector as shown in Fig. 9 with tip of rotation vectors directed through the line of sight. To quantify the relative eye and head contributions to gaze shifts, the ratio of vertical to horizontal (v/h) components of head in space and binocular eyes in head motor commands was calculated. The mean value of (v/h) for head in space was 0.80 with SD=$\pm$0.22, whereas the mean (v/h) ratio for left and right eyes in head was 2.23 (SD=$\pm$1.69) and 2.30 (SD=$\pm$2.45) respectively. These results are consistent with human results for large gaze shifts *i.e.*, the mean (v/h) ratio of head in space was 0.5$\pm$0.11(SD) for 90° eccentric target and 0.54$\pm$0.007(SD) for 70° target whereas mean (v/h) for eye in head was 1.42$\pm$0.27(SD) for 90° eccentric target and 2.51$\pm$0.26(SD) for 70° target [12]. Both the human and the robot results show that the horizontal components of head contribution were large compared to the vertical components while the opposite was true in case of eyes' components. Hence, the head contributed more along the horizontal meridian whereas the eyes contributed more vertically for large oblique gaze shifts. These results confirm a biological similar lawful relationship of eyes and head contributions along the gaze direction. However the resultant head position in space was not so scattered around the locations as was the case for the comparable human results (see Fig. 9).

3.3 Horizontal gaze and eye-head amplitude relationship

In biological studies, the effect of increasing the horizontal gaze shift amplitude on the eyes and the head gaze contribution has been intensively studied in primates [8,18,17,65]. The experimental protocol described in [8] was followed. In [8], eyes and head aligned movements were directed within $\pm$10° along the horizontal meridian in the tangential screen paradigm. The eyes initial position was centred in their orbits (*i.e.*, initial eyes position $\pm$5°). To ascertain the effect of incremental horizontal gaze amplitude on the eyes and head gaze contribution with the proposed network, the tangential screen target paradigm was used. In tangential screen paradigm targets can be placed perpendicular to the line of sight at any location in a 2-D plane subtended horizontally and vertically to $\pm$40°. For the robot experiments, visual targets were displaced along the horizontal meridian such that the target of interest was visible to at least one eye. Results were recorded for movements made when the eyes were initially at the centre of their orbits (*i.e.*, 0°). The robot head was ran-

domly positioned within ±10° range along the horizontal direction (*i.e.*, pan) with no initial motion along the vertical and torsional directions (*i.e.*, tilt and swing were both always kept at 0°). The gaze shifts were performed and onset and offset eyes position in head and head position in space were recorded for all trials. The results obtained from these experiments are shown with the comparable primate results in Fig. 10. The resultant gaze amplitude and head contribution were highly correlated, so that for small gaze amplitudes the head contribution was small and for larger gaze amplitudes the head contribution was large and showed almost linear relationship with large gaze amplitudes. The eye amplitude was also linearly related for small gaze amplitudes, however, for large gaze amplitudes eye amplitude was almost constant. These results are consistent with primates results [8] shown in Fig. 10b.

3.4 Effect of target displacement on movement amplitude

At gaze onset the visual axis and the position of head may not be the same, therefore the target displacement relative to gaze and the target displacement relative to the head will also be different. In this experiment the relationship of target displacement with gaze and head amplitude was investigated. The experimental procedure laid out in [8] was followed. In [8], the relationship between primary gaze shifts (without corrective movements) and displacement of the target relative to the direction of the line of sight (retinal error) was analysed with oblique gaze shifts. Experiments were carried out using the tangential screen paradigm with the oblique target randomly located within eccentricity of ±5° to ±20°. Then the robot eyes were posed at random initial position along the horizontal direction with restrained vertical initial position (*i.e.*, tilt=0°) and the head was positioned at random initial swing/torsion position (*i.e.*, pan=0° and tilt=0°) such that the target was at least visible to one eye. This initial position arrangement of eyes and head ensured that the gaze shift will always be performed in the oblique direction. The relationship between primary gaze shifts (*i.e.*, without corrective saccade) to visual targets and target displacement relative to visual axis (*i.e.*, retina error) directed through left eye was analysed and illustrated in Fig. 11. The first three steps of the eyes-head coordination strategy described in section 2.3 were followed for primary gaze shifts whereas the head and the left eye movement determined in the third and fifth step respectively were used as a measure of target displacement relative to gaze onset position (*i.e.*, gaze shifts with one corrective saccade). The relationship between horizontal and vertical components of gaze shifts' amplitude and horizontal and vertical components of target displacement is shown in Fig. 11. The primary gaze amplitude was better related to the target displacement as the ratio between gaze amplitude to target displacement was greater than 90% along the horizontal direction whereas it was greater than 80% in the vertical direction. This ratio also shows that the gaze shifts without correction and the gaze shifts with correction are closely related. The horizontal head amplitude shown in Fig. 11 shows a linear relationship

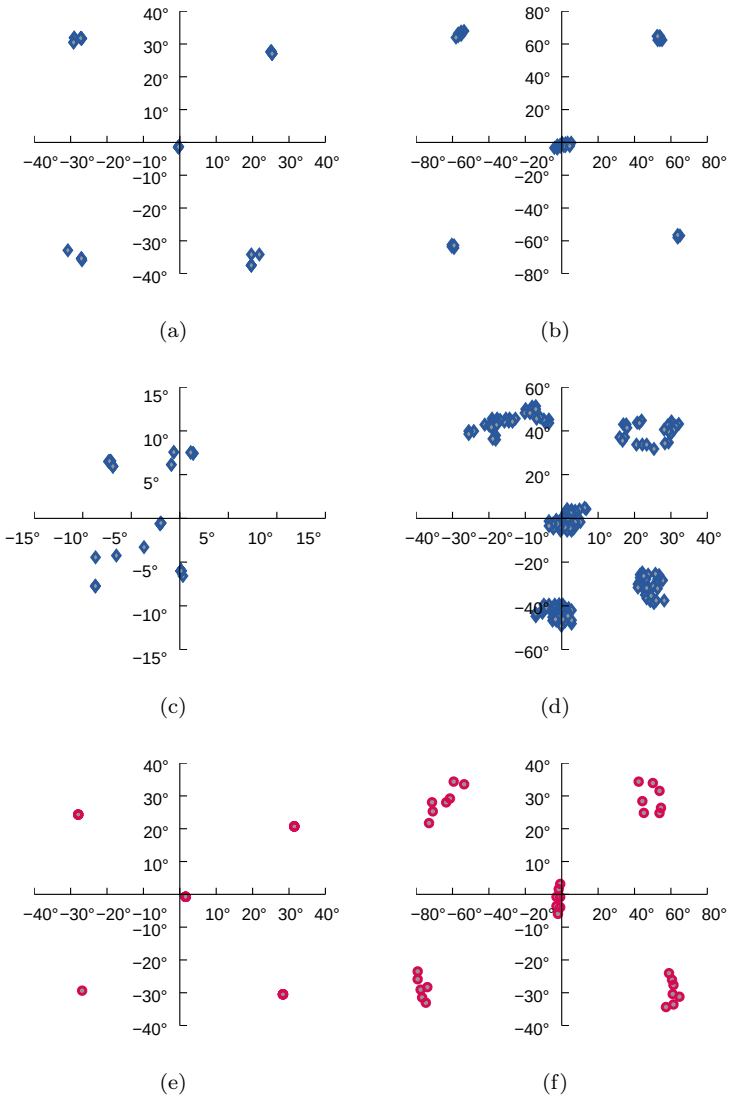
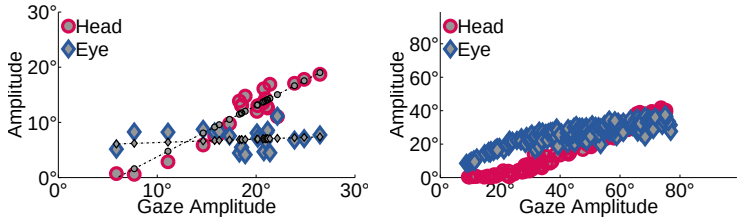


Fig. 9 Eye and head gaze shift contribution for visual targets arranged in square pattern paradigm. Figure (a) shows right eye position in space plotted with tip of rotation vectors using the results obtained from the trained eyes-head coordination network, whereas (b) represents right eye position in space for human data adapted from [12, Fig. 1(A)] for large gaze shifts. Figure (c) shows eye position in head with the proposed eyes-head coordination network, whereas (d) represents eye position in head for human data obtained from [12, Fig. 1(B)], (e) head in space with the eyes-head coordination network, (f) shows head in space for human gaze shifts [12, Fig. 1(C)].



(a) (b)

Fig. 10 Eye and head gaze shift contribution for horizontal gaze amplitude using the tangential screen paradigm. Figure (a) shows eye and head amplitude relationship with increasing horizontal gaze amplitude for the trained eyes-head coordination network. The eye and head contribution trend is shown with lines of best fit. Whereas figure (b) shows eye and head amplitude with increase in horizontal gaze amplitude for primate data adopted from [8, Fig. 6(F) and (D)].

with gaze amplitude as compared to the vertical component, furthermore the vertical head amplitude component was smaller compared to the vertical target displacement. Similar results were found in primates [8].

To further determine whether eyes and head movement amplitudes are better related to target displacement relative to gaze or target displacement relative to head. The data obtained from preceding experiment for target displacement relative to gaze was used for this analysis. The trials of the left eye and the head movements were selected for which target displacement relative to the head was relatively constant. The change in head position from the gaze onset to offset was used to determine the target displacement relative to the head. The trails were selected for two target displacements relative to head *i.e.*, 10° and 20°, however the target displacement relative to gaze was highly variable in each case. The results in Fig. 12 show that the eye amplitude has systematic relationship with the target displacement relative to gaze as compared to the target displacement relative to head. However the amplitude of head remained almost constant even with increasing target displacement relative to gaze. Thus head amplitude has a systematic relationship with the target displacement relative to head compared to the target displacement relative to gaze.

3.5 Effect of initial eyes position

Primate studies on eye-head coordination have shown that the initial eye position effects the relative contribution of eye and head movements to gaze shifts [9,8,28]. To assess the effect of initial eye position on eyes-head coordination using the proposed eyes-head coordination network, the tangential screen paradigm was used to place visual targets along the horizontal meridian. The robot eyes were positioned at the centre of their orbits or at two contralateral positions (10° and 20°) relative to the direction of the gaze shift,

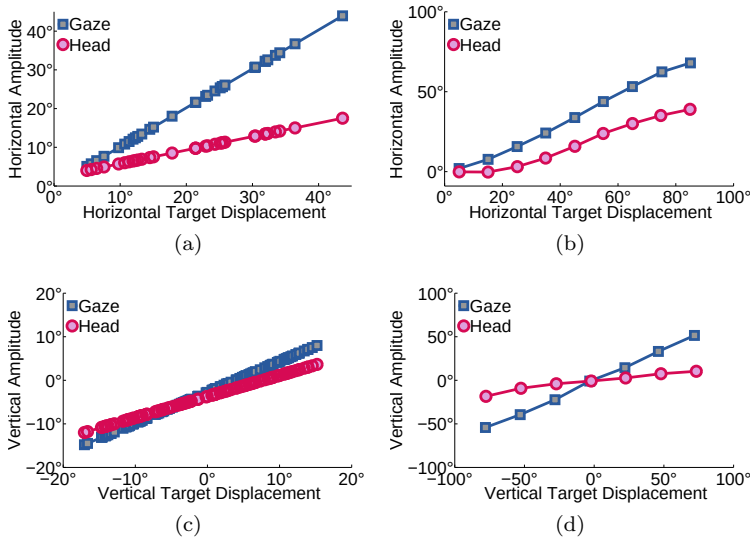


Fig. 11 Target displacement against gaze and head movement amplitude. The left column shows results obtained from the proposed network whereas the right column shows the primate results taken from [8, Fig. 4(A), (B), (C) and (D) : Monkey T]. The lines of best fit were drawn for each data pattern. Figure (a) shows a linear relationship of gaze and head amplitude with horizontal target displacement similar as primate results in (b). Figure (c) also shows linear relationship with target displacement however the slope of the data for head amplitude was reduced as in the primate data (d).

similar to the experimental procedure used by [8]. In [8], two set of gaze shifts were performed when the eyes were deviated in the orbits contralateral to the direction of movement. For the robot experiments, the head was oriented randomly along the horizontal meridian while initial head orientation along the vertical and torsional directions were restrained (*i.e.*, tilt and swing 0°). Before gaze onset and after gaze offset, eyes and head motor commands were recorded for 250 trails in each experiment. The eye and head contribution in case of different contralateral eye positions showed variability in gaze shift amplitude for each visual target. The results show that the slope of eye gaze amplitude increased with increasing contralateral eye position and the slope of head contribution accordingly reduced which are similar to the biological results [8] as shown in Fig. 13. The relationship between eyes-head contribution due to change in initial eyes position at gaze onset indicates that both eyes and head control circuits in the proposed network are independently controlled while having mutual interaction to adjust the amplitude of eyes-head gaze contribution. These results also show that the initial eyes position acts as one factor to resolve the eyes-head gaze contribution redundancy.

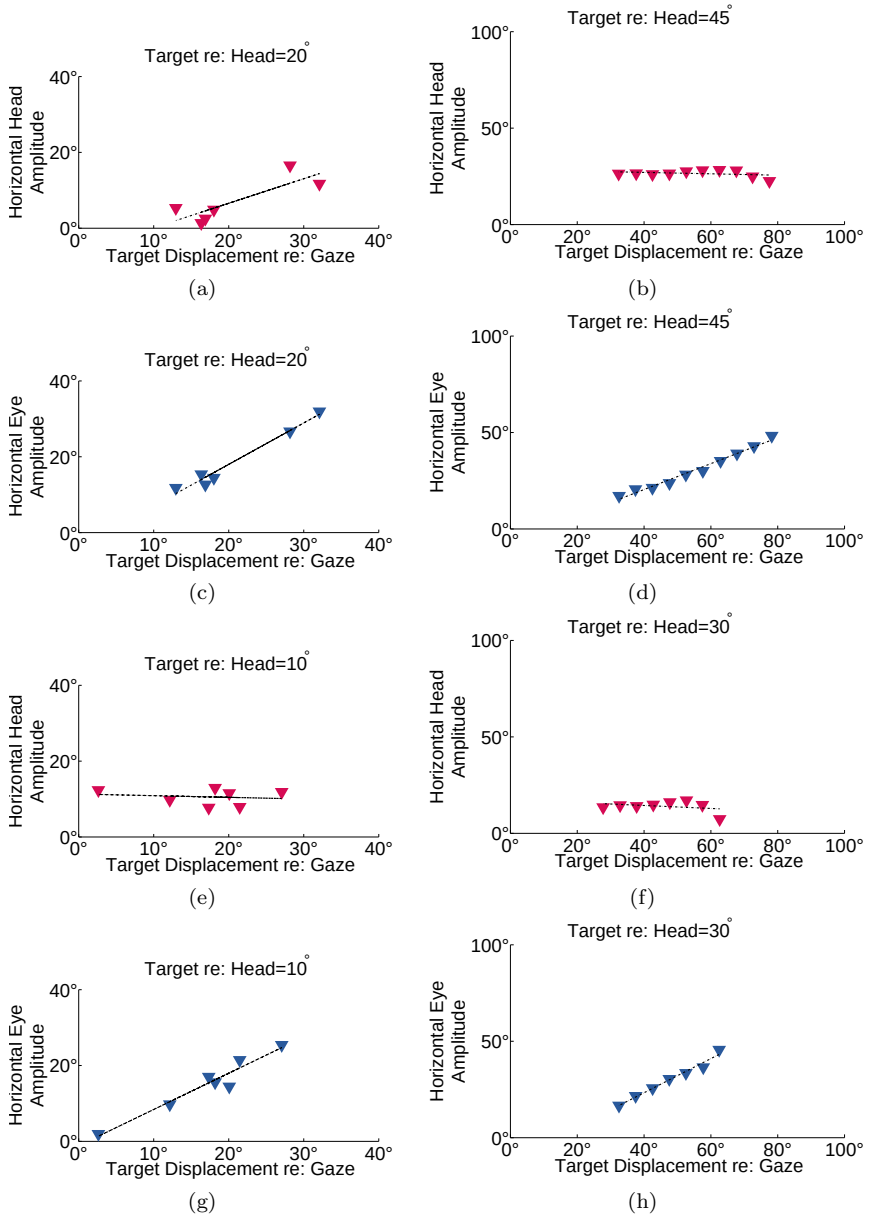


Fig. 12 Target displacement and horizontal eye-head amplitude relationship found using the tangential screen paradigm. The left column shows results obtained from the trained eyes-head coordination PC/BC-DIM network whereas right column shows the primate results taken from [8, Fig. 5(B), (C), (E) and (F) : Monkey T]. Figure (a) shows horizontal head amplitude against target displacement relative to gaze for 20° target displacement relative to head, whereas (b) shows primate head amplitude for 45° target displacement relative to head. Figure (c) shows eye amplitude against target displacement relative to gaze for 20° target displacement relative to head whereas (d) shows primate result for 45° target displacement relative to head. The figures (e) and (g) for 10° target relative to head using the trained PC/BC-DIM network whereas (f) and (h) show primate results for 30° target relative to head.

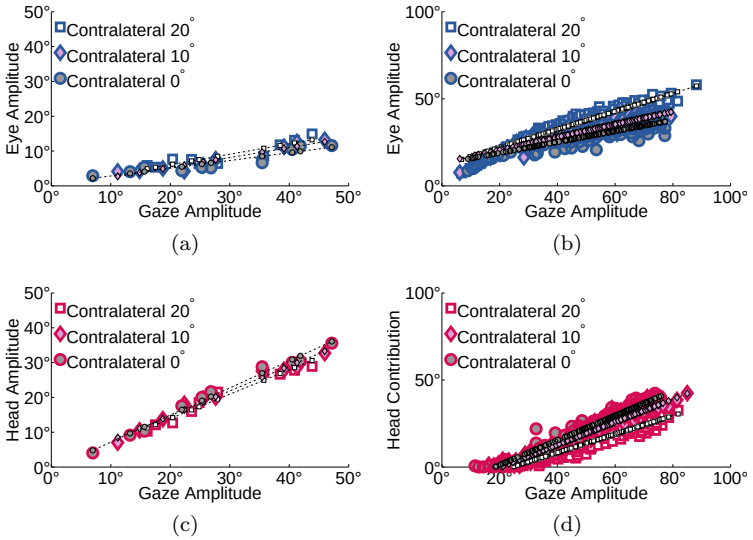


Fig. 13 The effect of contralateral eyes position on eye-head gaze contribution. The lines of best fit were plotted with data type similar markers as shown in the legend. (a) The magnitude and slope of eye gaze contribution increased with greater eccentricity of the eye relative to the target as shown with dashed best fit lines, whereas (b) shows similar increasing eye contribution in primates data taken from [8, Fig. 14(I), (J) and (K)]. (c) The slope of head contribution decreased with increasing contralateral eye position, whereas (d) show similar trend of head contribution in primates data.

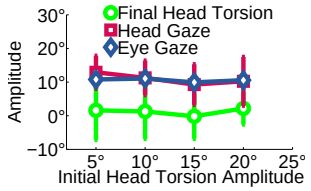


Fig. 14 The effect of initial head torsional position on eyes and head gaze contribution and final selected head torsional value. With change in initial head torsion position, eyes gaze amplitude showed no major effect, however, the head contribution changed so as to select a torsion value to bring the target near to the horizontal axis of both eyes (*i.e.*, a head torsion value near to zero).

3.6 Effect of initial head torsional position

The effect of initial head torsional position on eye and head gaze contribution was also investigated by changing the initial head torsion position in both clockwise and counter-clockwise directions. The head pose was set in a forward facing direction with restrained initial horizontal and vertical orientation (*i.e.*, pan and tilt 0°) and the eyes were positioned in the centre of their orbits (*i.e.*, pan and tilt 0°). The targets were displaced along the vertical meridian in the tangential screen paradigm such that the target of interest was visible to at

least one eye, and experiments were performed for 100 trails with each initial head torsion position. The change in head torsion value induced a change in the head gaze amplitude but produced no major change in the eye contribution. However the final selected head torsional value for each gaze shift changed the head gaze contribution in a way to bring target of interest near to the horizontal axes of both eyes and also kept final torsional value near to zero as shown in Fig. 14. These results show that the redundancy of head torsional position was resolved using described redundancy resolution procedure (see section 2.3).

4 Discussion

This article introduces an omni-directional basis function type neural network model for planning coordinated eyes-head gaze shifts. The proposed model comprises independent eyes and head control circuits engaged in mutual interaction for coordinated eyes-head gaze shifts. We showed using the eyes-head coordination strategy (section 2.3) how complex non-linear sensory-motor transformation can be achieved after transforming 4-D visual information to 7 DOFs eyes-head motor space while incorporating 7 DOFs efferent copy of eyes and head positions. The proposed eyes-head coordination network performed accurate large gaze shifts to targets of interest and convergent eyes movements to fixate on the targets with biological comparable accuracy. We compared several eyes and head coordination relationships with the primates data to evaluate the network performance. The eyes-head gaze direction relationship for large oblique gaze shifts was investigated using the proposed network. The results obtained from these experiments showed a lawful gaze contribution relationship with gaze direction since head contributed more along the horizontal direction and eyes along the vertical direction similar as in primates [8]. These experiments were performed with randomly sequenced gaze shifts between memory-based body-centred targets representations, and hence, had no effect of initial eyes and head positions. This implies that the gaze direction is a very important factor to determine eyes and head contribution during large gaze shifts. The investigation of horizontal eyes, head and gaze amplitude relationship introduced the gaze amplitude as another factor for eyes and head contribution. The eyes movement amplitude was large for small gaze shifts whereas for large gaze shifts it remained almost constant. In contrast head contribution for small gaze shifts was small and showed a linear incremental relation with large gaze shifts. The relationship of target displacement relative to gaze and target displacement relative to head with primary gaze shifts was investigated. The results showed a systematic relationship between target displacement and movement amplitudes. Furthermore the results showed that the target displacement relative to gaze was better related to eyes movement amplitude whereas the target displacement relative to head was related to head movement amplitude. The network showed primates comparable results provided in [8] for target displacement and movement amplitude relationship.

We also compared the effect of initial eyes position on gaze shift results. The increasing contralateral eye position relative to gaze direction introduced increase in eye contribution whereas the head contribution reduced accordingly which is similar to primate results [9,8,28]. This also confirms that both eyes and head control circuits are interacting with each other to amend gaze contribution amplitude in a close relation. The effect of initial head torsional position on eyes and head gaze contribution amplitude was examined which showed profound effect on the head gaze contribution. Furthermore the final selected head torsional value always remained near to zero as in primates [4, 12,52]. Therefore, the initial eyes and head position, the initial head torsional position and gaze direction form a basis to predict and select eyes and head gaze contribution for each gaze shift plan and to resolve the redundancies involved in shifting gaze to 3-D target of interest. Based on these initial parameters the network predicts and selects one gaze plan to resolve the gaze shift plan redundancy and one head torsional value to resolve the head torsional redundancy. It is planned to further exploit this ability of the network to build more comprehensive hierarchical networks in future work. We plan to learn eye-head-arm coordination using a body-centred representation, which can be used to develop a more comprehensive model of eyes, head and arm movement control for gaze shift and arm reaching to the same target of interest or gaze shift and pointing with arm to different body-centred visual targets or gaze shift to view the hand.

Acknowledgements This work was partially funded by Higher Education Commission Pakistan under grant No. PM(HRDI-UESTPs)/UK/HEC/2012.

References

1. Barnes, G.: Vestibulo-ocular function during co-ordinated head and eye movements to acquire visual targets. The Journal of Physiology **287**(1), 127–147 (1979)
2. Blakemore, C., Donaghy, M.: Co-ordination of head and eyes in the gaze changing behaviour of cats. The Journal of Physiology **300**(1), 317–335 (1980)
3. Constantin, A., Wang, H., Monteon, J., Martinez-Trujillo, J., Crawford, J.: 3-dimensional eye-head coordination in gaze shifts evoked during stimulation of the lateral intraparietal cortex. Neuroscience **164**(3), 1284–1302 (2009)
4. Crawford, J., Martinez-Trujillo, J., Klier, E.: Neural control of three-dimensional eye and head movements. Current Opinion in Neurobiology **13**(6), 655–662 (2003)
5. Crawford, J.D., Ceylan, M.Z., Klier, E.M., Guitton, D.: Three-dimensional eye-head coordination during gaze saccades in the primate. Journal of Neurophysiology **81**(4), 1760–1782 (1999)
6. De Meyer, K., Spratling, M.W.: Multiplicative gain modulation arises through unsupervised learning in a predictive coding model of cortical function **23**(6), 1536–67 (2011)
7. Freedman, E.G.: Interactions between eye and head control signals can account for movement kinematics. Biological Cybernetics **84**(6), 453–462 (2001)
8. Freedman, E.G., Sparks, D.L.: Eye-head coordination during head-unrestrained gaze shifts in rhesus monkeys. Journal of Neurophysiology **77**(5), 2328–2348 (1997)
9. Freedman, E.G., Sparks, D.L.: Coordination of the eyes and head: movement kinematics. Experimental Brain Research **131**(1), 22–32 (2000)
10. Galiana, H., Guitton, D.: Central organization and modeling of eye-head coordination during orienting gaze shiftsa. Annals of the New York Academy of Sciences **656**(1), 452–471 (1992)

11. Georgopoulos, A.P., Schwartz, A.B., Kettner, R.E.: Neuronal population coding of movement direction. Science **233**, 1416–9 (1986)
12. Glenn, B., Vilis, T.: Violations of listing's law after large eye and head gaze shifts. Journal of Neurophysiology **68**(1), 309–318 (1992)
13. Goossens, H.H., Van Opstal, A.: Human eye-head coordination in two dimensions under different sensorimotor conditions. Experimental Brain Research **114**(3), 542–560 (1997)
14. Gresty, M.: Coordination of head and eye movements to fixate continuous and intermittent targets. Vision Research **14**(6), 395–403 (1974)
15. Guitton, D.: Control of eyehead coordination during orienting gaze shifts. Trends in Neurosciences **15**(5), 174–179 (1992)
16. Guitton, D., Douglas, R., Volle, M.: Eye-head coordination in cats. Journal of Neurophysiology **52**(6), 1030–1050 (1984)
17. Guitton, D., Munoz, D.P., Galiana, H.L.: Gaze control in the cat: studies and modeling of the coupling between orienting eye and head movements in different behavioral tasks. Journal of Neurophysiology **64**(2), 509–531 (1990)
18. Guitton, D., Volle, M.: Gaze control in humans: eye-head coordination during orienting movements to targets within and beyond the oculomotor range. Journal of Neurophysiology **58**(3), 427–459 (1987)
19. Huang, Y., Rao, R.P.N.: Predictive coding. WIREs Cognitive Science **2**, 580–93 (2011). DOI 10.1002/wcs.142
20. Kardamakis, A.A., Moschovakis, A.K.: Optimal control of gaze shifts. The Journal of Neuroscience **29**(24), 7723–7730 (2009)
21. Klier, E.M., Wang, H., Crawford, J.D.: The superior colliculus encodes gaze commands in retinal coordinates. Nature Neuroscience **4**(6), 627–632 (2001)
22. Klier, E.M., Wang, H., Crawford, J.D.: Three-dimensional eye-head coordination is implemented downstream from the superior colliculus. Journal of Neurophysiology **89**(5), 2839–2853 (2003)
23. Laurutis, V., Robinson, D.: The vestibulo-ocular reflex during human saccadic eye movements. The Journal of Physiology **373**(1), 209–233 (1986)
24. Law, J., Shaw, P., Lee, M.: A biologically constrained architecture for developmental learning of eye–head gaze control on a humanoid robot. Autonomous Robots **35**(1), 77–92 (2013)
25. Lopes, M., Bernardino, A., Santos-Victor, J., Rosander, K., von Hofsten, C.: Biomimetic eye-neck coordination. In: Development and Learning, IEEE 8th International Conference on, pp. 1–8. IEEE (2009)
26. Maini, E.S., Teti, G., Rubino, M., Laschi, C., Dario, P.: Bio-inspired control of eye-head coordination in a robotic anthropomorphic head. In: Biomedical Robotics and Biomechatronics, The First IEEE/RAS-EMBS International Conference on, pp. 549–554. IEEE (2006)
27. Maurer, C., Mergner, T., Lücking, C., Becker, W.: Adaptive changes of saccadic eye–head coordination resulting from altered head posture in torticollis spasmodicus. Brain **124**(2), 413–426 (2001)
28. McCluskey, M.K., Cullen, K.E.: Eye, head, and body coordination during large gaze shifts in rhesus monkeys: movement kinematics and the influence of posture. Journal of Neurophysiology **97**(4), 2976–2991 (2007)
29. Medendorp, W., Melis, B., Gielen, C., Van Gisbergen, J.: Off-centric rotation axes in natural head movements: implications for vestibular reafference and kinematic redundancy. Journal of Neurophysiology **79**(4), 2025–2039 (1998)
30. Metta, G., Sandini, G., Vernon, D., Natale, L., Nori, F.: The icub humanoid robot: An open platform for research in embodied cognition. In: Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems, PerMIS '08, pp. 50–6. ACM, New York, NY, USA (2008). DOI 10.1145/1774674.1774683
31. Milighetti, G., Vallone, L., De Luca, A.: Adaptive predictive gaze control of a redundant humanoid robot head. In: Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference on, pp. 3192–3198. IEEE (2011)
32. Misslisch, H., Tweed, D., Vilis, T.: Neural constraints on eye motion in human eye-head saccades. Journal of Neurophysiology **79**(2), 859–869 (1998)
33. Muhammad, W., Spratling, M.W.: A neural model of binocular saccade planning and vergence control. Adaptive Behavior **23**(5), 265–282 (2015)

34. Munoz, D.P., Guitton, D.: Control of orienting gaze shifts by the tectoreticulospinal system in the head-free cat. ii. sustained discharges during motor preparation and fixation. Journal of Neurophysiology **66**(5), 1624–41 (1991)
35. Munoz, D.P., Guitton, D., Pelisson, D.: Control of orienting gaze shifts by the tectoreticulospinal system in the head-free cat. iii. spatiotemporal characteristics of phasic motor discharges. Journal of Neurophysiology **66**(5), 1642–1666 (1991)
36. Niebur, E.: Saliency map. Scholarpedia **2**(8), 2675 (2007)
37. Pelisson, D., Guitton, D., Munoz, D.: Compensatory eye and head movements generated by the cat following stimulation-induced perturbations in gaze position. Experimental Brain Research **78**(3), 654–658 (1989)
38. Pelisson, D., Prablanc, C., Urquizar, C.: Vestibuloocular reflex inhibition and gaze saccade control characteristics during eye-head orientation in humans. Journal of Neurophysiology **59**(3), 997–1013 (1988)
39. Phillips, J., Ling, L., Fuchs, A., Siebold, C., Plorde, J.: Rapid horizontal gaze movement in the monkey. Journal of Neurophysiology **73**(4), 1632–1652 (1995)
40. Proudlock, F.A., Shekhar, H., Gottlob, I.: Age-related changes in head and eye coordination. Neurobiology of Aging **25**(10), 1377–1385 (2004)
41. Rao, R.P.N., Ballard, D.H.: Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects **2**(1), 79–87 (1999)
42. Saeb, S., Weber, C., Triesch, J.: Learning the optimal control of coordinated eye and head movements. PLoS Computational Biology **7**(11), e1002,253 (2011)
43. Shibata, T., Vijayakumar, S., Conradt, J., Schaal, S.: Biomimetic oculomotor control. Adaptive Behavior **9**(3-4), 189–207 (2001)
44. Spratling, M.W.: Predictive coding as a model of biased competition in visual selective attention **48**(12), 1391–408 (2008)
45. Spratling, M.W.: Reconciling predictive coding and biased competition models of cortical function **2**(4), 1–8 (2008)
46. Spratling, M.W.: Learning posture invariant spatial representations through temporal correlations **1**(4), 253–63 (2009)
47. Spratling, M.W.: Classification using sparse representations: a biologically plausible approach **108**(1), 61–73 (2014)
48. Spratling, M.W.: Predictive coding as a model of cognition. Cognitive Processing (in press)
49. Spratling, M.W.: A neural implementation of bayesian inference based on predictive coding. submitted (sub.)
50. Spratling, M.W., De Meyer, K., Kompass, R.: Unsupervised learning of overlapping image components using divisive input modulation **2009**(381457), 1–19 (2009)
51. Srinivasa, N., Grossberg, S.: A head–neck–eye system that learns fault-tolerant saccades to 3-d targets using a self-organizing neural model. Neural Networks **21**(9), 1380–1391 (2008)
52. Straumann, D., Haslwanter, T., Hepp-Reymond, M.C., Hepp, K.: Listing's law for eye, head and arm movements and their synergistic control. Experimental Brain Research **86**(1), 209–215 (1991)
53. Takanishi, A., Matsuno, T., Kato, I.: Development of an anthropomorphic head-eye robot with two eyes-coordinated head-eye motion and pursuing motion in the depth direction. In: Intelligent Robots and Systems, 1997. IROS'97., Proceedings of the 1997 IEEE/RSJ International Conference on, vol. 2, pp. 799–804. IEEE (1997)
54. Thomson, D., Loeb, G., Richmond, F.: Effect of neck posture on the activation of feline neck muscles during voluntary head turns. Journal of Neurophysiology **72**(4), 2004–2014 (1994)
55. Tikhanoff, V., Cangelosi, A., Fitzpatrick, P., Metta, G., Natale, L., Nori, F.: An open-source simulator for cognitive robotics research: The prototype of the icub humanoid robot simulator. In: Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems, PerMIS '08, pp. 57–61. ACM, New York, NY, USA (2008). DOI 10.1145/1774674.1774684
56. Tomlinson, R.: Combined eye-head gaze shifts in the primate. iii. contributions to the accuracy of gaze saccades. Journal of Neurophysiology **64**(6), 1873–1891 (1990)
57. Tomlinson, R., Bahra, P.: Combined eye-head gaze shifts in the primate. i. metrics. Journal of Neurophysiology **56**(6), 1542–1557 (1986)

58. Tomlinson, R., Bahra, P.: Combined eye-head gaze shifts in the primate. ii. interactions between saccades and the vestibuloocular reflex. Journal of Neurophysiology **56**(6), 1558–1570 (1986)
59. Tweed, D.: Three-dimensional model of the human eye-head saccadic system. Journal of Neurophysiology **77**(2), 654–666 (1997)
60. Tweed, D., Glenn, B., Vilis, T.: Eye-head coordination during large gaze shifts. Journal of Neurophysiology **73**(2), 766–779 (1995)
61. Winters, J.M., Stark, L.: Muscle models: what is gained and what is lost by varying model complexity. Biological Cybernetics **55**(6), 403–420 (1987)
62. Zangemeister, W., Lehman, S., Stark, L.: Sensitivity analysis and optimization for a head movement model. Biological Cybernetics **41**(1), 33–45 (1981)
63. Zangemeister, W., Lehman, S., Stark, L.: Simulation of head movement trajectories: model and fit to main sequence. Biological Cybernetics **41**(1), 19–32 (1981)
64. Zangemeister, W., Stark, L.: Types of gaze movement: variable interactions of eye and head movements. Experimental Neurology **77**(3), 563–577 (1982)
65. Zangemeister, W.H., Stark, L.: Gaze latency: variable interactions of head and eye latency. Experimental Neurology **75**(2), 389–406 (1982)