4th International Conference on
Development and Learning and on Epigenetic Robotics
October 13-16, 2014. Palazzo Ducale, Genoa, Italy

TuDFP.2

Neural-Dynamic Architecture for Looking: Shift from Visual to Motor Target Representation for Memory Saccades.

Yulia Sandamirskaya*, Tobias Storck†

Institut für Neuroinformatik, Ruhr-Universität Bochum

Bochum, 44780, Germany

* Email: yulia.sandamirskaya@ini.rub.de, † Email: tobias.storck@ini.rub.de

Abstract—**Looking at objects is one of the most basic sensorimotor behaviours, which requires calibration of the perceptual and motor systems. Recently, we have introduced a neural-dynamic architecture, in which the sensorimotor transformations, which lead to precise saccadic gaze shifts, is initially learned and is autonomously updated if changes in the environment or in the motor plant of the agent require adaptation. Here, we demonstrate how the allocentric, gaze-direction independent memory representations may be formed in this architecture and how sequences of precise gaze shifts may be generated to memorised targets. Our simulated robotic experiments demonstrate functioning of the architecture on an autonomous embodied agent.**

I. INTRODUCTION

Looking, i.e. the ability to direct one's gaze to an object of interest, is an amazing behaviour, in which the neural system of an agent is able to generate a precise motor command to align the eye (and head) of the agent – and herewith the rest of its body – with the outside world. Human looking behaviour consists of fast gaze shifts, called saccades, separated by periods of fixation, or smooth pursuit. The saccades bring different portions of the environment into the highly sensitive fovea, where the visual inputs may be neurally processed during the fixation periods, creating an illusion of a stable world around us, available to acting upon it. The problem of discovering how the brain creates a stable and coherent representation of the immediate environment of an agent from the discrete fixations, separated by saccades, has been posed several decades ago [1], [2].

Several parts of the neural system, involved in generation of gaze shifts, have been examined since then, and both neurologically and behaviourally realistic models were developed. Despite the seeming simplicity of the looking system, the neural circuitry involved in saccades generation has an immense complexity [3], [4]. The mapping between the visual representation of the saccade target and the motor command has to be adaptive, since the variability in the neuromuscular system of the eye calls for a permanently running calibration process between the retinal representation of the target and the motor system [5], [6], [7]. Such calibration processes related to the control of gaze-shifts were found in cerebellar cortex and are hypothesised to be modulated by reward-related basal ganglia loops. Apart from this senso-rimotor mapping, the reference frame transformation from retinotopic to an allocentric (gaze-direction independent) coordinate frame [8] is needed to fuse information, collected during fixations between saccades, in order to generate a coherent representation of the visual scene. The need for such a representation is revealed in experiments on memory-based, double-step saccades. In these experiments, saccades are generated to two remembered targets in a sequence. Using retinotopic reference frame leads to an incorrect second saccade, since its retinal representation shifts during the first saccade. Developmental studies have shown that the ability to use a correct reference frame for generation of such saccades from memory develops over the first months of life [9]. It is obvious that other movements, directed at visually perceived objects, as, e.g., reaching movements, also cannot rely on retinotopic representations (only), since humans are able to generate such movements based on a remembered target position.

In this paper, we use the neural-dynamic framework of Dynamic Field Theory [10] to implement an embodied, dynamic, autonomous, and adaptive model for looking behavior. We have demonstrated the basic functionality of this model recently [11]. Here, we focus on the capability of the network to predict the motor outcome of the potential saccade, to use this prediction to create memories in a motor-based (non-retinal) reference frame, and, finally, to perform a sequence of saccades from memory. We show how errors, caused by the retinotopic saccade plans are corrected when motor-based reference frame becomes available. This allows to model a developmental pathway from retinocentric to allocentric memory saccades.

The properties and functioning of the model are demonstrated on an embodied (here, simulated) robotic agent. We connect the robotic camera and the pan-tilt motors of the camera head to the neural-dynamic architecture and demonstrate how gaze shifts may be initiated autonomously, how their errors maybe detected and trigger adaptation of the saccadic circuitry, leading to autonomous learning of sensorimotor maps and to memory formation. A transformation to a body-centred reference frame is performed by predicting the motor outcome of a planned, but delayed, saccade, to make memory independent of the current gaze

direction of the camera. We deal with exploration of a scene and inhibition of return, which allows to scan the scene with several objects with different visual features.

The embodied approach, used in our work, enables a tight integration between modelling and behaviour analysis, linking the neural, architectural level with the behavioural level. The architecture may also have technical applications, being a proof-of-concept demonstration of an autonomous self-calibrating robotic system, which learns to look at objects in its environment and to generate a representation of the visual scene, which may be used to generate object-directed actions without immediate perception and visual servoing (see [12], [13] for other examples of similar work).

II. METHODS: DYNAMIC NEURAL FIELDS

The neural-dynamic architecture for looking, presented here, is built within Dynamic Field Theory (DFT) – a mathematical and conceptual framework for modelling cognition using Dynamic Neural Fields (DNFs) [14], [10]. DNFs are activation functions defined over different perceptual and motor variables, which characterise the perceptual states and motor parameters of actions of an embodied agent.

The DNF activation follows the dynamics, Eq. (1), which has a characteristic attractor pattern: a localized activity 'peak'. The peaks are stabilised by lateral interactions in the DNF against decay and spread. Such peaks of activation are units of representation in DFT. Because of the stability and attractor properties of the DNF dynamics, cognitive models formulated in DFT may be coupled to real robotic motors and sensors and were shown to generate cognitive behavior in autonomous robots [15], [16]. In particular, activity peaks stabilise decisions about detection of a salient object in the visual input and selection among alternatives; the stabilised representations are critical to enable stable motor control and to perform cognitive operations on representations [17].

$$\tau \dot{u}(x,t) = -u(x,t) + h + \int f(u(x',t))\omega(x',x)dx' + I(x,t) \quad (1)$$

In Eq. (1), $u(x,t)$ is the activation of a dynamic neural field (DNF) at time t; x is one or several behavioral dimensions (e.g., color, space, position or velocity of an effector), over which the DNF is spanned. τ is the relaxation time-constant of the dynamics; h is the negative resting level, which defines the activation threshold of the field; $f(\cdot)$ is the sigmoidal non-linearity shaping the output of the neural field; the lateral connections in the DNF are shaped by a Mexican hat lateral interaction kernel, $\omega(|x'-x|)$, with a short-range excitation and a long-range inhibition parts; $I(x,t)$ is sum of the external inputs to the DNF, $I_i(x,t)$, where each of the inputs may either come from a sensor or from another DNF.

In the simplest case, when the DNF (e.g., $u_2(x,t)$), which provides input to the DNF $u(x,t)$, is defined over the same behavioural dimension, x, the input $I_{u_2}(x,t) = cf(u_2(x,t))$, where c is a scaling factor (Gaussian spreading function may be used instead). If the DNFs $u(x,t)$ and $u_3(x,y,t)$

have different dimensionality, input is extended or collapsed in the not-shared dimension. DNFs, which don't share a dimension may be coupled through a mapping function (similar to a weight matrix in standard neural networks).

The basic learning mechanism in the DFT is the formation of memory traces of positive activation of a DNF [18]. The memory trace – called *preshape* in DFT – is a dynamical layer, which receives input from the respective DNF and projects its output back to this DNF. The projection can be positive, facilitating activation of the DNF at previously activated locations, or negative, inhibiting the DNF at previously activated locations (such preshape may account for habituation and facilitate exploration). The preshape layer follows the equation (3), [17].

$$\tau_l \dot{P}(x,t) = \lambda_{build}\Big(-P(x,t) + f\big(u(x,t)\big)\Big)f\big(u(x,t)\big) - $$
$$-\lambda_{decay}P(x,t)\Big(1 - f\big(u(x,t)\big)\Big). \quad (2)$$

Here, $P(x,t)$ is the strength of the memory trace at site x of the DNF with activity $u(x,t)$, λ_{build} and λ_{decay} are the rates of build-up and decay of the memory trace. The build-up of the memory trace is active on the sites with a high positive output $f\big(u(x,t)\big)$, the decay is active on the sites with a low output. The memory trace $P(x,t)$ is an additive input to the DNF dynamics.

The adaptive model for looking, introduced next, consists of a number of interconnected DNFs of different dimensionality, a neuronal burst generator, linked to the motors of the robotic agent, a gain map, which scales the projections from the retinal representations of the saccadic targets to the motor command generator, as well as a motor command integrator and a difference command generator. The latter two are adaptations of standard mechanisms, which were developed in neural models of movement generation [19], other elements may be mapped on different parts of cortical and subcortical structures (e.g., superior calliculus for saccade target representation, or cerebellum for gain maps adaptation), involved in saccade generation, as we discussed recently [20].

III. THE MODEL

A. The DNF architecture for looking

Fig. 1 depicts the overall architecture for looking. This network of dynamic neural fields implements several behavioural functions along with generating gaze shifts towards visually perceived targets (e.g. exploring a scene, building a long-term memory of the scene, fixation dynamics and smooth pursuit, planning memory saccades, updating the sensorimotor mapping). Here, we explain the crucial elements of the network.

On the left of Fig. 1 (grey shaded rectangle), the camera and the motors of the pan-tilt unit are shown. The camera provides visual input to the *visual perception DNF* and the *center of visual field DNF*. The first of these two three-dimensional DNFs builds peaks over color-space
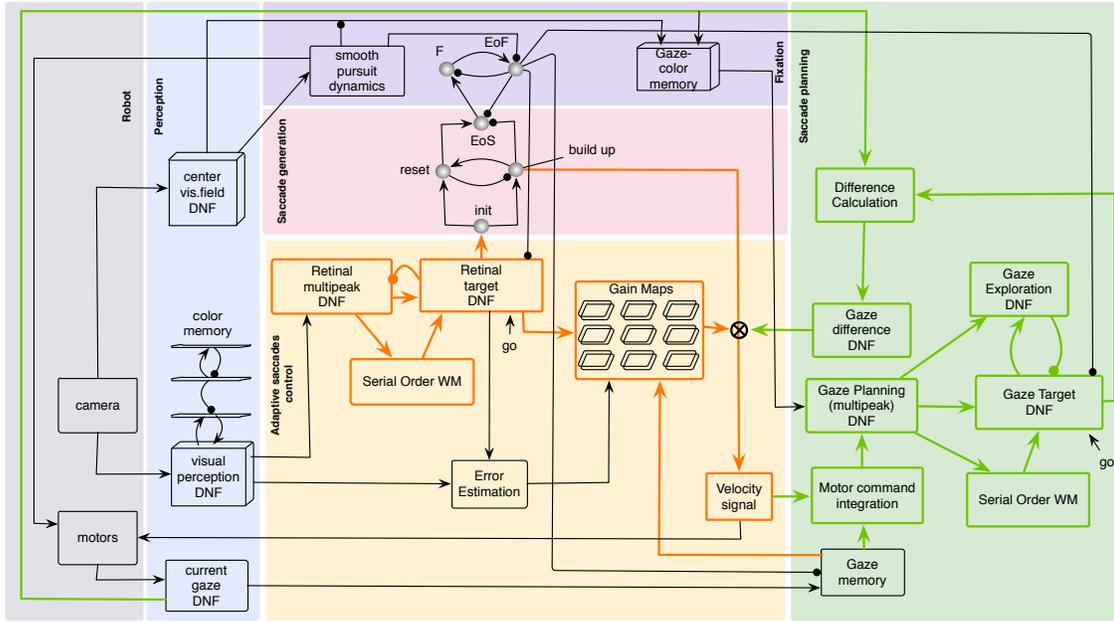
Fig. 1: The overall neural-dynamic architecture for looking. Arrows represent excitatory connectivity between DNFs, lines with filled circles show the inhibitory connections. See main text for details.

associations, which correspond to salient portions of the camera image. A cascade of one-dimensional *color-memory fields* with different time-scales of the memory dynamics keeps track of the colours, which were already brought into foreground, and temporally inhibit them in order to facilitate exploration of yet unseen colours. The *center of visual field DNF* is activated when the target object is centred in the visual field of the camera. The *current gaze DNF* builds stabilised representation of the state of the motors before the saccade. These DNFs constitute the perceptual system of the neural-dynamic architecture (blue shaded rectangle in the figure).

The *center of visual field DNF* activates the *smooth pursuit dynamics*, which tracks the object during active fixation. At the same time, the *gaze-color memory* DNF accumulates memory for the currently fixated object, storing its features and location in the gaze-space (i.e., the motor state is stored when the camera fixates the object in the centre of visual field). This memory is only formed if no motion is generated by the *smooth pursuit dynamics*. The *fixation* (F) and the *end of fixation* (EoF) nodes control activation and deactivation of the fixation phase of the camera (eye) movements. This fixation system is marked with a violet shaded region in the Fig. 1.

The pink shaded rectangle in Fig. 1 marks the saccade generation circuit, which consists of four dynamical nodes (zero-dimensional DNFs). The *initiation node* is activated by either the *retinal* or the *gaze-based target fields* (these DNFs will be presented further) and, in its turn, activates the *saccade burst node* and lifts the resting level of the *saccade reset node*. The later two nodes generate one oscillation of activation: the burst node is activated first, brings the reset node over the activation threshold, and is eventually

inhibited by the activated reset node. The reset node is left active by the input from the initiation node until the target's representation decays in the target DNFs (this time interval includes a fixation period, controlled by the *fixation* and the *end-of-fixation* dynamical nodes). The oscillation of the burst node's activation has a stereotypical (the same for all saccades) activation profile, which is scaled by the gain maps separately for the two motors of the system to result in a precise saccade. The scaled activation signal is sent to the motors and sets the velocity of the pan and tilt motors in our robotic implementation, or it corresponds to a signal, sent to motor neurons of biological looking systems.

The orange shaded rectangle in Fig. 1 depicts the retino-topic saccadic circuit, which implements an autonomous adaptation of the *gain maps* between the visual target representation and the motor commands. The gain maps generate correct amplitudes of the motor signal for the two motors of the robot (pan and tilt motors of the camera head), based on the location of the visual input in the *retinal target DNF* and the pose of the camera head before the saccade, stored in the *gaze memory DNF*.

An *error-estimation* module estimates the sign and amplitude of error of a saccadic gaze shift based on inputs from the *retinal target DNF*, which holds the initial position of the target in the retinal reference frame, and the *visual perception DNF*, which holds the position of the target in the same reference frame after the saccade. The identity of the target is determined by the color, stored in the color fields before the saccade, which biases the visual perception DNF to only be sensitive to objects of the stored color (this bias is weak, however). The error is estimated in terms of the saccade being too long or too short in the two directions of the visual space and respectively decreases or increases

the values, stored in the gain map (a simple neural-dynamic circuit performs this estimation, see section III-D for the formula). The gain map update is performed at locations, which correspond to the original retinal position of the target (link from the retinal target DNF) and the pose of the camera head before the saccade (link from the gaze memory DNF).

The main contribution of this paper are modifications to the parts of the architecture, marked orange and green in Fig. 1. These structures enable the system to generate a sequence of memory saccades and, moreover, switch from the retina-based memory saccades (which are typical for young infants) to allocentric, or spatiotopic, saccades, which characterise looking behaviour of adults (and lead to correct double-step saccades). Further, we describe these retinotopic and allocentric (gaze-based) pathways, whereas parts of the architecture introduced so far are shared by the two pathways.

B. The retinotopic pathway

In the retinotopic pathway (orange arrows and boxes in Fig. 1), the *retinal multipeak DNF* creates a multi peak representation of two or more objects, presented to the system sequentially (or selected sequentially by a saliency-based exploration mechanisms, not shown here). The *serial order working memory* preshape dynamics builds memory traces of the objects, represented in the retinal multipeak DNF. In particular, the traces of objects, stored earlier, have more time to build-up a memory trace and thus their representation is stronger in the serial order WM. This mechanism is computationally equivalent to item and order working memory, originally introduced by Grossberg [21]. Input from the serial order WM biases a sequential selection of the presented objects in the *retinal target DNF*. The representation of the selected object is stabilised and self-sustained in this DNF, representing the single target of the upcoming saccade during the eye movement and the fixation period.

For a retinotopic saccade, an activity peak in the *retinal target DNF* triggers a saccade generator circuit, activating the *initiate* node. When the first saccade is finished, the *retinal target DNF* is inhibited by the *end-of-fixation node* and the self-sustained peak in this field decays. A cascade of instabilities in several nodes follows (the initiation node, reset node, end-of-saccade node, fixation, and end-of-fixation nodes), which eventually releases inhibition on the *retinal target DNF*. The next target is selected based on the *serial order WM* and the *retinal multipeak DNF*. In the latter field, the representation of the first target object has ceased during the saccade because of the localised inhibition from the retinal target DNF (inhibition of return). The second saccade may be performed now based on the same principle, which results in a sequence of retinotopic saccades, the second of which misses the target, since its retinal representation is no longer valid after the first saccade.

C. The allocentric pathway

Since retinotopic saccades lead to errors when performing more than one saccade from memory, an alternative path gets recruited, in which the representation of the second target does not loose its validity after intermediate saccade(s). In our model, we use a gaze-based target representation to enable memory saccades. In particular, the locations of targets are stored in terms of the gaze-angles, which characterise the camera pose when the robot is fixating the objects. Using this representation, the saccade towards the target may be initiated from any initial pose, i.e. after any number of intermediate saccades. The same representation is used in our architecture to store long-term memory of the observed scene. However, to enable memory saccades (e.g., as performed in the double-step paradigm), the pose of the camera (eye) after the saccade has to be inferred, or predicted, without performing the actual gaze-shift movement.

The gaze-based target representation is shown in the green shaded region in Fig. 1. The *motor command integrator* integrates the velocity signal internally starting with the current gaze pose of the motors, instead of letting this command drive the motors. The integrated motor signal creates an activity peak in the *gaze planning DNF* at a location, which corresponds to the gaze angle of the camera head when it would fixate the object. This rate-to-space code transformation[1] is accomplished by a mechanism, equivalent to summing-up an input from a set of nodes, sensitive to different value ranges of the integrated motor signal (this transformation may also be subject to adaptation, which however, was not needed in our setting). The resulting gaze angles are stored in the *gaze planning DNF* for all objects, which the system will look at, whereas internally simulated "looking acts" are generated using the same circuitry as for the retina-based saccades (the retinal target DNF activated the saccade generating circuit here).

The *serial order WM* in the gaze-based pathway, similarly as in the retinal pathway, creates graded activation levels for objects, which are put into the *gaze planning DNF* in a sequence. The *gaze target DNF* selects the least recent object and stabilises this selection decision, while the *gaze exploration DNF* weakens the respective location for the next target selection (the *gaze target DNF* is inhibited after the actually executed saccade by the end of fixation node). The *gaze target DNF* builds the activity peak when the system is allowed to generate gaze shifts (*go* input). An activity peak in the *gaze target DNF* activates the *difference command generation* module, which subtracts the target gaze from the current gaze of the camera system. The generated difference value scales the saccade generator's output to generate a precise gaze-based saccade. Since the memory for saccade targets is held in motor coordinates, which are independent of the pose of the retina during planning, a sequence of correct saccades follows.

---

[1]From a single value of the integrator, which is equivalent to a rate-code representation in neuronal terms, to a DNF space-code representation, in which location of the (maximal) activation represents the encoded value.

D. Learning the gain maps and saccade generation: a mathematical formalization

The learning mechanism, which updates the *Gain maps* between the visual (retinal) representation of the saccadic target and the amplitude of the velocity profile, amounts to an adaptation equation, Eq. 3.

$$
\begin{aligned}
\tau_l \dot{G}^{h,v}(t;x,y,k,l) &= c_{EoS}(t)S_{err}^{x,y}(t) \cdot \\
&\cdot \left[f(u_{Tar}(t;x,y)) \times f(u_{gaze}^0(t;k,l)) \right] \quad (3)
\end{aligned}
$$

Here, $G^{h,v}(t;x,y,k,l)$ is a matrix of gain values, defined over the dimensions of the retinal space (x,y) and the space of initial gazes (k,l) (initial pan and tilt of the camera); t is time, τ_l is the time-constant scaling the learning progress. The two four-dimensional gain maps, $G^{h,v}$, scale amplitudes of the horizontal and vertical eye (camera) movements respectively. $c_{EoS}(t) = f(u_{EOS}(t))$ is the output of the *end-of-saccade* node, which signals accomplishment of a saccadic eye movement ("condition of satisfaction" of this movement, [22]); $f(u_{Tar}(t;x,y))$ is the output of the sustained target DNF, $f(u_{gaze}^0(t;k,l))$ is the output of the gaze memory DNF, which holds the memory of the eye position prior to the saccade. The (Kronecker) product of the outputs of these fields implements the selection of the region of the gain map, in which learning (update) is active: the region, where the DNFs' positive activations overlap. $S_{err}^{x,y}(t)$ is the (retinal) error signal, which is estimated in two directions based on the (stored in the sustained activation of the retinal target DNF) retinal position of the saccadic target and the position of this target after the saccade, according to Eq. (4) (only x component is shown here):

$$
\begin{aligned}
S_{err}^x(t) = c \cdot sign &\left[\int (x - x_{CF}) f(u_{Tar}(t;x,y) dxdy \cdot \right. \\
&\left. \cdot \int (x - x_{CF}) I_{perc}(t;x,y,) dxdy \right]. \quad (4)
\end{aligned}
$$

According to this equation, the error signal is positive, if activity peaks in the target, u_{Tar}, and perceptual, I_{perc}, DNFs are located in the same half of these DNFs (defined by the central position, x_{CF}), and are negative otherwise. The amplitude of the error signal is proportional to the offset of the activity peak in the perceptual DNF (e.g. retinal target location after the saccade) from the center of visual field. c is a scaling factor.

The velocity signal, $v^{h,v}(t)$ sent to the two motors of the robot is calculated based on Eq. 5:

$$
\begin{aligned}
v^{h,v}(t) = & Nf(u_{burst}(t)) \int_{x,y,k,l} \left((f(u_{Tar}(t;x,y)) \times \right. \\
& \times \left. f(u_{gaze}^0(t;k,l))) \circ G^{h,v}(t;x,y,k,l) \right) dxdydkdl \quad (5)
\end{aligned}
$$

Here, $f(u_{burst}(t))$ is the output of the burst node of the saccade generation circuit (see Fig. 1). $\circ$ denotes an entrywise product, the rest of the notation is the same as in Eq. (3). In Eq. 5, the velocity signal is scaled with the normalised sum of the values in the gain maps over

the region, defined by the positive activation in the retinal target DNF and the gaze memory DNF, which effectively corresponds to reading out an average value in the gain map in the region, defined by the positive activation in the u_{Tar} and u_{gaze}^0 DNFs. The normalisation factor $N = (\int \int f(u_{Tar}(t;x,y)) dxdy \int f(u_{gaze}^0(t;k,l) dkdl)^{-1}$.

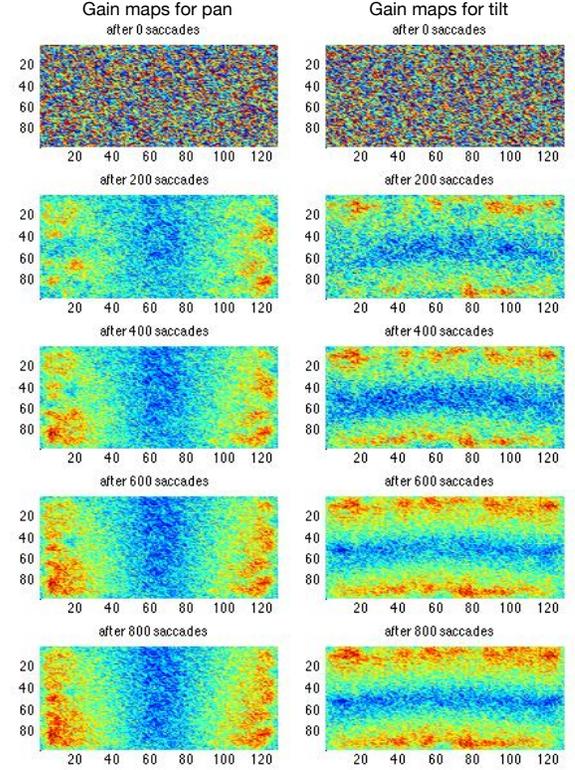## IV. RESULTS OF SIMULATED EXPERIMENTS

Fig. 2: A slice of each of the two 4D gain maps for the initial pose $(0, \frac{\pi}{4})$ at different time-points during learning. The color coding of the gain maps corresponds to [-0.5, 2.5].

A. Gain maps learning

Recently, we have demonstrated how the presented architecture is capable of learning to perform precise saccades and to adapt to changes in the environment or in the sensorimotor plant [11]. Here, we modified the learning processes by initialising the gain maps with small random numbers and simulating a more natural learning process, in which the maps are learned in a less controlled learning session. Fig. 2 shows the gain maps, learned by the modified system. To learn the maps, an object was randomly placed in front of the robot in such a way that over several trials the whole workspace was sampled: both the visual space and the space of initial poses of the camera head. When the target object was selected and stabilised in the retinal target DNF, a saccade was generated, which, however, in the beginning

of the learning process did not bring the target object to the center of visual field. The retinal error was detected and its sign and magnitude were determined in the error estimation module for each of the two dimensions of the retinal space. The gain maps were adjusted accordingly.

In the figure, you can see how the structure emerges in two of the gain maps (one for the pan and one for the tilt motors of the robot). The shown maps correspond to the following initial pose of the camera head before each saccade: pan = 0 and tilt = $\frac{\pi}{4}$ (the overall motor space was sampled in the range of pan = $[-\pi, \pi]$ and tilt = $[0, \frac{\pi}{2}]$). Note the curved shape of the mappings, which results from non-linearity of the sensorimotor transformation from the image space to the space of motor poses of the camera head.

B. Formation of motor-based representation

Snapshot 1: before first gaze

Snapshot 2: after gaze 1

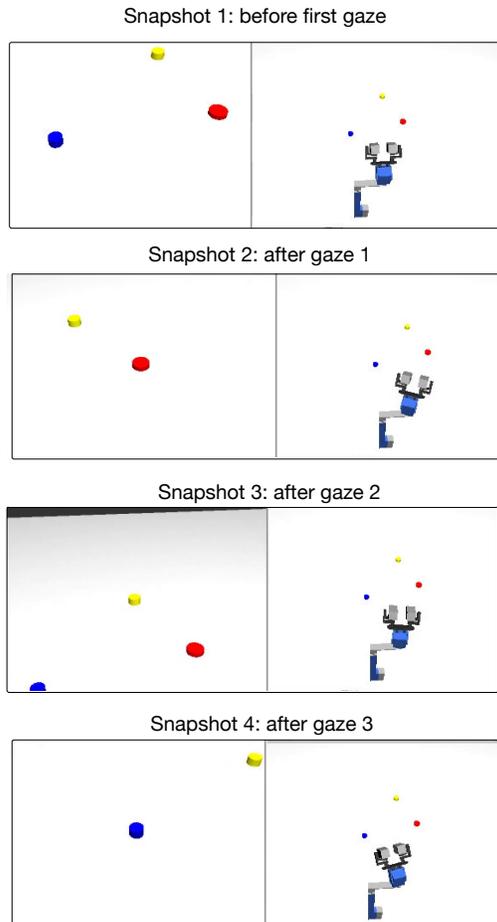Snapshot 3: after gaze 2

Snapshot 4: after gaze 3

Fig. 3: The simulated robotic setup during memory formation.

Fig. 3 shows the simulated robot and the scene, rendered by its camera. The three objects are fixated by the robot sequentially, driven by the saliency of the color blobs and a habituation dynamics, which makes the color of the previously fixated objects less attractive for further fixations. The allocentric (gaze-based) scene representation, stored in the gaze-color memory DNF, for this scene is shown in Fig. 4.

Only the spatial projection of the memory DNF is shown here. Each activity "bump" (red region in Fig. 4) is located in the third, color, dimension of this three-dimensional DNF according to the perceived color of the respective object.
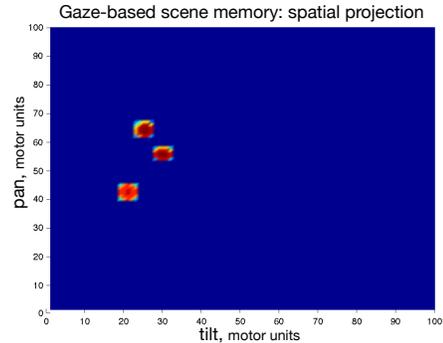
Fig. 4: The gaze-based memory for the visual scene.
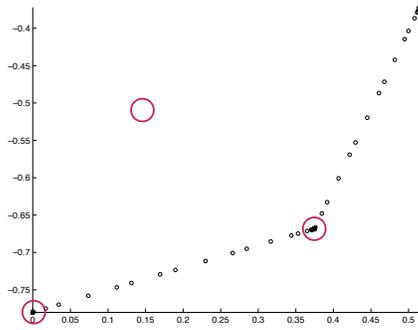
C. Double-step saccades

The retinotopic pathway (orange in Fig. 1), after learning the gain maps, leads to precise saccades to visually perceived targets (see Fig. 3). Even for these visually-driven saccades, the non-retinal information is used, since selection of the correct gain map relies on the proprioceptive information about the gaze direction prior to the saccade[2]. However, the output of the gain maps encodes the saccade amplitude implicitly, by defining the peak velocity of the movement (remember that the output of the gain map serves as a scaling factor for the oscillation, generated by the saccade generator and that the motor signal is the velocity signal). Thus, this representation does not automatically lead to correct saccades from memory. In particular, if two targets are stored in the retinotopic representation (i.e. in the retinal multipeak DNF in our model), the location of the second target has to be updated after the saccade to the first target.

The green pathway in Fig. 1 performs such an update. Moreover, in this pathway the targets of all saccades (including the first one) are stored in a gaze-based representation. The motor signal generated by the retinotopic pathway, in this case, does not go to the motors of the robot, but is instead integrated internally and the integrated value is used to predict the pose of the eye, which it would have if the eye movement were executed. This pose is stored in the gaze planning DNF. Moreover, a number of future targets may be stored in this field sequentially, as described in Section III-C. When the system is allowed to perform eye movements, the gaze-based target representation is used to derive the amplitude of the movement based on the sustained target representation in gaze coordinates and the current position of the eye, expressed in the same coordinates, as delivered by the proprioceptive inputs. The resulting activation is used to scale the output of the saccadic time-profile generator, the

---

[2]The gain maps in the Fig. 2 look differently for different initial gaze directions of the camera head because of the geometry of the robot; similar effect is caused by the anatomy of eye muscles in biological systems

initiation node of which is now activated by the gaze target DNF.

Fig. 5 shows traces of two saccades, performed using the retinal frame of reference (top), and using the gaze-based reference frame (bottom). The locations of objects, used in this experiment are shown as red circles in the figure.

(a) Retina-memory saccades ('young model').
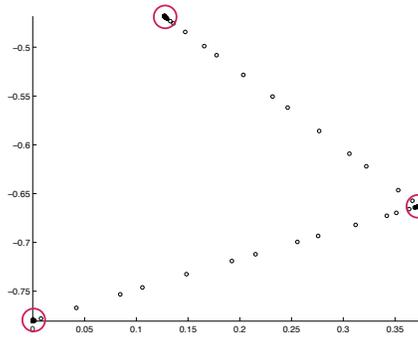
(b) Motor-memory saccades ('older model').

Fig. 5: Projection on the table-top surface of the gaze-angles of the camera during the two saccades in the simulated double-step experiment

V. CONCLUSION

In this paper, we have further explored capabilities of the recently introduced neural-dynamic model for generation of adaptive looking behaviour. The emphasis here was on formation of gaze-based memory and generation of memory saccades, which requires an allocentric representation of the visual targets. We have incorporated both retinotopic and gaze-based pathways in our model, which allows to use this model to study the developmental process, leading to more frequent and fast recruitment of the more allocentric representations in generation of saccadic eye movements and other object-directed actions. We used integration of the motor signal and creation of an anticipatory motor representation to create a gaze-direction independent target representation.

ACKNOWLEDGMENT

REFERENCES

[1] R. N. Aslin, "Perception of visual direction in human infants," *Visual perception and cognition in infancy*, pp. 91–119, 1993.

[2] J. J. Gibson, "The perception of the visual world." 1950.

[3] G. Gancarz and S. Grossberg, "A neural model of saccadic eye movement control explains task-specific adaptation," *Vision Research*, vol. 39, no. 18, pp. 3123–3143, 1999.

[4] B. Girard and A. Berthoz, "From brainstem to cortex: computational models of saccade generation circuitry." *Progress in neurobiology*, vol. 77, no. 4, pp. 215–51, Nov. 2005.

[5] M. Ebadzadeh and C. Darlot, "Cerebellar learning of bio-mechanical functions of extra-ocular muscles: modeling by artificial neural networks," *Neuroscience*, vol. 122, no. 4, pp. 941–966, Jan. 2003.

[6] J. J. Hopp and A. F. Fuchs, "The characteristics and neuronal substrate of saccadic eye movement plasticity." *Progress in neurobiology*, vol. 72, no. 1, pp. 27–53, Jan. 2004.

[7] D. Pelisson and N. Alahyane, "Sensorimotor adaptation of saccadic eye movements," *Neuroscience & Biobehavioral Reviews*, vol. 34, pp. 1103–1120, 2010.

[8] A. Gail and R. Andersen, "Neural dynamics in monkey parietal reach region reflect context-specific sensorimotor transformations." *The Journal of neuroscience*, vol. 26, no. 37, pp. 9376–84, Sep. 2006.

[9] R. O. Gilmore and M. H. Johnson, "Body-centered representations for visually-guided action emerge during early infancy," *Cognition*, vol. 65, pp. 1–9, 1997.

[10] G. Schöner, "Dynamical Systems Approaches to Cognition," in *Cambridge Handbook of Computational Cognitive Modeling*, R. Sun, Ed. Cambridge, UK: Cambridge University Press, 2008, pp. 101–126.

[11] C. Bell, T. Storck, and Y. Sandamirskaya, "Learning to Look: a Dynamic Neural Fields Architecture for Gaze Shift Generation," in *International Conference for Artificial Neural Networks, ICANN*, 2014.

[12] T. Elliott and N. R. Shadbolt, "Developing a Robot Visual System Using a Biologically Inspired Model of Neuronal Development," *Robotics and Autonomous Systems*, vol. 45, pp. 111–130, 2003.

[13] F. Chao, M. H. Lee, and J. J. Lee, "A developmental algorithm for ocular-motor coordination," *Robotics and Autonomous Systems*, vol. 58, no. 3, pp. 239–248, Mar. 2010.

[14] S. Amari, "Dynamics of pattern formation in lateral-inhibition type neural fields," *Biological Cybernetics*, vol. 27, pp. 77–87, 1977.

[15] Y. Sandamirskaya, S. K. Zibner, S. Schneegans, and G. Schöner, "Using Dynamic Field Theory to extend the embodiment stance toward higher cognition," *New Ideas in Psychology*, vol. 31, no. 3, pp. 322–339, Dec. 2013.

[16] W. Erlhagen and E. Bicho, "The dynamic neural field approach to cognitive robotics." *Journal of neural engineering*, vol. 3, no. 3, pp. R36–54, Sep. 2006.

[17] Y. Sandamirskaya, "Dynamic Neural Fields as a Step Towards Cognitive Neuromorphic Architectures," *Frontiers in Neuroscience*, vol. 7, p. 276, 2013.

[18] C. Wilimzig and G. Schöner, "The emergence of stimulus-response associations from neural activation fields: Dynamic Field Theory," in *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, Stresa, Italy, 2005, pp. 2359–2364.

[19] S. Grossberg and M. Kuperstein, *Neural dynamics of adaptive sensory-motor control: Ballistic eye movements.* Elsevier, 2011, vol. 30.

[20] Y. Sandamirskaya and T. Storck, *Artificial Neural Networks Methods and Applications.* Springer, 2014, in press, ch. Learning to look and looking to remember: a neural-dynamic embodied model for generation of saccadic gaze shifts and memory formation.

[21] S. Grossberg, "Behavioral Contrast in Short Term Memory: Serial Binary Memory Models or Parallel Continuous Memory Models?" *Journal of Mathematical Psychology*, vol. 17, pp. 199–219, 1978.

[22] Y. Sandamirskaya, M. Richter, and G. Schöner, "A neural-dynamic architecture for behavioral organization of an embodied agent," in *IEEE International Conference on Development and Learning and on Epigenetic Robotics (ICDL EPIROB 2011)*, 2011.