Modeling Perspective-Taking upon Observation of 3D Biological Motion

Fabian Schrodt*, Georg Layher†, Heiko Neumann† and Martin V. Butz*
*Cognitive Modeling, Department of Computer Science,
University of Tübingen, Sand 14, Tübingen, 72076 Germany
Email: tobias-fabian.schrodt@uni-tuebingen.de, martin.butz@uni-tuebingen.de
†Institute for Neural Information Processing, Faculty of Engineering and Computer Sciences,
Ulm University, James-Franck-Ring, Ulm, 89081 Germany,
Email: georg.layher@uni-ulm.de, heiko.neumann@uni-ulm.de

***Abstract*—It appears that the mirror neuron system plays a crucial role when learning by imitation. However, it remains unclear how mirror neuron properties develop in the first place. A likely prerequisite for developing mirror neurons may be the capability to transform observed motion into a sufficiently self-centered frame of reference. We propose an artificial neural network (NN) model that implements such a transformation capability by a highly embodied approach: The model first learns to correlate and predict self-induced motion patterns by associating egocentric visual and proprioceptive perceptions. Once these predictions are sufficiently accurate, a robust and invariant recognition of observed biological motion becomes possible by allowing a self-supervised, error-driven adaption of the visual frame of reference. The NN is a modified, dynamic, adaptive resonance model, which features self-supervised learning and adjustment, neural field normalization, and information-driven neural noise adaptation. The developed architecture is evaluated with a simulated 3D humanoid walker with 12 body landmarks and 10 angular DOF. The model essentially shows how an internal frame of reference adaptation for deriving the perspective of another person can be acquired by first learning about the own bodily motion dynamics and by then exploiting this self-knowledge upon the observation of other, relative, biological motion patterns. The insights gained by the model may have significant implications for the development of social capabilities and respective impairments.**

Keywords: Correspondence problem; mirror neurons; biological motion; perspective-taking; canonical views; recurrent neural networks.

I. Introduction

This paper addresses the question how we may be able to take the perspective of another person when we observe their bodily motion. The capability of learning by imitation has been attributed to the mirror neuron system [1]. However, preceding the activation of mirror neurons upon action observation, a transformation of the egocentric frame of reference to the observed person seems necessary to solve the correspondence problem [2] – which is a process that is hard-coded by most models on imitation learning. We propose an artificial neural network (NN) model that is able to solve this problem by deducing another point of view on the fly, which is a capability often referred to as perspective-taking [3]. This is accomplished by utilizing correlations and predictions about the own, embodied motion patterns. As a result, our model offers an explanation for the development of mirror neurons and their property to equate self-perception with observation.

The NN model we propose memorizes biological motion by encoding episodes of movements in a multiply-invariant space, which integrates positional and angular bodily features. It segments those episodes by events of significant motion non-linearities. Upon the recognition of a learned motion segment, the network predicts subsequent motion over a marginal time span to enable self-supervised learning and adaptation. Further, our model is able to learn multiple *canonical* perspectives on biological motion patterns. This is in accordance with neural areas that contribute to biological motion perception (such as STS and premotor areas), which were reported to show both view-dependent and view-independent neural responses [4], [5].

We assume that human imitative abilities are to some extent enabled by spatial visualizations of specific altercentric perspectives: On observation of a movement, a canonical perspective is taken, for which embodied associations to actions are available. This notably includes the perspective of the observed person. Comparably, specific canonical views of objects form attractors for mental rotation in object recognition [6]. We show that perspective-taking can be achieved by minimizing the divergence between observed and memorized motion patterns, which originally stem from embodied, visuo-proprioceptive associations: The continuous adaptation of the visual frame of reference is driven by patterns of *view-dependent relative positional* (visual) motion, while the recognition of biological motion is vigorously improved by the correspondence to *view-independent angular* (proprioceptive) motion.

Previously, we have shown that our model can realize continuous mental rotations towards canonical views of a simulated 2D arm motion [7]. Here, we enhance the model to 3D vision, adapt the prediction mechanism for faster convergence, and investigate the properties and performance in additional experiments using a simulated full body model.

†GL and HN have been supported by the SFB Transregio 62 funded by the German Research Foundation (DFG).

In the following, we first introduce the neural architecture for learning biological motion, building canonical views, and progressive perspective-taking. After we introduce our 3D simulation environment, we evaluate the model in several experimental setups (Section III) showing robust learning of one or multiple views on biological motion and the flexible adaptation of the internal perspective upon the presentation of novel views. In Section IV, we summarize the results, draw conclusions, and sketch-out future research perspectives.

II. Neural Network Model

The model consists of three successive stages illustrated in the overview given in Fig. 1. The first stage processes relative positional and angular values into mentally rotated direction sensitive population codes. The second stage performs a modulatory normalization and pooling of information. Stage III is a dynamic, self-supervised adaptive resonance model. It uses instar-learning to segment the pooled sensory stream given by Stage II memorizing recurring correlations, and outstar-learning to recall and predict the learned correlations. The predictive structures learned by the outstar process also enable the derivation of a prediction error. We detail the three stages and the involved techniques in the following sections.

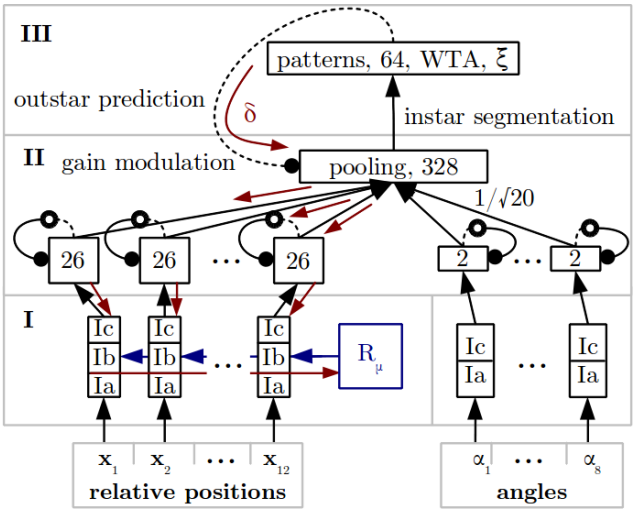

Fig. 1. Overview of the three-stage neural modeling approach in a 3D example with 12 positional and 8 angular features, resulting in $n = 20$ features. Boxes numbered with m indicate layers consisting of m neurons. Black arrows describe weighted forward connections, while filled circle arrowheads indicate modulations. Dashed lines denote recurrent, delaying connections. Blank circle arrowheads denote normalizing neurons. Red backwards connections indicate the flow of error back-propagation signals.

A. Stage I - Feature Preprocessing

The input to the network is driven by a rather arbitrary subset of relative visual positions of bodily landmarks and proprioceptive angles that jointly represent a body structure. Only features that vary during movements should be chosen, because the model will basically work in the motion domain. The network is initially driven by self-perceptions, bootstrapping the egocentric perspective. Further canonical perspectives can be learned upon observation of others, or by spatial visualization of altercentric perspectives on self-induced motion. At this time, we assume that the network's input signals can easily be recognized both during self-observation as well as during the observation of another person.

Fig. 2 (a) shows the model's feature processing for a single, relative, 3D feature position (e.g. the right wrist position relative to the center of the observed body): In interstage Ia, the relative position is transformed into a directional velocity by time-delayed inhibition, resulting in translation invariant representation. Interstage Ib implements a mental rotation of this directional velocity using gain field-like modulation [8] by a neural rotation module. The connectivity essentially realizes a 3D matrix multiplication

$$R_\mu \Delta\vec{x} = \Delta\vec{x}'$$

of a directional velocity $\Delta\vec{x} = (\Delta x, \Delta y, \Delta z)^T$ into a transformed directional velocity $\Delta\vec{x}' = (\Delta x', \Delta y', \Delta z')^T$ by an arbitrary rotation matrix R_μ. We chose $R_\mu = R_z R_y R_x$ to realize a Tait-Bryan rotation. Each rotation $R_{\{x,y,z\}}$ is implemented as a population of 3x3 neurons, which represent the elements of the 3D rotation matrix. It is driven by a mental rotation angle $\mu_{\{x,y,z\}}$, which is realized by a bias neuron. Multiplication of the rotation populations is again accomplished by gain-field modulation. In this sense, R_μ is driven by three variable mental rotation angles μ_x, μ_y, μ_z, which determine the degree of rotations about extrinsic axes. The same mental rotation R_μ is applied to all positional processing stages, by which multiple error signals can be merged at the model's rotation module. This simultaneous adaptation of the axes with respect to the integrated error allows the derivation of the shortest path rotation towards an error-minimal visual perspective. Interstage Ic implements directional convolution over time. Multiplying the rotated directional motion by a directional weighting matrix W converts the motion signals into a set of direction-selective neural activities. The weighing matrix is set up in a combinatorial fashion, as every single dimension of a D-dimensional input may increase, not change, or decrease, resulting in $3^D - 1$ direction-sensitive neurons (disregarding constant information).

The processing of each one-dimensional angular information, which is shown in Fig. 2 (b), is done analogously. A rotation mechanism for angles is not necessary and thus not applied. In summary, stage I provides a population of neurons for each feature of sensory processing, which is either sensitive to directional changes in a body-relative position (26 neurons for each position) or sensitive to directional changes in proprioceptive angles (2 neurons for each angle).

B. Stage II - Normalization and Pooling

Stage II first implements a separate normalization of activity in the direction-sensitive populations, by which the model becomes scale- and speed-invariant for each considered feature. That is, only the motion directions are regarded. Normalization of a layer's activity-vector can be achieved by

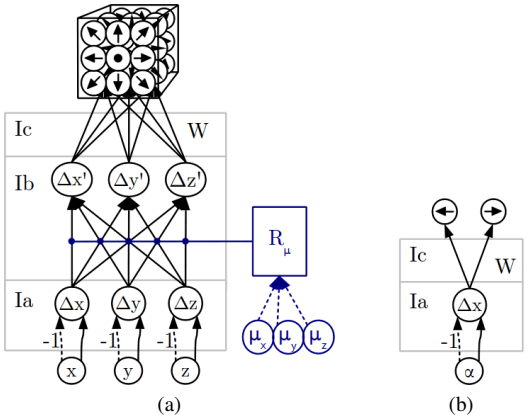

Fig. 2. Feature preprocessing stage I for (a) a single, visual, relative 3D position (rotation (in blue) is applied on all axes) and (b) a single, proprioceptive 1D angle.

gain modulation: The method we propose approximates a real-time normalization of a layer's output-vector to the Euclidean length 1. In our model, a common neuron indexed by j can formally be described by its input net$_j$, its activation function $f_j(\text{net}_j)$, its output o_j, some noise-term ξ_j (which we will address later), and the axonic modulatory factor a_j:

$$o_j = a_j \cdot f_j(\text{net}_j) \tag{1}$$

$$\text{net}_j = \xi_j + \sum_i w_{ij} \cdot o_i \ . \tag{2}$$

Normalization of a layer a can be put into execution by modulating all neurons j of that layer by the output o_a of a single, layer-specific normalizing neuron ($a_j := o_a$)[1], with

$$o_a(t) = \frac{o_a(t-1)}{\sum_j \overline{o_j}(t)^2} \ , \tag{3}$$

where $\overline{o_j}(t)$ denotes the moving average of $o_j(t-1)$.

Next, all normalized direction-sensitive fields are merged by one-to-one connections to a pooling layer (that is, without reducing the dimensionality), which serves as the input to the following Stage III. The connections are weighted by $1/\sqrt{n}$, where n denotes the number of features being processed, which ensures that the pooling layer input is normalized.

C. Stage III - Correlation Learning

Stage III realizes a clustering of the normalized and pooled information from Stage II (indexed by i) by Hebbian learning of weights fully connecting the pooling layer to a number of pattern-responsive neurons (indexed j). Each pattern neuron represents a constellation of positional and angular directional movements via its instar weight vector $\vec{w}_j$. The weights w_{ij} to a pattern j are trained by the instar learning rule [9]:

$$1/\eta \cdot \partial w_{ij}(t)/\partial t = \Delta w_{ij}(t) = o_j(t) \cdot (\text{net}_i(t) - w_{ij}(t)) \tag{4}$$

with learning rate η. We use winner-takes-all competitive learning [10] in the sense that only the most active pattern is adapting and predicting.

[1]No square root is necessary for the normalization to length 1.

Since the patterns to learn are initially typically unknown, we propose to bootstrap the weight vectors from scratch ($w_{ij}(t_0) = 0$) – which means that initially no sensory information is propagated to the pattern layer ($o_j(t_0) = 0$). For Hebbian learning to initially occur, we add standardized normal distributed neural noise $\xi_j = \mathcal{N}(0, \sigma)$ to the input net$_j$ of each neuron in the pattern layer (see Eq. (2)), such that pattern neurons are driven by the sum of signal and noise.

Yet to account for the prerequisite of normalized weight-vectors when using instar-learning, we assume that the excitability of a pattern neuron decreases proportional to its overall synaptic strength:

$$f_j(\text{net}_j) = \text{net}_j \cdot \min(||\vec{w}_j||^{-1}, r) \ , \tag{5}$$

where r denotes the initial or maximum responsiveness. In this way, the weight vector to a pattern neuron is virtually normalized. Also, on development of a pattern, its winning probability is magnified by the angle between the presented pattern and the memorized pattern (encoded by the instar weight vector), while the relative influence of neural noise decreases. Both the amount of neural noise – determined by σ – and the initial responsiveness r play crucial roles for the distribution of the network's pattern capacity: While σ influences the probability that a developed pattern is retrained, $\sigma \cdot r$ as well as the capacity of free patterns (which consists of the number of neurons with an instar weight-length below r^{-1}) determine the probability that an undeveloped pattern wins over a developed one and is thus consulted to increase the spatial resolution of this sensory episode. The transition performance from untrained to trained patterns and the recoding of trained patterns can be controlled by the learning rate η. This neural noise mechanism in combination with winner-takes-all learning can successfully avoid a "catastrophic forgetting" or constant recoding of patterns once learned without making prior assumptions about the input space.

Additional to the instar segmentation of information, we apply a predictive outstar learning and attentional gain control mechanism, by which Stage III becomes a self-supervised adaptive resonance model [11]. Outstar learning is realized by feedback connections from the pattern layer to the pooling layer, which are trained by:

$$1/\eta \cdot \partial w_{ji}(t)/\partial t = \Delta w_{ji}(t) = o_j(t-1) \cdot (\text{net}_i(t) - w_{ji}(t)) \ , \tag{6}$$

where neuron j is the winner of time step $t-1$. This means that the outgoing weight vector of a pattern neuron predicts the input of the pooling layer at the next time step.

The absolute outstar learning signal is also used on forward propagation from the winner of time step $t-1$ as modulatory gain in the pooling layer i:

$$a_i(t) := 1 - |\Delta w_{ji}(t)| \in [0, 1] \ . \tag{7}$$

By this modulation, the last winner inhibits the pooling layer's output (via Eq. 1): The larger the error in and the larger the reliability of the prediction, the stronger is the resulting inhibition (cf. Eq. 6), such that other patterns are more likely

to win in the next time step. In result, the pattern distinction is improved further.

On the other hand, the negative of the above outstar learning signal is backpropagated top-down through the network to adapt the mental rotation in an error-minimizing manner: An error signal δ_i – depicting the actual prediction error – is directly fed into each pooling neuron i by the last pattern winner j:

$$\delta_i(t) = -\Delta w_{ji}(t) \ . \tag{8}$$

This error signal is on backpropagation split by the feature specific population codes and finally merged at the rotational module, as shown in Fig. 1, where it is used to adapt the bias neurons online. This essentially minimizes the difference between the predicted and the perceived biological motion, while the adaptation is restricted to a 3D rotation via the gain-field modulation.

III. Experiments

To evaluate the model, we simulated a 3D stick figure walker and streamed relative joint and end-point locations as well as joint angles as sensory information into the NN model. We show that the model is able to segment this information when perceived from an egocentric perspective. We then show that the learned structure allows for a self-supervised, progressive view-point adaptation towards the learned egocentric perspective when similar biological motion is perceived from other perspectives. Moreover, we show that additional canonical views of motion can be memorized. Finally, we show that when several canonical views have been learned the model is transforming randomly oriented biological motion towards the closest canonical view.

For the following experiments, we chose $\eta = 0.01$ as instar/outstar learning rate, $\sigma = 0.002$ as pattern noise standard deviation, and $r = 100$ as maximum pattern responsiveness. On perspective adaptation, we trained the mental rotation bias neurons with learning rate 0.35 and used a momentum of 0.5. The reported results are averaged over 400 independent runs (training and evaluating the network starting with different random number generator seeds).

A. Simulation and Setup

We implemented a 3D simulation of a humanoid walking with 10 angular DOF (see Fig. 3). One step of the walker consists of 100 time steps, so that the movement is cyclic with a period of 200 time steps. As input to the model, the simulation provides the 3D positions of 12 joints or limb endpoints relative to the body's center $\vec{x}_1$ ... $\vec{x}_{12}$ as well as 8 joint angles α_1 ... α_8 (leaving out inner rotations). The view of the walker can be rotated for experimental purposes by an arbitrary rotation matrix R_ν, which is unknown to the model.

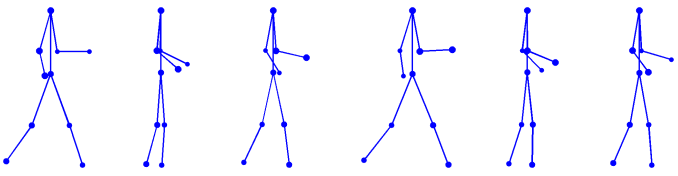

Fig. 3. 3D simulation of a humanoid walking to the right. Depth is visualized by the size of the points.

B. Learning of Egocentric Biological Motion

As a first step, we trained the network on the egocentric perspective of the simulated humanoid walking movement for 20k time steps (which equates to 100 repetitions of the movement). Using the parametrization above, on average, six patterns evolved from noise. The patterns form a cyclic series of winners over the repetition of the movement. Fig. 4 shows the activities of six pattern neurons for the first 1000 time steps of a representative run. It can be seen that the signal-to-noise ratio in the patterns' activities increases significantly during training, because increasing the length of a pattern's instar vector decreases its response to noise. Likewise, the winning probability of free patterns decreases when the activity of a trained pattern is high.

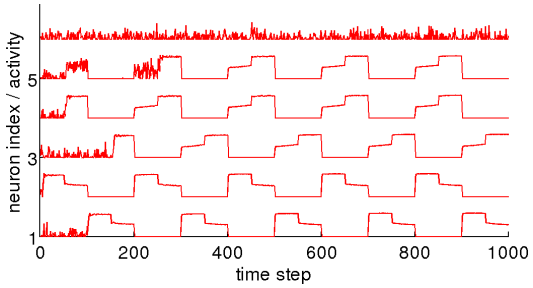

Fig. 4. Motion segmentation by the network. Neural noise initiates the development of patterns for each sufficiently diverse movement episode.

C. View-Independent Perspective-Taking

After learning the egocentric view on a self-induced motion as detailed above, the model is able to transform observed, similar motion into the learned frame of reference by adapting its visual perception. Because the model is generally scale and translation invariant, this adaption only includes the adjustment of yaw, pitch, and roll of its internal coordinate system, which we evaluated by examining the resulting model rotation matrix R_μ. In this experiment, an unfamiliar perspective on the learned movement was fed into the visual path of the model by rotating the walker arbitrarily using an uniform distribution in orientation space for the simulation's rotation matrix R_ν. The angular path of the network perceived the proprioceptive angles, assuming they can be determined sufficiently accurate from vision. This increases the pattern recognition ability of the network, even if proprioceptive information can only partially be derived visually. We let the model's rotation matrix R_μ adapt for 10k time steps according to the prediction error backpropagated to the visual rotation module. The progress of the resulting spatial visualization of the model is shown in Fig. 5.

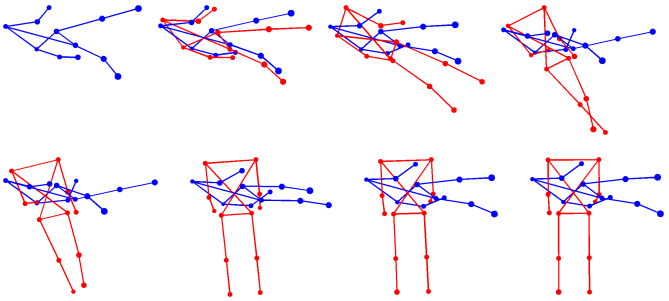

Fig. 5. From top-left to bottom-right: online adaptation of the visual perception. The model continuously rotates observed motion (shown in blue) with convergence to the orientation it was trained on (in red: perspective deduced by the model).

As a further measure of the model's performance, we analyzed the orientation difference (OD) between the model's derived inner view on the presented motion and the desired view, which was previously trained on. We define the OD as the minimal absolute angle of rotation needed about an arbitrary axis to transfer the deduced orientation (determined by R_μ) into the target orientation (determined by R_ν^{-1}). Fig. 6 illustrates the model's adaptation in terms of OD to the egocentric perspective: Altogether, 52% of all runs converged to the desired view of the movement. For these runs, the median OD fell below 1° after 43 time steps, which equals less than half a step of the simulated walker. After 10k time steps, the median of the OD was 0.468°. There were no outliers in our experiments.

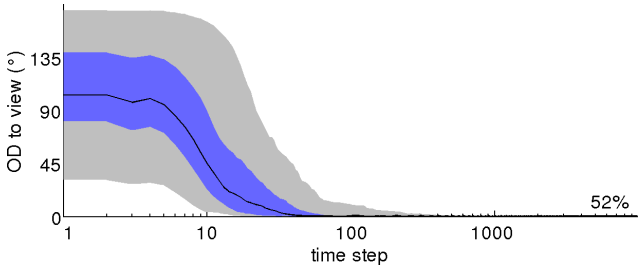

Fig. 6. The OD to the egocentric perspective decreases over time. Continuous box-plot notation: gray: whiskers (2.5 and 97.5% quantiles), blue: quartiles, black: median.

The residual 48% of the runs converged to an implicitly learned, top-down inverted perspective: Since each pattern predicts over a marginal time-span, it predominantly predicts the progress of motion in its own respective motion segment. Thus, the prediction is rather independent from the actual sequence of patterns, but considers each motion segment independently. Based on our choice of the input space, top-down inverted views on a cyclic movement can result in the inverse order of patterns with worse, but locally minimal prediction error. Thus, the model converges to a top-down inverted view on the movement primarily in cases where the initially observed rotation of the upright direction exceeds 90°. We evaluated the probability of convergence to the egocentric perspective in terms of initial OD in Fig. 7.

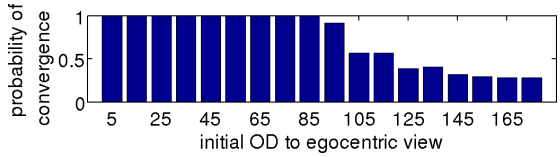

Fig. 7. Percentage of runs converging to the egocentric perspective dependent on the initial OD when a new perspective is shown. Bins: $x \pm 5°$ OD.

D. Learning Multiple Canonical Views

In the following, we investigate if additional canonical perspectives on biological motion can be learned and maintained by the model parallel to the egocentric perspective. We trained the network sequentially on the *egocentric* view as before, and additionally on the *facing* (180° vertical rotation), the *left* (-90° vertical rotation), and the *right* (90° vertical rotation) view on the walking. We trained each of the four perspectives for 25 repetitions of the movement and repeated the whole procedure four times, resulting in $80k$ iterations in total.

When training the network on additional data, new pattern neurons may be recruited or already trained but similar pattern neurons may be recoded (probabilities for both depend on the pattern noise parametrization). This applies for different perspectives as well as basically different movements. Using the foregoing parametrization, after learning, distinct groups of pattern neurons were responsible for the particular views as shown in Fig. 8, indicating that the different views could nicely be separated by view-dependent neurons.

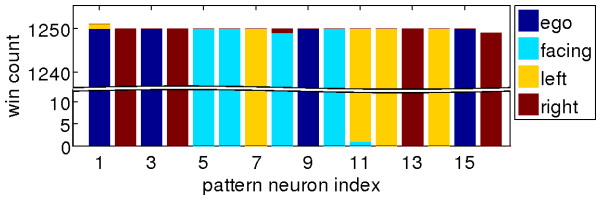

Fig. 8. Separate representation of canonical views. At the end of training, the four canonical views are represented in disjunct groups of pattern neurons. This can be seen by the number of time steps a pattern neuron is maximally active (winning) while a specific perspective is trained.

When monitoring the prediction error $\sqrt{\sum_i \delta_i^2}$ during learning the four canonical views, Fig. 9 shows a peak in the prediction error every time the perspective was changed. However, the magnitude of the peaks strongly decreases after the fourth view change, which is when the first view is shown for the second time, and it continues to decline further with increasing repetitions, since recoding and recruitment of new patterns reduces continuously. Overall, the error converged to a level below $0.1°$ OD.

E. Perspective-Taking of Canonical Views

After training the network on the four canonical perspectives, instar and outstar learning in the highest layer was disabled but the self-supervised, error-driven adaptation of the perspective was activated for 10k time steps. The properties of convergence are shown in Table I: It can be seen that the

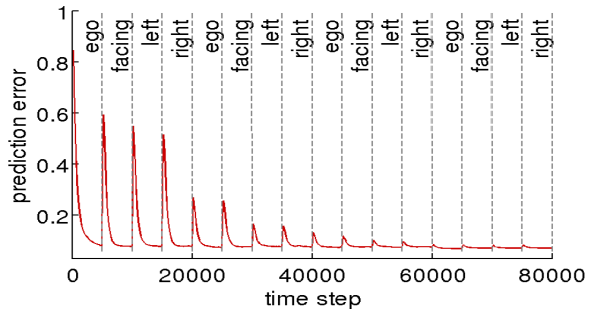

Fig. 9. Prediction error while learning from multiple canonical perspectives repeatedly.

TABLE I
PROPERTIES OF CONVERGENCE TO FOUR CANONICAL VIEWS.
MROD: MEDIAN REMAINING ORIENTATION DIFFERENCE
MCT: MEDIAN CONVERGENCE TIME

View	runs	MROD	MCT
Ego	10.25%	0.53283°	20
Facing	12.25%	0.53124°	23
Left	12.5%	0.53428°	24
Right	11.5%	0.53056°	28

learning of additional canonical views of biological motion sped up convergence, because the next canonical view was closer on average. The precision of perspective-taking was hardly influenced.

Again, there was an implicitly learned, locally error-minimal, top-down inverted attractor for each perspective trained on (not shown in the table). Because of this, approximately $1/8$ of the runs converged to each perspective (differences between individual or groups of perspectives are not statistically significant).

IV. CONCLUSION & FUTURE WORK

We have shown that our model is able to segment and memorize 3D biological motion in a scale- and translation-invariant space of visual and proprioceptive motion. We clarified that perspective-taking can account for orientation invariance in biological motion recognition: A spatial visualization of foreign motion in the egocentric or canonical frames of reference can thereby ensure the correspondence between self-induced and observed biological motion. Thus, our model offers an explanation for the (strictly congruent) mirror neuron property of both view-dependent as well as view-independent cells in the superior temporal sulcus (STS), which other theories on imitation, such as associative sequence learning [12], are unable to explain.

When the model is trained on data that produces a specific sequence of patterns, all reordered sequences of those patterns are implicitly learned simultaneously. This can result in the ability to recognize unknown movements, but also in the development of additional, especially top-down inverted perspective attractors on cyclic movements. Yet, difficulties on recognition of top-down inverted biological motion from point-light displays, as well as the unaffected recognition of reversed (or sped-up) motion determined in psychological experiments [13], [14] comply with our model. Further, the observation of top-down inverted motion without additional clues depicts a rather unusual situation in everyday life. Although pattern reordering may thus be an approach for generalizing motion perception and perspective-taking, further investigations have shown that explicit forecasting of linear motion segments can successfully ensure the adherence of biological motion sequences [15].

We assumed in this paper that all required visual and proprioceptive information is available both during self-perception as well as during observation. In future, visual features that are occluded as well as proprioceptive features that can not directly be inferred from vision when motion is observed, could be completed using variants of the predictive mechanisms described in this paper. Another open question is how mere feature positions can automatically be assigned to the specific network inputs without explicit labeling. We are currently investigating these and related issues to further improve the generality of our model as well as to rigorously test it on available data and neuro-psychological results.

REFERENCES

[1] G. Rizzolatti and L. Craighero, "The mirror-neuron system," *Annu. Rev. Neurosci.*, vol. 27, pp. 169–192, 2004.
[2] C. L. Nehaniv and K. Dautenhahn, "The correspondence problem," *Imitation in animals and artifacts*, p. 41, 2002.
[3] M. Hegarty and D. Waller, "A dissociation between mental rotation and perspective-taking spatial abilities," *Intelligence*, vol. 32, no. 2, pp. 175–191, 2004.
[4] V. Caggiano, L. Fogassi, G. Rizzolatti, J. K. Pomper, P. Thier, M. A. Giese, and A. Casile, "View-based encoding of actions in mirror neurons of area f5 in macaque premotor cortex," *Current Biology*, vol. 21, no. 2, pp. 144–148, 2011.
[5] T. Jellema and D. I. Perrett, "Neural representations of perceived bodily actions using a categorical frame of reference," *Neuropsychologia*, vol. 44, no. 9, pp. 1535–1546, 2006.
[6] M. J. Tarr and S. Pinker, "Mental rotation and orientation-dependence in shape recognition," *Cognitive psychology*, vol. 21, no. 2, pp. 233–282, 1989.
[7] F. Schrodt, G. Layher, H. Neumann, and M. V. Butz, "Modeling perspective-taking by correlating visual and proprioceptive dynamics," in *Proceedings fo the 36th Annual Conference of the Cognitive Science Society*, 2014.
[8] R. A. Andersen, G. K. Essick, and R. M. Siegel, "Encoding of spatial location by posterior parietal neurons," *Science*, vol. 230, no. 4724, pp. 456–458, 1985.
[9] S. Grossberg, "On the development of feature detectors in the visual cortex with applications to learning and reaction-diffusion systems," *Biological Cybernetics*, vol. 21, no. 3, pp. 145–159, 1976.
[10] D. E. Rumelhart and D. Zipser, "Feature discovery by competitive learning," *Cognitive science*, vol. 9, no. 1, pp. 75–112, 1985.
[11] S. Grossberg, "Adaptive pattern classification and universal recoding: Ii. feedback, expectation, olfaction, illusions," *Biological cybernetics*, vol. 23, no. 4, pp. 187–202, 1976.
[12] C. Heyes, "Where do mirror neurons come from?" *Neuroscience & Biobehavioral Reviews*, vol. 34, no. 4, pp. 575–583, 2010.
[13] V. A. Kuhlmeier, N. F. Troje, and V. Lee, "Young infants detect the direction of biological motion in point-light displays," *Infancy*, vol. 15, no. 1, pp. 83–93, 2010.
[14] M. Pavlova and A. Sokolov, "Orientation specificity in biological motion perception," *Perception & Psychophysics*, vol. 62, no. 5, pp. 889–899, 2000.
[15] F. Schrodt and M. V. Butz, "Modeling perspective-taking by forecasting 3D biological motion sequences," in *Cognitive Processing (Suppl. KogWis 2014)*, 2014.