A Biologically Inspired Model for Coding Sensorimotor Experience Leading to the Development of Pointing Behaviour in a Humanoid Robot

Ivana Kajić
Bernstein Center for Computational Neuroscience
Berlin, Germany
ivana.kajic@bccn-berlin.de

Guido Schillaci, Saša Bodiroža, Verena V. Hafner
Cognitive Robotics
Department of Computer Science
Humboldt-Universität zu Berlin
Berlin, Germany

ABSTRACT

Robots are gaining more attention in the neuroscientific community as means of verifying theoretical models of social skill development. In particular, humanoid robots which resemble dimensions of a child offer a controlled platform to simulate interactions between humans and infants. Such robots equipped with biologically inspired models of social and cognitive skill development might provide invaluable insights into learning mechanisms infants employ when interacting with others. One such mechanism which develops in infancy is the ability to share and direct attention of interacting participants. Pointing behaviour underlies joint attention and is preceded by hand-eye coordination. Here, we attempt to explain how pointing emerges from sensorimotor learning of hand-eye coordination in a humanoid robot. A robot learned joint configurations for different arm postures using random body babbling. Arm joint configurations obtained in babbling experiment were used to train biologically inspired models based on self-organizing maps. We train and analyse models with various map sizes and depending on their configuration relate them to different stages of sensorimotor skill development. Finally, we show that a model based on self-organizing maps implemented on a robotic platform accounts for pointing gestures when a human presents an object out of reach for the robot.

Keywords

cognitive robotics, HRI, self-organizing maps, sensorimotor coupling, body babbling, attentional mechanism, developmental learning

1. INTRODUCTION

A great deal has been learned from computational models of processes and mechanisms underlying human cognition and behaviour. Advances in neurosciences, in particular in brain imaging techniques, increased quantity and quality of experimental data which are invaluable for theoretical researchers to improve and adapt theoretical models. However, modelling the development of cognitive and social skills continues to be a daunting task due to several unresolved issues which modellers have to face. One of the issues is the integration of mechanisms required for the ongoing development of skills within cognitive models and cognitive architectures through the interaction with the environment [20].

Epigenetic robotics in combination with neural modelling offers an optimal starting point for this investigation. Humanoid robots equipped with neurally plausible models provide a controlled environment to analyse and steer learning mechanisms infants might employ when interacting with humans. On the other hand, it is of particular interest for the robotics community to study the development of social cognition and imitation learning skills in artificial agents. In fact, building robots as interactive learners could reduce the programming efforts needed by end-users for teaching them new skills or new tasks to accomplish.

One of the foundation for imitation learning and social cognition is the capability to understand, to manipulate and to coordinate attentional behaviours [26]. Joint attention, the capability to share the focus of attention between individuals, is fundamental in human social interaction and in human-robot interaction. Its development depends on the successive appearance of a number of underlying skills, such as the capabilities to detect the focus of the attention of the interacting partner, to detect and to maintain eye contact, and to manipulate the focus of attention of the interacting partner [11]. However, there are still several open questions about how such capabilities emerge in humans. For instance, it is still unclear how infants develop the capability to manipulate the attention of an interacting person through pointing gestures.

In humans, the first occurrence of pointing gestures starts around the age of nine months [2], when infants use such a gesture as a request for a certain object. The primitive forms of pointing, known as *imperative pointing*, may arise from failed reaching actions and may not carry any intentional meaning directed at the caregiver. In fact, developmental studies suggest that there is no relation between the production of primitive pointing gestures and the comprehension of pointing [6]. Around the age of twelve months, pointing starts to become declarative and to be used to draw the caregiver's attention to something which might also be out of reach for the adult [11]. Nonetheless, the development of pointing is preceded by the acquisition of early motor competences, such as hand-eye coordination [14].

Simulating similar developmental processes in robots could

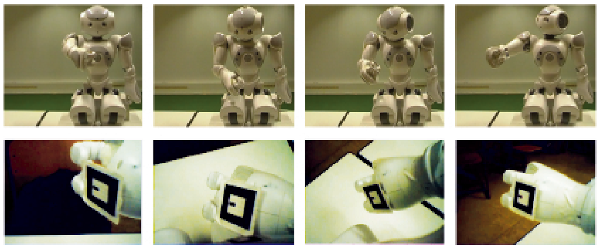

Figure 1: A sample sequence of random motor movements during a babbling phase on a Nao robot. The first row shows random hand movements, and the second row the corresponding frames captured by the onboard camera placed under the fake eyes of the Nao.

provide important insights in this investigation. In the robotics literature, interesting studies can be found on the development of motor competences in artificial agents. In [19], the author implemented on a humanoid robot an adaptive control system inspired by biological development of visuomotor coordination for the acquisition of orienting and reaching behaviours. Following a developmental paradigm, the system starts with moving the eyes only. At this point, control is a mixture of random and goal-directed movements. The development proceeds with the acquisition of closed loop gains, reflex-like modules controlling the arm sub-system, acquisition of an eye-head coordination and of a head-arm coordination map.

In [9], the development of pointing behaviours in a humanoid robot has been addressed. In particular, a humanoid robot (Aldebaran Nao) has been programmed to acquire early motor competences through an exploration behaviour, namely body babbling. Learning consisted in the robot exploring its arm movements while collecting sensorimotor data (hand and arm joint positions), thus building up an internal model of its own body. A simple predictive algorithm provided the robot with a mechanism for producing motor commands to reach desired hand positions. Pointing behaviours emerged when target points were presented outside the reach of the robot, thus strengthening the hypothesis that pointing may arise from grasping.

Here, a similar experiment is presented, where a humanoid robot acquires hand-eye coordination and reaching skills by exploring its movement capabilities through body babbling. As a result, similarly to [9], the robot shows pointing behaviours when required to approach targets outside its reach. Differently to [9], a biologically inspired model consisting of self-organizing maps (SOMs) has been used for modelling the hand-arm joints mapping. The model architecture is inspired by the Epigenetic Robotics Architecture presented in [20], where a structured association of multiple SOMs has been adopted for mapping different sensorimotor modalities in a humanoid robot.

This paper is structured as follows. Section 2 discusses the body babbling procedure that has been implemented in the humanoid robot Nao. Section 3 presents neurobiological support for the choice of the model and introduces a model consisting of SOMs trained with the information gathered during body babbling. Section 4 presents the experiment that has been carried out in this study to evaluate the model. In particular, a human subject held and moved an object in front of the robot. The trained model has been used for generating reaching actions towards the object. As a result, the robot followed the object with its head and arm, exhibiting pointing gestures when the object was outside its field of reach. Section 5 explains implications of the results and addresses the contributions of the work. Section 6 suggests future research directions.

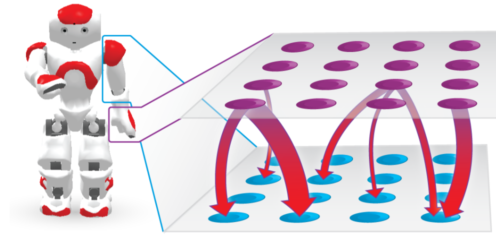

Figure 2: Scheme of the model architecture for learning hand-eye coordination with SOMs and connecting weights

2. LEARNING HAND-EYE COORDINATION THROUGH MOTOR BABBLING

Hand-eye coordination is an important motor skill acquired in infancy which precedes more complex behaviours. Infants acquire such motor skills through a self-exploring behaviour such as body babbling [17]. During body babbling, infants play with muscle movements which are then mapped to the resulting sensory consequences. This behaviour is fundamental in learning of limb postures and correlations between motor actions and resulting sensory input. In [10], the authors argued that the rise of new skills in infants can be analysed in terms of two developmental parameters: a social dimension and an intentional dimension. From both points of view, babbling falls at the zero-point, as it is a behaviour without social and intentional content.

Several robotics studies have been inspired by the infants' behaviour of body babbling. In [4] and in [24], exploration behaviours have been implemented in artificial agents for gathering evidence to form internal models of their bodily characteristics. In [5], the authors propose a way for combining knowledge through exploration and knowledge from others, through the creation and use of mirror neuron inspired internal models. In [1], an exploration mechanism driven by the active self-generation of high-level goals has been proposed. Such a mechanism allows active learning of inverse models in high-dimensional redundant robots.

We implemented acquisition of coordination skills through self-exploration via random motor babbling on the humanoid robot Nao from Aldebaran. The dimensions of a Nao resemble those of a child standing at a height of ca. 57 cm and simulating the real visual input perceived by a young human subject. A sample babbling sequence is shown in Figure 1.

3. THE MODEL

We aim to develop a model that mimics the formation of sensory maps required for the coordination of arm-hand movements at early stages of sensorimotor development. Sensory maps in the human brain contain neurons specialised in encoding specific modalities of sensory input. The plasticity in these areas is driven by the gradual formation of internal representations across the lifespan [23], [8]. The two distinct characteristics of sensory maps are the topological organization and the self-organizing property.

Topological structure is observed throughout the cortex such as somatotopic maps in the somatosensory cortex, tonotopic maps in the auditory cortex and retinotopic maps in the visual cortex. Although representing different sensory information, all these maps share the property that the similar input features are processed by the neighbouring patches of the brain tissue. For example, somatotopic maps follow the organization of the body, which means that spatially close sensory parts of the body are represented by the adjacent brain regions.

Self-organizing property becomes evident in sensory maps throughout the brain development and in response to pernicious bodily changes of sensorimotor system. For example, Farah [7] argues that the self-organization in the somatosensory maps takes place in the womb while the pose of the foetus imposes mutual touching of the face and the hands, as well as the feet and the genitals. She proposes that this might be the reason why these body parts, although not close to each other physically, are represented close to each other in the brain. However, Parpia provides arguments against causality between costimulation and somatotopy in sensorimotor cortex, suggested by Farah [22]. Nevertheless, it is important to note that both authors support that sensory information actively influences the organisation of the respective brain areas. Cortical self-organization is also apparent in case of amputated body parts. Merzenich [18] showed the reorganization of cortical maps in monkeys before and after the amputation of fingers.

Based on these insights into the way brain represents and manipulates sensory information, we decided to simulate sensory maps in our model using artificial neural networks (ANNs), which are computational algorithms inspired by the brain organization and structure. Expressed using mathematical vocabulary, an ANN is a graph whose nodes are neurons organised in a layered architecture and connecting edges among them are neural weights. Weights are computed using various learning rules such that they map the input values onto the desired output. ANNs do not capture the exhaustive level of information processing detail as observed in real biological systems, but rather attempt to approximate experimental data or phenomenologically simulate certain aspects of neural systems. The topology-preserving characteristic of a particular class of ANNs known as Kohonen networks [12] or self-organizing maps (SOMs) motivated the choice of the model. SOMs have been widely utilised in modelling of formations of different sensory modalities such as those in the auditory cortex [16], somatotopic maps [21] and orientation maps in the striate cortex [15]. One of the reasons for using a biologically-inspired model is to gain better understanding of the biological system through computational modelling. We use brain-inspired Hebbian learning paradigm to associate maps to simulate interaction between brain areas based on the interaction of the agent with the external world.

It has to be pointed out that the biologically inspired model has not been adopted just for the sake of reproducing a biological system into an artificial one. Rather, our aim is to provide an artificial system with capabilities such as autonomous learning, adaptability and plasticity. In fact, state-of-the-art robots still lack basic capabilities such as learning, adapting, reacting to unexpected circumstances, exhibiting a proper level of intelligence and autonomously and safely operating in unconstrained and uncertain environments.

Through the proposed model, a robot can autonomously build up an internal representation of its body, or of parts of it. In particular, the nature of the proposed model allows an artificial agent to build up and, eventually, to reshape its internal body representation through the interaction with the external world. In addition, a particular emphasis has been given to the developmental progression in the acquisition of motor and cognitive skills (such as attention manipulation through pointing gestures). We strongly believe that studying human development could give insights in finding those basic behavioural components that may allow for the autonomous mental and motor development in artificial agents. A robot capable of developing motor and mental skills autonomously can better deal with the aforementioned challenges related to real world scenarios.

3.1 SOMs and Hebbian learning

A SOM is constructed as a grid of neurons, where each neuron is represented as a n-dimensional weight vector $\mathbf{w_i}$. The number of dimensions of a weight vector corresponds to the dimensionality of input data. Each neuron approximates a certain region of data points in the input space yielding less units needed to represent the input.

Weights in the network are initially set to random values and then adjusted iteratively by presenting the input vector $\mathbf{x_p}$ randomly chosen from the input data. In each iteration, the winning neuron i is selected as a neuron whose weights are closest to the input vector in terms of the Euclidean distance:

$$\arg\min_i ||\mathbf{x_p} - \mathbf{w_i}|| \tag{1}$$

After selecting a winning neuron, the weights of all neurons are adjusted:

$$\mathbf{\Delta w_j} = \eta(t) h(i,j)(\mathbf{w_j} - \mathbf{x_p}) \tag{2}$$

The $\eta(t)$ parameter is a learning rate which defines the speed of change. The function $h(i)$ is a Gaussian neighborhood function defined over the grid of neurons as:

$$h(i,j) = e^{\left(\frac{\mathbf{w_i^2} - \mathbf{w_j^2}}{2\pi\sigma(t)^2}\right)} \tag{3}$$

The learning rate η and the spread of the Gaussian function σ are held constant for the first half of iterations, and afterwards are annealed exponentially. The function is centered around the winning neuron i and its values are computed for all neurons j in the grid. The spread of the function determines the extent to which neighbouring weights of a winning neuron are going to be affected in the current iteration. The topology of the network is preserved by pulling together neurons closest to the winning node. This underlies the assumption that the initial configuration of neurons in the network covers the space arbitrarily, and only upon iterative presentations of input data starts converging to the optimal state.

The activation function of a neuron, $A(\mathbf{x})$ is computed over the Euclidean distance between the neural weights and

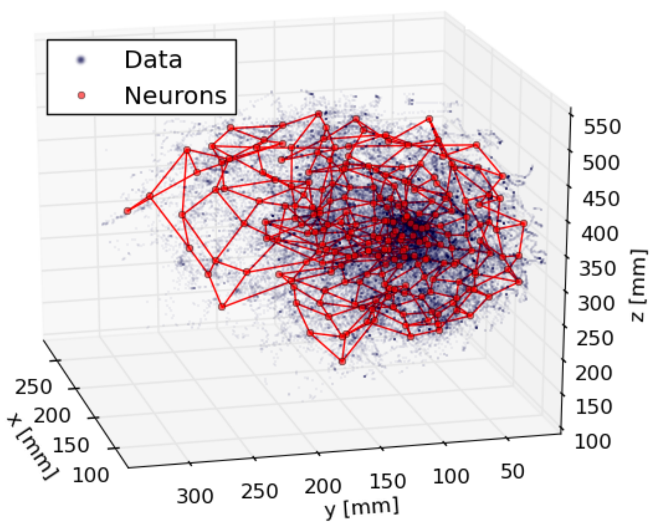

Figure 3: The SOM with 225 neurons approximating the left hand trajectory samples in the 15×15 model

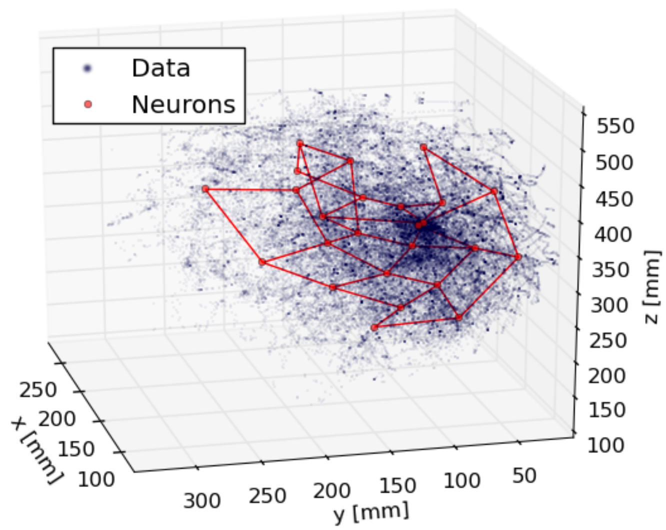

Figure 4: The SOM with 25 neurons approximating the left hand trajectory samples in the 5×5 model

the input vector, denoted with $\mathbf{x}$:

$$A(\mathbf{x}) = \frac{1}{1 + \tanh(\mathbf{x})} \tag{4}$$

It is a common practice in cognitive modelling to connect multiple SOMs using the associative links ([20], [27] and [13]). The Hebbian learning paradigm describes an associative connection between activities of two connected neurons. If a postsynaptic neuron j is always activated after the presynaptic neuron i, the connection between these two neurons is strengthened using the following rule:

$$\Delta w_{ij} = \eta_h A_i(\mathbf{x}) A_j(\mathbf{y}) \tag{5}$$

Initially, all weights between two SOMs are set to zero allowing for an activity-dependent role of structural growth in neural networks. The scaling factor η_h is chosen to be 0.01, to slow down the growth of weights.

3.2 Model architecture and training

The model consists of structured associations of two 2D SOMs with each SOM representing a different part of the left arm posture. This is schematically depicted in Figure 2 where the "blue" SOM is used to represent the elbow and shoulder positions, and the "red" SOM the hand positions.

The model was trained using data points gathered in the babbling experiment. The implementation of the babbling procedure was adapted from [25]. The robot has been provided with a simple behaviour based on sensorimotor coordination which allowed it to look at its own random arm movements. First, a motor command which is a desired angle position is sent to each joint of the left arm. When the hand of the robot, represented for simplicity by a fiducial marker[1], is detected, the joints of the neck are rotated in order to center the fiducial marker in the perceived visual input. The bottom camera placed in the robot's head has been used to capture the visual input. During the babbling process, information related to the estimated position of the marker is stored and mapped with the current configuration of the arm joints. The position of the marker is characterised by a horizontal, vertical and depth dimension. Four arm joints have been used: shoulder pitch, shoulder roll, elbow yaw and elbow roll. Together, the 3D marker position and 4D positions of joints form a 7D data point.

The babbling experiment lasted approximately 40 minutes and yielded 74,143 data points which were used to train the model. The training comprised adjustment of weights for all neurons in each SOM and it consisted of 20,000 iterations. In each iteration, a random input vector corresponding to one data point was chosen from the training set. The part of the vector describing the marker position was presented to the first SOM, and elbow and joint positions were presented to the second SOM. In both SOMs, neurons approximating inputs with lowest Euclidean distances were regarded as winning neurons. Weights were adjusted for a winning neuron and its neighbourhood. The learning rate η was set to 0.9 and the spread of the Gaussian neighbourhood function σ was 0.7. Both hyperparameters were kept constant for the first half of iterations, and afterwards annealed exponentially. After the local weight adjustment, the links between winning neurons in SOMs were computed using the Hebbian learning paradigm as explained in the previous section.

We trained two different instances of the model: one consisting of two SOMs with 50 neurons in total, where neurons in each SOM were arranged in 5 columns and 5 rows (the 5×5 model in Figure 4) and the other consisting of two SOMs with 450 neurons in total, where neurons in each SOM were arranged in 15 columns and 15 rows (the 15×15 model in Figure 3). The motivation for the two models underlies the assumption that the model containing more neurons represents the more advanced stage of pointing skill development. We assume that such stage is characterised by the increased number of specialised neurons and thus the 15×15 model might develop from the 5×5 model. Both models trained on the same data set using the same parameters. The 15×15 model was used to determine the robot's arm posture in the experiment with a human.

[1]We used the ARToolkit library for marker detection (http://www.hitl.washington.edu/artoolkit).

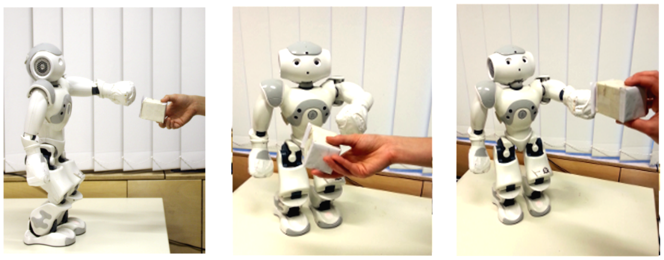

Figure 5: Pointing sequence in a human-robot interaction

We evaluate both models by comparing pointing precisions obtained in the experiment. Additionally, we theoretically analysed how the duration of babbling procedure influences the pointing precision under the assumption that shorter babbling phase corresponds to a shorter period of infant's engagement in sensorimotor self-exploration prior to pointing behaviour.

4. THE EXPERIMENT

The experiment consisted of human-robot interaction in a setting as shown in Figure 5. The human subject held an object tagged with a marker in front of the robot and moved it for approximately 2.5 minutes. Movements at varying speed were random and covered the space within and beyond the reach of the robot's hand. The robot followed the object with its head and arm.

Both the 15×15 and 5×5 models trained on babbling data were implemented on a robot. In the experiment, the 15×15 model was used to set the shoulder and elbow configuration. Although the same configurations determined by the 5×5 model were not used to set robot's joints, their values were saved and used for comparison with the values obtained by the 15×15 model. When the robot detected the object, the 3D coordinates were used to activate a neuron in the first SOM. Neuron with the strongest Hebbian link in the second SOM was chosen. Weight values of neurons were used to issue a motor command to set the configurations of the robot's joints.

4.1 Results

The 3D position of the marker was taken as a ground truth in the attempt to evaluate the precision of pointing. We plotted it against reached hand positions determined by the SOMs for both the 5×5 and the 15×15 models in Figure 6 (only for the horizontal dimension). It is important to notice that the robot's hand was in total 240 mm in length, and object positions beyond that length were not reached contributing to the high error. However, this error is not a trustworthy measure of pointing quality since we are interested in the direction of pointing along the axis rather than exact overlap of the trajectories. In the plot, one can see that the hand trajectory determined by both models approximately follows the position of a marker. The 15×15 model captures the trend with the higher precision.

As a simple quantification measure of pointing we introduced the pointing precision error, which is defined as the Euclidean distance between the 3D position of a marker and the 3D hand position. A paired-samples t-test was conducted to compare precision errors for the 5×5 model and the 15×15 model. There was a significant improvement in the performances from the 5×5 model ($Mean\ error = 99.93mm$, $Std.\ Dev. = 32.10$) to the 15×15 network ($Mean\ error = 80.32mm$, $Std.\ Dev. = 33.41$) conditions; $t(1400) = 76.47$, $p < 0.05$.

Table 1: Pointing precision (in mm) for different network sizes and different training conditions

	1% data	10% data	100% data
5×5	46.00 ± 37.75	29.82 ± 18.15	29.11 ± 14.99
15×15	39.45 ± 38.23	17.82 ± 15.78	14.18 ± 8.30

The influence of training data on pointing precision was analysed using babbling data sets of different sizes for the training of the 5×5 and the 15×15 model. We distinguished between three different cases: one training case used only the first 1% of the babbling data, the other case the first 10% of the babbling data and the last case the whole 100% of the babbling data. This amounts to six models in total whose precision accuracies were tested in a simulated experiment on a computer. After all six models have been trained, they were tested with a new babbling data set that was different from the one used for training. Data points in this set were acquired in random motor babbling experiment with the robot that lasted for approximately 30 minutes and it yielded 57,778 data points. Testing of a single model consisted in iterative presentation of the 3D hand position vector from the testing set to the first SOM, activation of a winning node and computation of Euclidean distances between the weights of a winning node and the hand position. Different training cases can be regarded as different learning processes with respect to the duration of the exploration in the random motor babbling phase. Precision errors for different training conditions and different models are presented in Table 1. The values differ from those observed in the pointing experiment, where the human subject was holding an object. This difference comes as a result of the first experiment using babbling data for both testing and training, while during the pointing experiment, testing was performed with an object which was often out of reach, resulting in a larger precision error, as discussed before.

5. DISCUSSION AND CONCLUSIONS

With this work an important open question "How can pointing emerge from grasping behaviour" (T2.3 from [11]) in developmental psychology and developmental robotics has been addressed. We identified and extended the informational and sensory prerequisites for realising pointing from motor babbling in a computational model. Compared to the similar model presented in [25], the advantage of our model is the biological plausibility, which aims to make a contribution towards the research direction that might be of interest for neuroscientists and roboticists. The modular organization of the model in terms of sensory maps and associated representations of sensory data via Hebbian learning allow for easier extension and identification of its components with the biological equivalents.

The results presented here enable us to draw several parallels with respect to the development of pointing skill in infants. First, the two instances of the model can be compared to the two different stages of hand-eye coordination

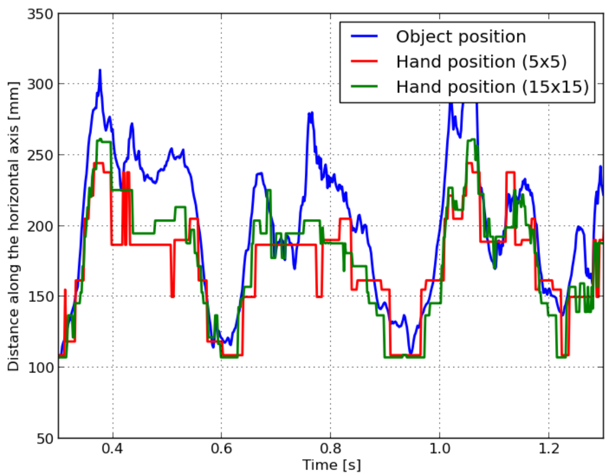

Figure 6: Left hand trajectory along the horizontal axis as predicted by the two networks and the ground truth value

development. Under the assumption that the more advanced stage is described by more specialised neurons which facilitate pointing, the model containing more neurons simulates the later stage by performing significantly better in terms of smaller pointing errors. Second, we explored the influence of data acquisition procedure through random motor babbling and showed that longer babbling phases yield better pointing accuracies. Following this line of arguments, one would expect that infants which explored the greater set of body configurations might acquire better manual coordination. This hypothesis can be further tested by analysing the motor coordination of children who were differently exposed to sports in their early childhood.

When trained on different data, the model for pointing can be used for learning other sensorimotor skills. For example, one could use the model to learn relations between observed pointing gestures and motor commands for moving into the pointed direction, similarly to [3]. Knowing and handling these relations is important, as situations requiring them occur in everyday life. For example, in a bookstore when we point to a book on the shelve we can not reach or on the street where we ask for an unknown path. Thus, for robots to exhibit human-like behaviour in such situations they need to be able to recognize the pointed position and to associate this position with a certain motor command which should be issued to reach that position. In order to achieve this behaviour, one would need to train one SOM to represent the space of joint configurations and associate this SOM with the motor commands in the second SOM. Depending on whether the robot should point or move, one should use one or the other SOM to determine kinematics of robot.

6. FUTURE WORK

The research presented here can be expanded into multiple directions. One might like to address the training procedure in a greater detail and analyse the influence of learning parameters on the pointing precision. Exciting questions with implications relevant for neuroscience might be tackled by introducing more complex neuron models that exhibit spiking behaviour. In that case, instead of simple Hebbian learning rule, SOMs can be coupled with more biologically realistic spike-timing-dependent plasticity rules. With this modifications, one would expect to strengthen the link between the model and the biological system.

A mechanism which reflects the ongoing development and learning in neural models is the ability of a model to gradually adapt to more complex inputs. The adaptation should be observed as the increase in the model complexity which, in return, should be observable in the roboot's behaviour. We speculate that one important aspect of such adaptive mechanism involves the increased number of neurons in a neural network. More complex SOM-based algorithms such as growing neural gas qualify as a starting point for simulation of ongoing development.

7. ACKNOWLEDGMENTS

This work has been supported by the Bernstein Center for Computational Neuroscience Berlin, by the EU-funded ITN INTRO (INTeractive RObotics research network) grant no. 238486 and by the Humboldt Graduate School (German Excellence Initiative).

8. REFERENCES

[1] A. Baranes and P.-Y. Oudeyer. Active learning of inverse models with intrinsically motivated goal exploration in robots. *Robot. Auton. Syst.*, 61(1):49–73, Jan. 2013.

[2] S. Baron-Cohen. *Mindblindness: An essay on autism and theory of mind*. The MIT Press, 1997.

[3] S. Bodiroža, A. Jevtić, B. Lara, and V. V. Hafner. Learning motion control for guiding a robot using gestures. In *4th Israeli Conference on Robotics (ICR2013)*, 2013.

[4] A. Dearden. *Developmental learning of internal models for robotics*. PhD thesis, Imperial College London, 2008.

[5] Y. Demiris and A. Dearden. From motor babbling to hierarchical learning by imitation: A robot developmental pathway. In *Proceedings of the Fifth International Workshop on Epigenetic Robotics: EpiRob*, pages 31–37, 2005.

[6] S. Desrochers, P. Morisette, and M. Ricard. Two perspectives on pointing in infancy. In C. Moore and P. Dunham, editors, *Joint Attention: its origins and role in development*, pages 85–101. Lawrence Erlbaum Associates, 1995.

[7] M. J. Farah. Why does the somatosensory homunculus have hand next to face and feet next to genitals?: An hypothesis. *Neural Computation*, 10(8):1983–1985, 1998.

[8] R. Feldman. *Development Across the Life Span: International Edition*. Pearson Education, Limited, 2013.

[9] V. V. Hafner and G. Schillaci. From field of view to field of reach - could pointing emerge from the development of grasping? *Frontiers in Computational Neuroscience*, 2011.

[10] B. Jansen, B. Boer, and T. Belpaeme. You did it on purpose! towards intentional embodied agents. In

F. Iida, R. Pfeifer, L. Steels, and Y. Kuniyoshi, editors, *Embodied Artificial Intelligence*, volume 3139 of *Lecture Notes in Computer Science*, pages 271–277. Springer Berlin Heidelberg, 2004.

[11] F. Kaplan and V. V. Hafner. The challenges of joint attention. *Interaction Studies*, 7(2):135–169, 2006.

[12] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 43(1):59–69, 1982.

[13] P. Li, X. Zhao, and B. Mac Whinney. Dynamic self-organization and early lexical development in children. *Cognitive Science*, 31(4):581–612, 2007.

[14] M. Lungarella and G. Metta. Beyond gazing, pointing, and reaching: A survey of developmental robotics. In *Epigenetic Robotics*, pages 81–89, 2003.

[15] C. Malsburg. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14(2):85–100, 1973.

[16] T. Martinetz, H. Ritter, and K. Schulten. Kohonens self-organizing maps for modeling the formation of the auditory cortex of a bat. In *SGAICO-Proceedings Connectionism in Perspective*, pages 403–412, Zürich, 1988.

[17] A. N. Meltzoff and M. K. Moore. Imitation of facial and manual gestures by human neonates. *Science*, 198(4312):75–78, 1977.

[18] M. M. Merzenich, R. J. Nelson, M. P. Stryker, M. S. Cynader, A. Schoppmann, and J. M. Zook. Somatosensory cortical map changes following digit amputation in adult monkeys. *Journal of Comparative Neurology*, 224:591–605, 1984.

[19] G. Metta. *Babyrobot – A study on sensori-motor development*. PhD thesis, 2000.

[20] A. Morse, J. de Greeff, T. Belpaeme, and A. Cangelosi. Epigenetic robotics architecture (era). *Autonomous Mental Development, IEEE Transactions on*, 2(4):325–339, 2010.

[21] K. Obermayer, H. Ritter, and K. Schulten. Large-scale simulation of a self-organizing neural network: Formation of a somatotopic map. In R. Eckmiller, G. Hartmann, and G. Hausk, editors, *Parallel Processing in Neural Systems and Computers*, pages 71–74. North-Holland, Amsterdam, 1990.

[22] P. Parpia. Reappraisal of the somatosensory homunculus and its discontinuities. *Neural computation*, 23(12):3001–3015, 2011.

[23] H. Ritter. Self-organizing maps for internal representations. *Psychological Research*, 52(2-3):128–136, 1990.

[24] R. Saegusa, G. Metta, G. Sandini, and S. Sakka. Active motor babbling for sensorimotor learning. In *Robotics and Biomimetics, 2008. ROBIO 2008. IEEE International Conference on*, pages 794–799, 2009.

[25] G. Schillaci and V. V. Hafner. Random movement strategies in self-exploration for a humanoid robot. In A. Billard, P. H. K. Jr., J. A. Adams, and J. G. Trafton, editors, *HRI*, pages 245–246. ACM, 2011.

[26] M. Tomasello. Joint attention as social cognition. In C. Moore and P. Dunham, editors, *Joint attention: its origins and role in development*, pages 103–130. Lawrence Erlbaum Associates, 1995.

[27] G. Westerman and E. R. Miranda. Modelling the development of mirror neurons for auditory-motor integration. *Journal of New Music Research*, 31(4):367–375, 2002.