5th International Conference on Development and Learning
and on Epigenetic Robotics
Aug 13-16, 2015. Providence, RI USA

The Role of Contrast Sensitivity in the Development of Binocular Vision: A Computational Study

Alexander Priamikov[1], Vikram Narayan[1] , Bertram E. Shi[2] and Jochen Triesch[1]
[1]Frankfurt Institute for Advanced Studies, Frankfurt am Main, Germany
[2]Dept. of Electronic and Computer Engineering, Hong Kong Univ. of Science and Technology, Hong Kong, China

Abstract— **The development of accurate binocular vision relies on the acquisition of disparity tuning and the calibration of vergence eye movements. Both processes are fundamentally limited by visual acuity, which increases only gradually during the first year of life. Next to limiting performance, however, early limitations of visual acuity may also aid rapid learning analogous to Newport's "less-is-more" hypothesis. Here we use computational modeling to assess the potential impact of early acuity limitations on the development of binocular vision. Our starting point is a previous model of the development of binocular vision, formulated in the Active Efficient Coding framework. We extend this model to incorporate the development of visual acuity between birth and 8 months. We find that the model does in fact learn faster if visual acuity starts out poor and increases with time, supporting the less-is-more hypothesis. Furthermore, we find that the speed of acuity improvement needs to be "just right", i.e., neither too rapid nor too slow for fastest learning. Overall, our model suggests that early limitations of visual acuity may aid infants in acquiring binocular vision skills and it provides a good starting point for computational modeling of developmental disorders of binocular vision.**

Index Terms—**binocular vision, contrast sensitivity function, disparity tuning, vision development, vergence eye movements, active efficient coding, efficient coding.**

I. Introduction

Human infants have limited perceptual, motor and cognitive abilities. These limitations have at least two possible explanations. An earlier one states that they are barriers that must be overcome if an infant wants to achieve adult function [1]. An alternative newer one states that these limitations are necessary for learning and development. It has been argued that such limitations reflect simple neural representations, which mature through development stages towards complex neural structures [2]. A popular concept supporting the stage-based idea is Newport's "less-is-more" hypothesis [3]. It states that children learn a language from small parts towards complex structures. The limited cognitive skills of children are useful in learning a language, because they help in identifying and recognizing language components. Such a "simple to complex" approach may also apply to visual development.

Early work in this direction was performed by Dominguez and Jacobs on the acquisition of binocular disparity sensitivity [4]. They showed that learning can benefit from a proper developmental progression. They studied four different models, which were trained to detect binocular disparities in image pairs. In three of these models the authors changed the visual input corresponding to different developmental progressions. These models received simple visual input early in training and the quality and richness of the input changed with progressing training. The visual input to the last model stayed unchanged during the whole training phase. Results of these simulations showed that developing models outperformed non-developmental ones.

Our study is based on a recently proposed computational model that learns binocular disparity representations jointly with eye movement control [5] in the recently developed Active Efficient Coding framework. In this approach a sparse coding model learns to encode the sensory data while a reinforcement learner directs the eyes so as to make the sparse coding model work as efficiently as possible. This approach leads to a fully self-calibrating sensory-motor learning of binocular vision. To study the effect of early perceptual limitations we incorporated the development of visual acuity between one and eight months. The visual input was filtered according to the measured contrast sensitivity functions at different ages [6,7]. Our results show that under such developmental conditions, the performance and speed of learning increases as compared to a non-developmental one. This supports the "less-is-more" hypothesis [3] and the results of [4]. Additionally, we found that the speed of the developmental progression should be properly configured to achieve the best performance. We therefore suggest that carefully scheduled limitations of visual acuity may aid infants in acquiring proper binocular vision.

II. Materials and Methods

In this section, we first explain the basic architecture of the model, then we describe how we model the development of visual acuity. Finally, we describe the OpenEyeSim simulator used to run the experiments.

A. Model architecture

The model consists of two main components (Fig. 1):

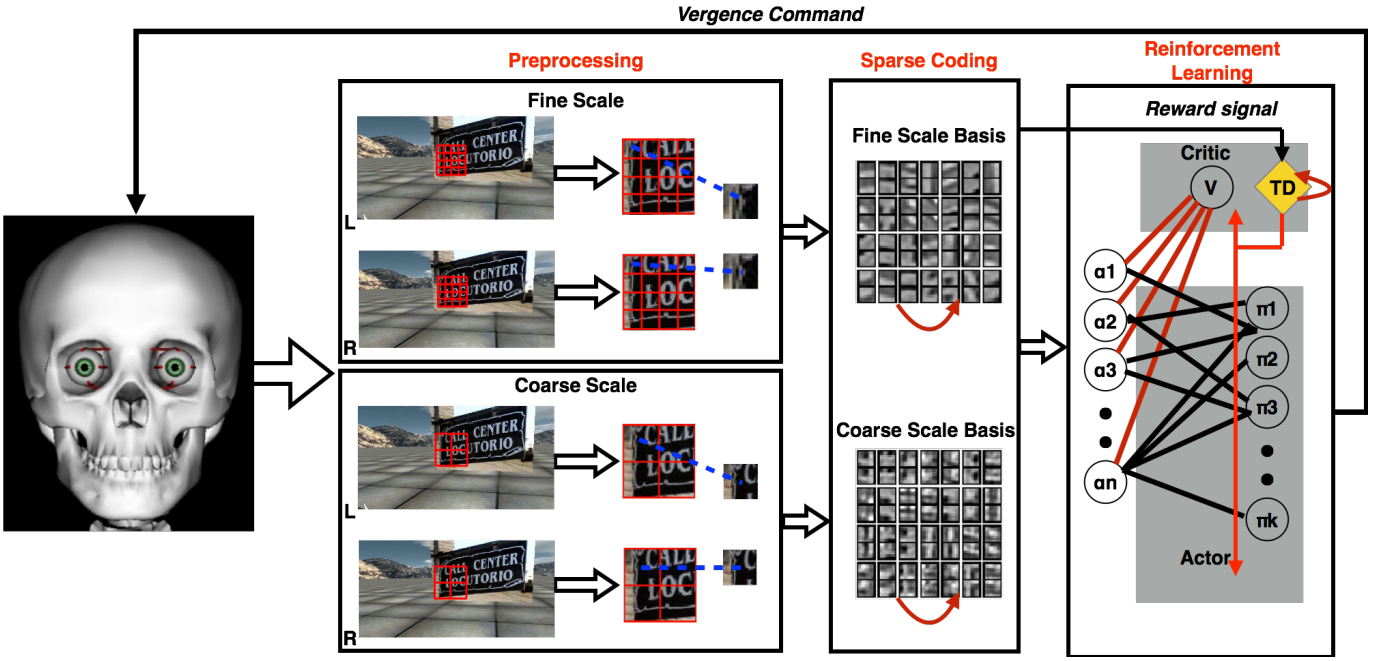

Figure 1. Detailed architecture of the proposed model. The binocular visual input, taken from the region around the center of the image, is cropped at both coarse and fine scales. Each scale is divided in a set of image patches (different size for coarse and fine scales). Such patches from corresponding spatial locations at a particular scale from left & right eyes taken together to obtain binocular patch. These binocular patches are encoded by a set of basis functions using the modified matching pursuit [8]. The reconstruction error is used as the reward signal to reinforcement learner, which performs actions on eye rotations based on encoded data coming from sparse coding models. The reinforcement component is represented as an actor-critic network and learns to minimize reconstruction error (maximize coding efficiency).

- **Sensory encoding by a sparse coding model**: The input images are decomposed into sets of binocular patches, concatenated from left and right eyes. Each patch is represented by a linear combination of basis functions. The basis functions adapt to the binocular patches and learn to represent binocular disparities.

- **Motor control by reinforcement learning**: The basis function activations are passed to the motor control component that learns to generate the motor commands, which adjust the eyes' vergence angle.

B. *Sparse Coding*

Similar to the model presented by Lonini et al. [5] we use a multi-scale approach - two sparse coding models for fine and coarse scales corresponding to foveal and parafoveal vision. The inputs are two RGB images with size 320x240 pixels. We convert each of these images into gray scale and extract fine and coarse scale windows. For the fine scale we use a 80x80 region around the center and for the coarse scale we use a 128x128 region. For the coarse scale we subsample the image by a factor of eight in a Gaussian pyramid hierarchy, for the fine scale we subsample the image by a factor of two. Subsequently patches of 8x8 pixels are extracted from each foveal window (81 for each scale). The patches are preprocessed to have zero mean and unit norm. Corresponding patches from left and right windows are concatenated into a stereo patch. Each stereo patch is vectorized and approximated through the sparse coding model by a linear combination of binocular basis functions ϕ_i:

$$X_k^{orig} = \begin{bmatrix} X_k^L \\ X_k^R \end{bmatrix} \approx \sum_{i=1}^{D} a_i \begin{bmatrix} \phi_i^L \\ \phi_i^R \end{bmatrix} = X_k^{rec} , \qquad (1)$$

where D is the total number of basis functions in a sparse coding model, $\boldsymbol{a_i}$ are the weighting coefficients, k is the patch number and X_k^L, X_k^R are the vectorized stereo patches taken from left and right image, respectively. X_k^{orig} are the original patches, X_k^{rec} are the reconstructed ones.

Each of the sparse coding models is trained to represent the original image as accurately as possible. The driving signal for this training is the total squared reconstruction error over all the stereo patches, normalized by the energy in the original image. This signal shows the loss of information due to the encoding and is defined in the same way for coarse and fine models:

$$e = \frac{1}{N} \sum_{k=1}^{N} \frac{\left\| X_k^{orig} - X_k^{rec} \right\|}{\left\| X_k^{orig} \right\|} , \qquad (2)$$

where N *(N=81)* is the number of stereo patches.

Learning the basis functions is a two-stage procedure in our model. For the first learning step we use a modified version of the matching pursuit algorithm [8], which selects a set of

coefficients a_i and basis functions ϕ_i from the basis dictionary to approximate the input patch. In order to ensure that all basis functions are used during learning, we add a homeostatic constraint to the matching pursuit [9]. Homeostasis acts at any time step, ensuring that the probability to choose a basis function is uniform across the dictionary. In the second step, the chosen bases are adapted through gradient descent on the reconstruction error.

At each training iteration we calculate a feature vector, ***s(t)***, by averaging the squared weighting coefficient over the *N* patches at each of the two scales:

$$\boldsymbol{s}(t) = \begin{bmatrix} \frac{1}{N}\sum_{k=1}^{N} a_1^{(k)^2}(t) \\ \vdots \\ \frac{1}{N}\sum_{k=1}^{N} a_D^{(k)^2}(t) \end{bmatrix}, \quad (3)$$

where $a_i^{(k)}$ is the coefficient of basis *i* for patch *k*, and *N* is the number of patches, and D is the size of the dictionary.

C. Reinforcement learning

The reinforcement learning (RL) agent maps the current state, as represented by the activations of basis functions ***s(t)*** of the sparse coding models, to a vergence command. A natural actor-critic algorithm is employed [10], where neural networks model the policy and value functions. The goal of the RL agent is to select actions to maximize the discounted cumulative future reward *R(t)*:

$$R(t) = \sum_{k=0}^{\infty} -\gamma^k [e_{t+k}^{coarse} + e_{t+k}^{fine}], \quad (4)$$

where $\boldsymbol{e}^{coarse}$ and $\boldsymbol{e}^{fine}$ are the reconstruction errors for the coarse and fine scale sparse coding models, respectively, and γ (γ=0.3) is a discount factor. We use a modified natural actor-critic algorithm with an additional regularization factor to keep the weights of the policy bounded [10]. We use two neural networks, which serve as actor and critic components. The critic component receives the input state ***s(t)*** and produces as output the value estimate ***V(t)*** of the current state:

$$V(t) = \boldsymbol{v}^T(t)\boldsymbol{s}(t), \quad (5)$$

where *v(t)* are the weights of the network at time *t*. The actor maps states to actions. Its output layer has as many neurons as the number of actions. We use a discrete set of 11 actions: *J = {0, ±0.5, ±1, ±2, ±4, ±8}*, which correspond to changes in vergence angles in degrees. The activation of the output neurons is calculated as:

$$q_j(t) = \boldsymbol{\theta}_j^T(t)\boldsymbol{s}(t), \quad (6)$$

where $\boldsymbol{\theta_j(t)}$ are the weights from the state ***s*** to an action $j \in J$ at time *t*. We use a softmax operation on the activation of output neurons to calculate the probabilities of choosing actions:

$$P_j(s(t)) = \frac{\exp(q_j(t)/T)}{\sum_{i=1}^{|J|}\exp(q_i(t)/T)}, \quad (7)$$

where the amount of exploration/exploitation is controlled by the temperature *T* (*T=1* during training and *T=0.01* during testing).

D. Contrast Sensitivity function

Contrast sensitivity is an aspect of the human visual system that is highly connected to visual acuity. It is a measure of the ability to discern between luminance levels in a static image. It is limited by two main factors: optical and neural.

Similar to a camera an eye has a lens, which can be characterized by a modulation transfer function. This function corresponds to a low-pass filter and has a cut-off at high spatial frequencies [11,12]. On the other side the neural component of contrast sensitivity provides a cut-off at low spatial frequencies [12]. This happens mainly because of processing in the retina and lateral geniculate nucleus, which can be modeled by isotropic spatial filters. Together they implement a band-pass filter, which has been characterized psychophysically by the contrast sensitivity function (CSF). The CSF gives the contrast detection thresholds for sine-wave gratings at different spatial frequencies. Such measurements can be performed by behavioral (forced-choice) or objective methods (measurement of visual evoked potentials).

In young infants, contrast sensitivities at all spatial frequencies are highly reduced compared to adults. The CSF tends to be a bandpass filter with a peak that shifts toward higher sensitivities and higher spatial frequencies with age [6,7]. In our studies we modeled the development of the CSF over the period from 1 to 8 months, where the most drastic changes occur. For modeling we used the fitting function presented by Kiorpes et al. [13]:

$$CSF(w) = f(wb)^d \exp(-cwb), \quad (8)$$

where *w* is the spatial frequency and *b,c,d,f* are free parameters, which control the gradient of the ascent and descent in the curve, as well as lateral and vertical shifts along the axes. Our CSFs are presented in Figure 2. The contrast sensitivity function has been measured for the set of ages of {1, 3, 6, 8} months. We interpolate the contrast sensitivity function in a piecewise-linear manner between these ages by changing the free parameters in (8).

To model the effect of the CSF, we transform the image to the frequency domain using the Fourier transform, multiply by the CSF and then take the inverse Fourier transform. Examples of such simulations are shown in Figure 3.

E. OpenEyeSim

We use OpenEyeSim [14] as a simulation platform in our studies. OpenEyeSim is a detailed three-dimensional model of human extraocular muscles that includes the rendering of 3D

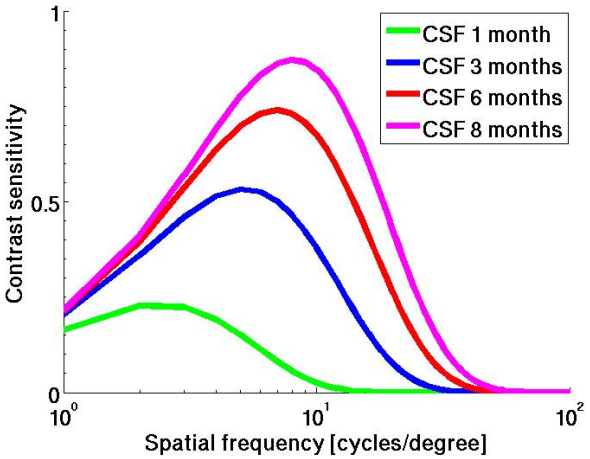

Figure 2. Contrast sensitivity functions from 1 to 8 months. We used a functional representation by Kiorpres [13] and fitted free parameters in order to fit the experimental data measured in infants at different ages [6,7].

Figure 3. Simulations of different contrast sensitivity functions. These pictures were obtained by using the fast Fourier transform and corresponding CSFs.

scenes from a virtual environment perceived by the eyes. The oculomotor plant is designed to mimic the nonlinear dynamics of the eye, based on a model of the extraocular muscles (EOMs). The simulator is able to work in two different regimes: kinematic and dynamic. In the dynamic regime, given EOM innervations, this model generates realistic gaze trajectories. In the kinematic regime, the simulator receives as an input eye positions and moves the eyes accordingly. Compared to the dynamic regime, this allows faster simulations for studies where the detailed movement dynamics can be neglected. In the present study we used the kinematic regime for the sake of simplicity and faster simulations.

III. Results

A. Experimental setup

A flat textured object is placed in front of the eyes at different depths and the agent learns to produce proper vergence eye movements. The object stays fixed at one depth for 10 iterations, after which the depth is changed according to a uniform distribution over the range [0.5, 2] meters. The textures applied to this object are natural images. The object size is such that it usually fills the fine and coarse scale windows.

For the evaluation of the learned policy we perform testing phases after every 10% of the total training time. During the testing phases we use a more greedy policy and do not perform any changes to the weights of the reinforcement learner or the sparse coding basis functions. For all of our experiments we set the whole training time to 50k iterations (100%). We vary the amount of training, where changes in contrast sensitivity happen. For example, for a 16% setting the contrast sensitivity develops from newborn values to 8-months-old values during the first 16% of the simulation and afterwards it stays fixed at the 8-months-old level. As a control experiment we used a setting where the CSF stayed static at the level of an 8-months-old during the whole training. A schematic representation of these experiments is shown in Figure 4.

B. Performance evaluation

The smallest non-zero action that the reinforcement learner is able to perform is a vergence change of 0.5 deg. For all four of the developmental schemes the error of the learned vergence policy eventually reaches a level close to this. However, the time at which this level is reached varies among the different developmental schemes. Figure 5 summarizes our findings on the performance of our system at its early stages under different developmental schemes. This figure shows the performance of the system averaged over 4 different trials for each of our experimental settings. After every 10% of training we performed testing phases of 1000 iterations where we measured the performance of the system in its ability to do proper vergence eye movements and focus on the object at different depths. The vergence error is defined as the absolute difference between the actual vergence angle and the ideal vergence angle that would put the object in the plane of fixation. During these testing phases we "froze" the sparse-coding and reinforcement learning models. Figure 5 provides clear evidence that using a developmental strategy leads to faster learning. Additionally, it shows that the speed of the development of visual acuity should be "just right" for maximizing learning speed. This can also be seen in Figure 6, where we plot the percentage of training time required for the performance to reach an average vergence error of 1 degree.

C. Analysis of the basis functions

It has been shown that receptive fields in the primary visual cortex have a Gabor-like structure [15]. We use the following function to fit the basis functions:

$$g(x,y)_{\Delta,\psi} = \exp\left(-\frac{x'^2+\beta^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda}+\psi\right) \quad (9)$$

$$\Delta = \{\lambda,\theta,\sigma,\beta,x_c,y_c\} \quad (10)$$

$$x' = (x-x_c)\cos\theta + (y-y_c)\sin\theta \quad (11)$$

$$y' = -(x-x_c)\sin\theta + (y-y_c)\cos\theta \quad (12)$$

where λ is the wavelength of the sinusoidal factor, θ represents the orientation, ψ is the phase offset, σ is the standard deviation of the Gaussian envelope, β is the spatial aspect ratio which specifies the ellipticity , and x_c , y_c are the centers.

We fit these Gabor functions to the left and right parts of the basis functions, which are constrained to only differ in phase. An illustration of obtained Gabor fits with reference to learnt

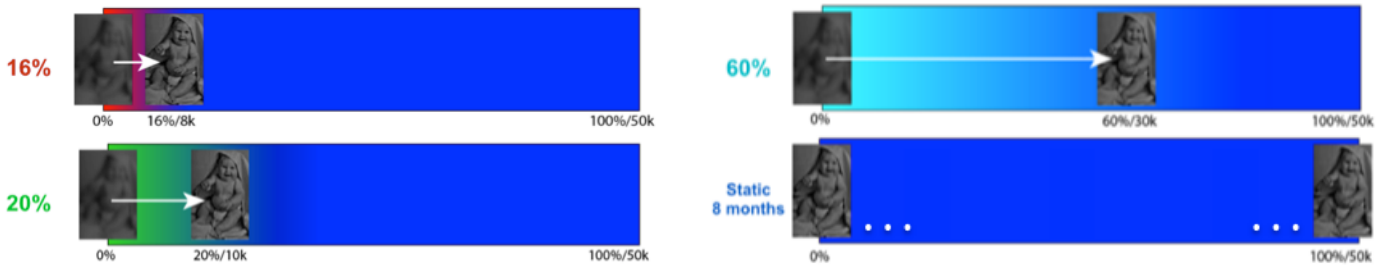

Figure 4. Schematic representation of the 4 experimental settings. In the bottom right one, which we call static 8 months, the contrast sensitivity corresponds to that of an 8-month-old during the entire training. In the other 3 settings our contrast sensitivity starts at the level of a 1-month-old and increases gradually to that of an 8-month-old with different speed of this improvement. In the "red" setting, the final contrast sensitivity of an 8-month-old is reached at 16% of the learning time, for green it is reached at 20% and for cyan at 60%. The same color labeling is also used for Figures 5 and 6.

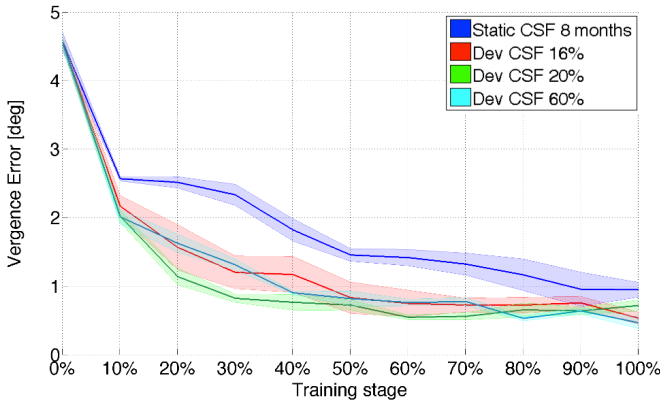

Figure 5. Learning performance under different experimental settings. Bold lines show the mean performance over 4 trials, shaded regions shows standard deviations. Performance of the system is measured every 10% of training.

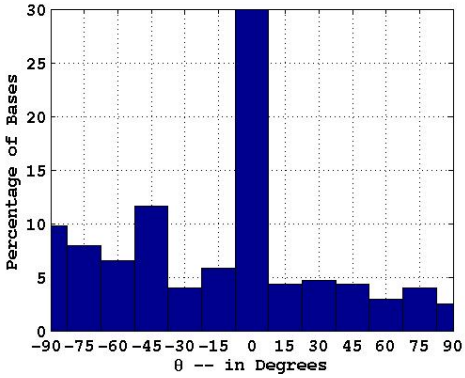

Figure 8. Histogram of orientation preferences (***fine*** scale). It can be observed that the most dominant orientation is vertical. This is in line with results reported in [14] which concludes that the majority of receptive fields are biased to either horizontal or vertical orientations.

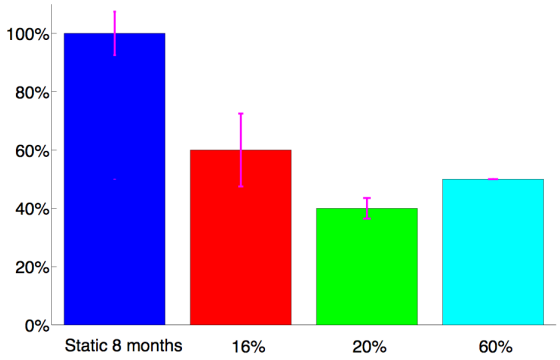

Figure 6. The percentage of training time required for reaching an average vergence error of 1 degree.

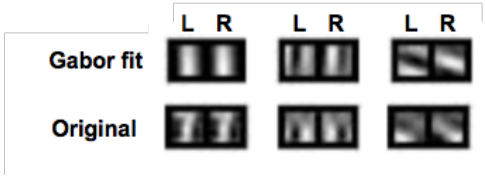

Figure 7. Qualitative evaluation of the Gabor fitting for three example basis fits from the fine scale. Results are similar for the coarse scale.

basis functions is given in Fig. 7 for the purpose of qualitative evaluation. Figure 8 shows the histograms of basis function orientations. The most dominant preferred orientation is vertical. This is in line with [16], which reports that more receptive fields have vertical or horizontal orientation preference compared to oblique ones.

We compare the changes of basis functions for two different experimental settings: with 8 months static filter used during the whole training time and with time-varying filter that changes in a piecewise linear fashion from 1 to 8 months of CSF development during 20% of the training time, and then stays constant. Figure 9 compares the preferred disparities at the beginning and the end of training for the fine scale.

The same behavior is observable also for the coarse scale. Initially, the system starts with a broad distribution of preferred disparities. During training, as the system achieves binocular coordination, most basis functions come to prefer zero disparity. Results show that for both coarse and fine scales, the cells develop their preferred tuning for zero disparity rapidly. For the fine scale at the end of training the number of cells with zero disparity preference is 1.7 times larger for the training under the development constraint.

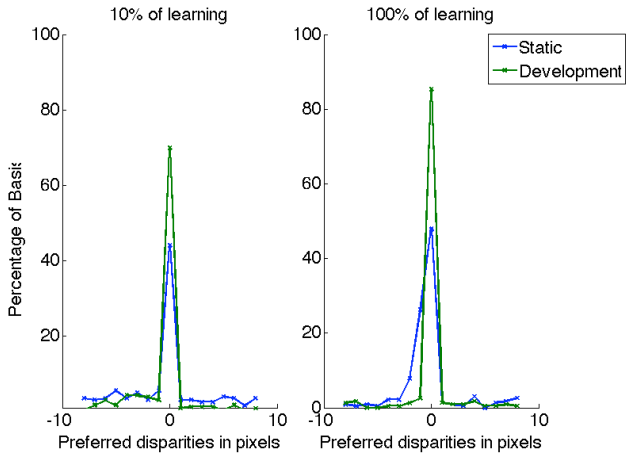

Figure 9. Preferred disparities for ***fine*** scale in the beginning and at the end of learning for experimental settings under the developmental constraint (green) and with static filter of 8-month-old infant (blue).

IV. Conclusion

We have analyzed the influence of a developmental constraint on the joint learning of visual representations and motor control. Our simulations were based on the Active Efficient Coding framework, an extension of the classic efficient coding hypothesis to active perception [17, 18]. We considered the contrast sensitivity of the visual system and based on available psychophysiological data of contrast sensitivity at different ages modeled the reduction of contrast sensitivity in young infants. Our findings support the "Less-is-More" hypothesis initially presented by Newport [3]. They also match the results of Dominguez and Jacobs, who also studied developmental constraints in binocular vision [4]. The learning under such a developmental constraint is faster compared to non-developmental learning. We found that the change in the development speed has to be "just right" to achieve fastest learning. We also showed that the developmental constraint influences not only the motor control part, but also the sensory representation. Under the developmental constraint the sensory representation specialized for close-to-zero disparities more quickly. Overall our study shows that developmental constraints on contrast sensitivity can speed up the self-calibration of sensory-motor loops.

Acknowledgments

This work was supported in part by the German BMBF under grant 01GQ1414, the Hong Kong Research Grants Council under Grant 618512, the German DAAD and the Quandt foundation.

References

[1] Piaget, J. (1952). The origins of intelligence in children. New York: International Universities Press.

[2] Turkewitz, G., & Kenney, P. A. (1982). Limitations on input as a basis for neural organization and perceptual development: A preliminary statement. Developmental Psychobiology, 15, 357–368.

[3] Newport, E. L. (1990). Maturational constraints on language learning. Cognitive Science, 14, 11–28.

[4] Dominguez, M., & Jacobs, R. A. (2003). Developmental constraints aid the acquisition of binocular disparity sensitivities. Neural Computation, 15(1), 161-182.

[5] Lonini, L., Zhao, Y., Chandrashekhariah, P., Shi, B. E., & Triesch, J. (2013, August). Autonomous learning of active multi-scale binocular vision. In Development and Learning and Epigenetic Robotics (ICDL).

[6] Atkinson, J., Braddick, O., & Moar, K. (1977). Development of contrast sensitivity over the first 3 months of life in the human infant. Vision Research, 17(9), 1037-1044.

[7] Kelly, J. P., Borchert, K., & Teller, D. Y. (1997). The development of chromatic and achromatic contrast sensitivity in infancy as tested with the sweep VEP. Vision Research, 37(15), 2057-2072.

[8] Mallat, S. G., & Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. Signal Processing, 41(12), 3397-3415.

[9] Perrinet, L. U. (2010). Role of homeostasis in learning sparse representations. Neural computation, 22(7), 1812-1836.

[10] Bhatnagar, S., Sutton, R. S., Ghavamzadeh, M., & Lee, M. (2009). Natural actor–critic algorithms. Automatica, 45(11), 2471-2482.

[11] Rovamo, J., Mustonen J., Näsänen R. (1994). Two simple psychophysical methods for determining the optical modulation transfer function of the human eye. Vision research, 34.19, 2493-2502.

[12] Barten, P. Contrast sensitivity of the human eye and its effects on image quality (1999), SPIE press Vol. 72.

[13] Movshon, J. A., & Kiorpes, L. (1988). Analysis of the development of spatial contrast sensitivity in monkey and human infants. JOSA A, 5(12), 2166-2172.

[14] Priamikov A., Triesch J. (2014). OpenEyeSim-A platform for biomechanical modeling of oculomotor control. In Development and Learning and Epigenetic Robotics (ICDL).

[15] Jones J., Palmer L. (1988) An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. Neurophysiology 58 (6): 1233–1258.

[16] Li B, Peterson MR, Freeman BD (2003) Oblique effect: a neural basis in the visual cortex. Journal of Neurophysiology 90: 204–217.

[17] Zhao, Y., Rothkopf, C. A., Triesch, J., & Shi, B. E. (2012). A unified model of the joint development of disparity selectivity and vergence control. In Development and Learning and Epigenetic Robotics (ICDL).

[18] Teulière, C., Forestier, S., Lonini, L., Zhang, C., Zhao, Y., Shi, B., & Triesch, J. (2014). Self-calibrating smooth pursuit through active efficient coding. Robotics and Autonomous Systems.