STDP Training of Hierarchical Spike Timing Model of Visual Information Processing

Petia Koprinkova-Hristova
Inst. of Inf. and Comm. Technologies
Bulgarian Academy of Sciences
Sofia, Bulgaria
pkoprinkova@bas.bg

Simona Nedelcheva
Inst. of Inf. and Comm. Technologies
Bulgarian Academy of Sciences
Sofia, Bulgaria
croft883@gmail.com

Nadejda Bocheva
Inst. of Neurobiology
Bulgarian Academy of Sciences
Sofia, Bulgaria
nadya@percept.bas.bg

Radolsava Kraleva
Department of Informatics
South West University
Blagoevgrad, Bulgaria
rady_kraleva@swu.bg

Velin Kralev
Department of Informatics
South West University
Blagoevgrad, Bulgaria
velin_kralev@swu.bg

Miroslava Stefanova
Inst. of Neurobiology
Bulgarian Academy of Sciences
Sofia, Bulgaria
mirad_st@abv.bg

Bilyana Genova
Inst. of Neurobiology
Bulgarian Academy of Sciences
Sofia, Bulgaria
b.genova@abv.bg

***Abstract*—We have developed a hierarchical spike timing neural network model in NEST simulator aimed to reproduce human decision in simplified simulated visual navigation task. The model consists of the following layers: retinal photoreceptors and ganglion cells (RGC); thalamic relay including lateral geniculate nucleus (LGN), thalamic reticular nucleus (TRN) and interneurons (IN); primary visual cortex (V1); middle temporal (MT) area; medial superior temporal (MST) area and lateral intraparietal cortex (LIP). All synaptic inter- and intra-layer connections of the initial model were static and structured according to the literature information. The present work extends the model with spike timing dependent plastic (STDP) synapses between MST and LIP layers. We investigated the possibility to train synaptic weights via STDP rule to mimic decisions taken by test subjects as well as to differentiate them according to their age.**

***Index Terms*—spike timing neuron model, spike timing dependent plasticity, visual system, decision making, saccade generation**

I. Introduction

In the process of evolution, in the mammalian brain have emerged areas with a specific type of functionality that can be regarded as a hierarchical structure processing the visual information coming through our eyes. The sensory area (retina) consisting of photo-receptive cells transforms the incoming light into electrical signal via retina ganglion cells (RGC). Next layer (lateral geniculate nucleus (LGN) and thalamic reticular nucleus (TRN)) has a role of relay transmitting the signal to the primary visual cortex (V1). The motion information processing, that is of primary interest in our investigation, is performed predominantly in the following two structures called middle temporal area (MT) that encodes the speed and direction of the moving objects and the medial superior temporal area (MST) that extracts information about the self-motion of the observer.

Project DN02-3-2016 "Modeling of voluntary saccadic eye movements during decision making" funded by the Bulgarian Science Fund.

Most of the existing motion information processing models are restricted to the interactions between some of mentioned areas like: V1 and MT in [1]–[4], V1, MT and MST in [5]; MT and MST in [6], [7]. Many models consider only the feedforward interactions (e.g. [4], [8]) disregarding the feedback connectivity; others employ rate-based equations (e.g. [9], [10]) considering an average number of spikes in a population of neurons. Spiking neural network (SNN) models that are spatially designed according to brain templates were recently generalized in [11].

In our preliminary research [12], as an attempt to simulate realistically the interactions between all described processing stages of encoding dynamic visual information in the human brain, we have developed a spike-timing neural network model that includes all mentioned above structures as well as a layer playing major role in the process of decision making based on perceived visual information and the preparation of a saccade (change of gaze direction) to the desired location called lateral intraparietal area (LIP). The spatial organization and connectivity of our model, although not fully anatomically analogous to the considered brain structures, is designed using available neuro-physiological knowledge about their spatial structure and connectivity. The model was implemented using NEST 2.12.0 simulator ([13]).

All synaptic inter- and intra-layer connections in our previous work [12] were static. In the present work we have extended the model with spike timing dependent plastic (STDP) synapses between MST and LIP layers in order to tune these parameters to reflect observed human decisions during experiment with visual stimuli simulating optic flow patterns of forward self-motion on a linear trajectory to the left or to the right of the center of the visual field with a gaze in the direction of heading. The subjects had to indicate the perceived direction of heading by saccade movement. The mean latency of the eye movements obtained in the behavior experiments was used as training signal fed into the output layer of the

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

model. We investigated the possibility to train our model to mimic decisions taken by different test subjects as well as to differentiate the humans by age based on tuned model parameters.

The paper is organized as follows: section II describes briefly the overall model structure and parameters; next we describe briefly experimental set-up and data processing; section IV presents results from STDP training of the model and obtained parameters typical for mean behaviour of each one of the three tested age groups; the concluding section comments obtained results and determines directions for our future work.

II. Model Structure

The hierarchical model structure from [12] is shown on Fig. 1.

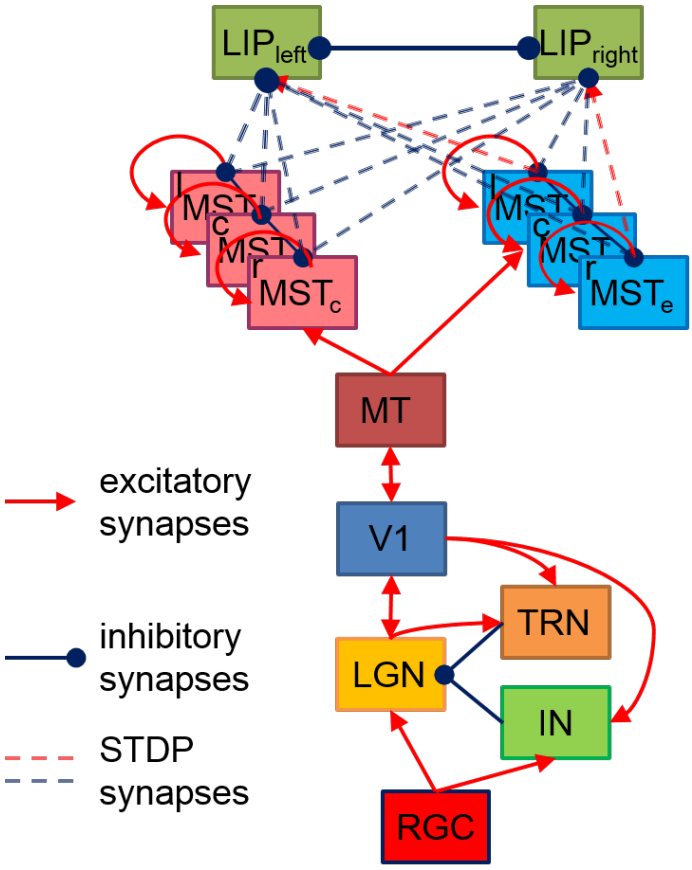


Fig. 1. Model structure.

The model connectivity is based on the available literature information. Each layer consists of neurons positioned in a regular two-dimensional grid. The receptive field of each neuron depends on the function of the layer it belongs to, on its spatial position within its layer and on the layers it is connected to.

The reaction of RGC to light changes is simulated by a convolution of a spatiotemporal filter with the images falling on the retina, following models from [14], [15]. Its spatial component has a circular shape modeled by a difference of two Gaussians (DOG) while the temporal component has a bi-phasic profile determined by the difference of two Gamma functions. The model contains two layers of ON and OFF RGC and their corresponding LGN and IN/TRN neurons, having identical relative to visual scene positions and opposite ("on-center off-surround" (ON) and "off-center on-surround" (OFF)) receptive fields placed in reverse order like in [14]. Each layer consists of totally 400 neurons, positioned on 20×20 grid. The continuous current generated by RGC is injected into LGN and IN via one-to-one connections. The structure of direct excitatory synaptic feedforward connectivity between LGN and V1 is also adopted from [14]. LGN also receives inhibitory feedback from V1 via IN and TRN according to [16].

As in [14], the neurons in V1 are separated into four groups – two excitatory and two inhibitory, having a ratio of 4/1 excitatory/inhibitory neurons (400/100 in our model) and connected via corresponding excitatory and inhibitory lateral connections. All excitatory neurons are positioned at 20×20 grid while the 10×10 inhibitory neurons are dispersed among them. Being orientation sensitive, V1 neurons have elongated receptive fields defined by Gabor probability function as in [17]. The "pinwheel structure" of the spatiotemporal maps of the orientations and phases of V1 neurons receptive fields was generated using a relatively new and easily implemented model ([18]). Lateral connections in V1 are determined by Gabor correlations between the positions, phases, and orientations of each pair of neurons. As in [14], neurons from inhibitory populations connect preferentially to neurons having a receptive field phase difference of around $180°$. In our model, the frequencies and standard deviations of Gabor filters for lateral connections were chosen so that all neurons in the layer have approximately circular receptive fields.

MT has identical to V1 size and structure and its lateral connections are designed in the same way while the connections from V1 cells depend on the angle between the orientation preferences of each two cells like in [19]. The orientation and phase maps of this layer were generated in the same way as those of V1.

The MST consist of two layers, each one containing 400 neurons positioned on 20×20 grid, sensitive to expansion (MSTe) and contraction (MSTc) movement patterns respectively, like in [20]. Each MST cell from both layers has randomly assigned an expansion/contraction connection template having a circular shape with specific width and a focal point having three possible positions - left, right or at the visual scene center respectively that are shown as three separate groups on Fig. 1. All MST neurons have on-center receptive fields. Each MST neuron collects inputs from MT cells corresponding to its pattern template. Both layers have intra- and interlayer excitatory/inhibitory recurrent connections between cells having similar/different sensitivity. The lateral connections are determined based on neurons' positions and template similarities. All neurons have Gaussian receptive fields. Connections within expansion/contraction layers are excitatory or inhibitory in dependence on their focal points

similarity. Connections between expansion and contraction layers are all inhibitory and depend both on similarities of their positions and focal points.

Since our model aims to decide whether the expansion center of a moving dot stimulus is left or right from the stimulus center, we proposed a task-dependent design of excitatory/inhibitory connections from MST expansion/contraction layers to the two LIP sub-regions whose increased firing rate corresponds to two taken decisions for two alternative motor responses - eye movement to the left or to the right. Both LIP areas are modeled by two neurons receiving excitatory input from MST expansion layer neurons having focal points corresponding to their decision responses (left or right) and inhibitory input from all other MST neurons. There are also lateral inhibitory connections between both LIP areas (see Fig. 1).

More details about connectivity design can be found in our previous work [12].

For the neurons in LGN conductance-based leaky integrate-and-fire neuron model as in [21] (iaf_chxk_2008 in NEST) was adopted. For the rest of neurons, leaky integrate-and-fire model with exponential shaped postsynaptic currents according to [22] (iaf_psc_exp in NEST) was used. All connection parameters are the same as in the cited literature sources.

III. Experimental Set-up

A. Behavioral experiment and data collection

The stimuli used for the stimulations are taken from a behavioral experiment with human subjects from three age groups young (between 20 and 34 years old), elderly (from 57 to 84 years old) and middle aged group (between 36 and 52 years old).

A detailed description of the experiment can be found in [23]. Here we used only the stimuli from the one of the conditions. In it, out of 50 dots, 36 dots moved away from a common center, while the rest 14 dots moved on random trajectories. All dots have a limited life time of 100 ms after which they were randomly displaced to a different position preserving their previous trajectory. Only one-third of the dots changed position on every frame. The speed of dot motion was $3.6deg/s$. The virtual center of the patterns is defined by the average position of the centers of all frames. The horizontal position of this center could be located at one of 7 different positions to the left or to the right of the pattern center. The horizontal shifts were in the range from 0.67 cm to 4.67 cm from the pattern center. For each position of the virtual center we generated 10 different replicates. All dot patterns (with 14 different shifts and 10 replicates) were presented in random order in a single experimental session i.e. 140 patterns in total. The stimuli were shown on a gray background with mean luminance of $50cd/m^2$.

Two screen shots for left and right expansion center stimuli are shown on Fig. 2.

The tested subjects sat at 57 cm from a monitor screen. Their task was to observe the patterns with steady fixation and to determine whether the center was to the left or to the right from the middle of the screen. After making a decision they have to make a saccade movement to its perceived center and to press a mouse button corresponding to their decision. If the subjects could not make a decision during stimulus presentation (3.3 s for 100 frames) the screen turned gray until the subjects made a response.

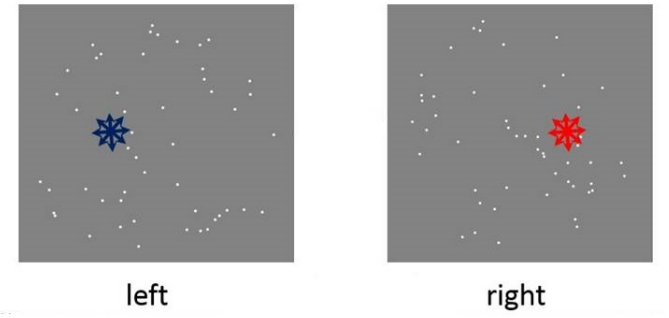


Fig. 2. Example of stimuli screen shots. Blue and red arrows show positions of imaginary expansion centers.

The stimulus presentation was controlled by a custom program developed under Visual C++ and OpenGl. The stimuli were presented as white dots on a NEC MultiSync LCD monitor with Nvidia Quadro 900XGL graphic board at a refresh rate of 60 Hz and screen resolution 1280×1024 pixels. The stimulus presentation was preceded by a fixation point with duration of 500 ms positioned at the center of the screen.

The eye movements of the participants in the experiment were recorded by a Jazz novo eye tracking system (Ober Consulting Sp. Z o.o.) with temporal precision of 1 kHz.

B. Data processing

The raw data was collected in a relational database [24]. It allowed us to process all sensors data in order to extract only the records from the presentation of a stimulus on the screen to the mouse button press. The data between the stimulus presentations were excluded since it is not relevant to the eye movements during task performance.

Observations of the processed eye movement records were refined by removing the outliers and a drift diffusion model of mean response time of each one of the age groups for all four experimental conditions was derived [23].

Based on the identified mean reaction times from [23] we've created training signals as generating currents I_{left} and I_{right} for the left and right LIP neurons respectively as follows:

$$I_{left/right} = A/(1 + \exp(k_{left/right}t)) \quad (1)$$

Amplitude $A = 200$ defines maximal input current (in pA) while $k_{left/right}$ determines settling time of the exponent that corresponds to the mean reaction time determined from experiments for each age group and experimental condition. Fig. 3 shows an example of generating currents for motion experimental condition of three age groups.

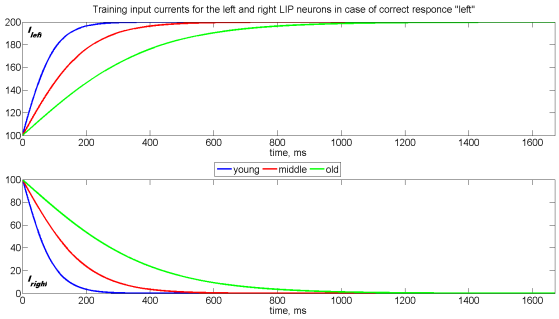


Fig. 3. Training signal.

Parameters of training signal for the considered case (correct response "left") are shown in Table I

TABLE I
TRAINING SIGNAL PARAMETERS

I	Age Group		
	Young	*Middle*	*Old*
k_{left}	-0.02	-0.01	-0.005
k_{right}	0.02	0.01	0.005

IV. SIMULATIONS

The overall model was tested using visual stimulation simulating an observer's motion on a linear trajectory with eyes fixed in the heading direction. Example stimuli used in the behavioural study were used.

Spike trains generated by both LIP neurons (left and right) having static connections from MST area in response to the stimuli with center displacement to the left with 20 pixels moving for a duration of 1670 ms are presented on Fig. 4.

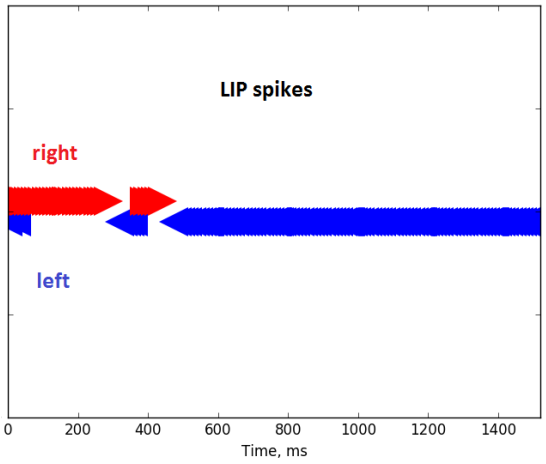


Fig. 4. Initial model simulation. Blue arrow shows spikes from the left LIP while red arrows - spikes from the right LIP neuron.

Next we have simulated the model with STDP synapses between MST and LIP areas and without training signal. Although the resulting weight connections were slightly changed, the spike trains generated from both LIP neurons were almost identical to the case with static synapses.

Simulations with training signal (1) however achieved behaviour with correct reaction (left in this case) with age-specific delay as it is shown on figures 5, 6 and 7 respectively for the three age groups. The results suggest differences between the three age groups in the time needed to achieve correct decision to respond similar to the behavioural data.

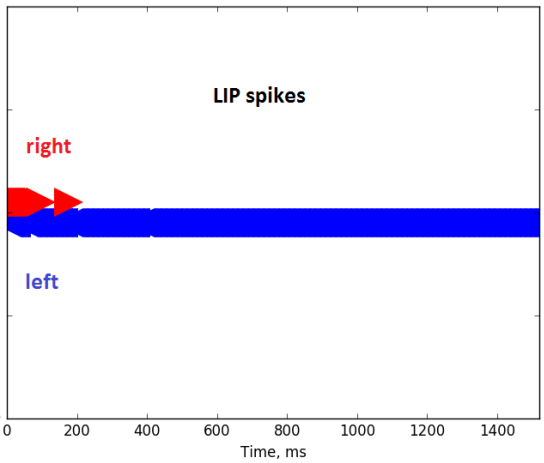


Fig. 5. Model reaction after training with signal for young age test group. Blue arrow shows spikes from the left LIP while red arrows - spikes from the right LIP neuron.

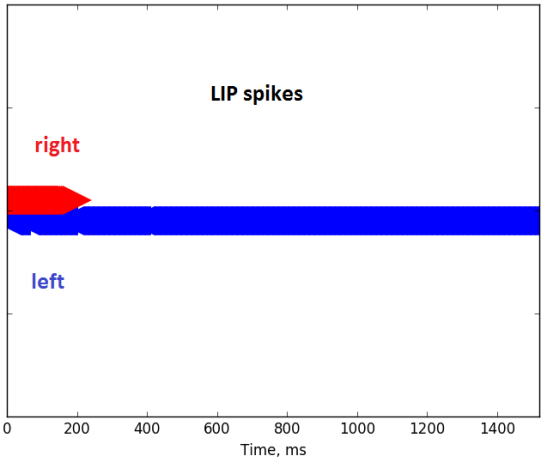


Fig. 6. Model reaction after training with signal for middle age test group. Blue arrow shows spikes from the left LIP while red arrows - spikes from the right LIP neuron.

Achieved after training STDP synaptic connections are shown on figures 8, 9 and 10 respectively.

Since from these figures connection weight differences are hardly visible, we also represent them on figures 11 - 14

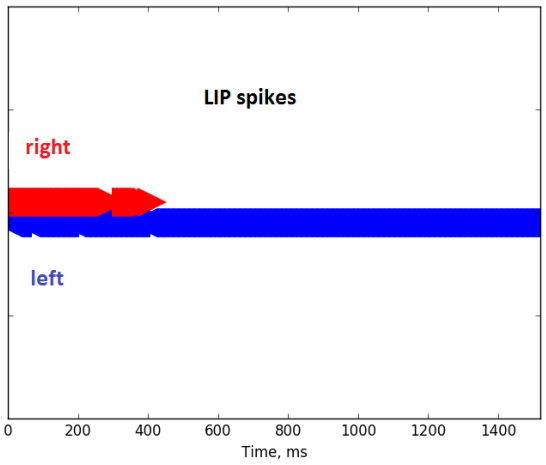


Fig. 7. Model reaction after training with teaching signal for elderly test group. Blue arrow shows spikes from the left LIP while red arrows - spikes from the right LIP neuron.

for all three age groups together. It is clear that in the case of connectivity matrices having positive as well as negative values (connections from MSTe to both LIP areas), only excitatory weights were changed while in case of only negative values (connections from MSTc to both LIP areas) changes were more significant for inhibitory connections from MSTc to the left LIP area and less for those from MSTc to the right LIP area.

Tables II and III represent mean and variance of connection weights that were changed.

TABLE II
CONNECTION WEIGHTS MEAN VALUES.

Connections	Age Group		
	Young	***Middle***	***Old***
MSTe to LIPleft positive	0.9225	0.9293	0.9138
MSTe to LIPright positive	0.9298	0.9237	0.9236
MSTc to LIPleft negative	-0.9456	-0.9430	-0.9424
MSTc to LIPright negative	-0.9999	-0.9998	-0.9746

TABLE III
CONNECTION WEIGHTS VARIANCES.

Connections	Age Group		
	Young	***Middle***	***Old***
MSTe to LIPleft positive	$6.18E^{-5}$	$8.20E^{-5}$	$3.22E^{-5}$
MSTe to LIPright positive	$8.258E^{-5}$	$1.41E^{-5}$	$1.22E^{-5}$
MSTc to LIPleft negative	0.0052	0.0057	0.0058
MSTc to LIPright negative	$1.80E^{-10}$	$7.33E^{-8}$	0.00123

From figures 11 - 14 it is obvious that the biggest differences in connection weights for different age groups are observed for connections from MST area to left LIP area (that corresponds to the correct response). Although mean weight values are quite similar for the three groups of tested subjects (see Table II), mean absolute values of connection

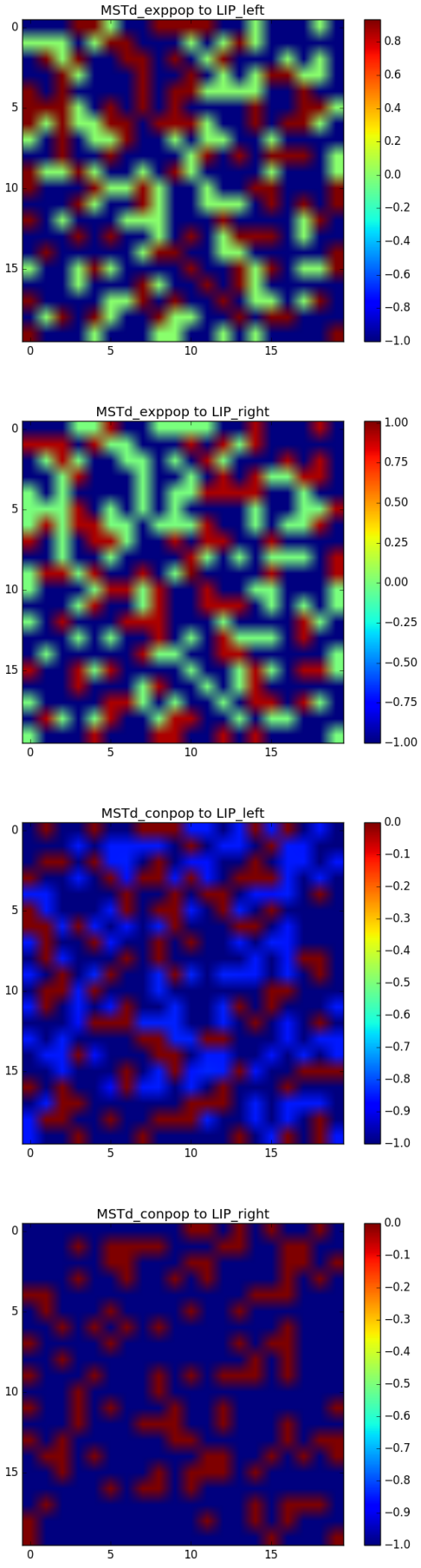
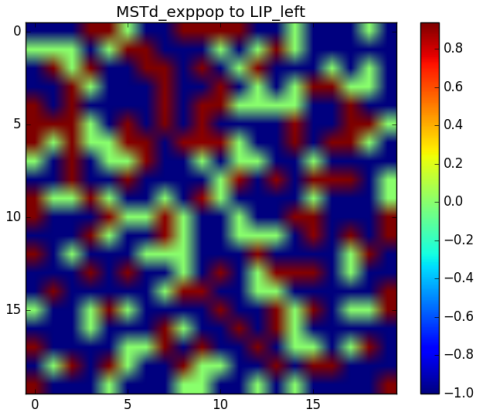
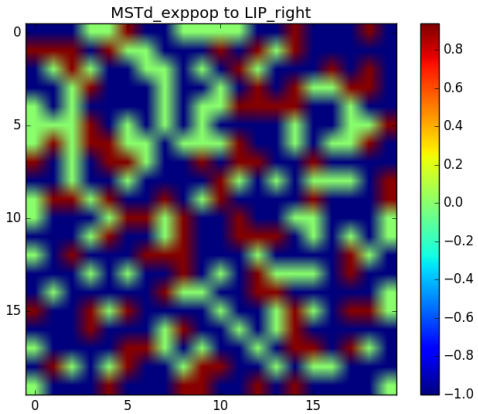
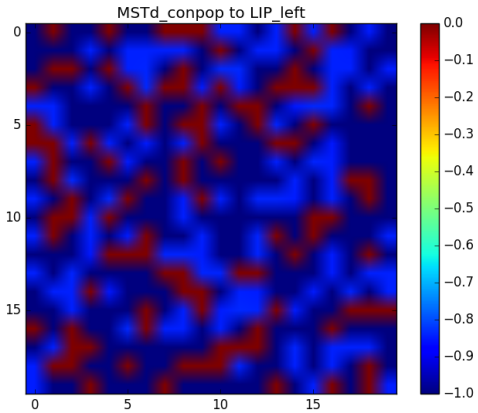


Fig. 8. Trained model parameters for young group.







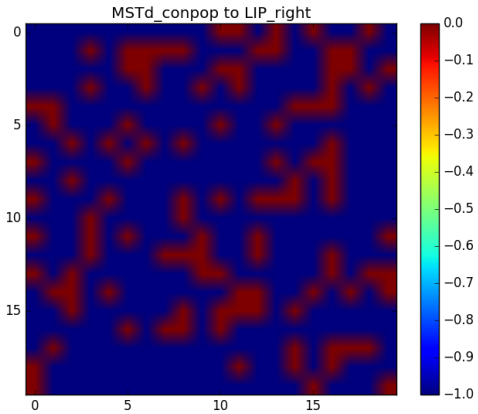
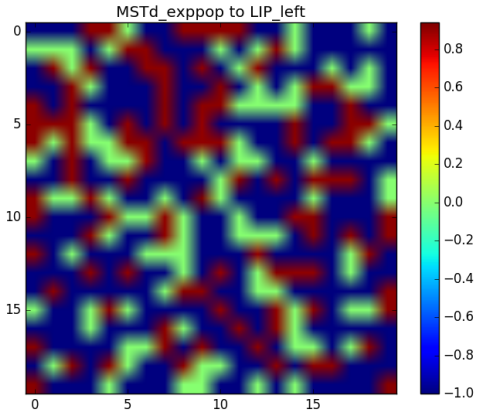
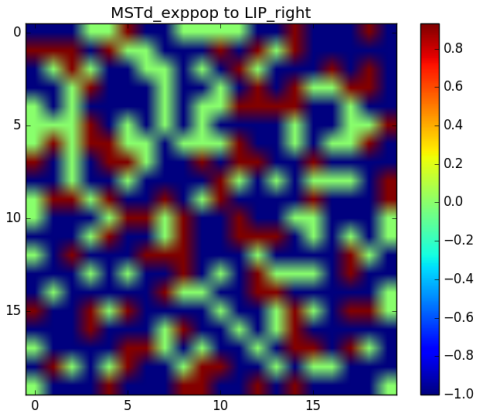
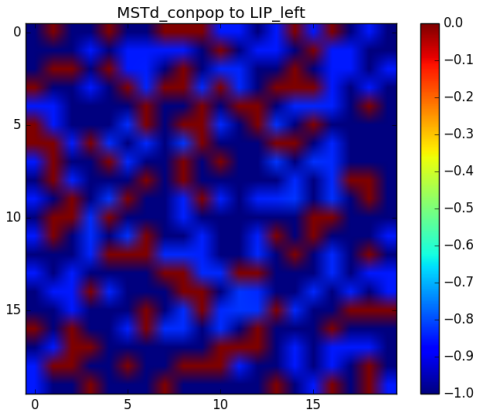


Fig. 9. Trained model parameters for middle group.







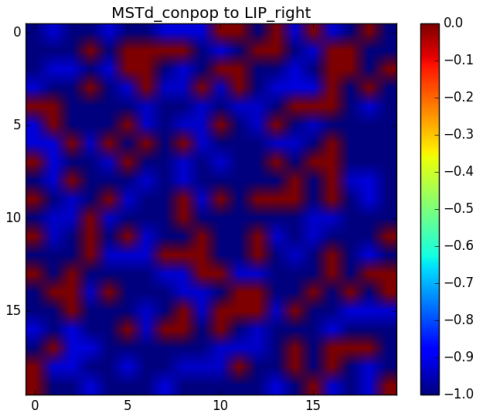


Fig. 10. Trained model parameters for old group.

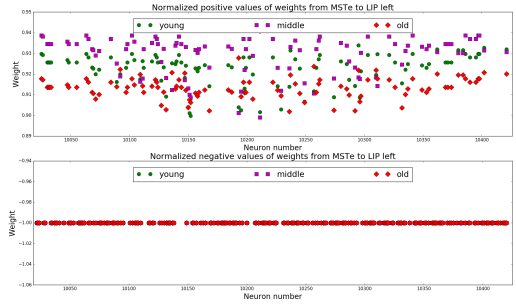


Fig. 11. Normalized values of connection weights from MSTe to LIP left area.

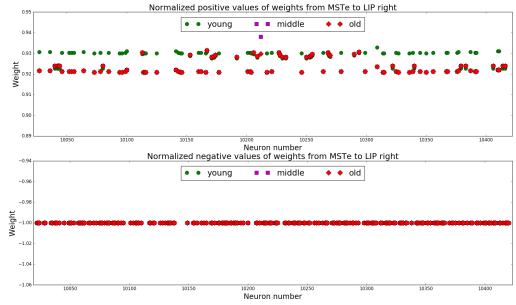


Fig. 12. Normalized values of connection weights from MSTe to LIP right area.

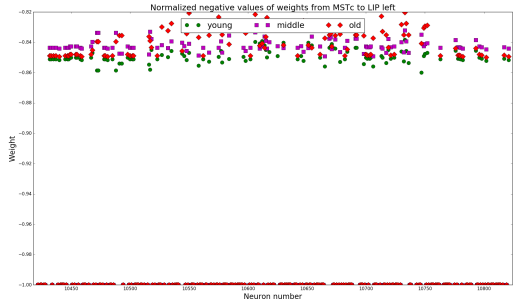


Fig. 13. Normalized values of connection weights from MSTc to LIP left area.

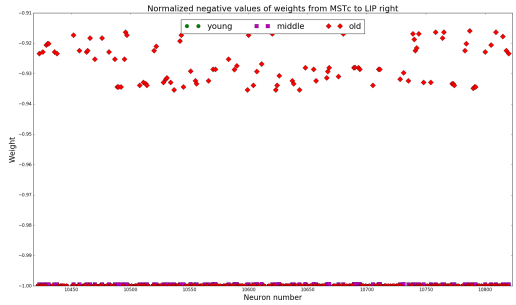


Fig. 14. Normalized values of connection weights from MSTc to LIP right area.

weights for elderly group are smaller, while their variances (see Table III) tend to be bigger for inhibitory connections and smaller for the excitatory once. This effect is most significant for inhibitory connections from contraction template MSTc neurons to the LIP right area (corresponding to wrong decision in considered case). Hence, our results led to conclusion that aging decreases overall connectivity strength and provokes diverse inhibitory connections. They also imply a change in the balance between the weight of the excitatory and inhibitory connections between the expansion and contraction layers of MST and the LIP area with age. This change might induce instability in the network relating these two brain structures reducing the accuracy of the responses and increasing the decision time. The lessened weight of the connections in the two layers of MST could be considered as indicator of brain re-organization with ageing. The results also imply that the decision choice in LIP is determined not only by the changes in the connection weights between the templates corresponding to the correct response but by the changes in the weights between all connections between MST and LIP.

V. Conclusions

These are our first results on STDP training of the developed hierarchical visual information processing model. They concern only connectivity between last two layers of the model (from MST to LIP area). Further improvement of the results could be achieved by allowing dynamic (STDP) connectivity between lower layers in the hierarchical structure as well.

Observed parameter differences are not so significant at this stage since the training signal design was based on mean latency of eye movements for each one of the tested age groups. We hope that further enrichment of training signals using individuals' latency as well as the experimental results from all four stimuli conditions will allow to refine the model parameters in future.

Even these preliminary results demonstrated that STDP plasticity in combination with behavioral experimental data rather than with electro-physiological recordings directly from the brain could simulate logical changes of synaptic strength with aging.

References

[1] P. Bayerl and H. Neumann, "Disambiguating visual motion through contextual feedback modulation," Neural Computation, vol. 16 (10), pp.2041-–2066, 2004.

[2] P. Bayerl, A Model of Visual Perception, PhD Thesis, Ulm University, Germany, 2005.

[3] M. Chessa, S. Sabatini, and F. Solari, "A systematic analysis of a V1–MT neural model for motion estimation," Neurocomputing, vol. 173, pp. 1811-–1823, 2016.

[4] E. Simoncelli and D. Heeger, "A model of neuronal responses in visual area MT," Vision Research, vol. 38, pp. 743-–761, 1998.

[5] F. Raudies, E. Mingolla, and H. Neumann, "Active gaze control improves optic flow-based segmentation and steering," PLoS ONE, vol. 7 (6), pp. 1—19, 2012.

[6] S. Grossberg, E. Mingolla, and C. Pack, "A neural model of motion processing and visual navigation by cortical area MST," Cerebral Cortex, vol. 9 (8), pp. 878-—895, 1999.

[7] J. Perrone, "A neural-based code for computing image velocity from small sets of middle temporal (MT/V5) neuron inputs," Journal of Vision, vol. 12 (8), doi: 10.1167/12.8.1, 2012.

[8] F. Solari, M. Chessa, K. Medathati, and P. Kornprobst, “What can we expect from a V1-MT feedforward architecture for optical flow estimation?,” Signal Processing: Image Commuication, vol. 39, Part B, pp. 342–354, 2015.
[9] S. Grossberg, E. Mingolla, and L. Viswanathan, “Neural dynamics of motion integration and segmentation within and across apertures,” Vision Research, vol. 41 (19), pp. 2521—2553, 2001.
[10] F. Raudies and H. Neumann, “A neural model of the temporal dynamics of figure–ground segregation in motion perception,” Neural Networks, vol. 23 (2), pp. 160-–176, 2010.
[11] N. Kasabov, Time-Space, Spiking Neural Networks and Brain-Inspired Artificial Intelligence, Part of the Springer Series on Bio- and Neurosystems, vol. 7, Springer-Verlag Berlin Heidelberg, 2019.
[12] P. Koprinkova-Hristova, N. Bocheva, S. Nedelcheva, and M. Stefanova, “Spike timing neural model of motion perception and decision making,” Frontiers in Computational Neuroscience, vol. 13, art. no. 20, doi: 10.3389/fncom.2019.00020, 2019.
[13] S. Kunkel et al., NEST 2.12.0. Zenodo, doi: 10.5281/zenodo.259534, 2017.
[14] J. Kremkow, L. U. Perrinet, C. Monier, J.-M. Alonso, A. Aertsen, Y. Fregnac, and G. S. Masson, “Push-pull receptive field organization and synaptic depression: Mechanisms for reliably encoding naturalistic stimuli in V1,” Frontiers in Neural Circuits, vol. 10, doi: 10.3389/fncir.2016.00037, 2016.
[15] T. W. Troyer, A. E. Krukowski, N. J. Priebe, and K. D. Miller, “Contrast invariant orientation tuning in cat visual cortex: thalamocortical input tuning and correlation-based intracortical connectivity,” J. Neurosci., vol. 18, pp. 5908-–5927, 1998.
[16] M. Ghodratia, S.-M. Khaligh-Razavic, and S. R. Lehky, “Towards building a more complex view of the lateral geniculate nucleus: Recent advances in understanding its role,” Progress in Neurobiology, vol. 156, pp. 214–255, 2017.
[17] S. Nedelcheva and P. Koprinkova-Hristova, “Orientation selectivity tuning of a spike timing neural network model of the first layer of the human visual cortex,” in: Advanced Computing in Industrial Mathematics, K. Georgiev, M. Todorov, and I. Georgiev (Eds.), Studies in Computational Intelligence, vol. 793, pp. 291-–303, 2019.
[18] S. Sadeh and S. Rotter, “Statistics and geometry of orientation selectivity in primary visual cortex,” Biol. Cybern., vol. 108, pp. 631-653, 2014.
[19] M.-J. Escobar, G. S. Masson, T.·Vieville, and P. Kornprobst, “Action recognition using a bio-inspired feedforward spiking network,” Int. J. Comput. Vis., vol. 82, pp. 284–301, 2009.
[20] O. W. Layton and B. R. Fajen, “Possible role for recurrent interactions between expansion and contraction cells in MSTd during self-motion perception in dynamic environments,” Journal of Vision, vol. 17 (5), pp. 1–21, 2017.
[21] A. Casti, F. Hayot, Y. Xiao, and E. Kaplan, “A simple model of retina-LGN transmission,” J. Comput. Neurosci., vol. 24, pp. 235-252, 2008.
[22] M. Tsodyks, A. Uziel, and H. Markram, “Synchrony generation in recurrent networks with frequency-dependent synapse,” The Journal of Neuroscience, vol. 20 (RC50), pp. 1–5, 2000.
[23] N. Bocheva, B. Genova, M. Stefanova, “Drift diffusion modeling of response time in heading estimation based on motion and form cues,” Int. J. of Biology and Biomedical Engineering, vol. 12, pp. 75-–83, 2018.
[24] R. Kraleva, V. Kralev, N. Sinyagina, P. Koprinkova-Hristova, and N. Bocheva, N., “Design and analysis of a relational database for behavioral experiments data processing,” International Journal of Online Engineering, vol. 14 (2), pp. 117–132, 2018.