Comenius University in Bratislava

Faculty of Mathematics, Physics and Informatics

Department of Applied Informatics

REPRESENTATION OF OBJECT POSITION IN VARIOUS FRAMES OF REFERENCE USING A ROBOTIC SIMULATOR

Master Thesis

Student: Marcel Švec

Supervisor: doc. Ing. Igor Farkaš, PhD.

June 2013

Introduction

- We are able to accurately reach for the object that we see.

- In the brain, the information about object position is represented by populations of neurons.

- Neurons in early visual pathways represent spatial information relative to the retinal position
  → they use eye-centered frame of reference.

Blohm et al. (2008)

- The visual information changes with every eye- or head-movement, but we still perceive the world as stable. Therefore the brain has to use also posture signals (gaze direction, head tilt…) in order to create other representations suitable for the given task, e.g. for reaching it might be useful to use representation in hand-centered frame of reference.

- We ask: what are the computational principles underlying these transformations?
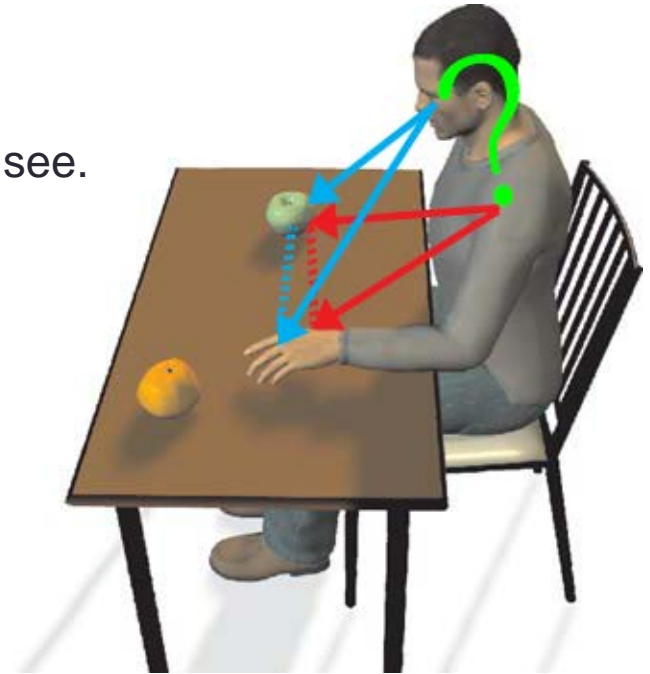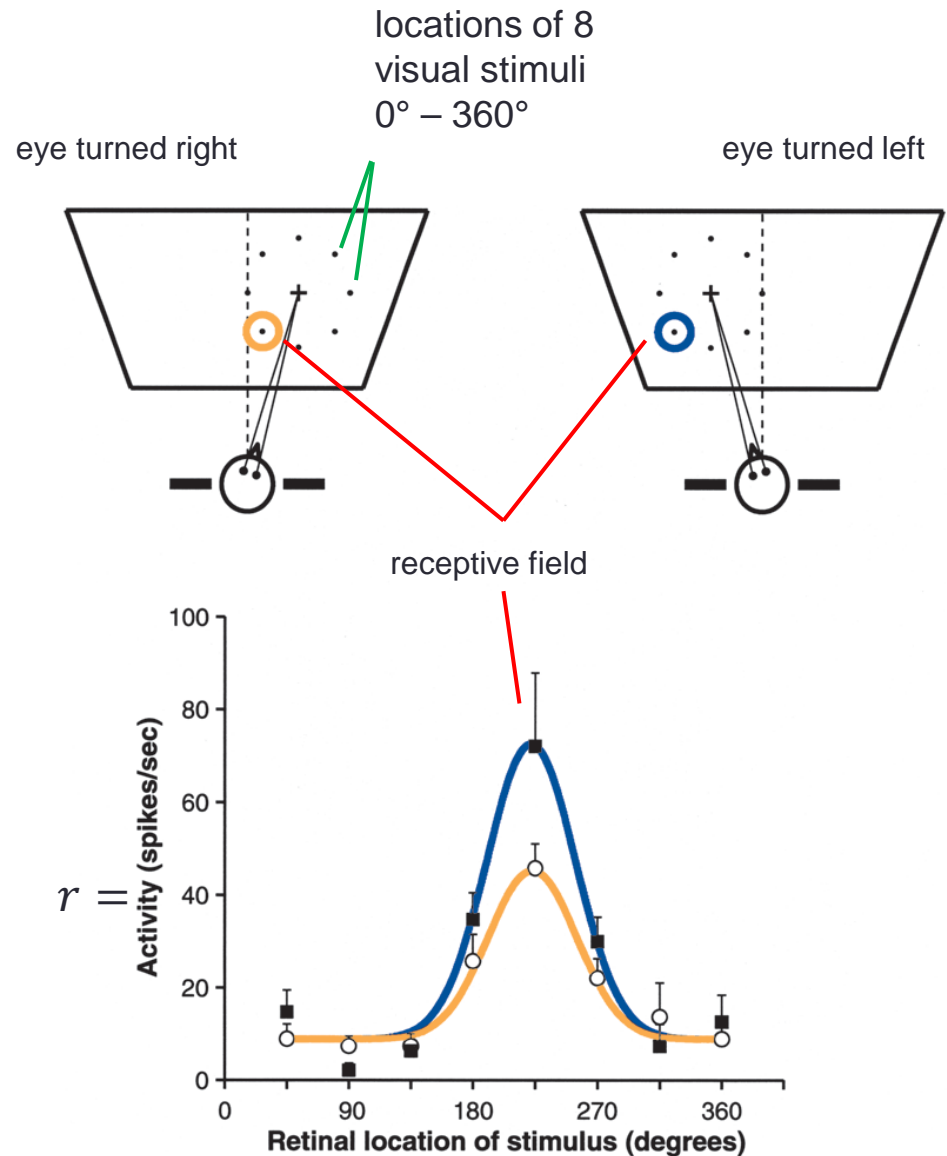
Content

- Reference frames

- Gain modulation and gain fields

- Feed-forward and basis-functions neural network models

- Our experiment using data generated in the iCub simulator

- Conclusions

Frames of reference

- Egocentric vs. allocentric reference frame

- We focus on:
  - Eye-centered
  - Head-centered
  - Hand-centered (body-centered)

- Examples:
  - Eye-centered: The organization of neurons in the primary visual cortex (V1) is topographic, meaning that receptive fields of adjacent neurons represent points nearby in visual space. *(not inevitable in general)*

  - Head-centered: Neuron's activity does not change with eye-movement (assuming the same visual stimulus), but do change along with the head-movement.

Gain modulation

- Nonlinear combination of information from two modalities.

- The sensitivity is modulated by one modality (e.g. postural) without changing the selectivity to the other modality (e.g. sensory).

- Example:
  - Neuron's visual responses are gain-modulated by gaze angle.

  - The response function changes amplitude (gain), but the preferred location and shape remain.

- Computing with gain fields:
  - $r = f(x_{target} - a)g(x_{gaze})$
  - $R = F(c_1 x_{target} + c_2 x_{gaze})$

locations of 8 visual stimuli $0° - 360°$

eye turned right

eye turned left

receptive field

$r =$

Activity (spikes/sec)

Retinal location of stimulus (degrees)

Salinas and Sejnowski (2001)

Computing with gain fields – coordinate transformation
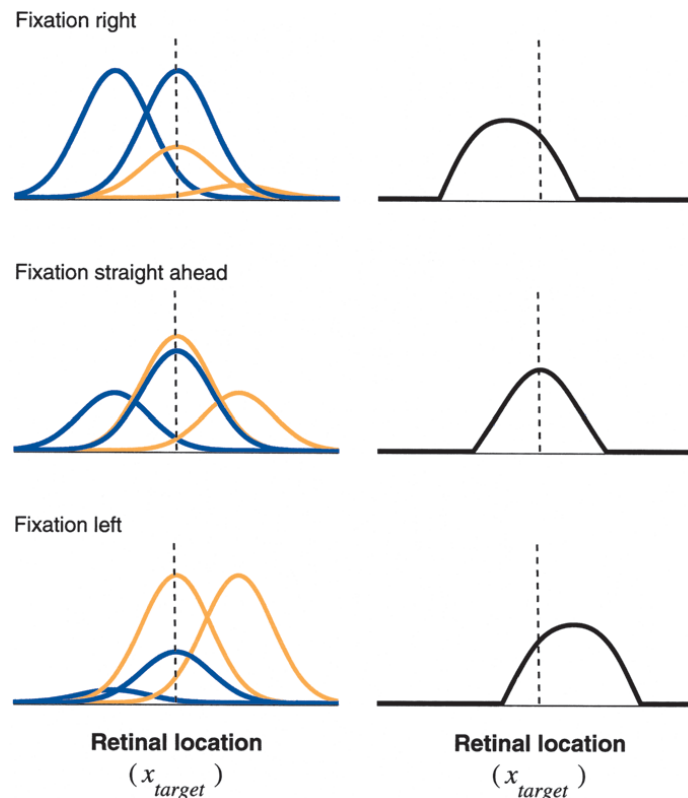
- Neuron's response:

$$r = f(x_{target} - a)g(x_{gaze})$$

- Population of downstream neurons may thus represent:

$$x_{target} + x_{gaze}$$

Salinas and Sejnowski (2001)

- Response of the downstream neuron

$$R = F(c_1 x_{target} + c_2 x_{gaze})$$
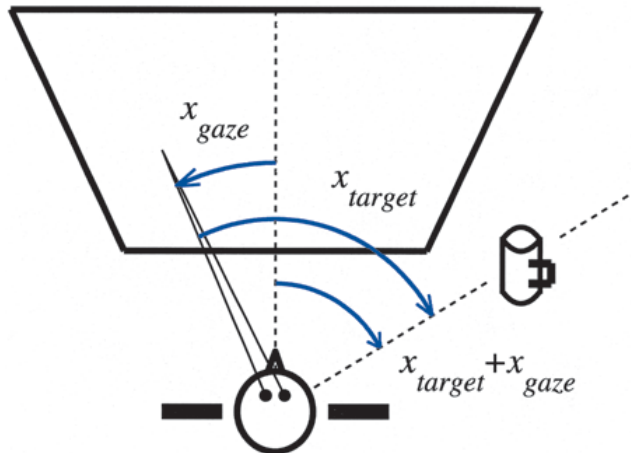
- Fixed visual stimulus, different eye fixations:

Fixation right

Fixation straight ahead

Fixation left

Retinal location
(x_{target})

Retinal location
(x_{target})

Receptive field shifts – indicates different reference frame,
e.g. head- or body-centered

Gain fields in neural networks

- Zipser and Andersen (1988) trained 3-layer feed-forward neural network to compute head-centered target position from eye-centered visual stimulus and gaze direction (2D).
- Hidden neurons developed gain-fields similar to what had been observed in PPC of macaque monkeys.


Zipser and Andersen (1988)
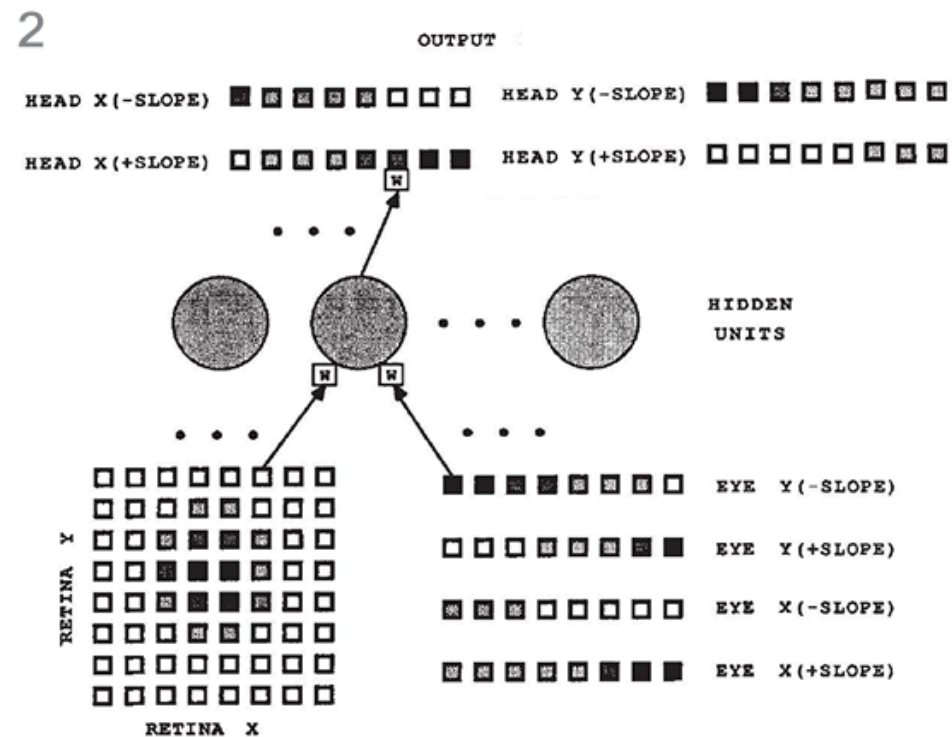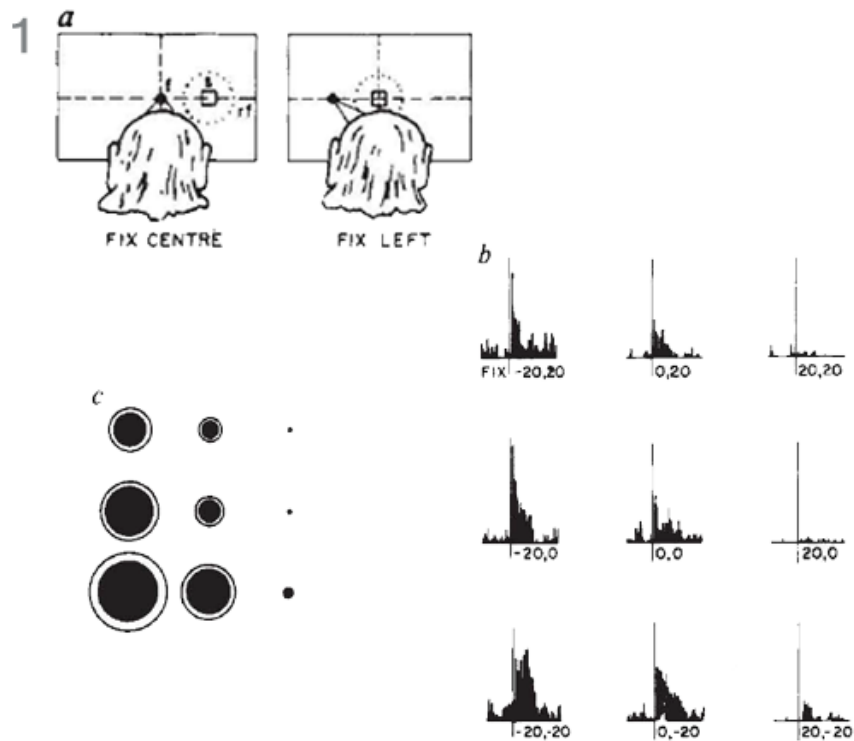
Advanced feed-forward model for reaching in 3D

- 4-layered feed-forward network

- Input:
  - eye-centered hand and target positions and disparities
  - eye position
  - head position
  - vergence

- 2-hidden layers

- Output (read-out) layer – desired reaching vector (3D)


Blohm et al., (2009)
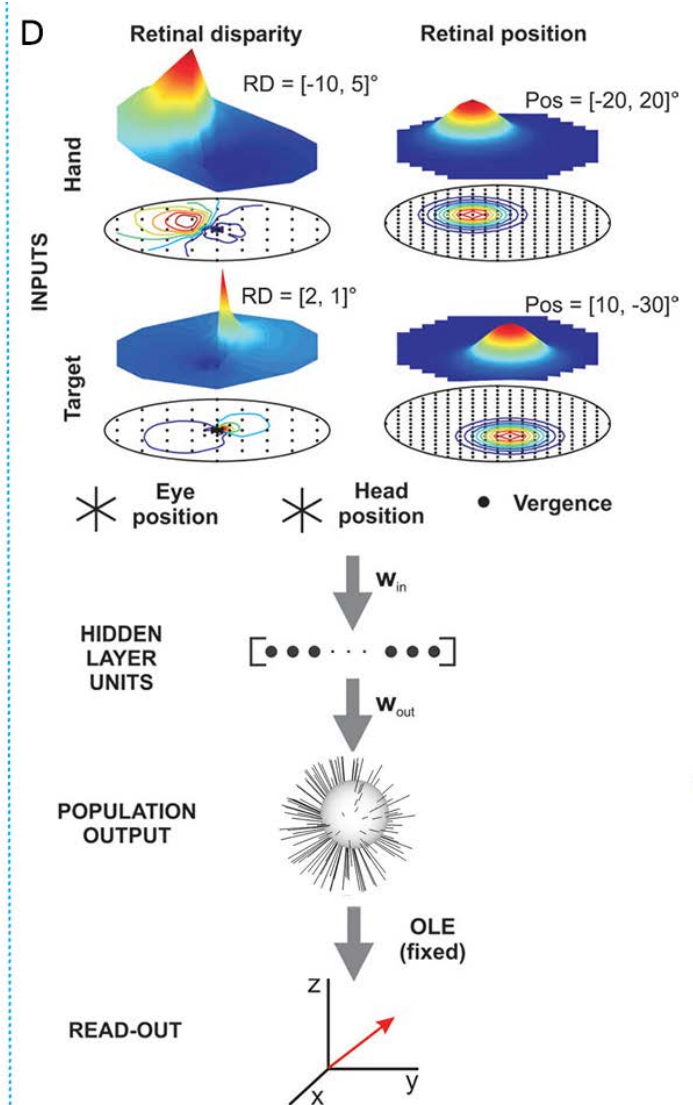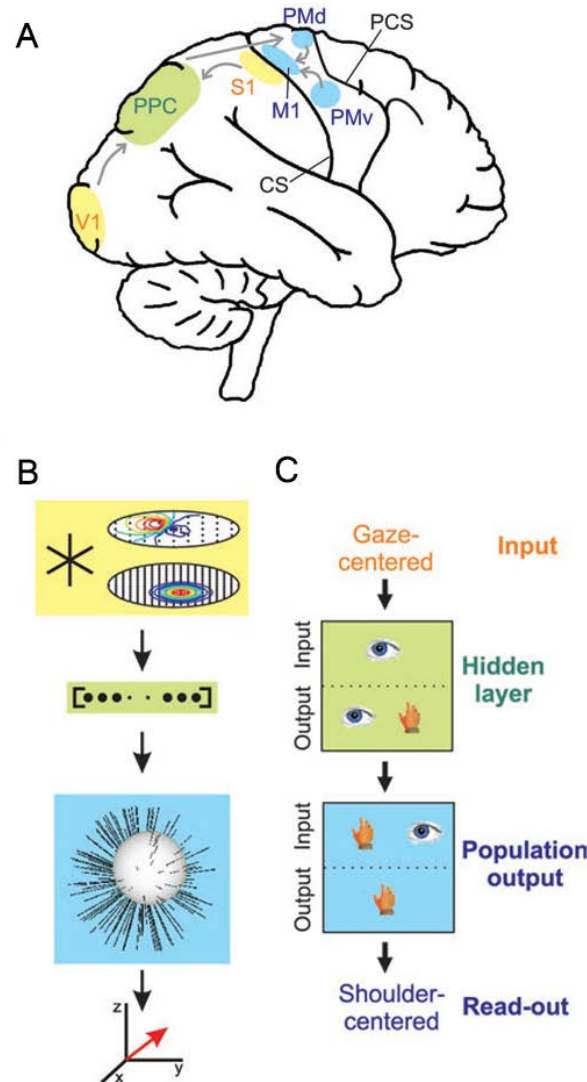
Basis functions networks

- Parietal neurons behave like the basis functions of the input signals.
- The same basis functions can be used to compute many motor plans.
- Recurrent connections enable computations in any direction and solve statistical issues.
- Basis functions are learned in unsupervised manner.
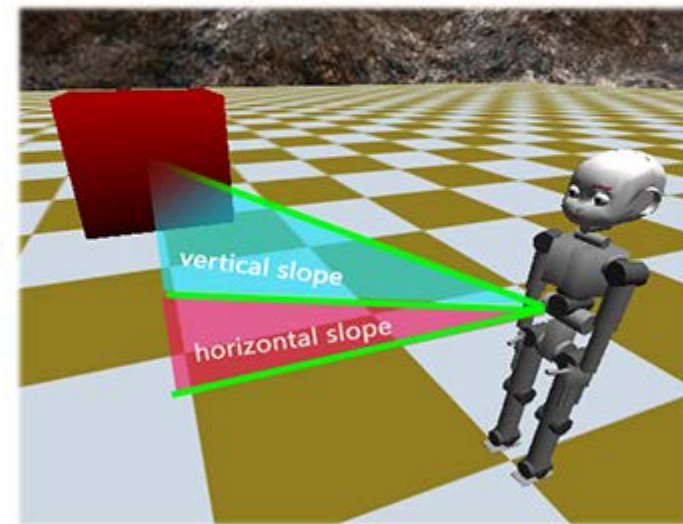- Course of dimensionality.

Pouget and Snyder (2000)
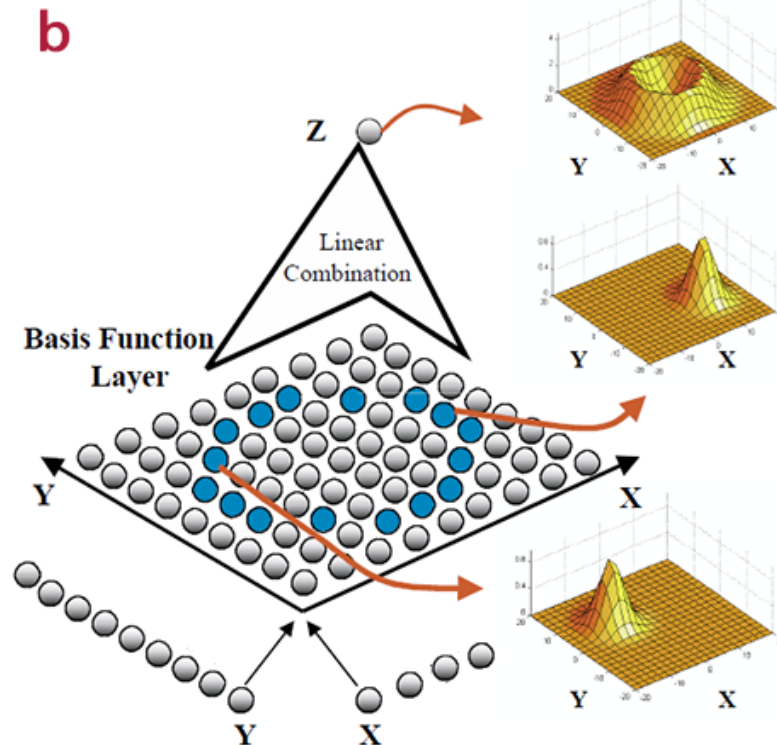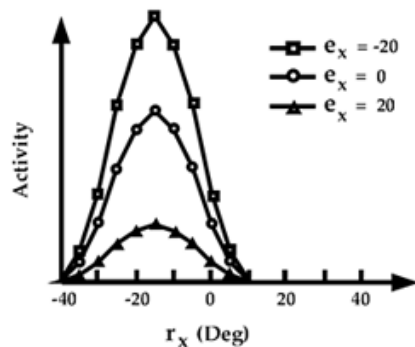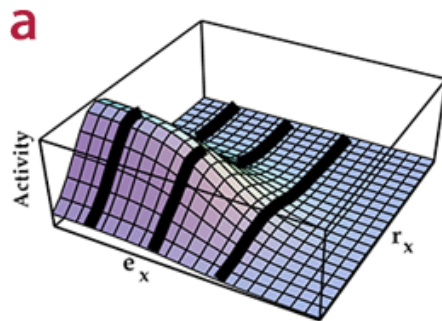
Experiment

- Input:
  - eye position – vertical and horizontal orientation (angle)
  - visual stimulus – images from the left and right eye (processed)
- Output:
 - body-referenced target position represented by horizontal and vertical slope
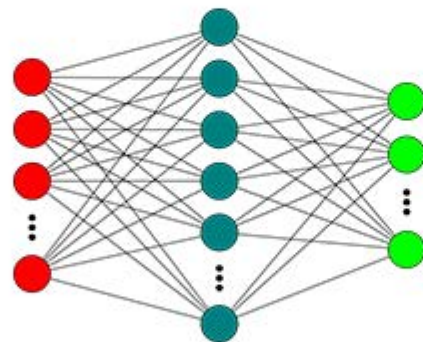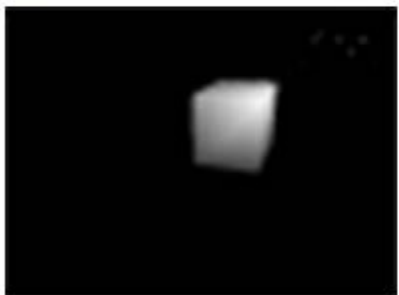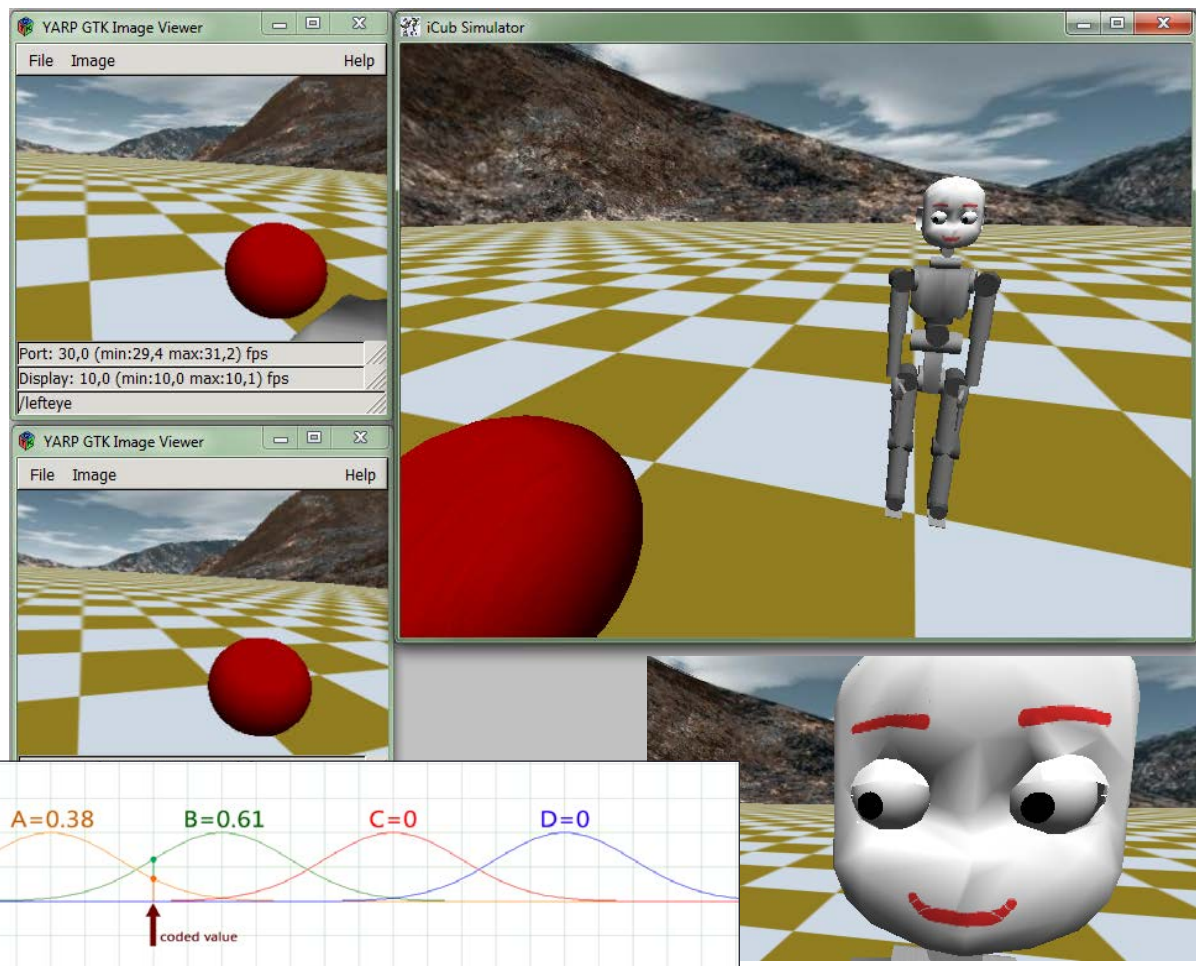

3-layered feed-forward network

Experiment – generating dataset in iCub simulator

- iCub cameras – pinhole projection, resolution 320x240 px, eyes limits are $\langle -35°, 15° \rangle$ vertically and $\langle -50°, 50° \rangle$ horizontally

- Objects at random locations (but in FOV).
- Random sizes (but in some limits with respect to perspective)
- Random shapes: sphere, cylinder, box
- 1500 patterns
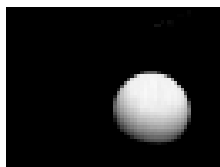
Processed image

Experiment – network model

- Input layer – **6176** neurons
  - *eye_tilt* + *eye_version* + *left_eye_image* + *right_eye_image* = 11 + 21 + 64*48 + 64*48
  - Width of tuning curves: tilt $\sigma = 5$, version $\sigma = 7$

- Hidden layer – **64** neurons
  - limited performance with less than 40 neurons
  - activation function – sigmoid, $gain = 0{,}05$, balancing retinal and eye-position inputs:
  - $net = r \cdot \sum_j^{N_r} w_i r_i + e \cdot \sum_j^{N_e} w_j e_j$ $\qquad f(net) = 1/e^{-gain*net}$
  - $r = \frac{R \cdot (N_e + N_r)}{N_r \cdot (R+E)}$ $\qquad e = \frac{E \cdot (N_e + N_r)}{N_e \cdot (R+E)}$ $\qquad \text{R:E} = 2:1$

- Output layer – **38** neurons
  - x-slope + y-slope = 19 + 19. Every 10 degrees in interval $\langle -90°, 90° \rangle$, $\sigma = 10$
  - activation function – sigmoid, $gain = 0{,}1$

Experiment – training, results

- Training
  - FANN – Fast Artificial Neural Network Library (C, many bindings)
  - Backpropagation, RPROP, quickprop, momentum
  - 1000 patterns
- Results
  - Mean squared error $MSE < 5 \cdot 10^4$
  - Backpropagation with learning rate $\alpha = 1.5$
    and momentum term $\mu = 0.9$
  - Accuracy for dataset with spheres of the same size was $2°$ (mean and standard deviation), for complex datasets $4°$
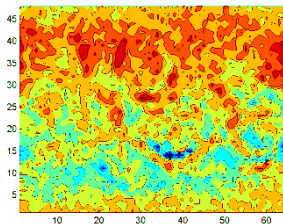
Distribution of errors over 500 testing patterns:
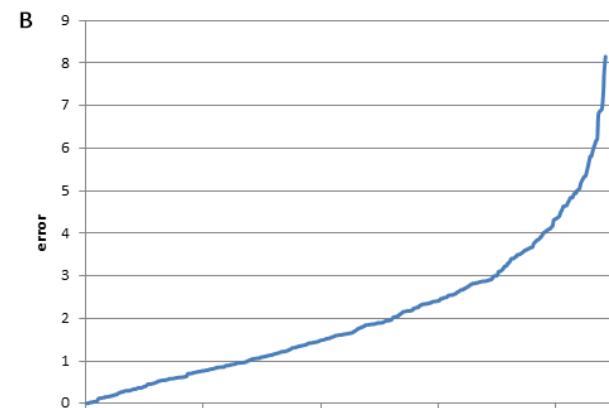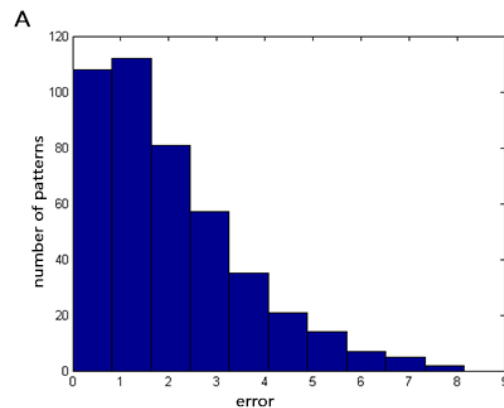
Experiment – hidden layer analysis – receptive fields

- The majority of units developed continuous receptive fields for particular area in visual space.
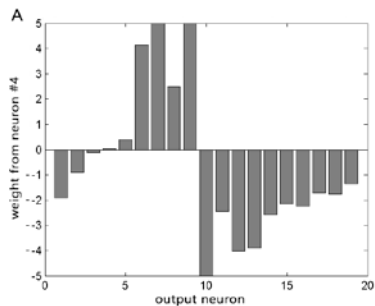
- A – 41, B – 15, C – 8 units
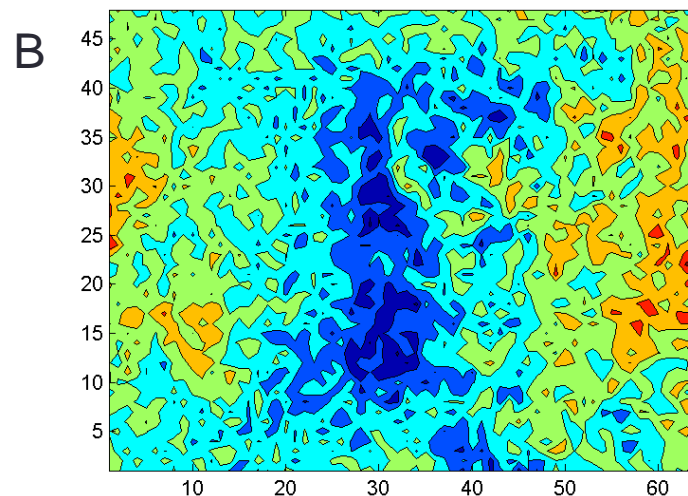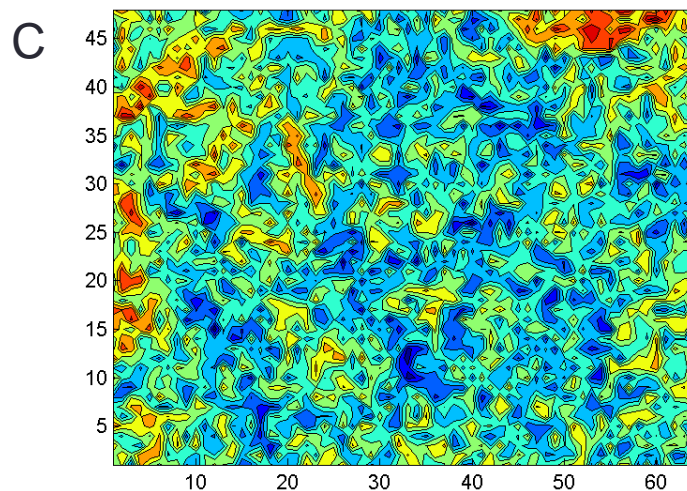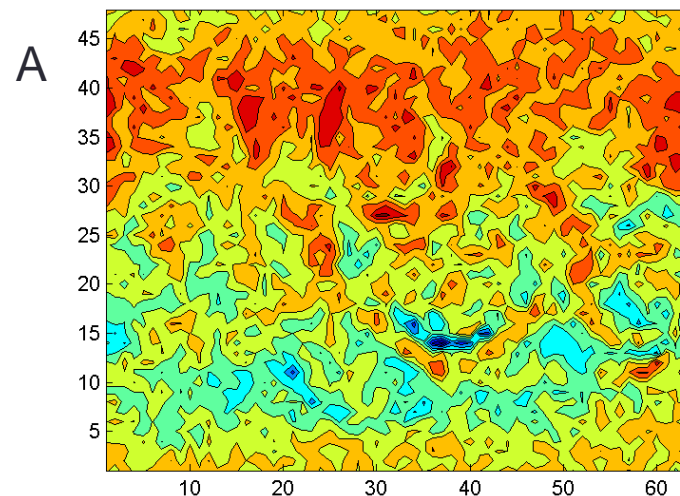
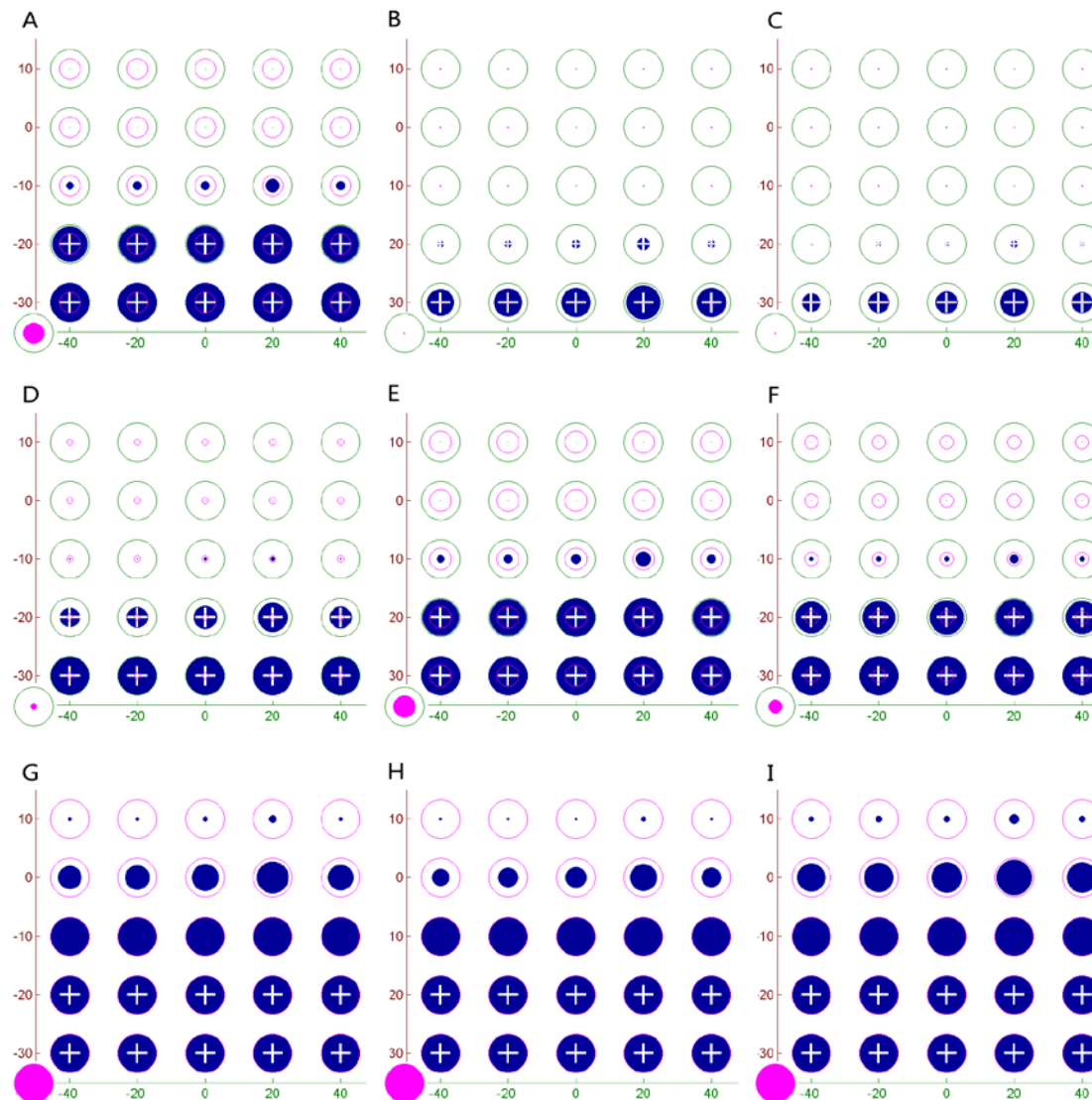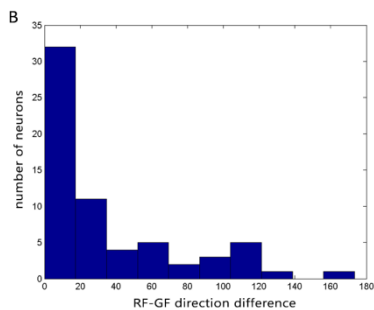# Experiment – hidden layer analysis – gain modulation 1/2

Response field of hidden unit #4 for various visual stimuli and gaze direction:

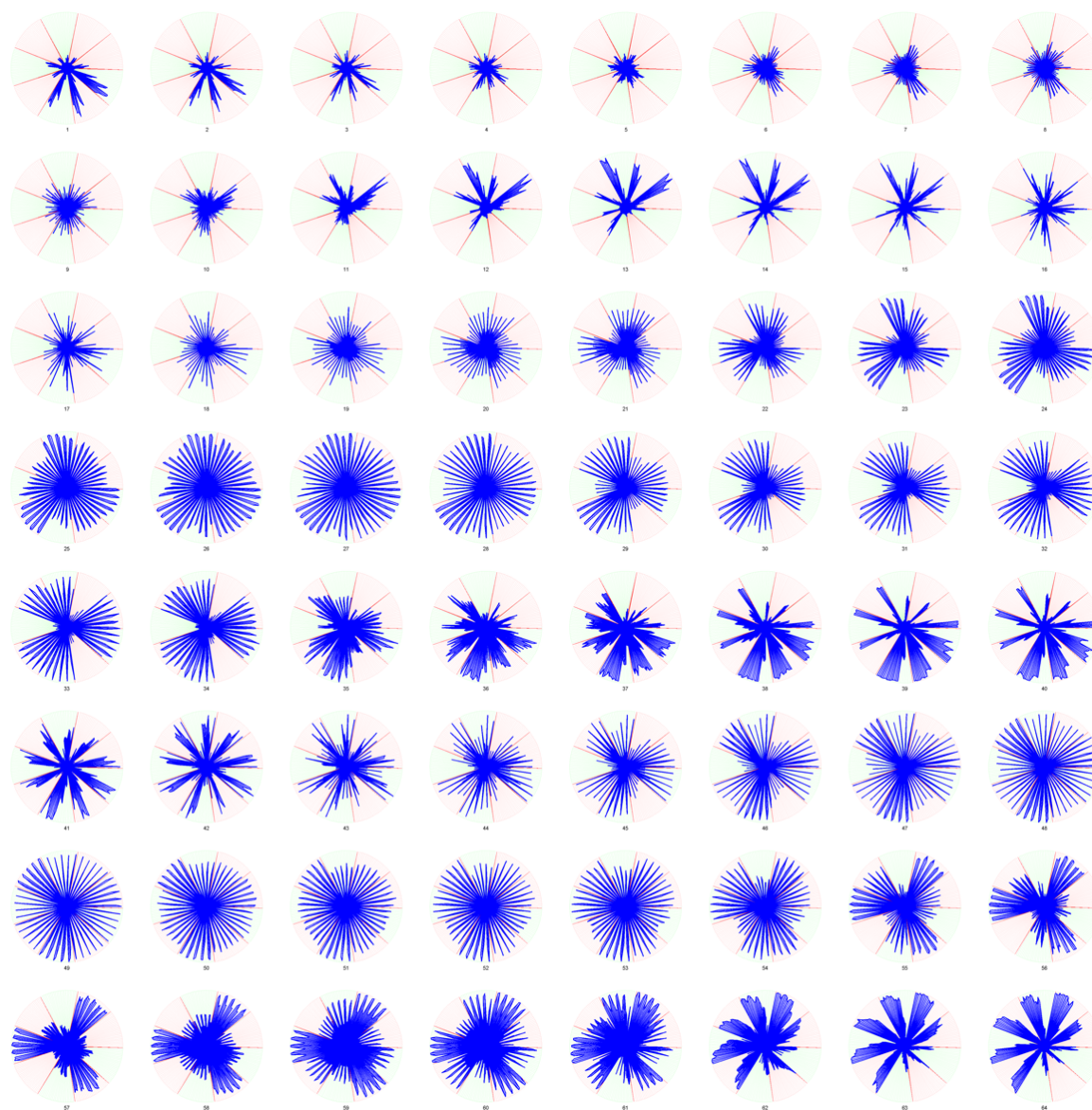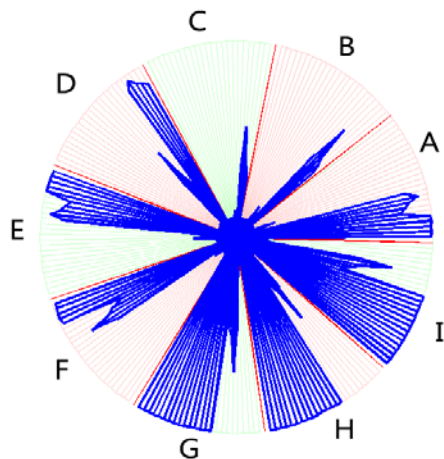Weights to vertical output units:

Histogram of differences between the directions of receptive fields and gain fields:
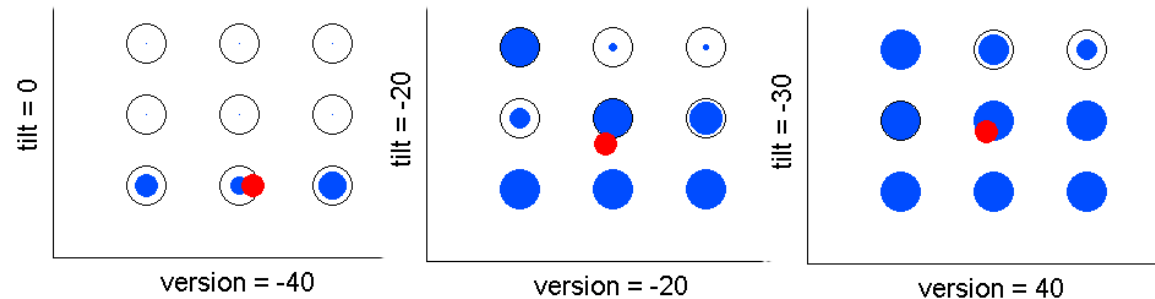
Experiment – hidden layer analysis – gain modulation 2/2

Star-plot
visualisation of
response fields of
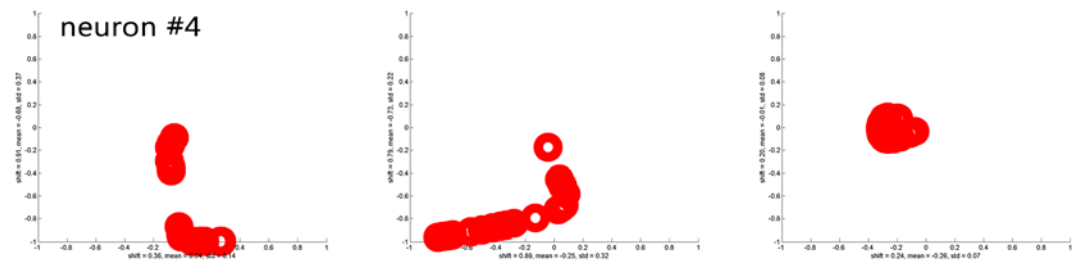all hidden units
sorted by 1-D SOM

Experiment – hidden layer analysis – reference frames
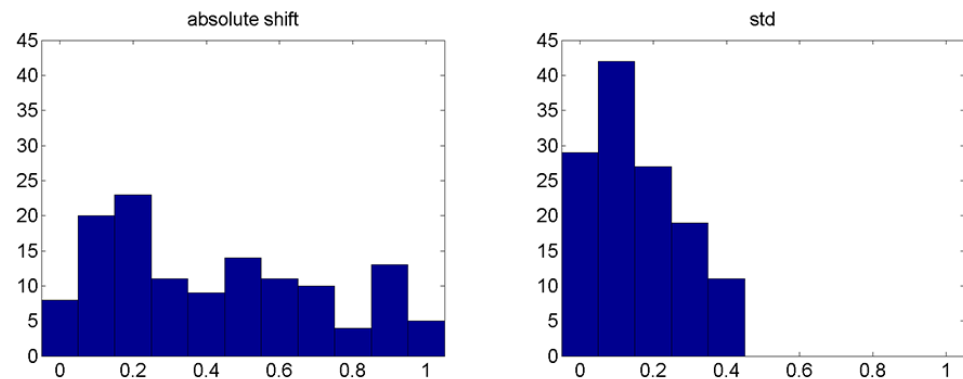
Analysis of shifts of receptive fields for hidden unit #4:

Examples of RF shifts:

Histogram of RF shifts for all hidden units:

Conclusions

- The notion of frame of reference is central to spatial representations in neural networks

- Gain modulation is a crucial and widespread mechanism for multimodal integration (coordinate transformations)

- There are network models for spatial transformations based on feed-forward and basis-function networks

- We used iCub simulator for generating data for 3-layer feed-forward neural network that was trained to perform transformation from eye- to body-centered reference frame using the information about gaze direction. Main advantage: this approach accounts for body geometry without the need for the additional mathematical models.

- Accuracy of the network was $\approx 4°$

- Several visualisation techniques revealed the effect of gain modulation. Reference frame analysis indicates that the hidden layer uses intermediate reference frame.

- Possible future work: experiment with the distance and head movements.

svec.marcel@gmail.com

Thanks for your attention

References:

- Blohm et al. (2008). *Spatial transformations for eye–hand coordination*, Volume 9, pp. 203–211. Elsevier Inc.

- Blohm, G., G. Keith, and J. Crawford (2009). Decoding the cortical transformations for visually guided reaching in 3D space. *Cerebral Cortex 19* (6), pp. 1372–1393.

- Pouget, A. and L. H. Snyder (2000). Computational approaches to sensorimotor transformations. *Nature Neuroscience 3 Suppl*, pp. 1192–1198.

- Salinas, E. and T. J. Sejnowski (2001). Gain modulation in the central nervous system: Where behavior, neurophysiology, and computation meet. *The Neuroscientist 7*, pp. 430–440.

- Zipser, D. and R. A. Andersen (1988). A backpropagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature 331*, pp. 679–684.