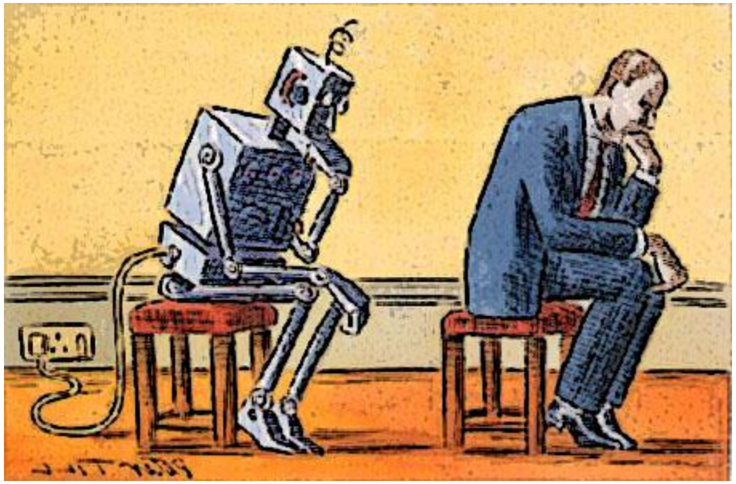
Integrated Neural Symbolic Knowledge Technologies for Action

Stefan Wermter
Knowledge Technology
Dept. of Computer Science
University of Hamburg

http://www.informatik.uni-hamburg.de/WTM/

Motivating Questions

- Cognitive functions like language processing, vision, navigation, learning…
- Human brain performance often far superior compared to computational models



Acting in Cluttered Environments

- We and animals have astonishing abilities of perception and action in cluttered environments

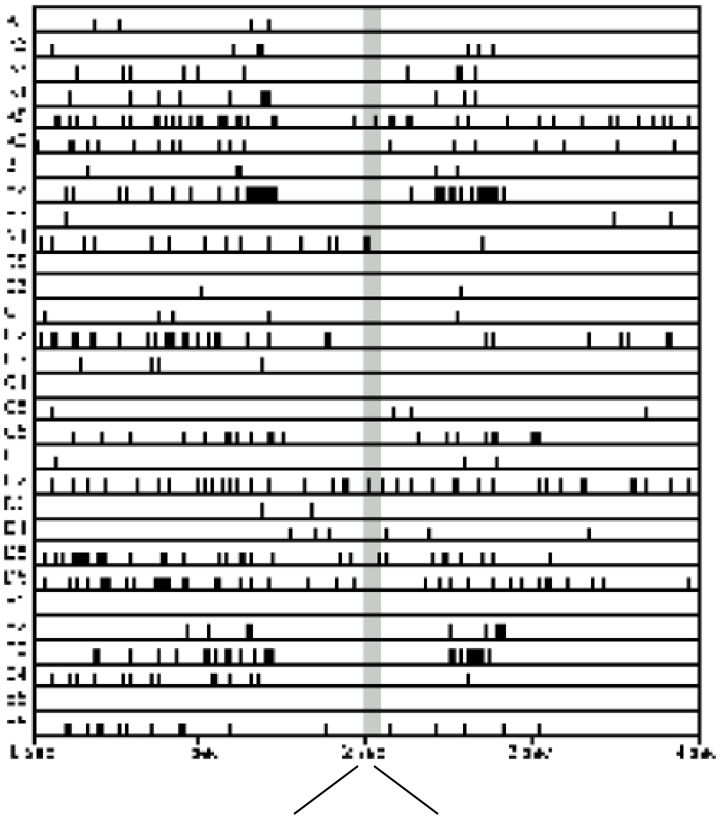
Biological inspiration:

- How does the human brain do the job?
- How can we learn from it for robot perception and action?

Motivating Questions

- How is it possible to bridge the large gap between neural network processing in the brain and intelligent performance of humans?
- How is it possible to build more effective systems which integrate neural techniques into intelligent systems?

How do Neurons communicate and act?

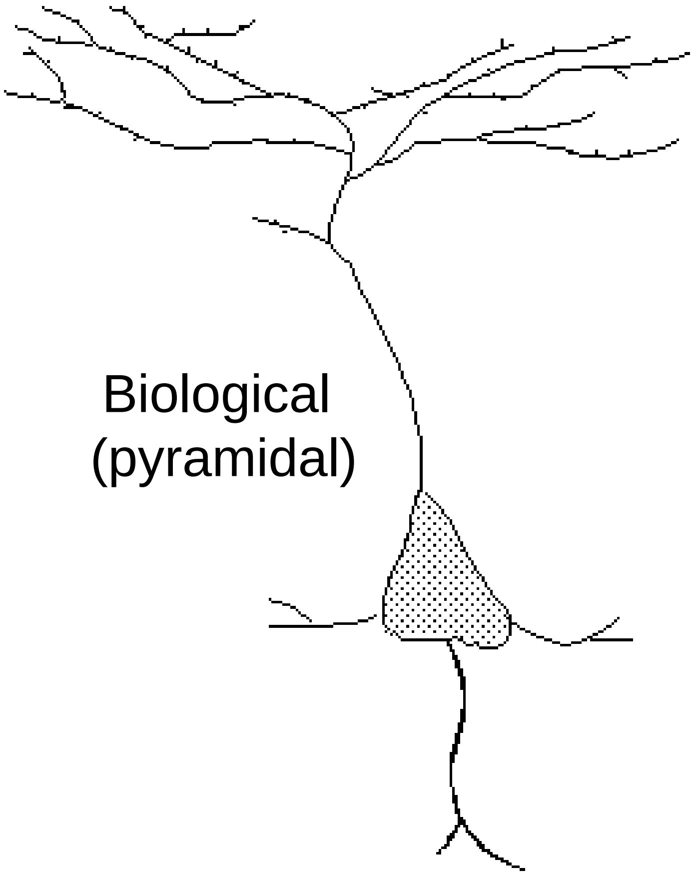


100 ms

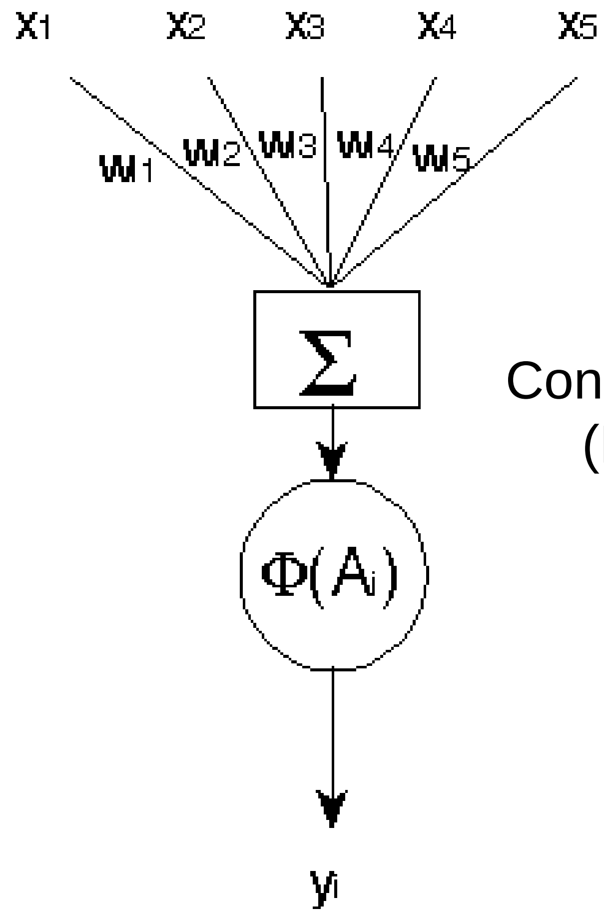
Firing rate → Connectionist neurons

Firing time → Spiking neurons

Biological and Artificial Neurons



Biological (pyramidal)



Connectionist neurons (McCulloch-Pitts)

direction of signal transmission

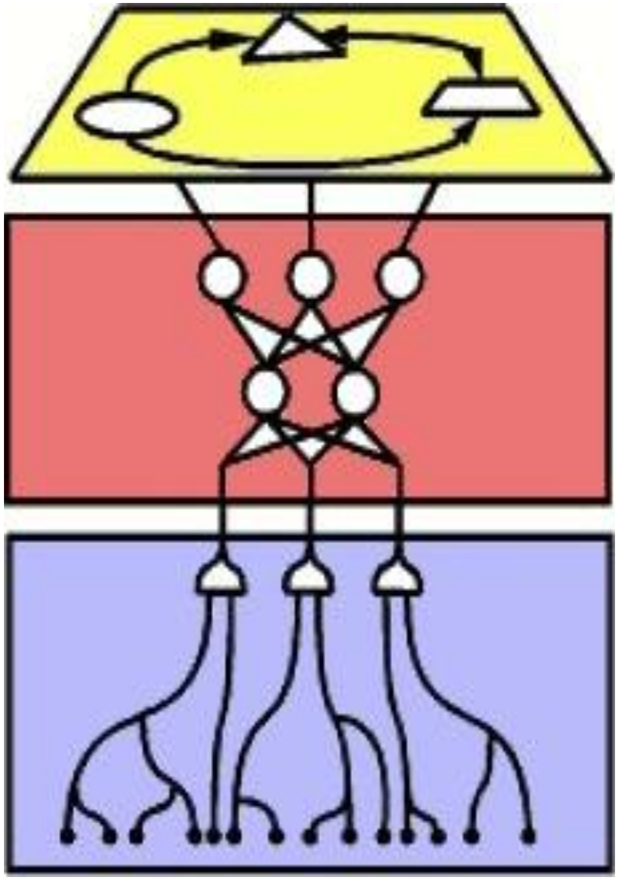
$$y_i = \Phi(A_i) = \Phi\left(\sum_{j}^{N} w_{ij} x_j - \vartheta_i\right)$$

NEST: NEural Symbolic Technology architecture

- Symbolic knowledge and action understanding

- Neural/symbolic knowledge: Connectionist learning, crossmodal integration

- Sensory and neural input from several modalities

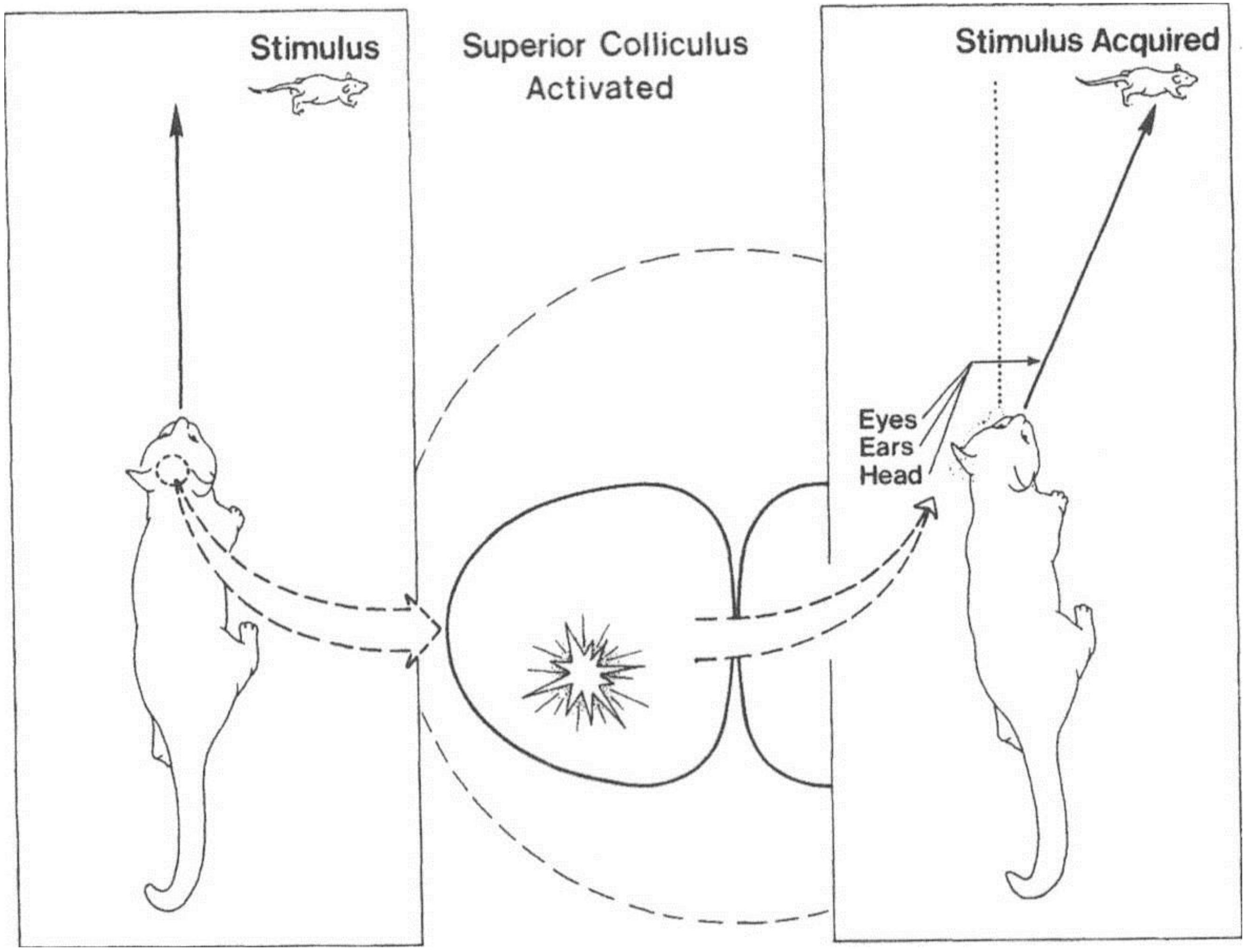


1. How do Multiple Modalities in the Brain Inform Action Understanding?

- Representation should follow some organizational principles of the brain

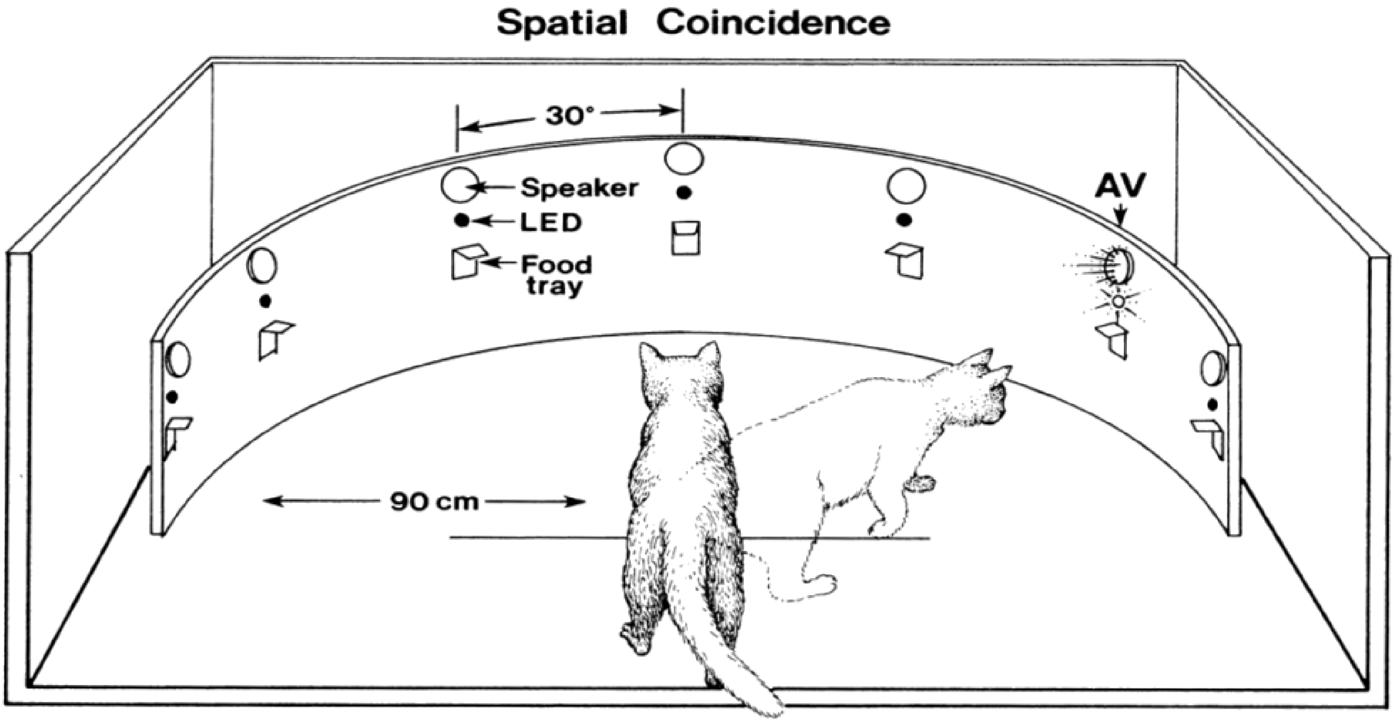
- ***Midbrain: Superior colliculus*** plays a crucial role
  - Contains unimodal visual, auditory, somatosensory and multisensory neurons

Multimodal Integration for Action in the Midbrain

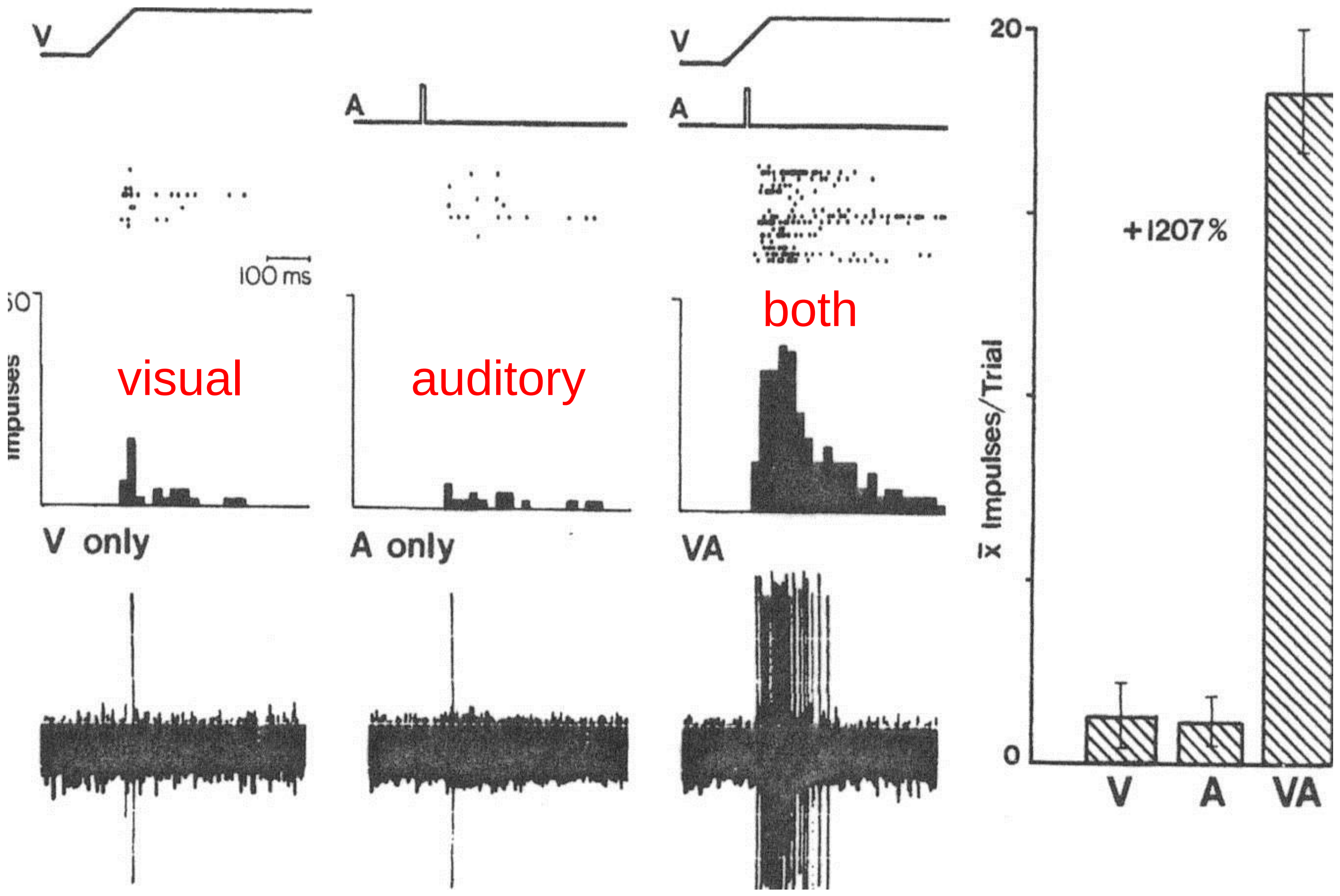


Stein, Meredith et al.

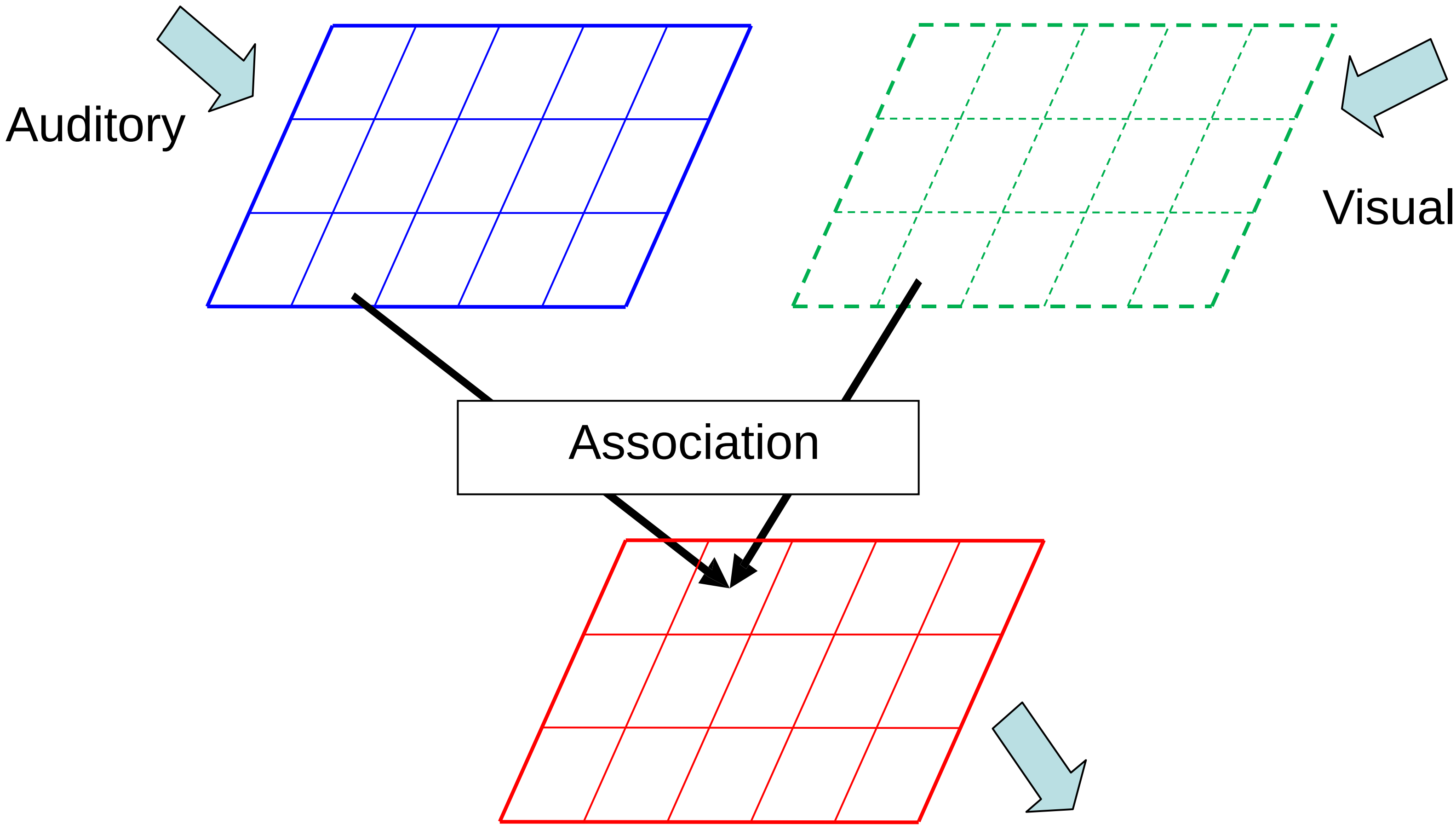
Behavioral Paradigm to Study Spatial Coincidence



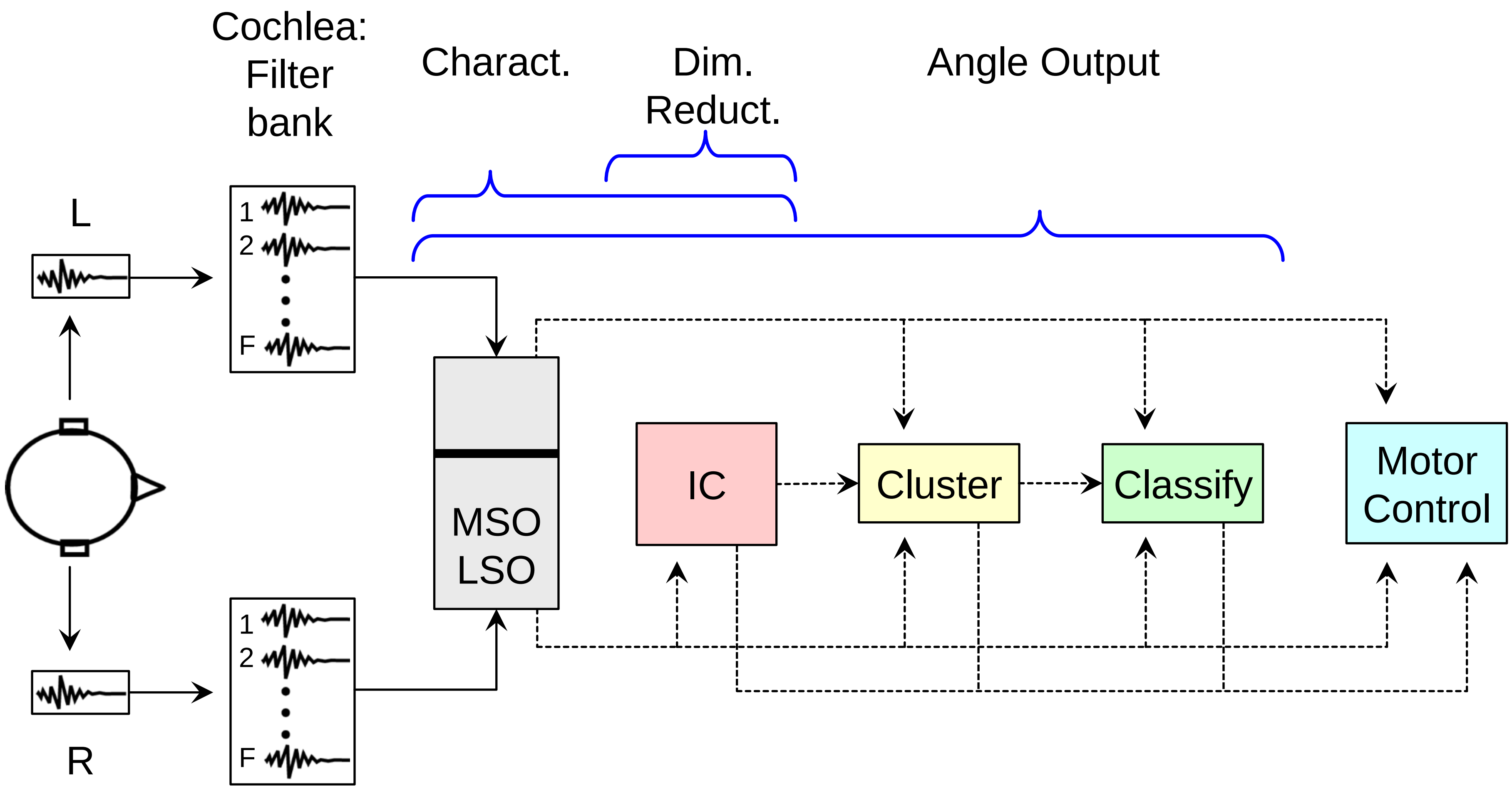
Visual Auditory Response Enhancement



Towards a Computational Neural Architecture: Map Alignment



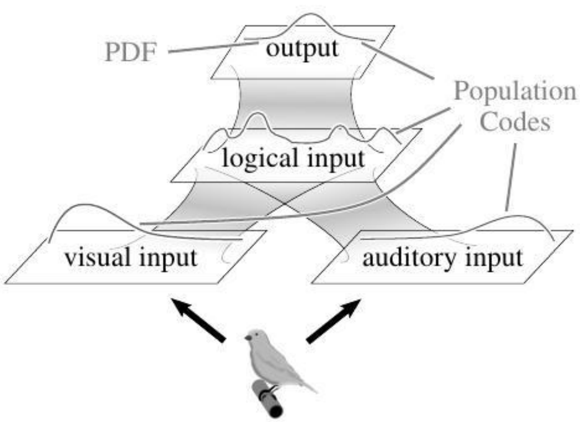
Auditory Localization with Spiking Maps

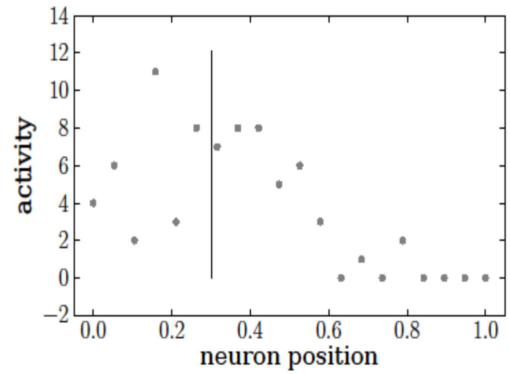


MSO: Medial Superior Olive
LSO: Lateral Superior Olive
IC: Inferior Colliculus

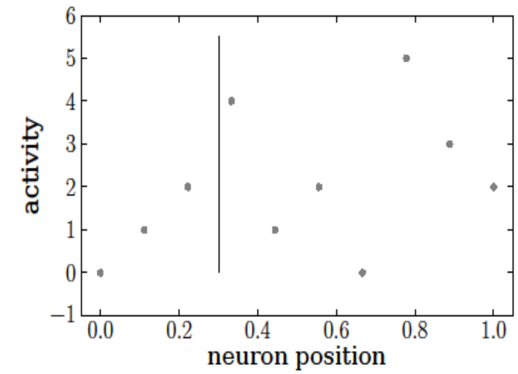
Chacon, Liu, Magg, Wermter 2012,13

Integration of visual and auditory Input with Population Responses

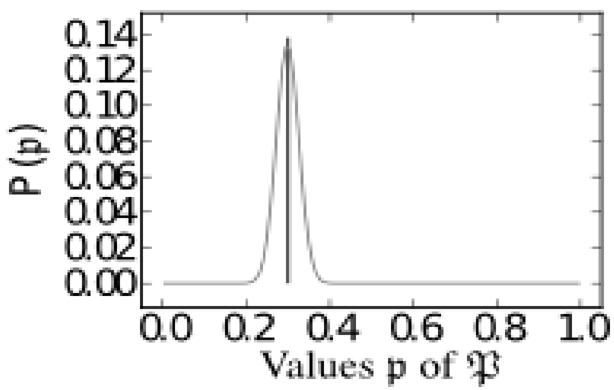


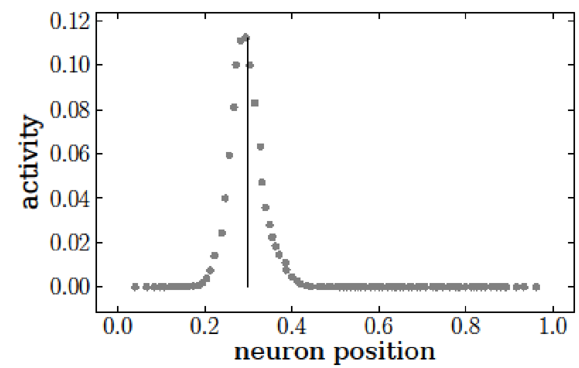


'visual' input



'auditory' input



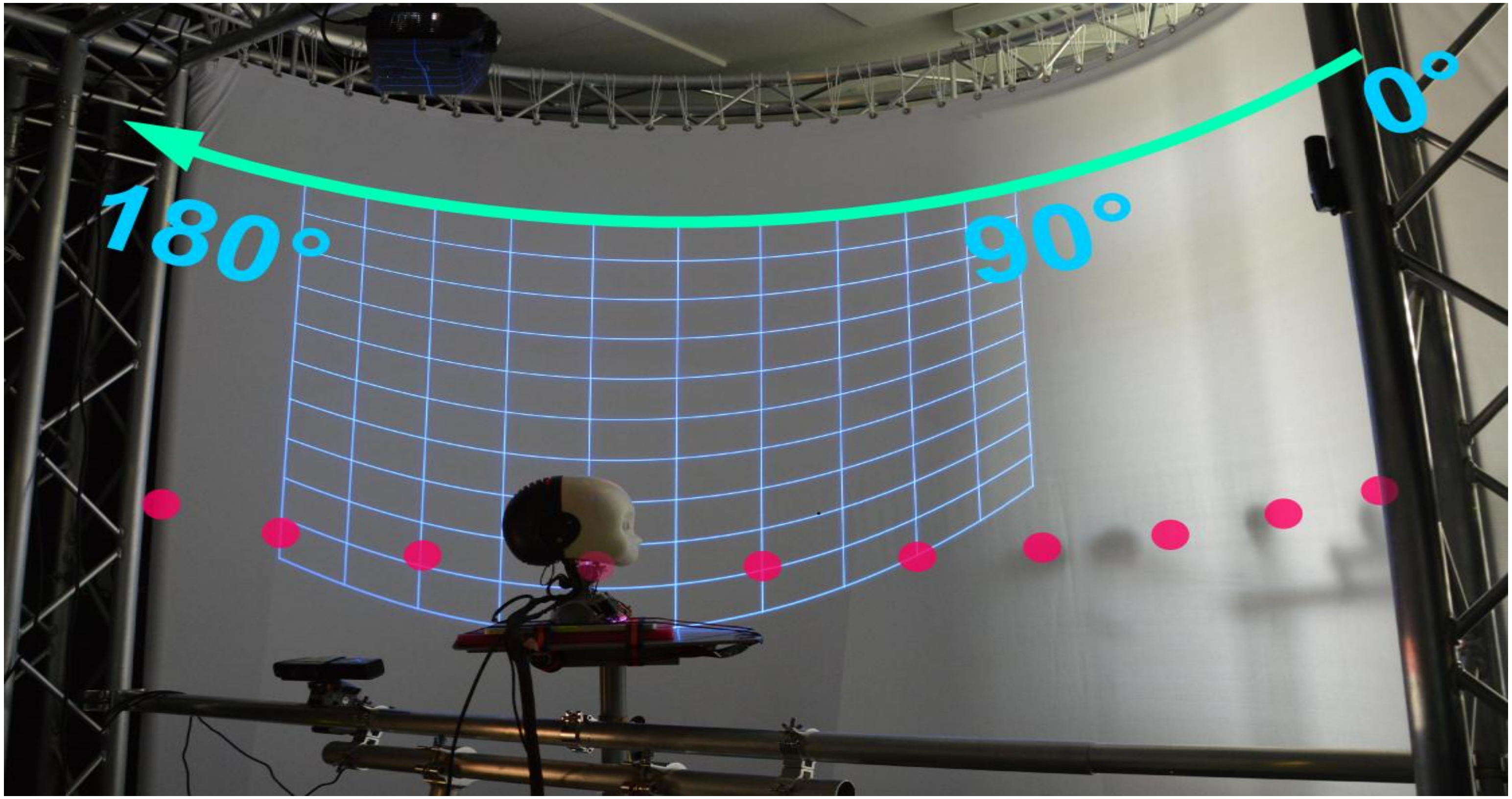


Bauer, Wermter 2013

desired Probability Density function

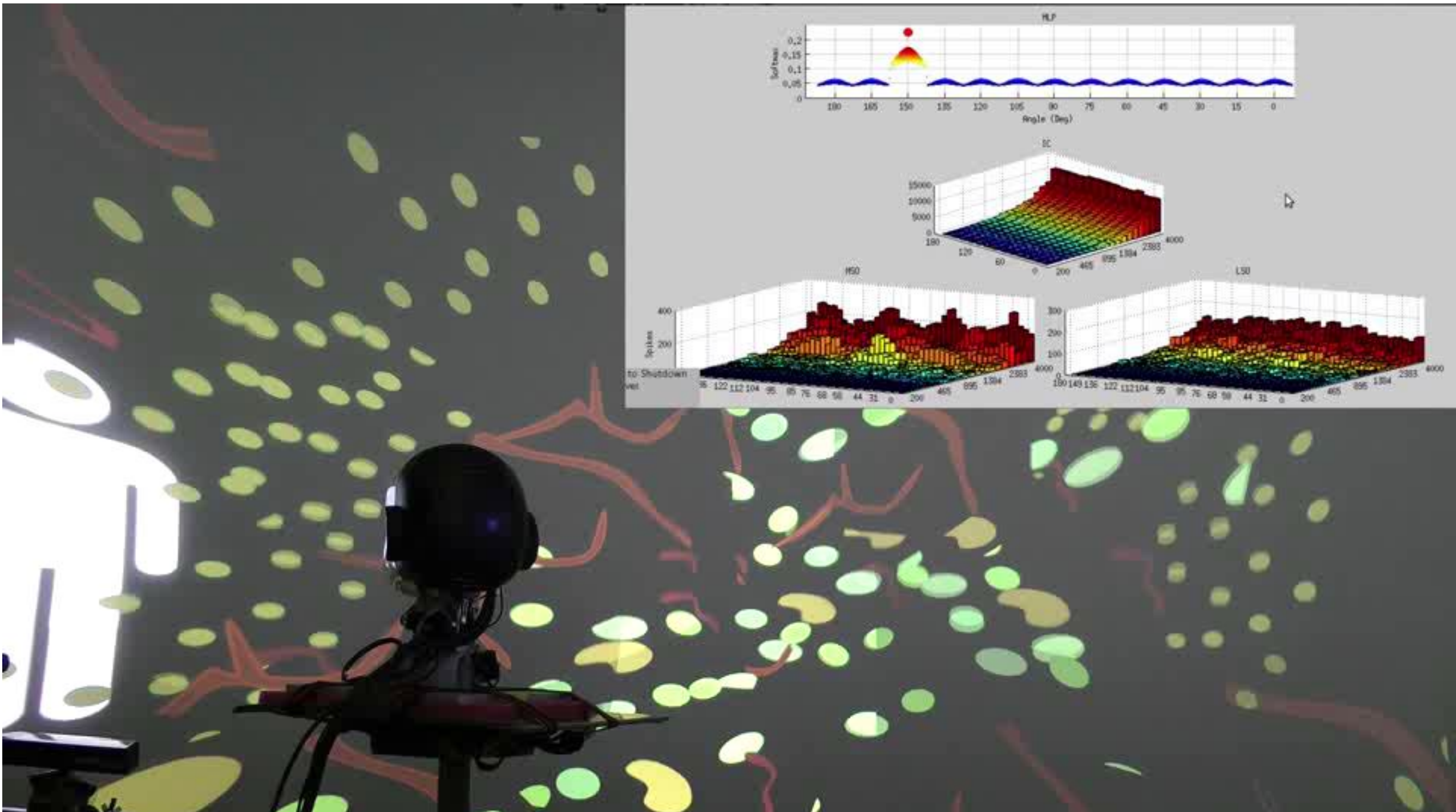
network output

Multimodal Human Robot Interaction Lab



- Audio-visual virtual reality setup with four projectors
- 13 speakers with distance 15°
- iCub humanoid robotic head

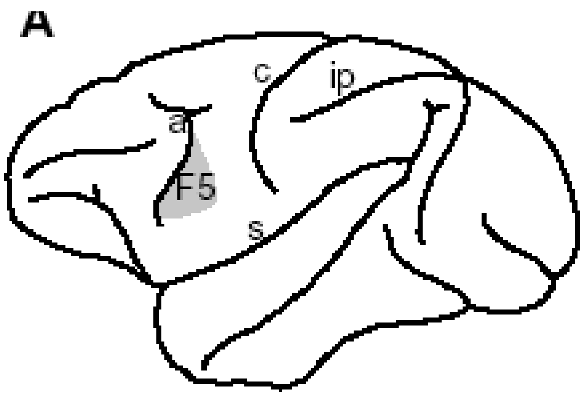
Sound Localization for Action Selection

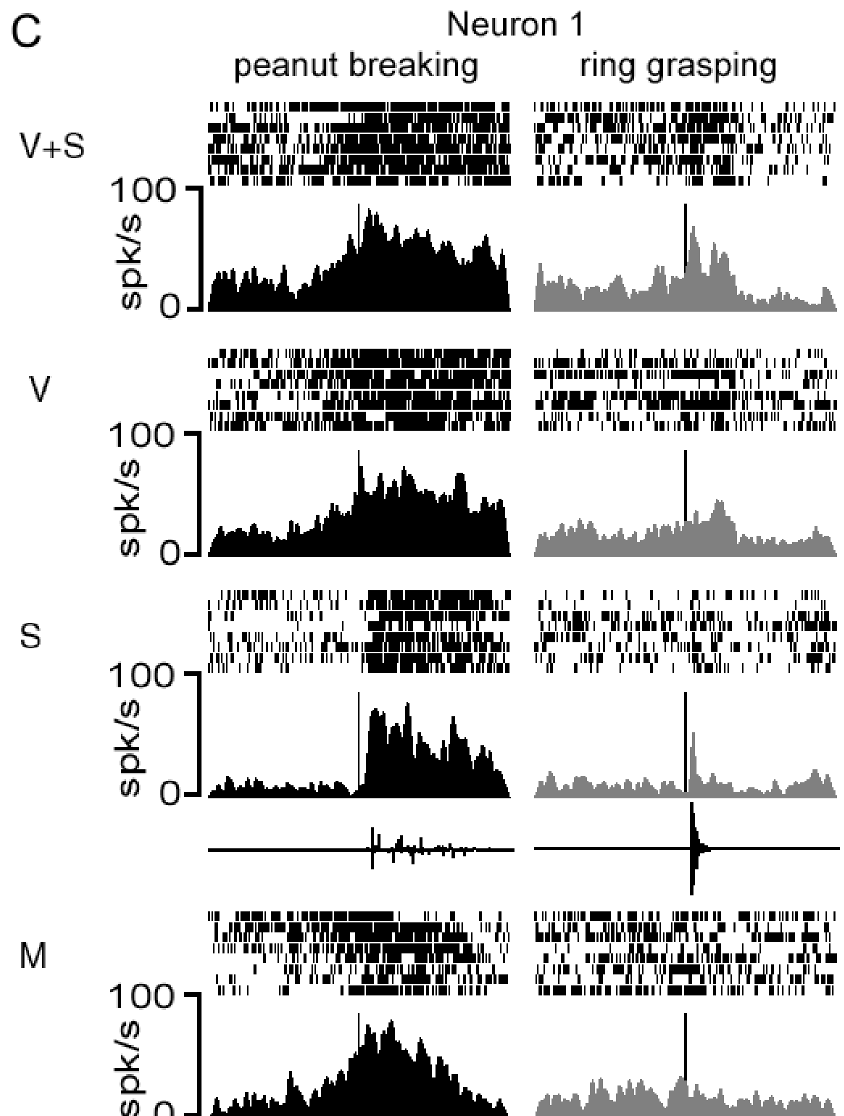


2. How do Multiple Modalities in the Brain Inform Action Understanding?

- Superior colliculus plays a crucial role
  - Contains unimodal visual, auditory, somatosensory and multisensory neurons
- ***Cortical areas*** play a crucial role
  - Contain for instance mirror neurons as smallest entities for multimodal cortical integration

Specific Responses of a F5 Audio-Visual Mirror Neuron





vision sound

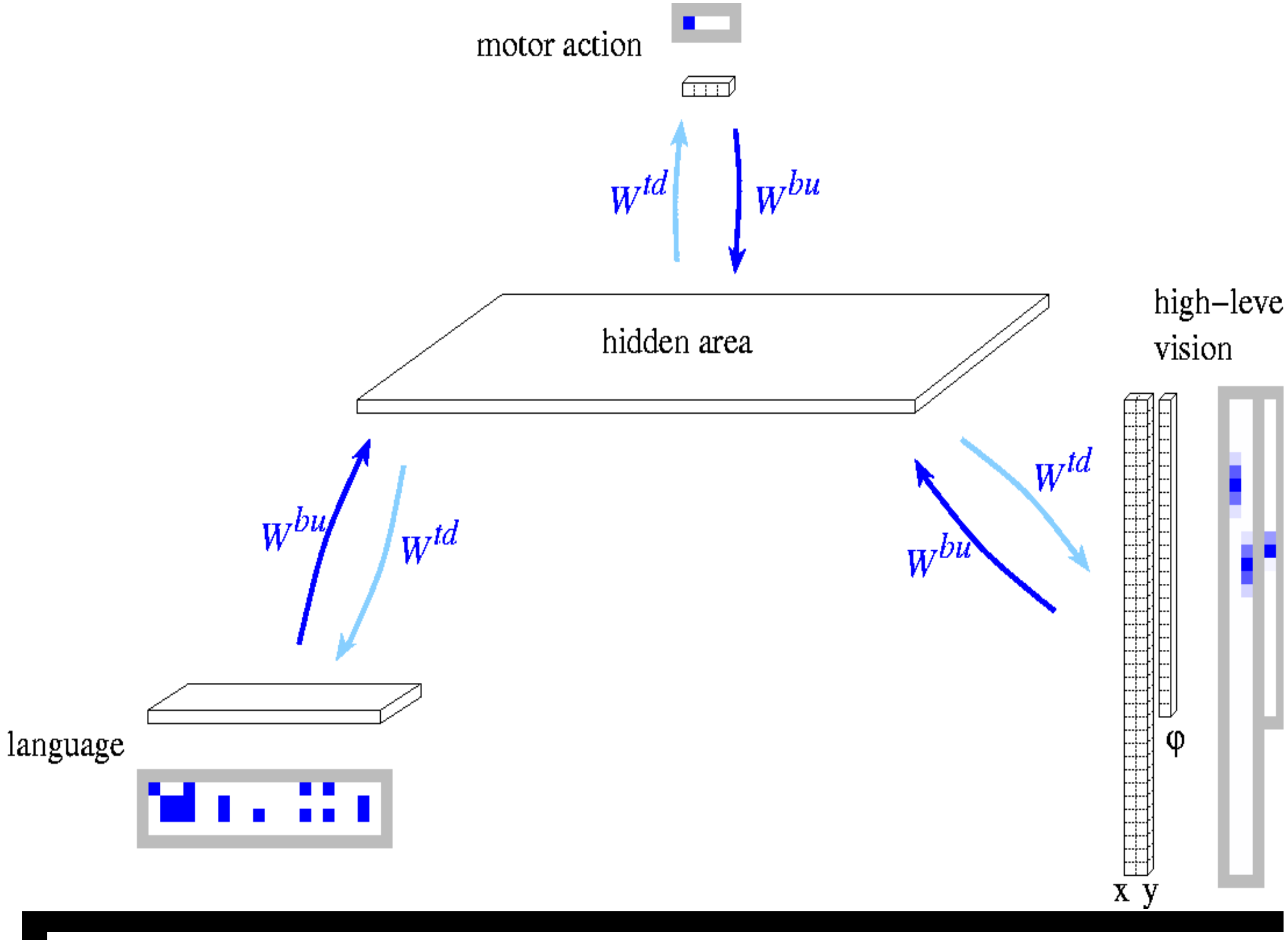
visual

sound

motor

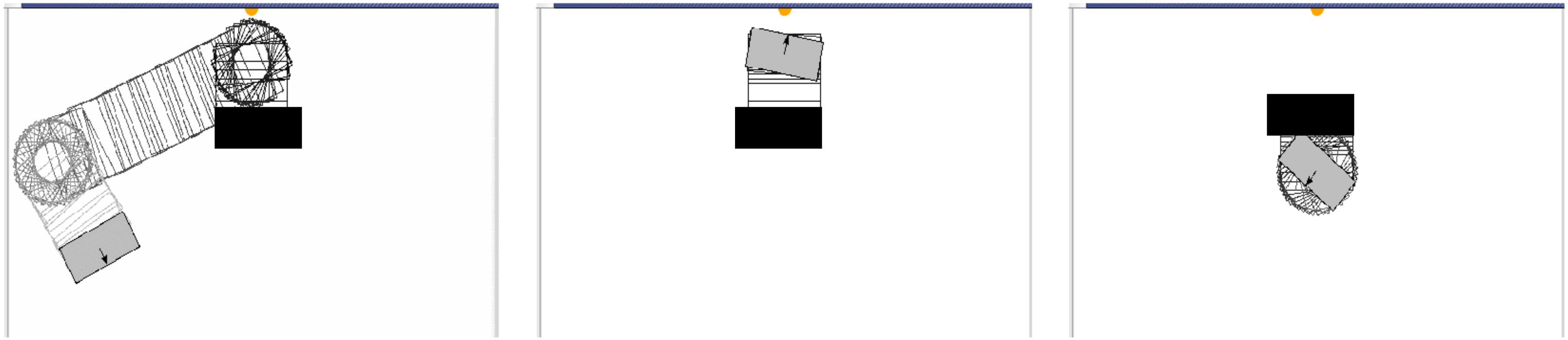
Keysers et al. Exp. Brain R...

Association Network for Vision, Motor, and Language Representation



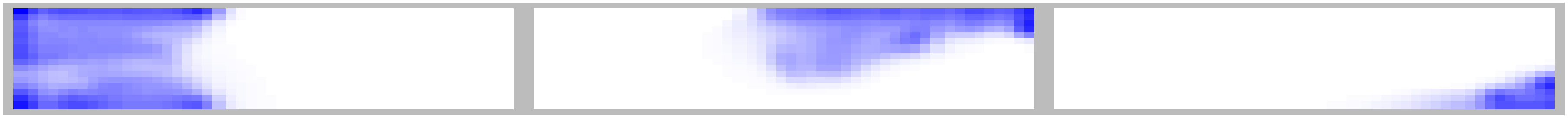
Language Instructed Behavior

'go' 'pick' 'lift'



Receptive fields as weight representations

'go' 'pick' 'lift'

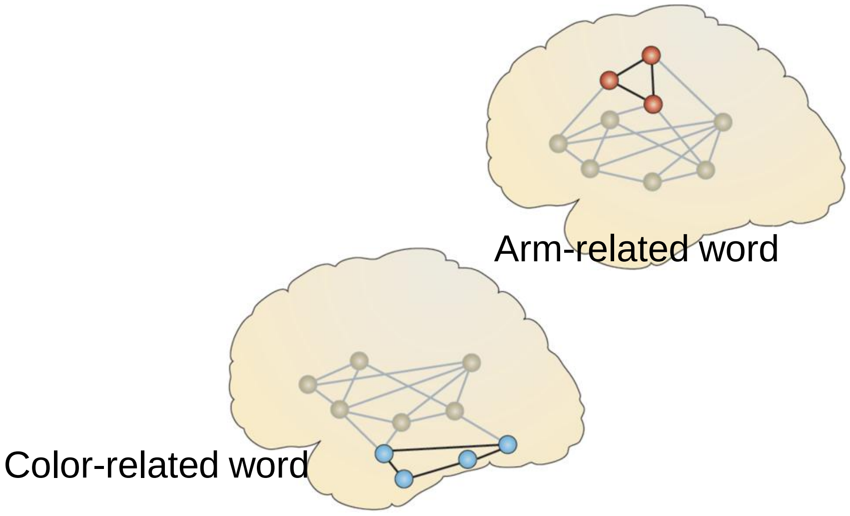


Reinforcement learning actor-critic approach, Weber Wermter Elshaw et al. 2003, 2004

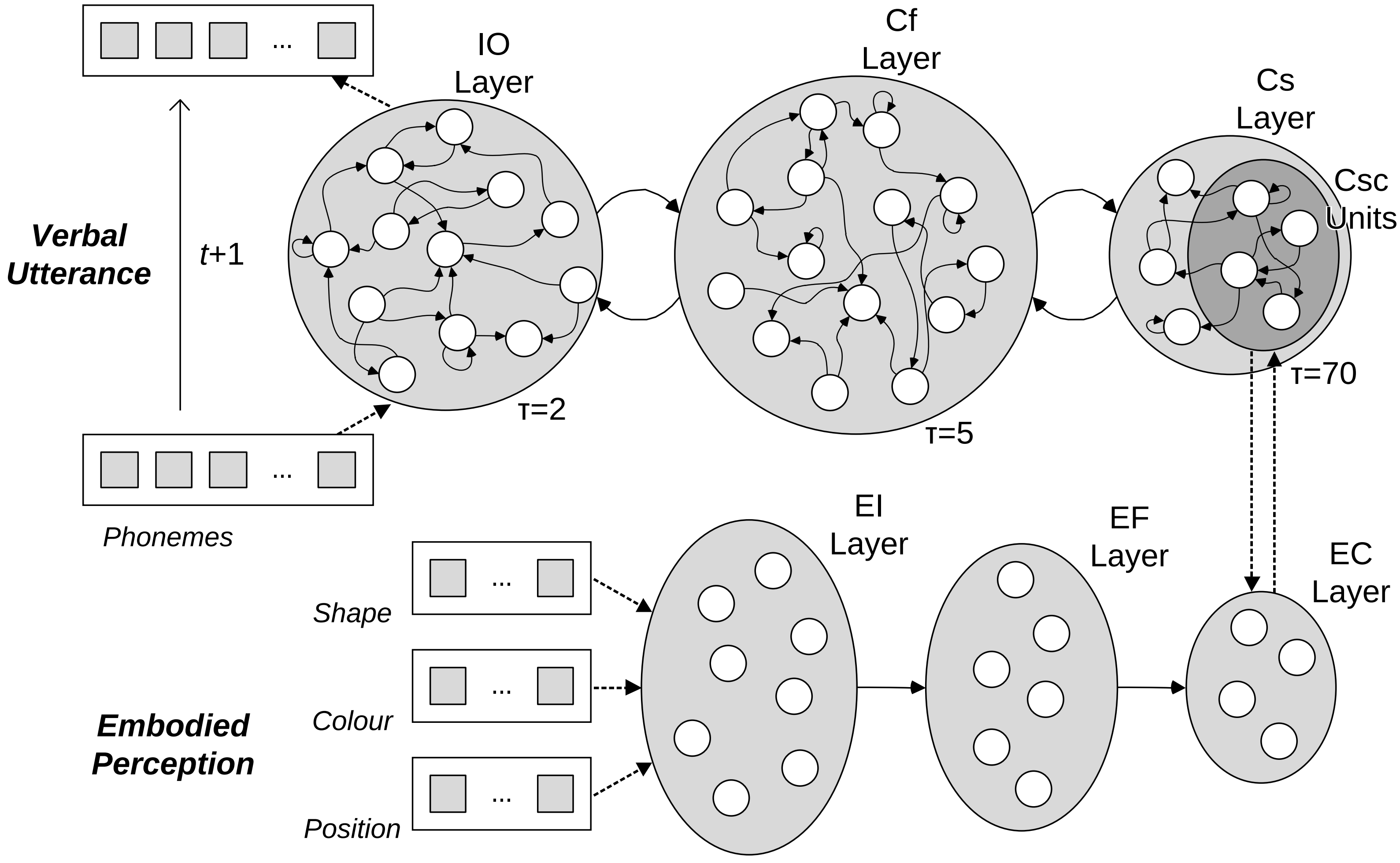
Integrating Language and Vision with a Multiple Timescale Recurrent Neural Network

- Core representations in cognition ***are not*** amodal symbols and structures [Barsalou 2008]

- Action-perception circuits are necessary for, and make an important contribution to, ***semantic*** processing [Pulvermüller 2006, 2010]



Extended MTRNN Model



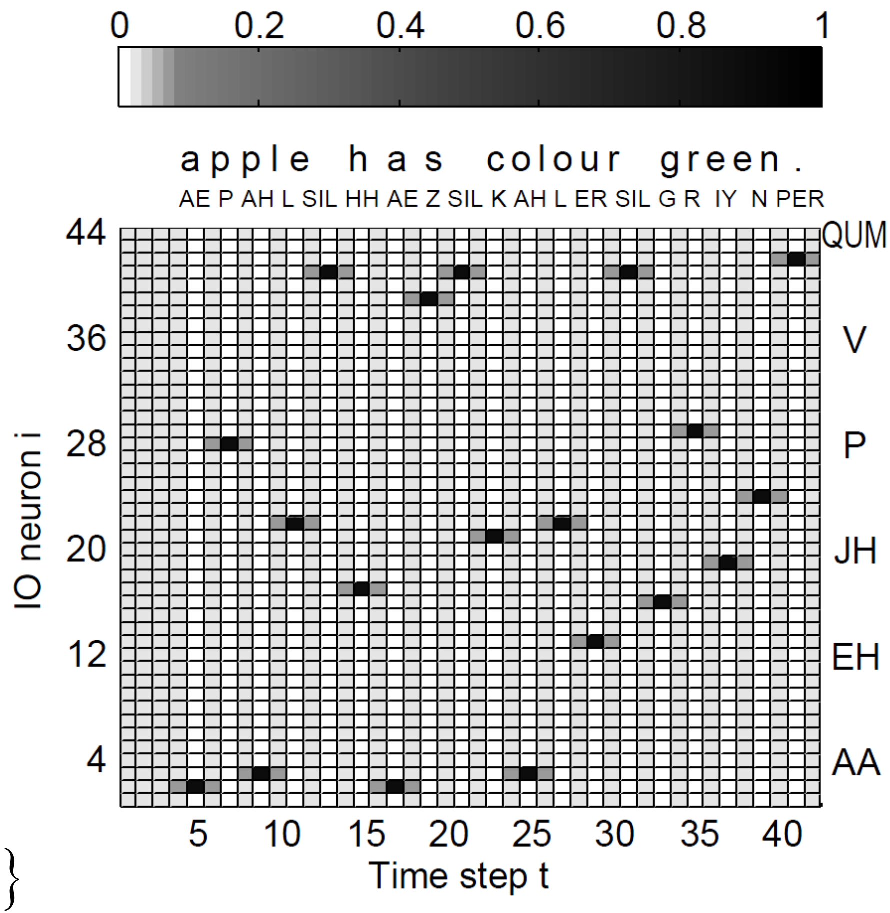
Heinrich , Weber, Wermter 2013

Verbal Utterance Representation

```
S → INFORM
INFORM → POS is a OBJ.
INFORM → OBJ has colour COL.
OBJ → apple | banana | dice | phone
POS → above | below | left | right
COL → blue | green | red | yellow
```

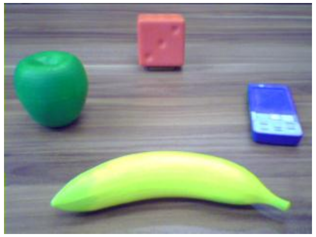
- Small symbolic grammar
- Transferred to ***phonetic*** utterances
  - Based on ARPAbet

$$\Sigma = \{'AA',\ldots,'ZH'\}$$
$$\cup \{'SIL','PER','EXM','QUM'\}$$

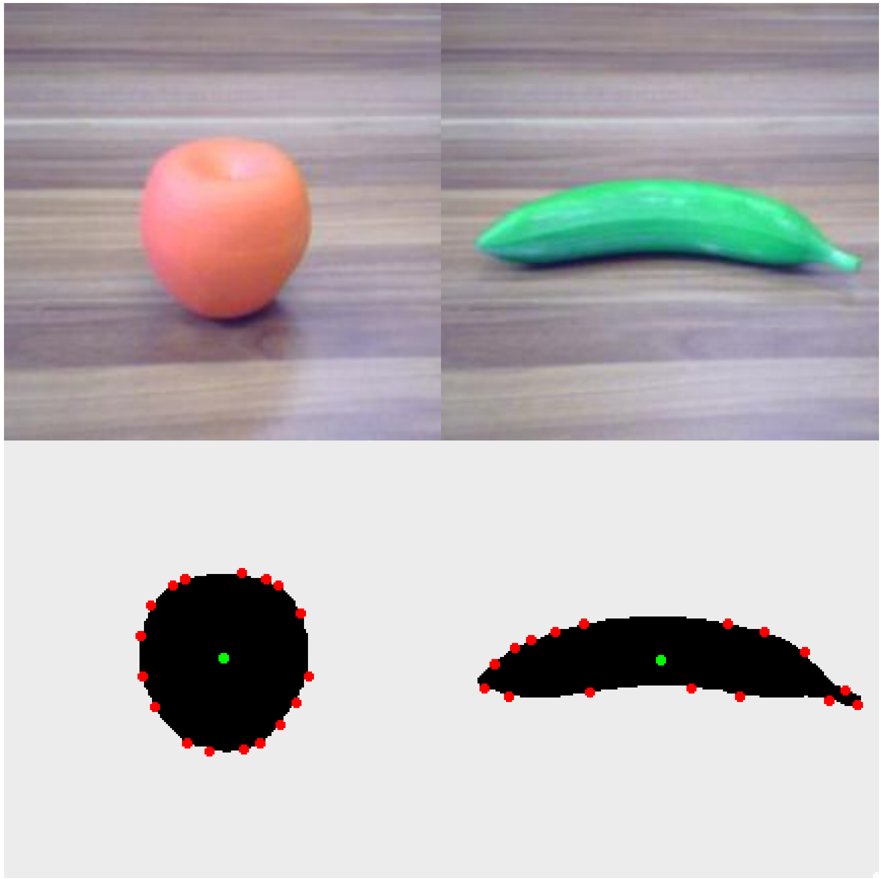


Sample encoded utterance

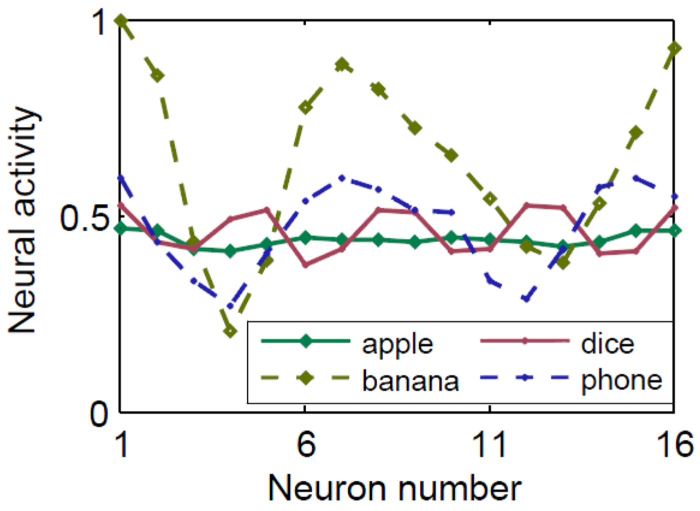
Visual Perception Representation



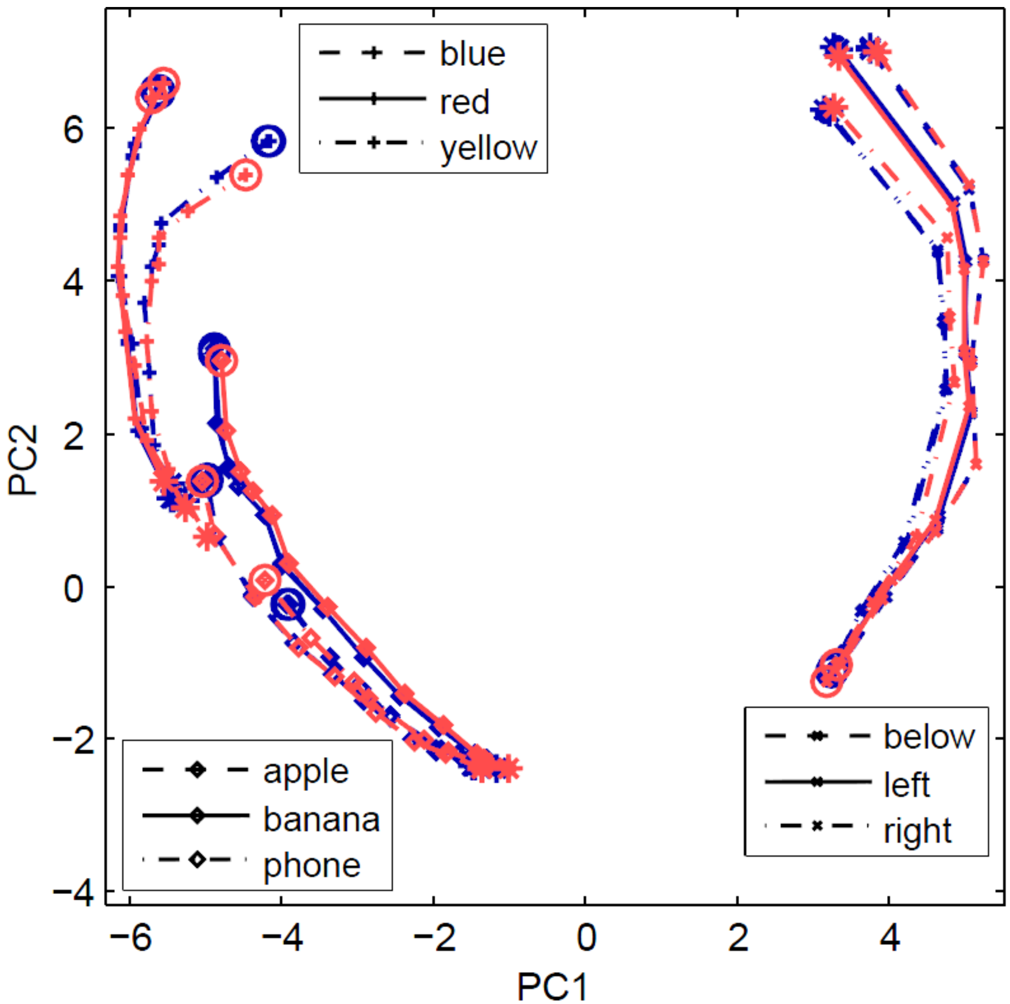
- Shape: Capture ***salient features*** around the objects



- Segmentation with mean shift
- Object discrimination with Canny edge & Suzuki contour
- Determined center of mass & 16 ***distances*** to salient points



Network Behaviour and PCA Activity over Time



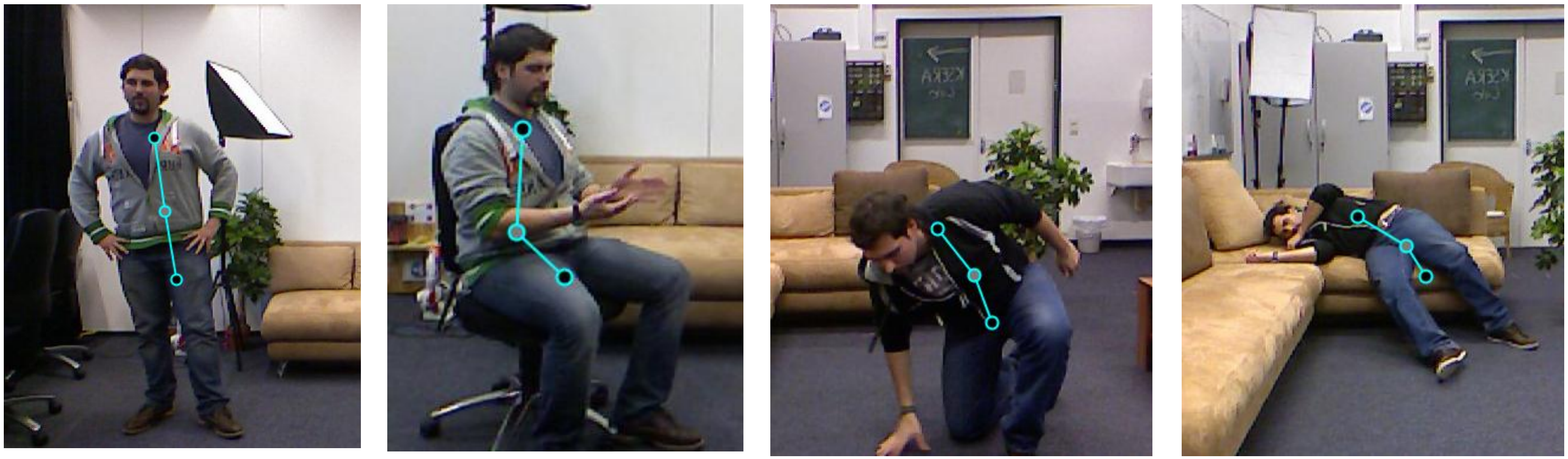
- Neural activity in Cf layer is similar for same word class

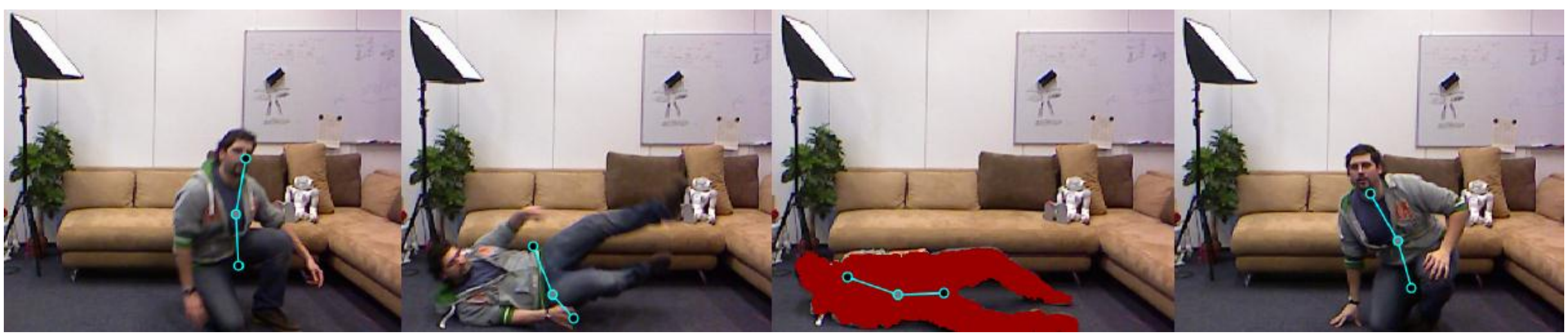
3. Learning, Recognizing and Naming Actions

- Human 3D motion tracking
  - Extraction of ***spatio-temporal properties*** from moving targets
  - Use of depth and color information

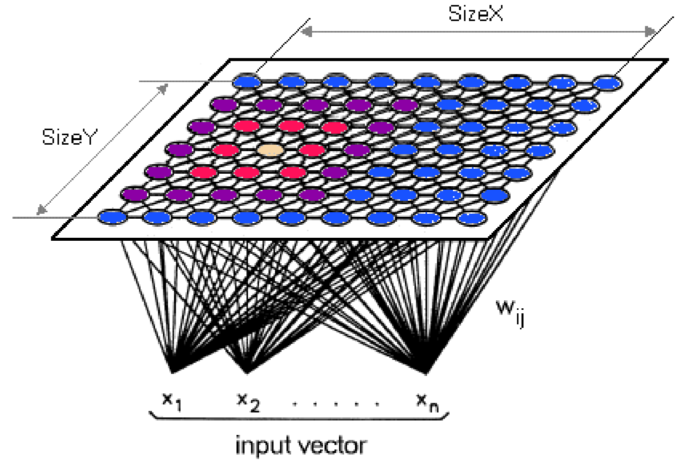
- Unsupervised novelty detection
  - Neural-statistical architecture
    based on self-organizing maps (SOM)
- Evaluation:
  - Robust to changes in light conditions
  - Highly occluded targets

Motion Representation



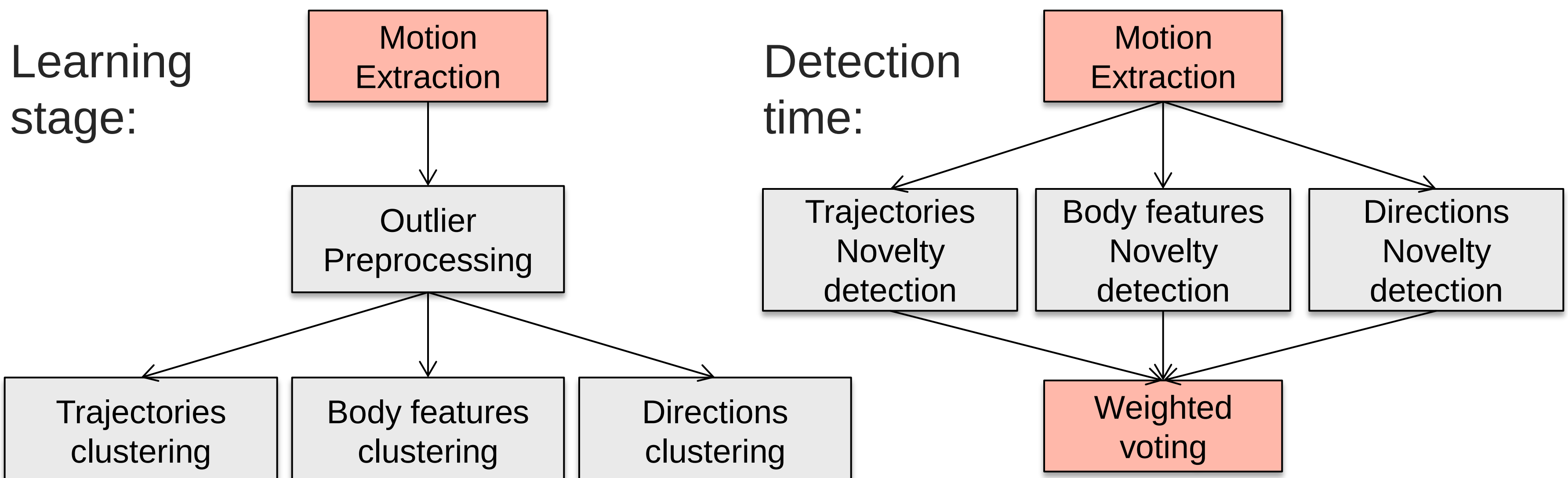


Modular Neural Architecture



- SOM-based neural architecture

- What is a normal action?
- P-value is smaller than the given threshold, the observation is reported as abnormal



Experimental Conditions

- Training data
  - Depth video sequences from monitored home-like environment
  - Frame rate: 30 Hz
  - VGA resolution of 640x480
  - 20 minutes of indoor domestic actions
    - Walking
    - Sitting
    - Picking up objects

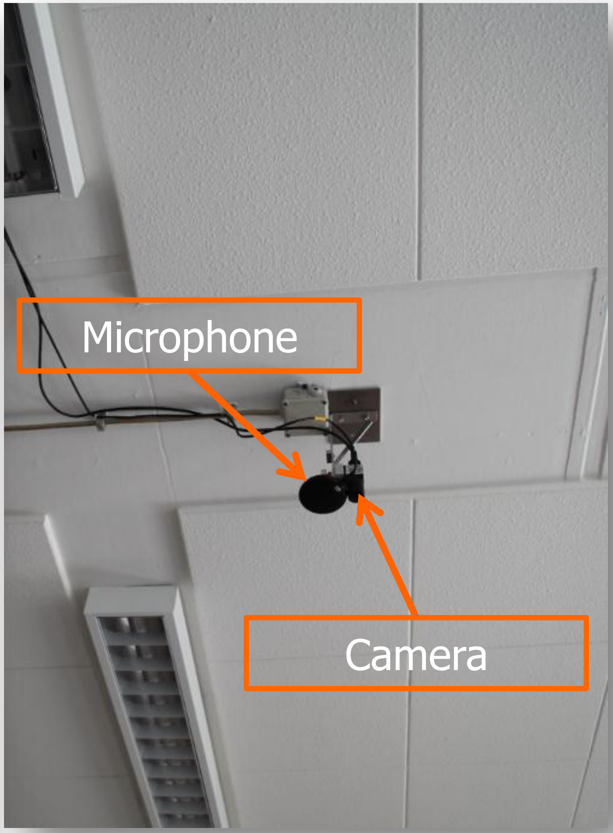
- SOM networks
  - Input vectors: 34.560
  - Distance: Euclidean
  - Neighborhood function: Gaussian
  - Initialization: Random
  - Batch training algorithm

4. Learning a Cognitive Map for Robot Navigation

- Achieve complex cognitive task such as fetching object after instruction
- Anchoring the visual appearance features in the cognitive map for navigation
- Pro-active obstacle avoidance
- Semantics of action is grounded in the navigational map

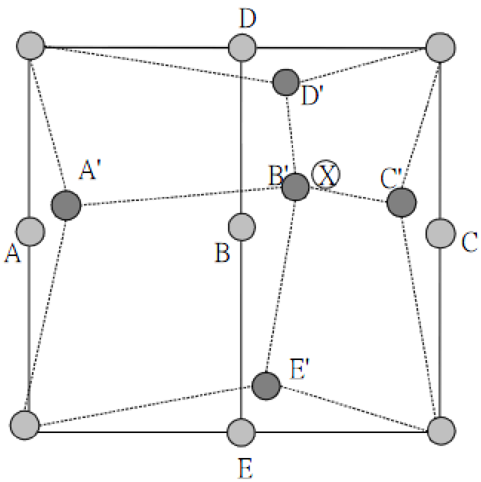
Environment

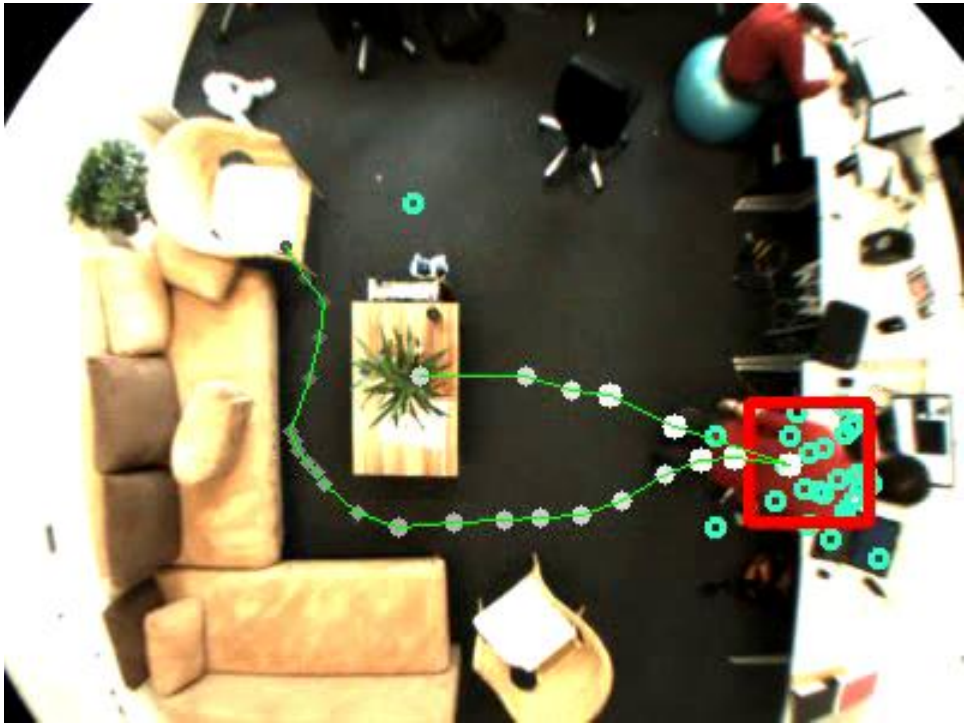
- Ceiling-mounted Camera & Microphone



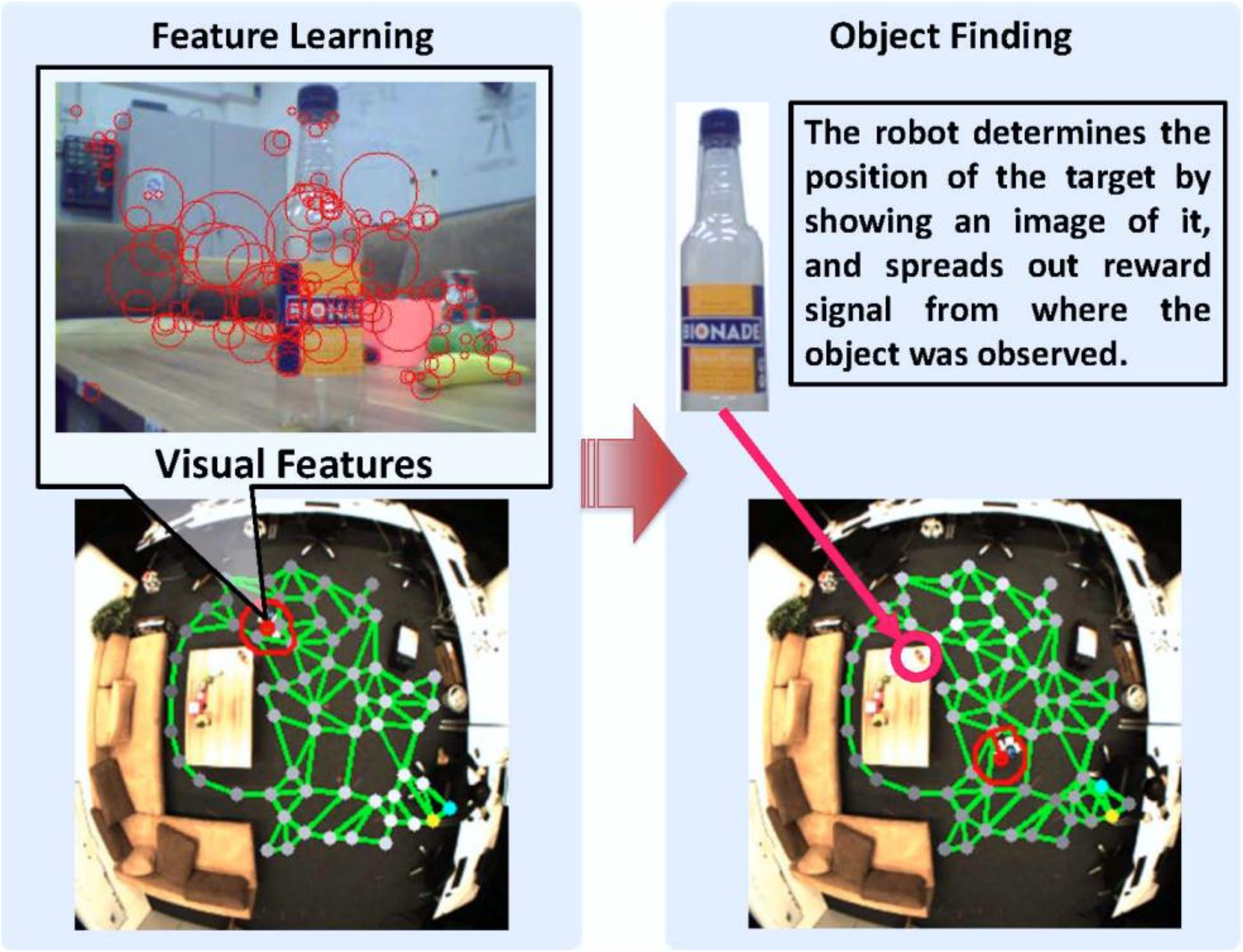
Integration of Color, Shape, & Movement Cues

- Tracking movement of user and robot for plan navigation
- Growing neural gas algorithm for cognitive map learning

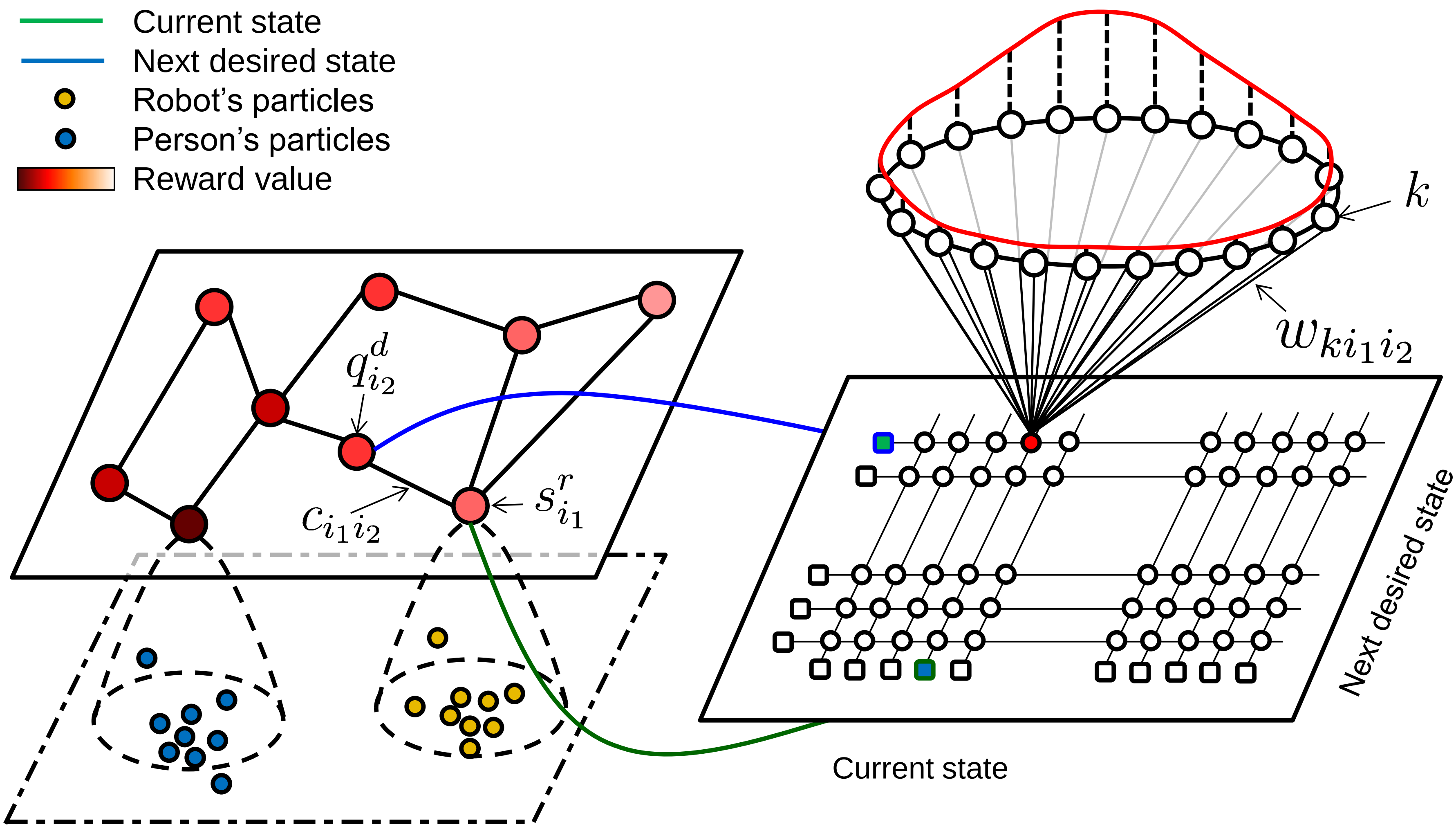




Anchoring Appearance Features at Map Nodes

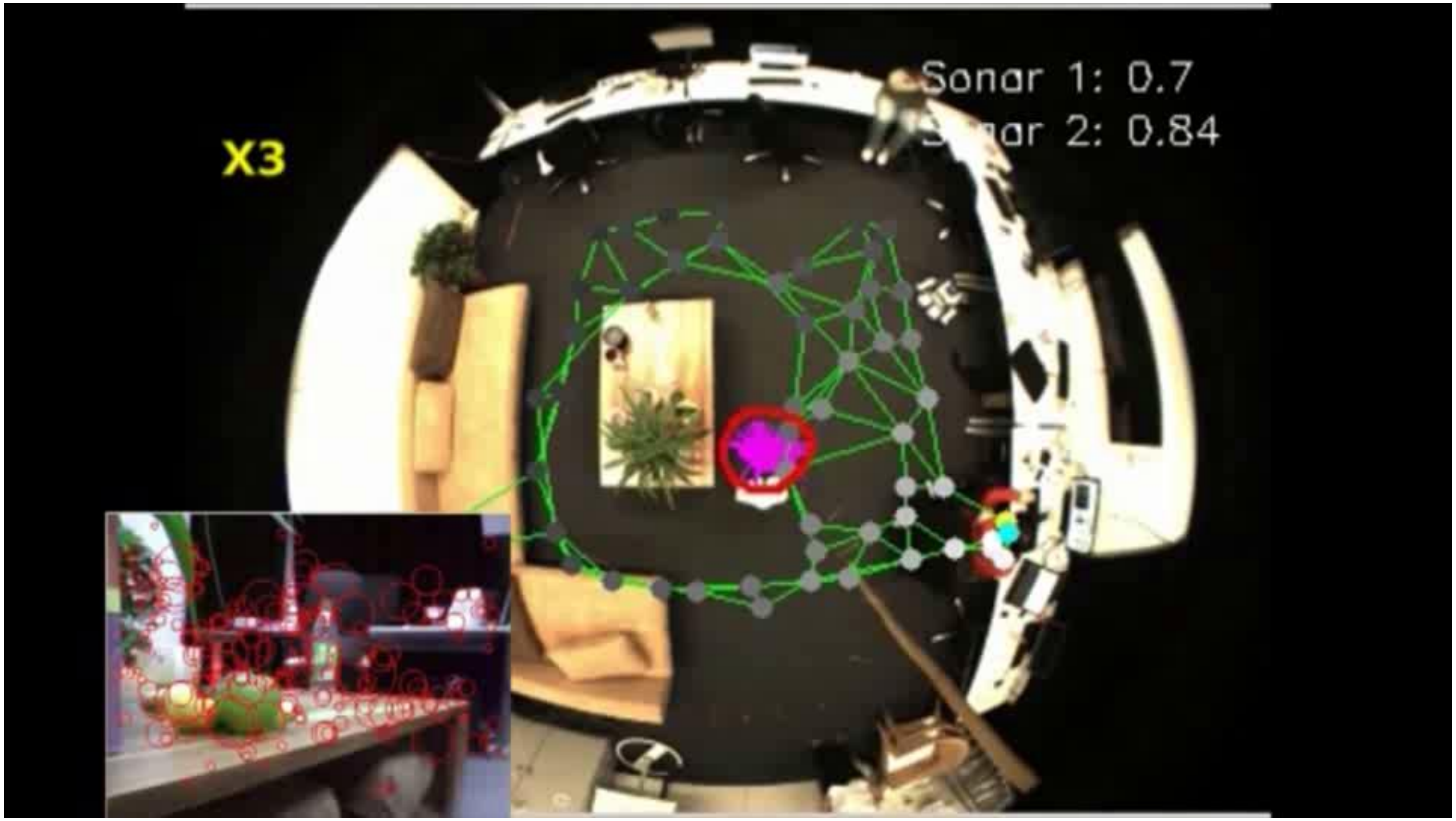


Architecture: Neural Gas and Neural Fields

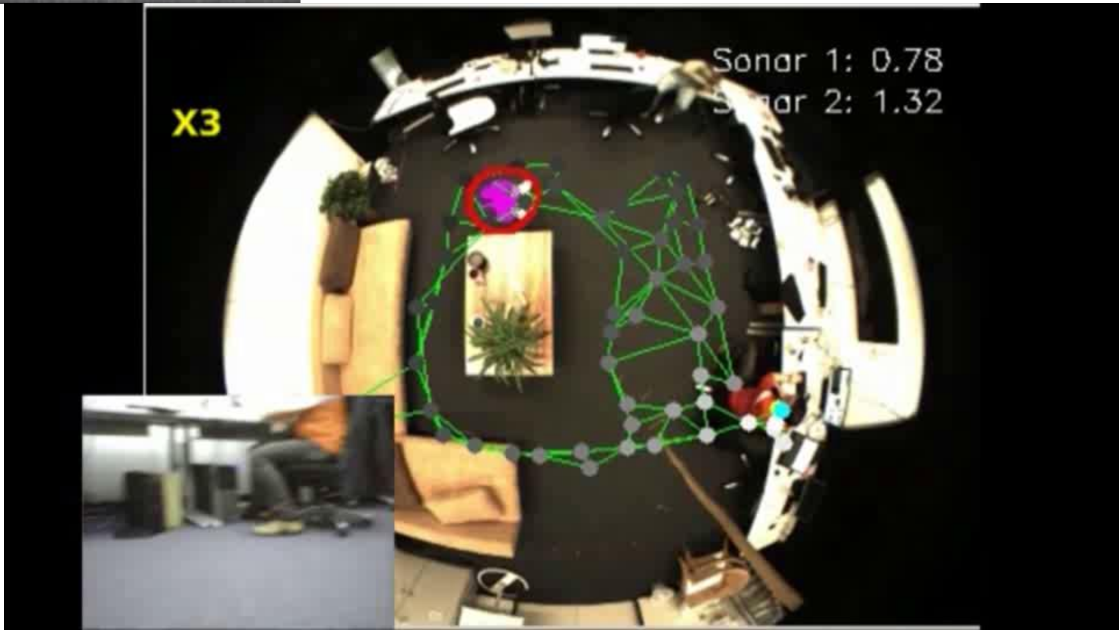


Yan, Weber, Wermter 2012 2013

Building the Map and Storing the Features at Neurons of the Map



Grasping Bottle and Bringing to Person



Summary and Conclusions

- Need to understand human language and action architectures of the brain
- Neurocognitive approaches of grounding action understanding
- NEST: Neural Symbolic Technology architecture
- Computational models need neural, statistical and symbolic representations at different levels for integration
- http://www.informatik.uni-hamburg.de/WTM/

References

- Bauer Weber Wermter IJCAI13 IJCNN12
- Chacon Liu Magg Wermter IJCNN13 ICANN12
- Heinrich Weber Wermter ICANN13 ICANN12
- Parisi Wermter IJCNN13
- Yan Weber Wermter IJCNN12
- Elshaw Weber Wermter (Mirrorbot project)
- Wermter et al. Biomimetic Neural Learning 2005
- http://www.informatik.uni-hamburg.de/WTM/