PERGAMON

AVAILABLE AT www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

Neural Networks 16 (2003) 691–699

Neural Networks

www.elsevier.com/locate/neunet

2003 Special issue

Learning robot actions based on self-organising language memory

Stefan Wermter*, Mark Elshaw

Centre for Hybrid Intelligent Systems[1]*, School of Computing and Technology, University of Sunderland, St Peter's Way, Sunderland, SR6 0DD, United Kingdom*

Abstract

In the MirrorBot project we examine perceptual processes using models of cortical assemblies and mirror neurons to explore the emergence of semantic representations of actions, percepts and concepts in a neural robot. The hypothesis under investigation is whether a neural model will produce a life-like perception system for actions. In this context we focus in this paper on how instructions for actions can be modeled in a self-organising memory. Current approaches for robot control often do not use language and ignore neural learning. However, our approach uses language instruction and draws from the concepts of regional distributed modularity, self-organisation and neural assemblies. We describe a self-organising model that clusters actions into different locations depending on the body part they are associated with. In particular, we use actual sensor readings from the MIRA robot to represent semantic features of the action verbs. Furthermore, we outline a hierarchical computational model for a self-organising robot action control system using language for instruction.

Keywords: Mirror neurons; Robot; Language Memory

1. Introduction

In recent years, there has been much progress in understanding the underlying perceptual and neurocognitive processes in the brain as well as the techniques for these studies (lesion studies, fMRI, MEG and EEG Binder et al., 1997; Braitenberg, 1997; Dogil et al., 2002; Friedman et al., 1998; Joliot et al.,1999; Melcher, 2000; Pulvermüller, 1999; Taylor, 1999). Recent advances in MEG and fMRI techniques have clarified the dynamics of these processes, but it is still a long way for translating these results directly into computational architectures.

Other work has suggested the grounding of visual and auditory perception in neural representations (Dorffner & Prem, 1993; Feldman et al., 1996; Harnad, 1990). However, these results have not been integrated with behavioural studies to any great extent. Recently, a class of neurons has been found in the rostral part of the ventral premotor cortex (area F5) in monkeys that are active both when a monkey handles an object and when it observes an experimenter performing similar actions (Rizzolatti & Arbib, 1998). More recently, PET studies have implicated these 'mirrorneurons' in the gesture recognition system of humans. This system involves Broca's area, a language area in humans, which is generally believed to be the human homologue of area F5 in monkeys. Therefore, we explore the role of mirror neurons and cell assemblies for multimodal integration of action, vision, and language in the MirrorBot project.

Recently, there has been a growing interest in learning for robotics. However, these approaches rarely use neural networks or language instruction. Furthermore, the robots are restricted in their general autonomous behaviour and only learn what has been preprogrammed. Even the 'Talking Heads' approach that incorporates the emergence of language in robots (Steels, 1998) gives little consideration to neuroscience-inspired learning in humans.

Some robots like the tour-guide robot Rhino (Burgard et.al., 2000) have been quite robust in terms of their localization and navigation behaviour; however, they do not interact via language. Although the conversation office robot jjj-2 (Asoh et al., 1997) can be instructed to navigate to certain landmarks and the Minerva tour-guide (Thrun et al., 1999) interacts by using simply preprogrammed speech, they are restricted in their ability to learn. Furthermore, the Kismet interactive robot (Breazeal & Scassellati, 1999) can recognise and represent emotions using a static sophisticated head but does not understand or generate real language.

* Corresponding author. Tel.: +44-191-515-3279; fax: +44-191-515-3553.

E-mail address: stefan.wermter@sunderland.ac.uk (S. Wermter).

[1] http://www.his.sunderland.ac.uk

0893-6080/03/$ - see front matter
doi:10.1016/S0893-6080(03)00100-X

This paper for robot control using language incorporates some neuroscience evidence related to the architectural and processing characteristics of the brain (Wermter, Austin, & Willshaw, 2001). In particular it focuses on the neurocognitive evidence on cortical assemblies and regional modularity for language areas in the brain (Pulvermüller, 2003).

2. Regional distributed modularity

Various distributed neural networks in diverse regions in the brain carry out processing in a parallel distributed manner to perform specific cognitive functions (Reggia, Shkuro, & Shevtsova, 2001). The brain consists of a group of collaborating networks, none of which can deal with a complex task alone (Wermter et al., 2001).

For over a century researchers have considered how language is processed by the brain. As a response therefore there are various models of modular language processing. The classical model of language processing comes from lesion studies in the brain and is based on Broca's and Wernicke's language modules linked via the arcuate fasciculus (Bear, Connors, & Paradiso, 1996).

However, brain regions that are involved in language also include those outside the traditional language areas (Gazzaniga, Ivy, & Mangun, 1998; Purves, 1997). For instance, speech comprehension and information recollection has been observed to involve four regions in the left hemisphere of the cerebral cortex (Binder et al., 1997). Semantic language operations involve the superior temporal sulcus, middle temporal gyrus, angular gyrus and lateral frontal lobe (Friedman et al., 1998).

Recently, cortical assemblies have been identified in the cortex that activate in response to the performance of motor tasks at a semantic level (Pulvermüller, 1999; Rizzolatti, & Arbib, 1998). This evidence supports that these neurons are involved in actions, observing actions and communicating actions. The neurocognitive evidence of Pulvermüller (1999, 2002 and 2003) supports that cell assemblies are activated in different regions of the brain dependent on the word type being processed. This evidence offers the basis for our approach. Pulvermüller 1999 noted that activation was found in both hemispheres of the brain for content words and for vision words in the perisylvian and in the parietal, temporal and/oroccipital lobes. However, function words that have a grammatical role were limited to the perisylvian cortex. For action words that involve moving one's own body the perisylvian cell assemblies were associated with motor, premotor, and prefrontal cortices. Assemblies that depict vision words were found in the perisylvian and visual cortices in parietal, temporal and/or occipital lobes.

It is important to relate the neurons that represent the word with those neurons associated with perception and actions that reflect the semantic information of a word. Hence, if a word is repeatedly presented with a stimulus, the representation of this stimulus is incorporated into the representation for the word. For content words the semantic features that influence the cell assemblies come from various modalities and include the complexity of activity performed, facial expression or sound, the type and number of muscles involved, the colour of the stimulus, the object complexity and movement involved, the tool used, and whether the person can see herself doing this activity.

When examining the processing of action verbs that relate to the leg, face and arm Pulvermüller, Hare, and Hummel (2000) found that cell assemblies are associated through semantic information with the appropriate body part. They found that the average response times for lexical decisions was faster for face-associated words than for arm-associated words and the arm-associated response times faster than the leg ones. There was a significant difference for the prefrontal region and occipital regions and above the motor and premotor cortex.

The prefrontal area was found to be associated mainly with arm verbs and the occipital visual areas for face verbs. Furthermore, it was noted by Rizzolatti, Fogassi, and Gallese, (2001) that when subjects were required to observe actions made by the mouth, hand and foot that the foot was located dorsally and mouth and hand ventrally in the brain. This neurocognitive evidence motivates our approach for self-organising associative memory in multiple regions of the brain.

3. Self-organisation

Self-organising networks offer an unsupervised associative memory approach (Fig. 1). Self-organising networks consist of an input and an output layer, with every input neuron linked to all the neurons in the output layer

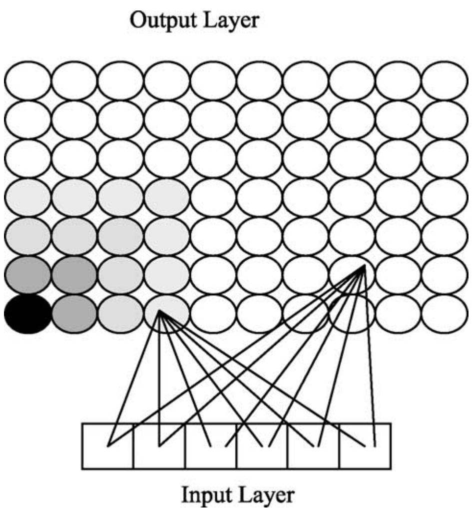

Fig. 1. Architecture of a self-organising network: the darker the neuron on the output layer the higher the activation.

(Kohonen, 1997). The output layer creates a topographical representation that clusters similar inputs together in a two-dimensional neural layer.

A typical self-organising network algorithm has an input vector represented as $\mathbf{i} = [i_1, i_2, \ldots, i_n]$. The input vector is presented to every output unit of the network. The weights between the links in the network are represented as

$$w_j = [w_{j1}, w_{j2}, \ldots, w_{jn}] \quad (1)$$

where j identifies unit j in the output layer and n is the nth element of the input. The output of unit j is the weighted sum of its inputs, given by:

$$o_j = \sum w_{jk} i_k = \mathbf{w}_j \cdot \mathbf{i} \quad (2)$$

The weights are usually initialised randomly and hence a unit of the network will react more strongly than others to a specific input representation. The weight vector of this unit as well as the eight neighbouring units are altered based on the following update rule:

$$\Delta \mathbf{w}_{jk} = \alpha(i_k - w_{jk}) \text{ and } w_{jk}(t+1) = w_{jk}(t) + \Delta w_{jk} \quad (3)$$

where α is the learning rate that is usually set between 0.2 and 0.5.

4. Architecture

A structured architecture of self-organising networks is under development to perform robot control based on associating language and actions. This system learns to associate the semantic features that are found in the sensor readings that represent the action with a representation of the word.

As can be seen at the bottom of Fig. 2 the architecture firstly contains a self-organising network to associate the action sensor readings with the appropriate body part by clustering the verbs in different regions. At the next processing level there is a self-organising network for each body part that uses the sensor reading vectors to associate the actual action verbs with different regions. To the right in the architecture, the words that are represented using their phonemes are clustered in a self-organising network. The upper-most self-organising network associates the actions with their appropriate words. Hence, by associating the action representation with the word the robot can describe the action with a word when it receives only the action representation and vice versa perform the action when it is given the word only.

In this system the input is used to produce the output by recreating the action from the sensor readings. The sensor readings provide information on the action such as the velocity of the separate wheels, the gripper activities and how the constituent subactions relate to the states of sensors such as break-beam and table sensors.

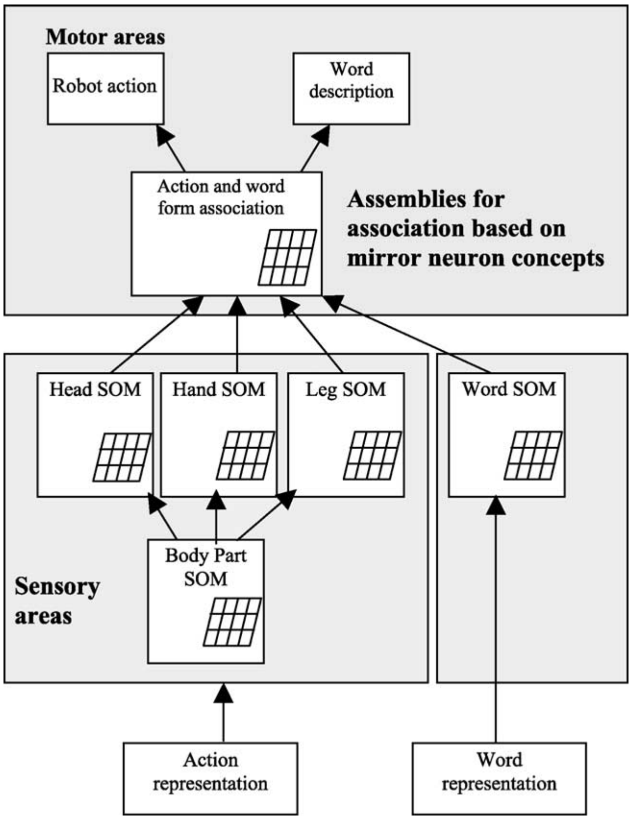

Fig. 2. Overall architecture based on modular distributed and hierarchical self-organising memory.

If the robot receives the ‘put’ action sensor reading representation, it would be introduced into the trained body part network and activate the hand region of the output layer. The hand self-organising network would then position the sensor readings input in the put region of the output layer. As the robot is describing the word there is no necessary input from the word self-organising network into the association self-organising network. However, as the network has previously learned to associate this action with the appropriate word the put region of the network is activated. The robot will then state using its language synthesis that the action semantic features provided are those for put. This describes the pathway from the internal action representation via the association area to the language description. In a similar, but opposite pathway the word input representation can lead via the association area to the sensory robot action.

This approach offers some brain-inspired regional modularity by having multiple self-organising networks, each performing a subtask of the overall task. These networks are linked in a distributed overall memory organization. Furthermore, at the higher functional level as can be seen from Fig. 2, this architecture includes components that are analogous to brain regions at a higher level. For instance, the SOMs that take the action representations and cluster these are related to the sensory motor cortical areas of the brain. The approach also takes into account the neurocognitive evidence of Pulvermüller et al. in that cell assemblies in different regions are associated with specific action verbs as a functional unit, with the association being based on the action verbs relationship with the appropriate body part.

This architecture links some concepts of the mirror neuron theory. The relationship of mirror neurons to language was pointed out by Rizzolatti and Arbib (1998) who found that neurons located in the F5 area of a primate's brain were activated by both the performance of the action and its observation. The recognition of motor actions comes from the presence of a goal and so the motor system does not solely control the movement (Gallese & Goldman, 1998). The role of these mirror neurons is to associate action representations with vision or language representations. The mirror neuron system was a critical discovery as it shows the role played by the motor cortex in action depiction (Rizzolatti, Fogassi, & Galles 2001). By using the sensor readings as input the mirror neuron concept is considered since the understanding of the action can come from either performing the action or a stored representation is linked to observing the action.

5. Self-organisation on the robot

In our experiments we examine the body parts self-organising map, and the head, hand and leg self-organising maps of the modular system. In our previous study (Elshaw & Wermter, 2001) it was possible to identify that a self-organising network was able to cluster action verbs related to body parts. However, this approach relied upon subjective encodings for the applicable features for the action verbs. In order to have greater objectivity and to incorporate self-organising maps into a robot control system, sensor readings were taken from the MIrror-neuron Robot Agent (MIRA) (Fig. 3). Now sensor readings represent semantic features to describe the action verbs such as the degree of motion and object manipulation. We will now describe the experimental context for testing the architecture.

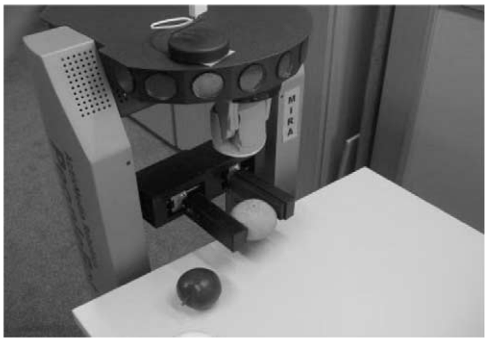

Fig. 3. The MIRA performing the 'pick' action.

5.1. Experimental method

The MIRA robot is based on a PeopleBot platform, and has a PC, microphone and speakers and a PC104 audio board. Wireless communication between the robot and a computer is used. The robot has an adjustable 120-degree pan-tilt camera and fixed-field IR sensors to sense the underside of the table. The robot also has a 2-degree gripper that contains break-beam sensors to detect the object. By using MIRA we take advantage of speech, motion and vision interfaces and explore the development towards a novel neural architecture.

MIRA was set up to perform various actions that are associated in humans with the leg, head or hand. The leg verb actions were 'turn left', 'turn right', 'forward' and 'backward'; head action verbs were 'head up', 'head down', 'head right' and 'head left' and finally the hand verbs were 'pick', put, 'lift', 'drop' and 'touch'. One action can be made of several basic actions. For instance, the hand verb action pick included the following subactions (i) slowly move forward to the table; (ii) tilt camera downwards to see table, (iii) lift gripper to table height; (iv) open gripper; (v) close gripper on object; (vi) stop forward motion; and (vii) lift gripper. This sequence of subactions corresponds in principle (although not in detail) to motor schemata since a complex action is represented as a sequence of basic actions. Sensor readings were taken for such sequences of basic actions.

In order to provide sufficient and varied training and test data the actions were repeated 20 times (15 training and five test) under diverse conditions. For instance, the speed the robot was travelling at and the angle that the camera was tilted or panned to were varied. The sensor readings were taken 10 times a second while MIRA performed these actions including the state of the gripper, the velocity of the wheels and the angle that the robot's camera was at. The full list of the sensor readings is given in Table 1.

To reduce the size of the input for the self-organising network to a manageable level, 10 sets of the readings were taken over time to represent the action. This was achieved by taking the first, last and eight equi-distant sets of readings and combining them to create a single input for a sample. Preprocessing was performed on the data to make it suitable for introduction into the neural networks. As self-organising networks require the input values to be represented numerically 'yes' was represented as 1 and 'no' 0. The gripper break-beam state values were represented as 'no beams broken' 0.25, 'inner broken' 0.5, 'outer broken' 0.75 and 'both broken' 1.

There was a need to normalise the sensor readings for such variables as velocity of left wheel, velocity of right wheel, x coordinate of robot, y coordinate of robot, and the pan and tilt of the camera. Normalisation was done by taking the sensor readings for the specific feature for all samples across the 10 sets of readings and positioning the values

Table 1
Sensor readings taken by robot during actions

Sensor reading	Value
Velocity of left wheel	Real number
Velocity of right wheel	Real number
x coordinate of robot	Real number
y coordinate of robot	Real number
Break-beam state of gripper	No beams broken, inner broken, outer broken, both broken
Gripper state	Gripper fully open, closed, between open and closed
Gripper at highest or lowest position	No Yes
Gripper moving upwards or downwards	No Yes
Table sensors activated	No Yes
Gripper opening or closing	No Yes
Pan of camera	Integer
Tilt of camera	Integer

between 0 and 1 dependent on its relative size. For example, the x coordinate values were normalised using the Eq. (4).

$$\frac{x - \min(x)}{\max(x) - \min(x)} \text{ for } all\ x \tag{4}$$

5.2. Unsupervised learning

For the self-organising network to cluster actions based on the appropriate body part the input layer had 120 units, one for each of the preprocessed sensor readings. The output layers had various sizes (from 8 by 8 units to 13 by13 units) and the networks were trained between 50 and 500 epochs at intervals of 50 epochs. Fig. 4 provides an example self-organising network showing the input and output for a 'put' action training. The number of training and test samples for each of the 13 actions were 15 and 5, respectively. There were 260 samples in total, 195 for training and 65 for testing. The location of each of the training and test samples on the self-organising maps were identified based on the units that had the highest activation.

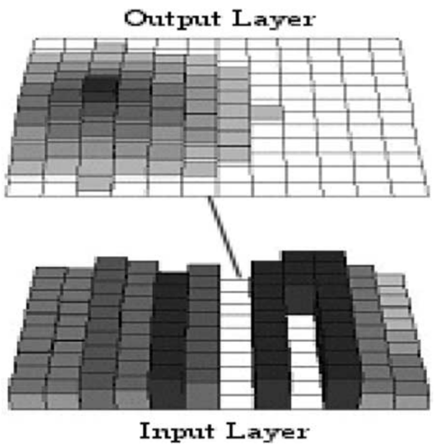

Fig. 4. Example self-organising network showing the input and output for a 'put' training sample.

For the hand, head and leg self-organising networks the inputs were the preprocessed sensor readings, the output layers were varied between 8 by 8 units and 10 by 10 units and they were trained for between 25 and 200 epochs. However, the networks received only sensor readings input for the appropriate body part.

6. Results and discussion

When considering self-organising networks for clustering the actions into their body parts with output layers of between 8 by 8 and 11 by 11 units the networks were only able to produce a split between hand actions and the other two classes. These networks were not able to cluster the leg and head actions into different regions. However, it did indicate an ability to produce a split between simple actions such as 'forward' or 'head right', and more complex actions such as 'put' or 'pick'.

Figs. 5 and 6 show a self-organising network that was 12 by 12 units before training. Before training, it was not possible to differentiate between the hand and leg action samples. Over 90% of the test and training samples for the head actions had the highest activation for the unit that also had the highest activation for 80% of the leg actions. Furthermore, the hand actions were spread out and did not cluster into a single region. However, once this network architecture was trained for 50 epochs there was a clear clustering into the three body parts (Figs. 7 and 8). The hand action words such as 'pick', 'touch', 'lift' were at the bottom of the training and test output layers in the hand body part region, with the head actions slightly below and to the right of the leg region. Although one unit within the head region

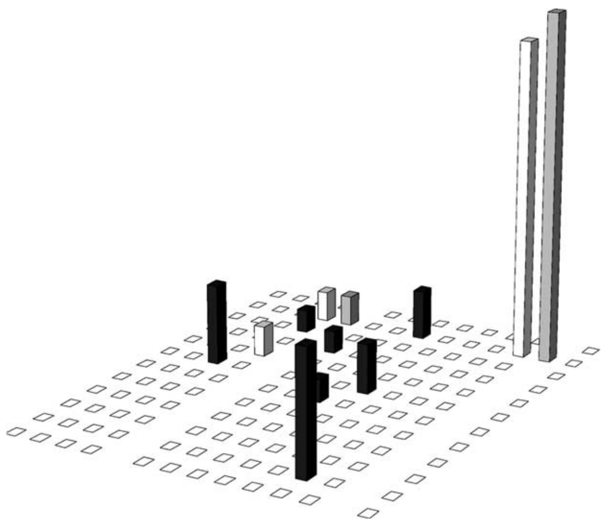

Fig. 5. The percentage of the training samples for the body parts that have highest activation for each unit on a 12 by 12 units network before training. (Black—Hand, White—Head, Grey—Leg).

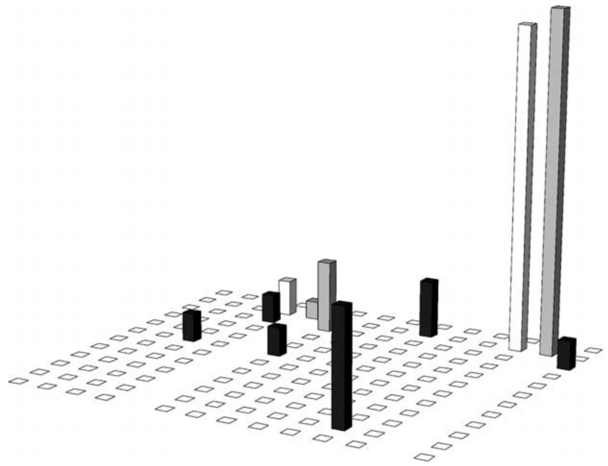

Fig. 6. The percentage of the test samples for the body parts that have highest activation for each unit on a 12 by 12 units network before training. (Black—Hand, White—Head, Grey—Leg).

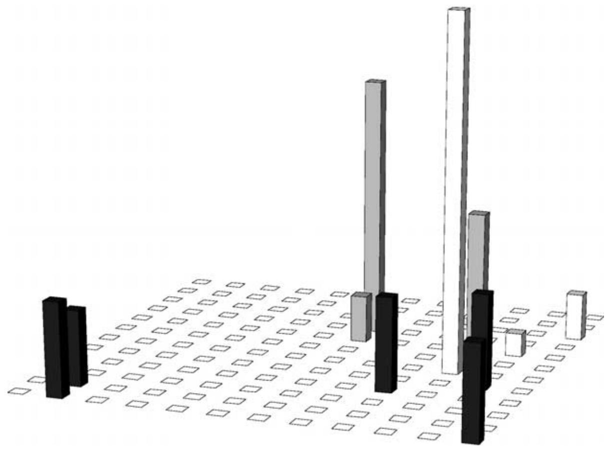

Fig. 8. The percentage of the test samples for the body parts that have highest activation for each unit on a 12 by 12 units network after a training time of 50 epochs. (Black—Hand, White—Head, Grey—Leg).

contained both head and leg action samples with the highest activation, the percentage for head samples was much higher on both test and training data. For the training and test data the percentage of head action samples with the highest activation for that unit was over 60% for training samples and 70% for test samples. Due to the major difference between the head and leg action percentages for this unit, only the head percentage is shown on Figs. 7 and 8.

As can be seen from Table 2 for the training data 100% of the hand and head fell in the appropriate region and 88% of the leg data. For test data the percentage was even better with 100% for hand and head and 90% for leg. It is interesting to note that within the hand verb region there was a good division into the actual action classes. In Figs. 7 and 8 'pick' was located in the lower right of the map, 'put' in the lower left, 'drop' in the unit above 'pick', 'touch' at the top of the hand region and most of the 'lift' samples were located in a unit just below 'touch'. For the other two classes there was some splitting into the individual actions but not on the scale of the hand class.

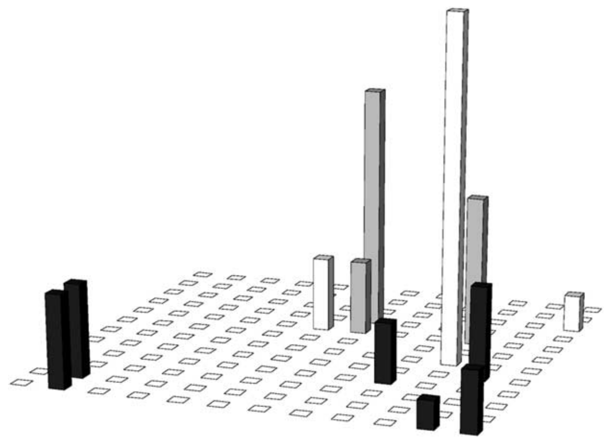

Fig. 7. The percentage of the training samples for the body parts that have highest activation for each unit on a 12 by 12 units network after a training time of 50 epochs. (Black—Hand, White—Head, Grey—Leg).

Hence, such a network can in principle realise the findings of Pulvermüller et al. on the processing of action verbs with different clusters representing the specific body parts. The network was able to identify the semantic features from the actual sensor readings for the individual action verb classes that were specific to the appropriate body part. These features were likely to include the degree of move, whether there was an object involved and the type and number of motors used.

For such an architecture on both training and test data the clusters were in very similar positions on the output layer, which points to the ability of the network to generalise on data it has not seen before. When considering the percentage of test data that fell in the regions identified by the training data the percentages were very high. For the hand actions 100%, head actions 95% and leg actions 88% of the test data fell into the appropriate training region.

Therefore, if the self-organising network was used in the control of a robot it can perform successfully in an online manner clustering semantic features of the action to the appropriate region of the output layer.

Turning to the hand, head and leg self-organising networks, when considering the clustering of the specific body part actions for all three types of action, the size of network that performed best was 8 by 8. For the hand network the training time that produced the best clustering was 50 epochs, for the head network it was 150 epochs and

Table 2
Percentages for the training and test samples that were associated with the appropriate body part clusters for the 12 by 12 units network

Body part	Training samples on training clusters (%)	Test samples on test clusters (%)	Test samples on training clusters (%)
Hand	100	100	100
Head	100	100	95
Leg	88	90	88

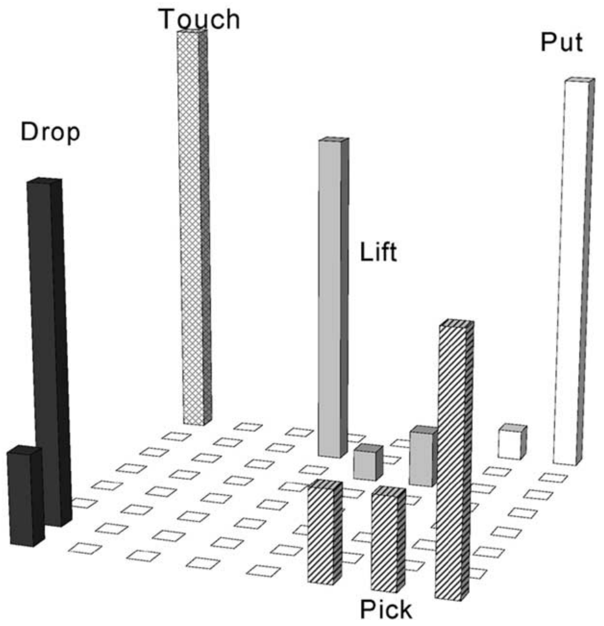

Fig. 9. The percentage of the training samples for the specific hand actions that have the highest activation for each unit on a 8 by 8 units network with a training time of 50 epochs.

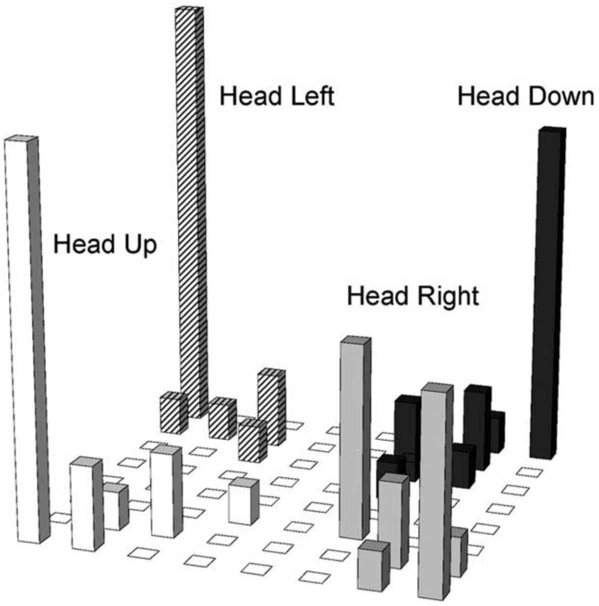

Fig. 11. The percentage of the training samples for the specific head actions that have highest activation for each unit on a 8 by 8 units network with a training time of 100 epochs.

for the leg self-organising network it was 100 epochs. As can be seen from Figs. 9–14 there was clear clustering into different regions for the hand, head and leg actions. For the hand actions on both the training and test data 'touch' was in the top left corner, 'put' in the top right corner, 'drop' in the bottom left corner, 'pick' in the bottom right and 'lift' was in the upper center. Turning to the head self-organising network in Figs. 11 and 12 'head left' was in the upper left region, 'head down' in the upper right, 'head up' in lower left and 'head right' in the lower left region. For the leg self-organising network 'forward' was in the upper left of the map, 'turn left' in the upper right, 'turn right' in the lower left and 'backward' in the lower right corner of the map.

The good performance of the hand, head and leg networks can be observed from Tables 3–5. For the hand network 100% of the 'drop', 'pick', 'lift' and 'touch' and 93% of the 'put' training samples fell into the appropriate

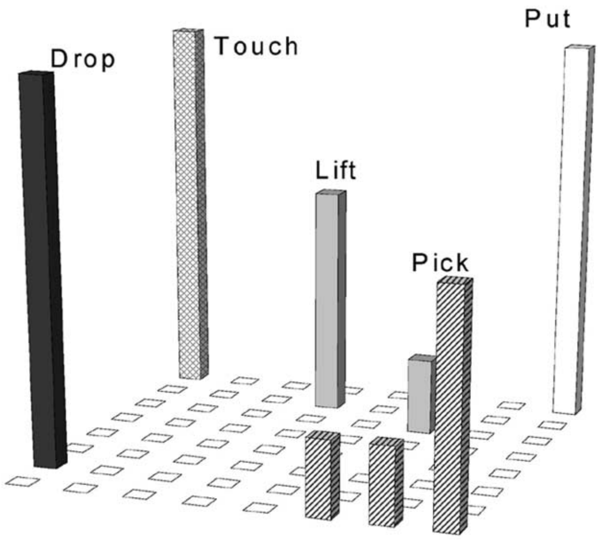

Fig. 10. The percentage of the test samples for the specific hand actions that have highest activation for each unit on a 8 by 8 units network with a training time of 50 epochs.

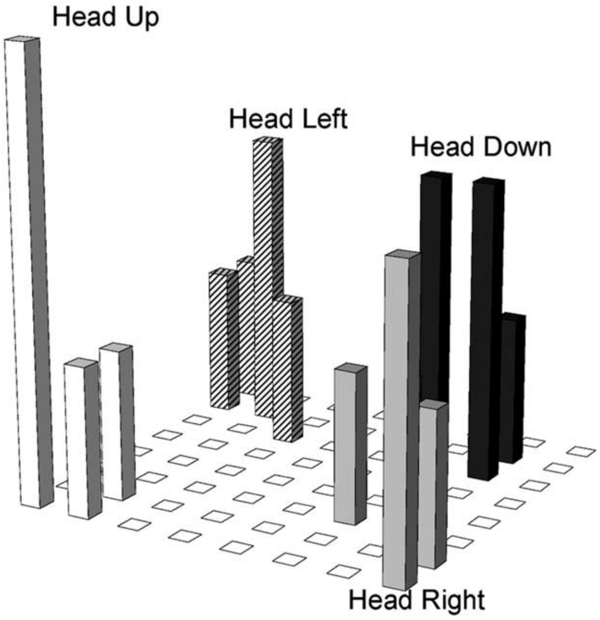

Fig. 12. The percentage of the test samples for the specific head actions that have highest activation for a each unit on a 8 by 8 units network with a training time of 100 epochs.

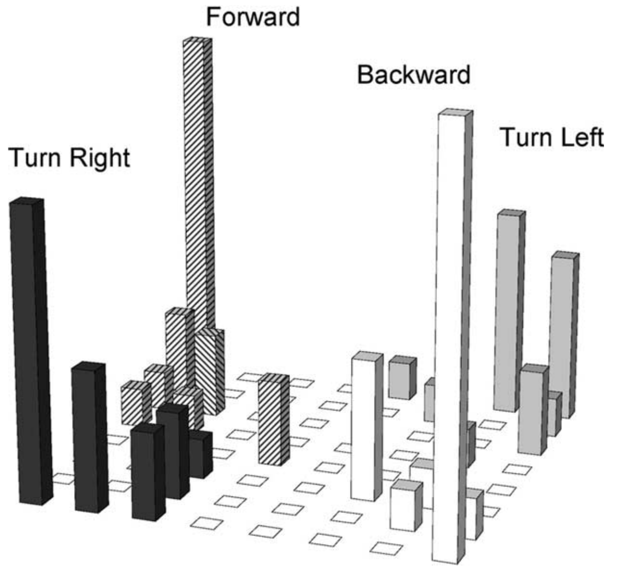

Fig. 13. The percentage of the training samples for the specific leg actions that have the highest activation for each unit on a 8 by 8 units network with a training time of 100 epochs.

training cluster. For the test data it was 100% for all the hand actions. For all head actions they achieved a performance of 100%. When considering the percentage of test data that fell into the regions identified by the training data, the percentages for the hand and head networks was 100%. Finally, as can be seen from Table 5 the leg actions 'backward' and 'turn left' achieved 100% for the training samples in the appropriate training clusters, the test samples on the test clusters and the test samples in the training clusters. Although 'turn right' and 'forward' did not achieve 100% on all three conditions,

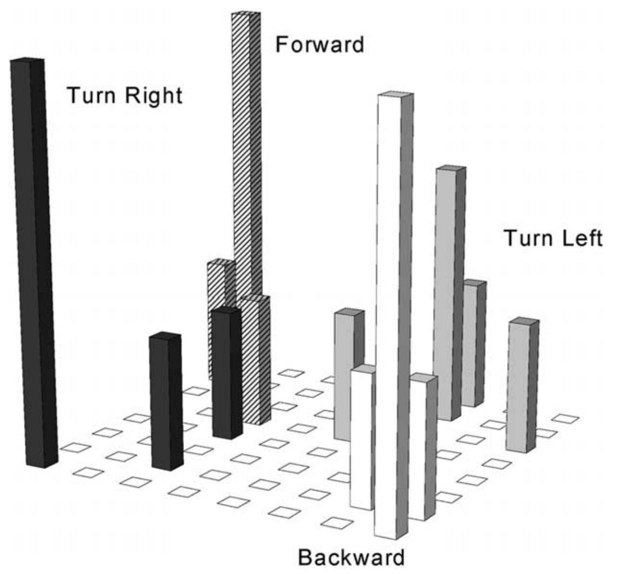

Fig. 14. The percentage of the test samples for the specific leg actions that have highest activation for each unit on a 8 by 8 units network with a training time of 100 epochs.

Table 3
Percentages for the hand training and test samples that were associated with the appropriate clusters for the 8 by 8 units network

Hand actions	Training samples on training clusters (%)	Test samples on test clusters (%)	Test samples on training clusters (%)
Put	93	100	100
Drop	100	100	100
Pick	100	100	100
Lift	100	100	100
Touch	100	100	100

Table 4
Percentages for the head training and test samples that were associated with the appropriate clusters for the 8 by 8 units network

Head actions	Training samples on training clusters (%)	Test samples on test clusters (%)	Test samples on training clusters (%)
Head down	100	100	100
Head left	100	100	100
Head right	100	100	100
Head right	100	100	100

Table 5
Percentages for the leg training and test samples that were associated with the appropriate clusters for the 8 by 8 units network

Leg actions	Training samples on training clusters (%)	Test samples on test clusters (%)	Test samples on training clusters (%)
Backward	100	100	100
Turn right	100	80	80
Turn left	100	100	100
Forward	87	100	80

the performance was still reasonably good. On the training samples on the training clusters 'forward' achieved 87% and 'turn right' 100%, for the test samples on the test clusters 'forward' achieved 100% and 'turn right' 80%, and for test samples on training clusters both actions achieved 80%.

Hence, the performance of the head, leg and hand self-organising networks are in principle suitable for use in a robot control system based on language instruction. This is because it is likely, based on the clear clustering demonstrated, that the sensor reading input will be accurately represented and mapped to the appropriate network region. As this location is the basis for the association between the action and the word this will contribute to the successful identification of the action and its description.

7. Conclusion

We described a model for a robot control system based on language instructions. This model considers that cell assemblies in different regions of the brain are used to process action verbs based on their association with appropriate body parts. By using self-organising networks for each of the three body parts considered it was possible in nearly all cases to cluster the individual actions and to identify them. We directly used sensor readings to represent low level semantic features. While we have not intended to implement a full mirror neuron system in this paper, this paper describes a self-organising approach that controls a robot using language instructions. This approach is based on distributed regional modularity and neurocognitive evidence on clustering action verbs. Such an approach uses sensor readings as the input to the robot and also as the basis for the robot's behaviour. In this sense, we have presented a self-organising language memory for an environmentally grounded robot.

Acknowledgements

This work is part of the MirrorBot project supported by the EU in the FET-IST programme under grant IST- 2001–35282. Thanks to Peter Watt for assistance with data collection and robot experiments.

References

Asoh, H., Huyamizu, S., Isao, H., Motomura, Y., Akaho, S., & Matsu, T (1997). Socially embedded learning of office-conversant robot jjjo-2. *Proceedings of the International Joint Conference on Artificial Intelligence Nagoa*.

Bear, M., Connors, B., & Paradiso, M. (1996). *Neuroscience: Exploring the Brain*.

Binder, J., Frost, J., Hammeke, T., Cox, R., Rao, S., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance images. *The Journal of Neuroscience*, *17*(1), 280–288.

Braitenberg, V. (1997). Searching for language mechanisms in the brain. *Cybernetics and Systems*, *28*, 187–213.

Breazeal, C., & Scassellati, B (1999). A context-dependent attention system for a social robot. In: *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI99)* (pp. 1146–1151). Stockholm, Sweden.

Burgard, W., Cremers, A. B., Fox, D., Hähnel, D., Lakemeyer, G., Schulz, D., Steiner, W., & Thrun, S. (2000). Experiences with an interactive museum tour-guide robot. *Artificial Intelligence*, *114*, 1–2.

Dogil, G., Ackermann, H., Grodd, W., Haider, H., Kamp, H., Mayer, J., Riecker, A., & Wildgruber, D. (2002). The speaking brain: a tutorial introduction to fMRI experiments in the production of speech, prosody and syntax. *Journal of Neurolinguistics*, *15*(1), 59–90.

Dorffner, G., & Prem, E (1993). Connectionism, symbol grounding, and autonomous agents. In: *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 144–148). Boulder.

Elshaw, M., & Wermter, S (2002). A neurocognitive approach to self-organisation of verb actions. In: *Proceedings of the International Joint Conference on Neural Networks* (pp. 24–29). Honolulu, USA.

Feldman, J. A., Lakoff, G., Bailey, D. R., Narayanan, S., Regier, T., & Stolcke, A. (1996). L_0—The first five years of an automated language acquisition project. *AI Review*, 8.

Friedman, L., Kenny, J., Wise, A., Wu, D., Stuve, T., Miller, D., Jesberger, J., & Lewin, J. (1998). Brain activation during silent word generation evaluated with functional MRI. *Brain and Language*, *64*, 943–959.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Science*, *2*(12), 493–501.

Gazzaniga, M., Ivry, R., & Mangun, G. (1998). *Cognitive neuroscience: The biology of the mind*. New York: W.W. Norton and Company Ltd.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, *42*, 335–346.

Joliot, M., Papathanassiou, D., Mellet, E., Quinton, O., Mazoyer, N., Courtheoux, P., & Mazoyer, B. (1999). fMRI and PET of self-paced finger movement: Comparison of intersubject sterotaxic averaged data. *Neuroimage*, *10*, 430–447.

Kohonen, T. (1997). *Self-organizing maps*. Germany: Springer.

Melcher, J. (2000). Human auditory function studied with fMRI. In: *Proceedings of the Workshop the Nature of Speech Perception*.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*(2), 253–336.

Pulvermüller, F. (2002). A brain perspective on language mechanisms: From discrete neuronal ensembles to serial order. *Progress in Neurobiology*, *67*, 85–111.

Pulvermüller, F. (2003). *The neuroscience of language: On brain circuits of words and serial order*. Cambridge: Cambridge University Press.

Pulvermüller, F., Hare, M., & Hummel, F. (2000). Neurophysiological distinction of verb categories. *Cognitive Neuroscience*, *11*(12), 2789–2793.

Purves, D. (1997). *Neuroscience*. Sunderland, MA: Sinauer.

Reggia, J., Shkuro, Y., & Shevtsova, N. (2001). Computational investigation of hemispheric specialization and interactions (pp. 68–82). In S. Wermter, J. Austin, & D. Willshaw (Eds.), *Emergent Neural Computational Architectures Based on Neuroscience.* Heidelberg: Springer.

Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. *Trends in Neuroscience*, *21*(5), 188–194.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Review*, *2*, 661–670.

Steels, L. (1998). The origins of syntax in visually grounded robotic agents. *Artificial Intelligence*, *103*(1-2), 133–156.

Taylor, J. (1999). Towards the networks of the brain: From brain imaging to consciousness. *Neural Networks*, *12*(7-8), 943–959.

Thrun, S., Bennewitz, M., Burgard, W., Dellaert, F., Fox, D., Haehnel, D., Rosenberg, C., Roy, N., Schulte, J., & Schulz, D (1999). MINERVA: A second generation mobile tour-guide robot. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'99).*

Wermter, S., Austin, J., Willshaw, D., & Elshaw, M. (2001). Towards novel neuroscience-inspired computing (pp. 1–9). In S. Wermter, J. Austin, & D. Willshaw (Eds.), *Emergent Neural Computational Architectures Based on Neuroscience*. Heidelberg: Springer.