A Modular Approach to Self-organisation of Robot Control Based on Language Instruction

STEFAN WERMTER, MARK ELSHAW and SIMON FARRAND

Centre for Hybrid Intelligent Systems, School of Computing and Technology, University of Sunderland, St Peter's Way, Sunderland, SR6 0DD, UK
email: [Stefan.Wermter][Mark.Elshaw][Simon.Farrand]@sunderland.ac.uk
web: http://www.his.sunderland.ac.uk

Abstract

In this paper we focus on how instructions for actions can be modelled in a self-organising memory. Our approach draws from the concepts of regional distributed modularity and self-organisation. We describe a self-organising model that clusters action representations into different locations dependent on the body part they are related to. In the first case study we consider semantic representations of action verb meaning and then extend this concept significantly in a second case study by using actual sensor readings from our MIRA robot. Furthermore, we outline a modular model for a self-organising robot action control system using language for instruction. Our approach for robot control using language incorporates some evidence related to the architectural and processing characteristics of the brain (Wermter et al. 2001b). This paper focuses on the neurocognitive clustering of actions and regional modularity for language areas in the brain. In particular, we describe a self-organising network that realises action clustering (Pulvermüller 2003).

1. Introduction

Despite the growing interest in robotic learning many approaches fail to take account of neural networks, cognitive evidence or language instruction. While some robots have some learning capability, they are still restricted in their general autonomous behaviour. Even the Talking Heads approach that incorporates the emergence and evolution of language in robot's language development gives little consideration to neuroscience-inspired learning in humans (Steels 1998).

Some robots like the tour-guide robot Rhino (Burgard et al. 2000) have been quite robust in terms of their localisation and navigation behaviour, however they do not interact via language. People interact with Rhino by pushing coloured buttons so that the robot can respond with information concerning individual exhibits. When considering robots from the language interface perspective, the conversation office robot jijo-2 (Asoh et al. 1997) can be instructed to navigate to certain landmarks. Furthermore there is Kismet (Breazeal and Scassellati 1999) an interactive robot that is a static sophisticated head with moving eyes, ears, lips and face that can recognise emotions but does not understand or generate real language. Finally, there is Minerva, a navigating tour-guide with two cameras on a neck. Minerva can interact by either synthesising a pre-

programmed text for speech output, generate a few phrases or smiling and frowning as necessary (Thrun et al. 1999). Currently there is no neural learning or coupling of language with other modalities.

In our approach we would like to study the integration of language and robot action based on brain-inspired neural networks. The brain is often viewed as various distributed neural networks in diverse regions which carry out processing in a parallel distributed manner to perform specific cognitive functions such as language processing (Dodel et al. 2001, Reggia et al. 2001, Knoblauch and Palm 2001). According to Reilly 2001 the brain performs as a group of collaborating specialists that perform language processing by splitting this task into smaller subtasks. Although it is not fully understood why certain regions of the brain are associated with language processing, the parallel performance achieved would not be possible without some form of distributed modular organisation (Reggia et al. 2001). Furthermore, Hicks and Monaghan 2001 note that a feature of regional modularity is the division of the activities between different hemispheres of the brain to perform a cognitive function. This form of regional distributed modularity is the basis of our approach as various self-organising networks are to be used in a modular manner.

2. Modular assemblies in the brain

Until recently distributed regional modularity in language processing was identified through brain examination after lesions that are responsible for observed language deficits (Dogil et al. 2002). Although the mapping of various language functions to specific brain regions has been done, this lesion approach is very limited and technical advances led to the introduction of brain imaging techniques. The four most common brain imaging techniques are positron emission tomography (PET), functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG) and electroencephalogram (EEG) (Braitenberg 1997, Dogil et al. 2002, Taylor 2001, Pulvermüller 2003, Joliot et al. 1999). PET and fMRI both examine the neural activity within the brain to create an image of the regions associated with a cognitive task. For PET studies those regions that are considered responsible for a cognitive function have the largest blood flow and for fMRI it is those regions with the highest blood oxygen level (Gazzaniga et al. 1998, Dogil et al. 2002). EEG measures voltage fluctuations produced by regional brain activity through electrodes positioned on the surface of the scalp and MEG uses variations in the magnetic field to measure brain activity using sophisticated superconducting quantum devices (Gazzaniga et al. 1998, Taylor 2001).

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 and Arbib 1998, Rizzolatti et al. 2001). Neurocognitive evidence of Pulvermüller 2003 on action verb processing also provides a new architectural approach for language processing, which is based on the use of cell assemblies and synfire chains. The associative learning and memory theories of Hebb 1949 suggest that cognitive representations depend on associations between neurons in an individual population and between distributed cortical neuronal populations. A population or circuit of neurons that forms a functional unit for a cognitive representation task is known as a cell assembly (Pulvermüller 1999, Huyck 2001). Cell assemblies rely on a connectivity structure

between neurons that support one another's firing and hence have a greater probability of being co-activated in a reliable fashion (Treves and Rolls 1994, Spitzer 1999, Wermter et al. 2001a, Wermter and Panchev 2002). Synfire chains are formed from the spatiotemporal firing patterns of different associated cell assemblies and rely on the activation of one or more cell assemblies to activate the next assembly in the chain. A synfire chain can represent words, with the representation depending on which and when the cell assemblies are activated (Pulvermüller 1999, Pulvermüller 2003).

When looking at the cell assemblies that represent particular word types, Pulvermüller 1999 observes that activation is found in both hemispheres of the brain for content words. Perisylvian assemblies and posterior cortex neurons represent perception words such as odour or taste. Nouns that relate to animals activate the inferior temporal or occipital cortices. However, function words that have a grammatical role are limited to the perisylvian cortex. For action words that involve moving one's own body the perisylvian cell assemblies are associated with motor, premotor, and prefrontal cortices. Assemblies that depict vision words are 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 form with those neurons associated with perception and actions that reflect the semantic information of a word's meaning. Hence, if a word is repeatedly presented with a stimulus the depiction 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 themselves doing this activity. Words are therefore depicted via regions historically known as language regions and additional regions connected with the word's meaning. This neurocognitive evidence directs our approach through the use of meaning and associative memory in multiple regions of the brain.

To support their findings on Hebbian learning and synfire chains Pulvermüller et al. 2000 and Pulvermüller et al. 2001 examine the processing of action verbs. Associative learning suggests distinctions between categories of action verbs based on the part of the body that performs the action. Based on this distinction Pulvermüller et al. 2001 hypothesise that as different muscles are used for the each body part, clear differences are observed in the regions of the cortex activated. The association of action verbs with these regions of the brain comes from the language acquisition stage when an actual verb is spoken before, during or after the action.

Pulvermüller et al. 2001 perform three experiments on brain processing of action verbs to test their hypothesis. In the first two experiments different groups of subjects are given leg-, arm- and face-related action verbs and pseudo-words. Then the subjects indicate whether the stimulus is a word or pseudo-word. In the third experiment subjects are required to use a rating system to answer questions on the cognitive processes a word arouses. EEG electrodes at various points along the scalp produce neurophysical recordings. These experiments identify differences between the action verbs based on the body parts they relate to. The average response times for lexical decisions are faster for face-associated words than for arm-associated words, and the arm-associated words are faster than leg ones. There is a significant difference for the three body parts for the

prefrontal and occipital regions, and the motor and premotor cortex. The prefrontal area is found to be associated mainly with arm-related verbs and the occipital visual areas for face-related verbs. Hence these differences in latencies and location of activation points of action verbs depend on the related body part. These differences are based on the semantic representations by cell assemblies of the action verbs. By using this cognitive evidence for our language system grounded in robot action control we envisage a new architectural approach. This neurocognitive evidence motivates our approach for self-organising associative memory in multiple regions of the brain, since we use the concept of action verbs being processed dependent on the region of the body they relate to and the modularity in the brain.

3. From neurocognitive evidence to modularity in robots

As the overall aim of the research is to combine neurocognitive evidence and self-organisation to examine language control of robot action, we consider some past approaches to robot systems that incorporate modular self-organising networks. There is a significant amount of research on modular self-organising network robot systems (Owen and Nehmzow 1996, Nolfi 1997, Tani and Fukumura 1997, Calabretta et al. 1998, Oka et al. 1998, Nehmzow 1999, Voegtlin and Verschure 1999, Bryson and Stein 2001). Here we will consider just a few representative examples to offer an overview of these approaches. We focus in particular on different neural architectures.

3.1 Modular and self-organising robot architectures

Considerable interest exists in a modular design for improving the performance of robots (Calabretta et al. 1998). Based on the work of Nolfi 1997, Calabretta et al. 1998 examine a modular neural network approach for the control of a robot to perform litter collection. This approach is described as an emergent modular architecture as it enables behaviour to be split into sub-elements that match diverse neural modules as a response to evolutionary adaptive procedures. The task of litter collection is split into various sub-activities: (i) to examine the environment; (ii) to identify a piece of litter and position itself so it can be picked up; (iii) pick up the litter; (iv) move toward a wall while avoiding other objects; (iv) identify the wall and allow the litter to be dropped outside the area; and (v) drop the litter outside the area.

Three different architectures are considered for this task: a simple feedforward network, the hardwired modular architecture that allows the required behaviour to be controlled by different neural modules, and finally the duplication-based modular architecture where the modules are not hardwired but added during the evolutionary process. A genetic algorithm is used to evolve the connection weights of the neural networks. It is found that the modular architectures outperform those with a basic network structure. For the hardwired modular architecture the evolved individuals develop a single module to control the left motor, the pick-up process and use two competing neural modules for the right motor. For the duplication-based modular approach the typical evolved individual uses both neural modules to control the left motor, the right motor, the pick-up procedure and the release process. This individual uses different modules depending on the environmental conditions.

Similar to our approach Owen and Nehmzow 1996 and Nehmzow 1999 use self-organising networks, however they use them to enable a robot to perform route learning and vision processing. Route learning tackles the problem of the division between higher level symbolic processing and lower level information gathering from sensors. Self-organising networks produce a topographical mapping of the environment from sensor stimuli. Clear regions of the output layer are associated with locations in the environment, where the robot maps its own environment in an unsupervised manner. Regions of the network are seen as perceptual landmarks in the robot environment, and so navigation is achieved through the association of actions with landmarks. The robot uses the association between locus and action for route learning through the inclusion of the action in the weight and input vector. During route recall the action part of the input is set to zero, and the appropriate action comes from the weight vector of the winning unit of a pattern completion task.

The additional vision processing by Nehmzow 1999 clusters vision data from a camera in an autonomous manner to differentiate between images that include boxes and those that do not. Various pre-processing activities are performed including edge detection and generation of a binary image to produce a 100-element input vector for a 10 by 10 unit self-organising network. By using the self-organising network to process the robot's sensory signals, distinct sensory perceptions are mapped onto clear areas of the network, with close perceptual patterns clustering together in an area. The self-organising network is able to differentiate between static pictures containing boxes and those not containing boxes, and the robot is guided to these boxes using this approach.

The approaches considered here are of interest to us as they clearly show the suitability of robot behaviour control based on a modular self-organising architecture. However, they fail to consider significantly the benefits offered by language processing in the control process. Our approach incorporates both a neural architecture and language processing to control robot behaviour. The use of both the robot internal state and language in our general approach matches the multi-modal input and processing associated with human behaviour.

3.2 Self-organising networks

Self-organising networks such as the one shown in figure 1 consist of an input and an output layer, with every input neuron linked to all the neurons in the output layer (Kohonen 1997). The output layer creates a topological representation that clusters similar inputs together in a two-dimensional neural layer.

Self-organising neural networks have an input vector represented as $i = [i_1, i_2, ..., i_n]$. The input vector is presented to every output unit of the network; the weights between the links in the network are provided by:

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

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

$$o_j = \sum_{k=1}^{n} w_{jk} i_k = w_j \bullet i \qquad (2)$$

The weights are 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 neighbouring units are altered based on the following equation:

$$\Delta w_{jk} = \alpha(i_k - w_{jk}) \text{ and } w_{jk}(t+1) = w_{jk}(t) + \Delta w_{jk} \qquad (3)$$

where α is the learning rate parameter that is usually between 0.2 and 0.5. The shape and units in the neighbourhood depends on the neighbourhood function used. The number of units in neighbourhood usually drops gradually over time. The weights of the units in the neighbourhood are updated differently depending on how far they are from the winning unit.

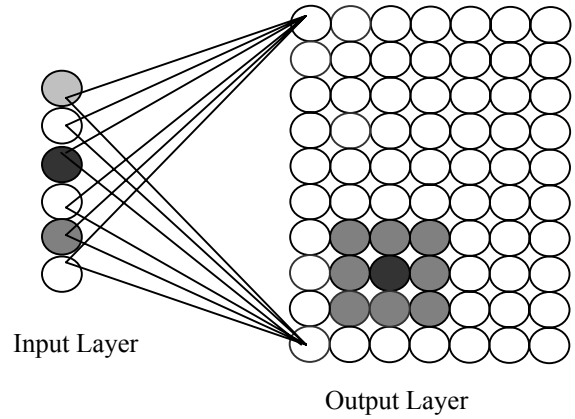


Figure 1. A representation of a self-organising network: The darker the neuron the higher the activation.

4. Self-organising for action verb processing

The first step in a language processing system for robot action was to ensure that self-organising networks were suitable for realising the findings of Pulvermüller 1999, Pulvermüller 2003, Pulvermüller et al. 1999, Pulvermüller et al. 2000, Pulvermüller 2001 and Pulvermüller et al. 2001 on action verb processing based on the body part they relate to. There was also an aim to use simple but real language material. This was done through a systematic experiment that involved the creation of a data set and training and testing of self-organising networks of various map sizes.

4.1 Experimental method

We intended to use actual spoken language that involved a degree of training to perform actions. We also intended to start with rather simple sequences. Hence, a data set was extracted from the transcripts of the speech between a child and adults from the Childes Corpus (Bloom et al. 1974, Bloom et al. 1975). The Childes Corpus is an on-line

database of children's utterances (MacWhinney 1995). The actual data set that was used consisted of questions/requests and responses between the child and adults. Interactions were extracted if they included action verbs that were associated with the hand, head or leg. The actual action verbs that were used in this experiment and example interactions can be seen in table 1 and table 2.

Table 1. Action verbs associated with body parts.

Leg	Head	Hand
Stand	See	Get
Walk	Eat	Put
Come	Drink	Push
Ride	Sing	Write
	Ask	Draw

Table 2. Typical body related interactions.

Head	Adult	you SEE who is in this car?
	Child	right there
Hand	Child	can I PUSH it?
	Adult	not yet.
Leg	Adult	COME here.
	Child	right.

From these interactions, the utterances containing the listed action verbs were taken. Once this was performed the action verb and the first two nouns were extracted to produce a three word phrase. A representation was devised based on semantic information of the action verb meaning and the nouns to introduce the three-word phrase into self-organising neural networks. The semantic information was based on an approach similar to McClelland and Kawamoto 1986 and used various semantic features and their values. For example, for the action verbs one such feature was the level of physical effort required when performing the action and the possible values were 'small', 'medium' or 'large'. For the full set of features for the actions verbs and possible values that were used see table 3. The noun semantic features included whether the noun was 'human or part of a human' and the possible values were similar to those for the verbs (see table 4). These representations were selected with the expectation that they can be replaced by some sensor readings and regarded as a start point.

Table 3. Semantic features for action verb meaning.

Semantic Feature	Responses
Level of self movement	Small Medium Large
Physical movement of object	Small Medium Large
Physical exertion involved	Small Medium Large
Requires precision operations with object	No Yes
Physical interaction with person	No Yes
Level of communication	Low Medium High
Produces understandable statements	No Yes
Physical impact on object	None/Low Reposition Consume

Table 4. Semantic features for noun meaning.

Semantic Features	Responses
Human/part of human	Yes No
Edible	Yes No
Living or copy of a living thing	Yes No
Toy	Yes No
Consistency	Gas Liquid Solid

As the input to the neural networks was in numeric form the responses for the action verbs and the nouns were represented with numeric values. For both the action verbs and nouns 'small' was represented by 01, 'medium' 10, 'large' 11, 'yes' 1 and 'no' 0. In addition for the action verbs the representation for the response to the physical impact on the object was 'none/low' 01, 'reposition' 10, 'consume' 11. Table 5 provides the full representation of the action verbs considered in this experiment. The semantic representation for the noun Peter was human, not edible, living, not a toy and solid which gave a numeric representation of 101011. For the pronouns 'you', 'I', 'he' and 'she' it was possible to identify the semantic reference and allocate a representation. However, for the pronoun 'it' an arbitrary created number of 000101 was allocated as it is not always possible to determine what the pronoun 'it' refers to. If there were not two nouns in the three-word phrase the missing nouns were represented as 000000.

Table 5. Representation of the action verbs.

Action Verbs	Representation	Action Verbs	Representation	Action Verbs	Representation
Walk	1101110001001	See	0101010001001	Get	1011101001010
Stand	0101110001001	Eat	0101010001011	Put	1011101001010
Come	1101110001001	Drink	0101010001011	Push	1011101001010
Ride	1111110001001	Sing	0101010111101	Write	1010011001101
		Ask	0101010111101	Draw	1010011001101

4.2 Unsupervised learning

In the experiment the input layer to the self-organising networks was fixed to 25 units. 13 units were used for the action verb and 6 for each of the two possible nouns. The output layers had various sizes (7 units by 7 units, 9 by 9, 10 by 10 and 12 by 12). The networks were trained for up to 100 epochs at 25 epoch intervals using the three word phrases. In total there were 525 training and 284 test samples. Training samples were presented to the networks in a random order and the weights updated at the end of each epoch. The location of each of these training phrases on the self-organising output layers was identified based on the units that had the highest activation. The trained networks were tested for their ability to generalise by identifying the location on their self-organising map for unseen action verb phrases. The objective was to determine the network architecture and training time that achieved the clustering of action verb phrases into regions related with the appropriate body part and to generalise on unseen samples.

As the number of training and test samples were different for each body part and between the action verbs (see tables 6 and 7), percentage values rather than absolute values were determined for the units. These percentages were based on the number of

training and test samples for specific action verbs and the action verbs related with specific body parts whose activation was highest for each network unit. For example, if 30 out of 50 'get' action verb training samples had the highest activation for the top left hand unit, the unit had a value of 60% for this verb. To remove the impact of outliers only those units that contained 5% for a specific body part and 10% for a specific action verb were considered.

Table 6. Sample numbers by body part.

Body Part	Training Samples	Test Samples
Hand	284	143
Head	134	79
Leg	107	62
Total	525	284

Table 7. Sample numbers for action verbs.

Action verb	Training Samples	Test Samples
Get	50	27
Put	152	74
Push	11	5
Write	58	27
Draw	13	10
See	56	32
Eat	32	18
Drink	6	5
Sing	18	11
Ask	22	13
Walk	27	14
Stand	27	16
Come	42	23
Ride	11	9
Total	525	284

4.3 Results and discussion

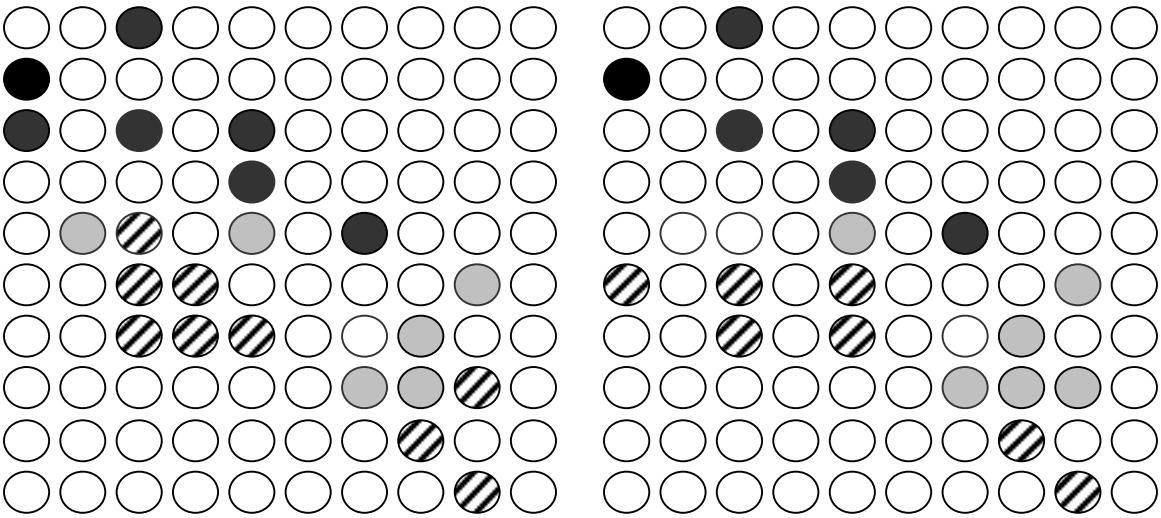
For self-organising networks of output layer sizes 7 by 7 units and 8 by 8 units there were no clear clustering according to the body parts when considering percentages for the individual action verbs and for the body parts. However, when turning to networks that had 9 by 9 units and 10 by 10 units and thus less restricted memory it was possible to see clear clusters based on body parts. This indicated the importance of network size when performing a task. For example, from figure 2 for a 10 by 10 units network trained for 75 epochs the hand action verbs were clustered in the upper left region and below these the head action verbs were clustered. In figure 2 the black units are hand units, diagonal lines units are head and grey units are leg. The white units are those units that had less than 5% of the samples for specific body parts that had the greatest activation, however in most cases this was 0%. It was also found that for one of the units in the leg region of the output map it had the highest activation for over 5% for both leg and head action verbs. Nevertheless, leg had the highest percentage value for both the training and test data and

so the unit is shown as leg in figure 2. Hence, there was a clear clustering and consistency between the maps of the training and unseen test samples.

When considering the location of the specific action verbs, head action verbs such as 'eat', 'drink' and 'see' were grouped together in one region of the output layer and head action verbs such as 'sing' and 'ask' in another. Furthermore, for the 10 by 10 units network although there was a single area for hand action verbs those related with communication activities such as 'write' and 'draw' were to the right of this area and the other hand action verbs to the left.

When considering the percentage of action verbs that were located in the appropriate body part cluster on training for this network the performance was better for the hand verbs compared with head and leg verbs. For the training samples that fell in the appropriate training clusters the results were 89% for hand, 73% for head and 62% for leg. Turning to the test samples, 91% of the hand, 71% of the head and 68% of the leg were in the appropriate cluster (see table 8).

For the training and test data the clusters were in very similar positions on the output layer, which indicated the ability of the networks to generalise on unseen data. For the test action verbs that were located in the appropriate training data clusters hand achieved 90%, head achieved 66% and leg achieved 55%. Although the performance on the leg data was not as good as the other two there was a number of samples that fell just outside these regions for the 10 by 10 units network. The percentage of all the test samples that were located in the appropriate training body part cluster was 75%.



Training Data Test Data

Figure 2. The units on the training and test data for a network with 10 by 10 units with learning over 75 epochs that had 5% or more samples for the specific body parts whose activation value was the greatest. (Black – Hand, Diagonal lines – Head, Grey – Leg, White – Units that had less than 5% of samples for specific body parts that had the greatest activation.)

Table 8. Percentages for the training and test samples that were located within the appropriate body part clusters for the 10 by 10 units network.

Body part	Training samples on training clusters (%)	Test samples on test clusters (%)	Test samples on training clusters (%)
Hand	89	91	90
Head	73	71	66
Leg	62	68	55

5. Modular Architecture for robot control

The self-organising network considered so far modelled findings of Pulvermüller by relating action verbs with their appropriate body part using different self-organising regions. This supports Pulvermüller et al.'s 2001 hypothesis that the clustering of the action verbs on the part of the body categories resulted from differences in semantics features contributing to their meaning. The split in head action verbs between 'ask' and 'sing', and 'eat', 'drink' and 'see' may have come from the self-organising network's identification of the distinctive features of these subcategories. Hence, while these two subcategories contained common features related to the body part, the 'sing' and 'ask' subcategory incorporated communication-related features that were distinct from the 'eat', 'drink' and 'see' subcategory.

As a response to the first experiment that identified that a self-organising network was able to model the findings of Pulvermüller, a new robot-grounded modular architecture of self-organising networks was developed. This system was developed to perform robot control based on sensor readings rather than hand-coded semantic features. Therefore this experiment is fundamentally different since it uses features embodied directly in the motors and sensors of the robot and thereby eliminating the hand-coding of feature values.

Our architecture explores the use of internal sensors as action descriptions in a self-organising memory. As can be seen from figure 3 the architecture contains a self-organising network to relate 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 cluster the actual action verbs in different regions. Once the action is clustered on the higher-level body part network the input can be 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 wheels, the gripper state and how the readings for various sensors relate to the components of the overall action.

Our approach offers 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 organisation. Note that we do not claim that the self-organising map is a biologically valid cortex model. Rather we demonstrate how regional clustering of actions emerges based on grounded sensor/motor features. The approach takes into account the evidence of Pulvermüller in that different regions of self-organising networks are related with specific action verbs based on their association with the appropriate body part.

There are certain links with concepts of the mirror neuron theory. Rizzolatti and Arbib 1998 point out the relationship between mirror neurons and language as neurons located

in the F5 area of a primate's brain are 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 movement (Gallese and Goldman 1998). 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 the stored representation linked to observing the action.

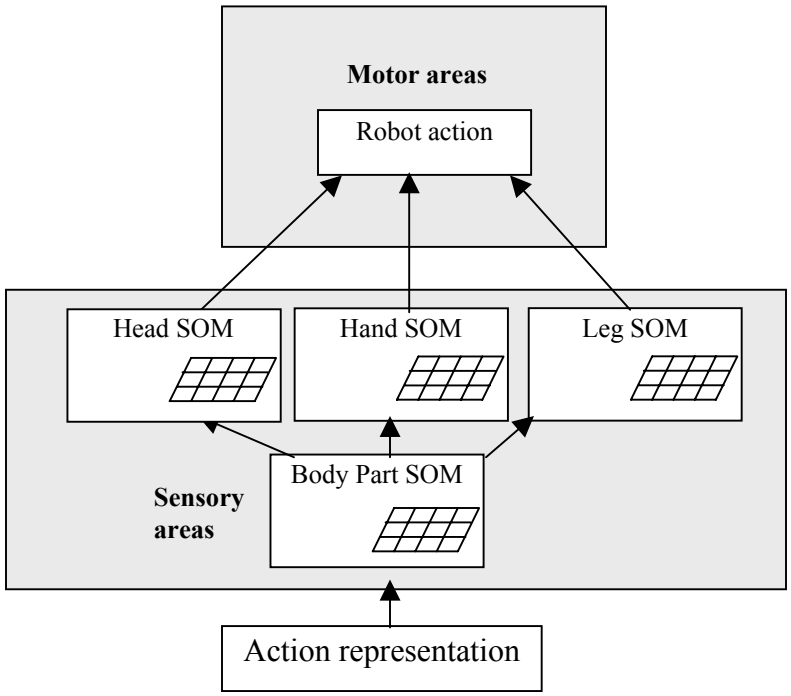


Figure 3. Overall architecture based on modular distributed and self-organising memory.

6. Learning to cluster robot actions

In order to have greater objectivity in the modular architecture for robot control based on language instruction sensor readings were taken from the MIRA robot. Such sensor readings represent semantic features to describe the actions.

6.1 Experimental method

Our MIRA robot in figure 4 is based on a PeopleBot platform, and has an onboard PC, microphone and speakers and a PC104 audio board. The robot's I/O interface allows for the recording of data from numerous internal and external sensors. Wireless communication between the robot and a computer is used. The robot has an adjustable 120-degree pan-tilt camera to look up to see faces or down to objects and infrared sensors to sense the underside of the table. The robot also has 2-degree gripper that contains break-beam sensors to detect whether an object is available to be picked up. MIRA was designed particularly to support human robot interaction. We take advantage of speech, motion and vision interfaces while at the same time exploring and developing a novel neurally-inspired architecture.

The MIRA robot performed various actions that are related 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 consist of several basic subactions. For instance, the hand verb action 'put' 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) stop forward motion; and (v) open gripper to put object on table. 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.

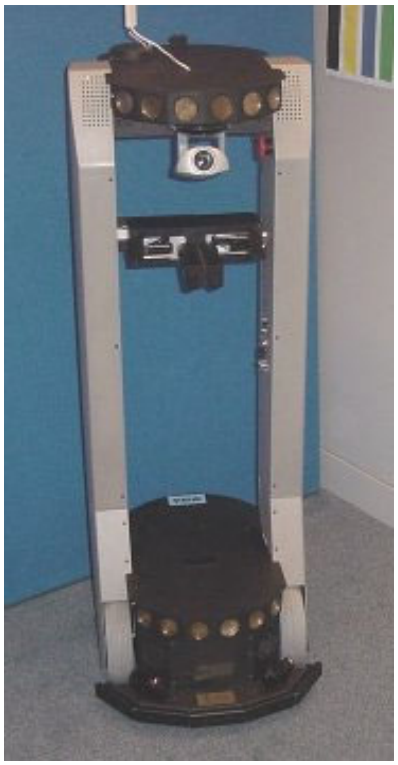


Figure 4. MIRA (MIrror neuron Robot Agent) based on the PeopleBot robot platform.

In order to provide sufficient and varied training and test data the actions were repeated 20 times under diverse conditions. For instance, the angle the robot turned to and the camera tilted and panned were varied. The sensor readings that were taken every tenth of a second while MIRA performed these actions included whether the gripper was moving, the location of the robot and the table sensor state. The full list of the sensor readings is given in table 9.

To reduce the size of the input for the self-organising networks to a manageable level, 10 sets of the readings were taken over time to represent the action. Taking the first, last and eight equi-distant sets of readings and combining them to create a single input for a sample achieved this. Pre-processing was performed on the data to make it suitable for feeding 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 scale 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. Scaling was done by taking the sensor readings for the specific feature for all samples across the 10 sets of readings and positioning the values between 0 and 1 dependent on its relative size using equation (4).

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

Table 9. 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 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

6.2 Unsupervised Learning

For the self-organising networks to cluster actions based on the appropriate body part the input layer had 120 units, one for each of the pre-processed sensor readings. The output layers had various sizes (from 8 by 8 units to 13 by 13 units) and the networks were trained for up to 500 epochs at internals of 50 epochs to find the optimum architecture. The number of training and test samples for each of the 13 actions were 15 and 5 respectively (260 samples in total). 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.

For the hand, head and leg self-organising networks the input was the pre-processed sensor readings, the output layers were varied between 8 by 8 units to 10 by 10 units and they were trained at 25 epoch intervals up to 200 epochs. However, the networks received only the sensor readings input for the action verbs related to the appropriate body part.

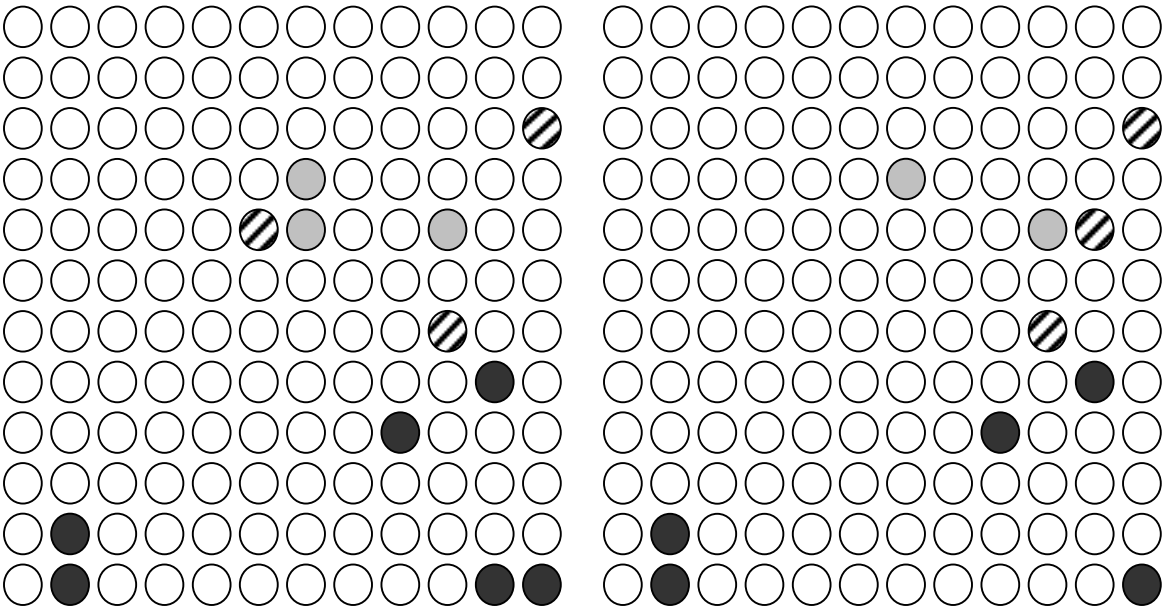
6.3 Results and Discussion

When considering self-organising networks with output layers of between 8 by 8 and 11 by 11 units they were only able to produce clustering between hand actions and the

other two classes. This however indicated an ability to produce a split between simple leg and head actions, and the more complex hand actions.

Turning to a 12 by 12 units network trained for 50 epochs there was fairly clear clustering into the three body parts (see figure 5). The hand action words 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 contained both head and leg action samples with the highest activation for the training and test samples, as the percentage for the head samples was over 60% higher for both types of sample the unit is shown as a head one in figure 5. Within the hand verb region in figure 5 there was a good division into the actual action classes. For instance, 'put' was in the lower left, 'pick' was located in the lower right of the map, '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'. Table 10 shows for the training data 100% of the hand and head actions fell in the appropriate region and 88% of the leg data. For test data the percentages were 100% for hand and head and 90% for leg.

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. Despite using actual sensor readings to represent semantic features the model realised the findings of Pulvermüller on the processing of action verbs. However, we would like to note that we do not claim that the self-organising map topology directly corresponds to brain regions but it is the case that similar actions were found in similar clusters.



Training Data Test Data

Figure 5. The units on the training and test data for a network with 12 by 12 units with learning over 50 epochs that had samples for the specific body parts whose activation value was the greatest. (Black – Hand, Diagonal lines – Head, Grey – Leg, White – Units that had no samples for specific body parts that had the greatest activation.)

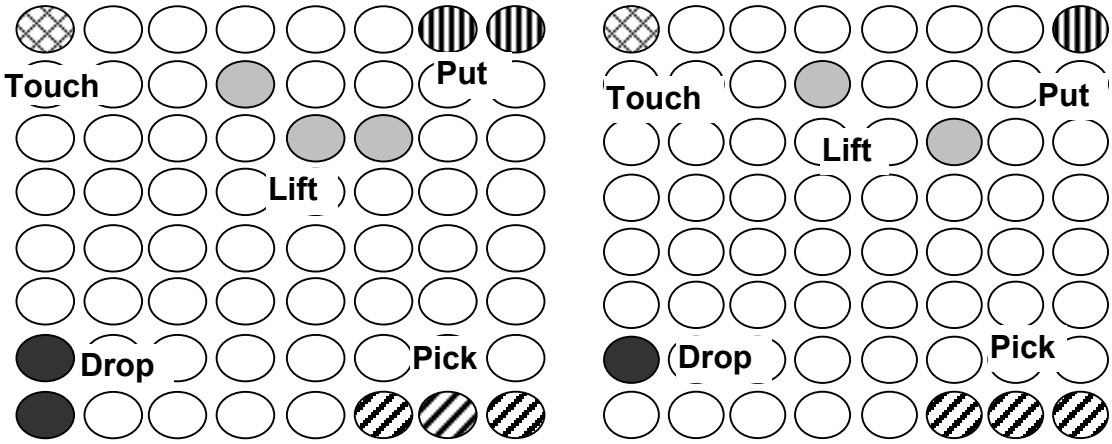
Table 10. Percentages for the training and test samples that were located within the appropriate body part clusters for the 12 by 12 units network.

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

For the hand, head and leg self-organising networks, when considering the clustering of the actions verbs for the specific body part, the network size 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 for the leg self-organising network it was 100 epochs. As can be seen from figure 6 to figure 8 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 figure 7 '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 right 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 Table 11 to Table 13. For the hand network, 100% of the 'drop', 'pick', 'lift' and 'touch' and 93% of the 'put' training samples fell into the appropriate 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 hand and head networks achieved 100%. Finally, as can be seen from Table 13 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, 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 can be considered suitable for 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 and so the appropriate robot action produced.

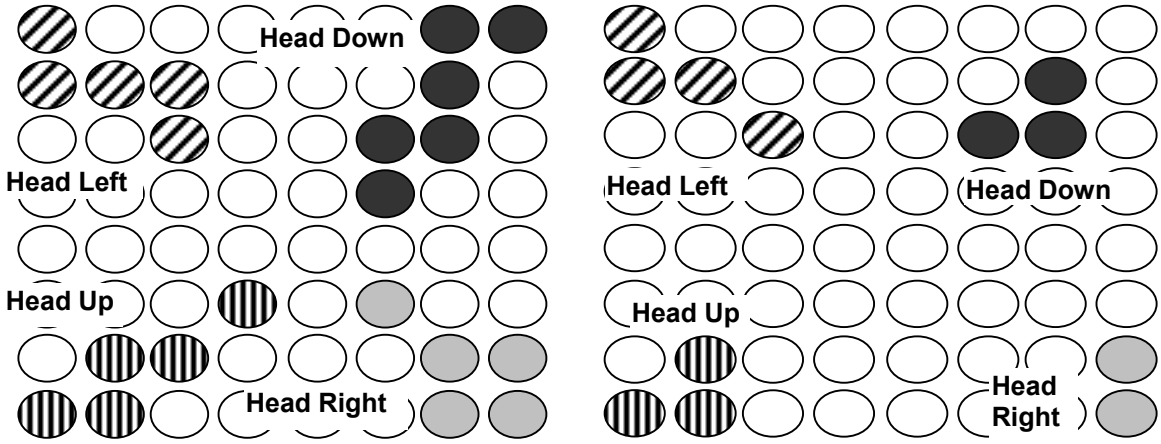


Training Data Test Data

Figure 6. The units on the training and test data for a network with 8 by 8 units with learning over 50 epochs that had samples for specific hand actions whose activation value was the greatest. (White units are those units that had no samples for the specific hand actions that had the greatest activation.)

Table 11. Percentages for the hand training and test samples that were located within 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

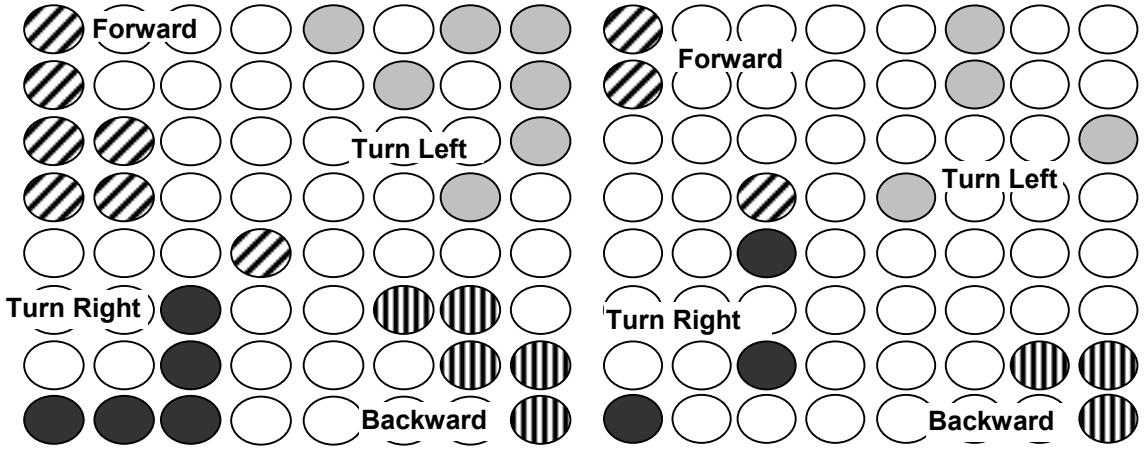


Training Data Test Data

Figure 7. The units on the training and test data for a network with 8 by 8 units with learning over 150 epochs that had samples for specific head actions whose activation value was the greatest. (White units are those units that had no samples for the specific head actions that had the greatest activation.)

Table 12. Percentages for the head training and test samples that were located within 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 Up	100	100	100



Training Data Test Data

Figure 8. The units on the training and test data for a network with 8 by 8 units with learning over 100 epochs that had samples for specific leg actions whose activation value was the greatest. (White units are those units that had no samples for the specific leg actions that had the greatest activation.)

Table 13. Percentages for the leg training and test samples that were located within 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

7. Discussion of our modular self-organising approach and related research

In this paper we have shown that the modular self-organising approach associates actions with instructions. Despite this only being one major step in the MIRA robot's overall development into an interactive robot, the architecture was designed to take account of future developments. For instance, it is possible to simply include additional modules to perform different functions based on different neural learning algorithms into the model. These learning algorithm modules could be used to create the actions based

on the original sensor reading inputs and perform object recognition, navigation etc. This combination of diverse learning approaches to perform goal-directed behaviour in a modular manner is found in the brain. According to Doya 1999 the cerebellum incorporates mainly supervised learning, the basal ganglia mainly reinforcement learning and the cerebral cortex mainly unsupervised learning. The supervised learning of the cerebellum provides an internal model of the world; the reinforcement learning of the basal ganglia considers the current state and selects the appropriate behaviour based on this; and finally the unsupervised learning of the cerebral cortex represents the external world and the internal context. For instance, it is envisaged that the creation of the actions based on the sensor reading inputs will use an auto-associative learning approach to relate the inputs with the conditions of the motors at a specific time phase.

Our modular self-organising approach also offers the benefit of robustness since if one of the body part self-organising maps is damaged, then certain functionality of the model will persist. Hence, if the leg self-organising network is damaged the approach can still correctly identify hand and head-related actions and get the robot to perform them. In a similar manner the distributed architecture enables the brain to continue to perform most cognitive functions even if a restricted cortex region is damaged. For instance in language processing, subjects with damage to Broca's area of the brain lack certain capabilities such as the creation of fluent spoken responses, action naming, understanding of reversible sentences, performance of certain verb operations and understanding of 'which-questions' (Purves 1997, Marshall et al. 1998, Brendt and Caramazza 1999). However, they can comprehend language, deal with non-reversible sentences, perform object and verb recognition and comprehension, understand 'who-questions' and identify semantic and verb errors with little difficulty.

The modular self-organising approach enables the introduction of a neural multi-modal approach that can associate actions, language and vision. For instance the robot could produce the appropriate action if it sees the human operator performing the action. Furthermore, the observation of a particular object such as a cup could create the association actions that could be performed on such an object such as 'pick' or 'put' or statements such as 'That is a cup, I can pick it up'. What constitutes the subactions that make up the action could depend on the current context.

Related work to our model is that of Bailey 1995 and Bailey et al. 1998 who investigate the neurally plausible grounding of action verbs in motor actions, such that an agent could execute the action of a verb it has learned. They develop a system called Verb Learn that could learn motor-action prototypes for verbs such as 'slide' or 'push' that allows both recognition and execution of a learned verb. Verb Learn learns from examples of verb word/action pairs and employs Bayesian Model Merging to accommodate for different verb senses where representations of prototypical motor-actions for a verb are created or merged according to a minimum length description criterion. Bailey's work differs from ours in that verbs are mapped to motor actions in a simulated environment. Furthermore, Verb Learn is able to cope with multiple word inputs and multiple senses of an action. Nevertheless, our approach makes use of a robot rather than an animator and uses the actual sensor readings of the robot to develop representations for actions.

Further research that is related to our modular self-organising approach includes that of McGuire et al. 2002 who develop a robot to perform grasping operations based on

language, gestures and vision through interactive demonstrations. In this architecture the speech processing and attention mechanism provide the language, vision and gesture inputs that converge in an integration module. The attention system uses a spatial organisation of visual clues and a layered structure of neural networks for combining low-level feature vectors into a focus of attention for the camera. Once the system decides on the object to manipulate, control moves to the robot arm that performs the grasping action using a finite state automata based on inputs from a wrist camera and fingertip sensors. Current advantages of this approach over ours are that it connects vision and gestures to actions and deals with spontaneous speech in a form of dialog rather than a single word action instruction. However, our approach by combining language and multiple sensor readings offers multi-modality and a greater indication of the robot internal state than McGuire et al.'s approach that relies on combining language, vision and fingertip sensor readings. It should be noted that vision is to be incorporated into our approach in the future. Moreover, our approach incorporates more diverse body-part-related actions by using modular self-organising learning that is less dependent on supervised feedback to activate switches between subactions.

Further related research is the intelligent ASIMO robot of Sakagami et al. 2002 that also combines multi-modal vision and language processing to perform navigation and human interaction. In this system there are algorithms for face recognition and Bayesian statistics for gesture recognition. A map of the robot's environment is produced based on visual inputs that the robot uses to follow a pre-defined route. This approach also relates language and vision to actions, but does not make use of the internal sensor readings from the robot. The ASIMO robot can perform actions that appear more human like than our approach by recreating the behaviour of a receptionist. In doing so it locates a human's face, greets a visitor, checks whether a person has an appointment and takes the visitor to the appropriate room. However, ASIMO is restricted mainly to leg related actions to navigate around its environment, with our approach offering actions associated with three body parts using the neurocognitive findings by Pulvermüller 2003. Although our approach is based on unsupervised learning the ASIMO robot makes little use of learning by using a fixed map of its environment, pre-defined routes and a lookup database of faces. As Sakagami et al.'s approach does not rely on learning it can produce behaviour that appears complex and intelligent but may be very brittle. For instance if a person visits who is not in the database or the furniture is rearranged in the office the performance of the robot will suffer.

Roy 1999 and Roy and Pentland 2002 develop a robotic system called CELL that equipped with a camera and microphone, could derive a lexicon of shape and colour words from speech input. CELL can learn associations to objects seen by a camera, and thereby ground the words in sensory data. Similarly, we intend to extend MIRA's learned association capacity to combine vision, language and motor actions.

In the future it is anticipated that the MIRA robot will offer more intuitive and user-friendly spoken interactions (Prodanov et al. 2002). Currently the approach relies on a microphone and speakers to allow human-robot communication. Although such an approach can achieve recognition rates of over 80% and is more natural than robots that rely on typed inputs (Rhino robot, Burgard et al. 2000), spoken communication in humans relies on facial expressions. As noted by Miwa et al. 2002 a robot head is critical for successful spoken communication between humans and robots. Miwa et al. 2002 in

their WE-4 robot create six expressions of emotion by altering the lips, jaw, face colour and voice. The ability of a robot to interact with an untrained person and learn from this interaction may also be important for the performance of robots in general (Breazeal 2003).

8. Conclusion

A model was described for robot control that uses the concept that different neural regions process action verbs based on their relation with appropriate body parts. We are not claiming that the clustering on the self-organising networks are identical to the somatotopic organisation found in the motor system of the cortex. However, by using modular self-organising networks for each of the three body parts considered it was possible to cluster and identify the individual actions. The neurocognitive approach by Pulvermüller states that assemblies in different regions represent a word based on the semantic input associated with that word. In our experiment based on the semantic inputs from various modalities action verbs were represented based on common features that related them to the relevant body part and diverse features.

We described a self-organising approach that controls a robot using language instructions. We moved from an approach that used the semantic features of action verb meaning to directly using sensor readings of actions to represent low level semantic features. This approach used sensor readings as the input to the robot and also as the basis for the robot's behaviour. We believe that such a self-organising language memory has a lot of potential for environmentally grounded robots in the future.

Acknowledgement

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 who helped with the sensor reading acquisition on the robot. This paper is a substantially extended version of Elshaw, Wermter and Watt 2003 a six-page conference paper published at the International Joint Conference on Neural Networks 2003.

References

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

Bailey, D., Chang, N., Feldman, J. and Narayanan, S. 1998, Extending embodied lexical development. *Proceedings of the Nineteenth Annual Meeting of the Cognitive Science Conference*.

Bailey, D. 1995, Getting a grip: A computational model of the acquisition of verb semantics for hand actions. *International Cognitive Linguistics Association Conference, (Albuquerque, USA)*.

Bloom, L., Hood, L. and Lightbown, L. 1974, Imitation in language development: If, when and why. **Cognitive Psychology**, 6, 380-420.

Bloom, L., Lightbown L. and Hood, L. 1975, Structure and variation in child language. **Monographs of the Society for Research in Child Development**, 40 (Serial No. 160).

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

Breazeal, C. 2003, Towards social robots. **Robotics and Autonomous Systems**, 1040, 1-9.

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

Brendt, R. and Caramazza, A. 1999, How 'regular' is sentence comprehension in Broca's aphasia? It depends on how you select the patients. **Brain and Language**, 64(2), 231-256.

Bryson, J. and Stein, L. 2001, Modularity and specialized learning: Mapping between agent architectures and brain organization. *In Emergent Neural Computational Architectures based on Neuroscience*, edited by Wermter, S., Austin, J. and Willshaw, D., (Heidelberg, Germany: Springer-Verlag), pp. 93-113.

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

Calabretta, R., Nolfi, S., Parisi, D. and Wagner, P. 1998, Emergence of functional modularity in robots. *In Proceedings of Artificial Life VI, (Los Angeles, USA)*, edited by Adami, C., Belew, R., Kitano, H. and Taylor, C., (Cambridge, USA: MIT Press), pp. 68-82.

Dodel, S., Herrmann, M. and Geisel, T. 2001, Stimulus-independent data analysis for fMRI. *In Emergent Neural Computational Architectures based on Neuroscience*, edited by Wermter, S., Austin, J. and Willshaw, D., (Heidelberg, Germany: Springer-Verlag), pp. 39-52.

Dogil, G., Ackermann, H., Grodd, W., Haider, H., Kamp, H., Mayer, J., Riecker A. and 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.

Doya, K. 1999, What are the computations of the cerebellum, the basal ganglia, and the cerebral cortex. **Neural Networks**, 12, 961-969.

Elshaw, M., Wermter, S. and Watt, P. 2003, Self-organisation of language instruction for robot action. *Proceedings of the International Joint Conference on Neural Networks, (Oregon, USA)*.

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

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

Hebb, D. 1949, *The Organization of Behaviour*, (New York: Wiley).

Hicks, J. and Monaghan, P. 2001, Explorations of the interaction between split processing and stimulus types. *In Emergent Neural Computational Architectures based on Neuroscience*, edited by Wermter, S., Austin, J. and Willshaw, D., (Heidelberg, Germany: Springer-Verlag), pp. 83-98,

Huyck, C. 2001, Cell assemblies as an intermediate level model of cognition. *In Emergent Neural Computational Architectures based on Neuroscience*, edited by Wermter, S., Austin, J. and Willshaw, D., (Heidelberg, Germany: Springer-Verlag), pp. 383-397.

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

Knoblauch, A. and Palm, G. 2001, Spiking associative memory and scene segmentation by synchronization of cortical activity. *In Emergent Neural Computational Architectures based on Neuroscience*, edited by Wermter, S., Austin, J. and Willshaw, D., (Heidelberg, Germany: Springer-Verlag), pp. 407-427.

Kohonen, T. 1997, Self-organizing Maps. (Heidelberg, Germany: Springer Verlag).

MacWhinney, B. 1995, *The CHILDES Project*, (Lawrence Erlbaum Associates), pp. 272-331.

Marshall, J., Pring, T., and Chait, S. 1998, Verb retrieval and sentence production in aphasia. **Brain and Language**, 63(2), 159-183.

McClelland, J. and Kawamoto, A. 1986, Mechanisms of sentence processing: Assigning roles to constituents of sentences. *In Parallel Distributed Processing Vol. 2*, edited by McClelland, J. and Rumelhart, D., (Cambridge, USA: MIT Press), pp. 272-331.

McGuire, P., Fritsch, J., Steil, J., Röthling, F., Fink, G., Wachsmuth, S., Sagerer, G. and Ritter, H. 2002, Multi-modal human-machine communication for instructing robot grasping tasks. *Proceedings of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems, (Lausanne, Switzerland)*, pp. 1082-1088.

Miwa, H., Okuchi, T., Takanobu, H. and Takanishi, A. 2002, Development of a new human-like head robot WE-4. *Proceeding of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems (Lausanne, Switzerland)*, pp. 2443-2448.

Nehmzow, U. 1999, Meaning through clustering by self-organisation of spatial and temporal information. *In Computation for metaphors, analogy and agents*, edited by Nehaniv, C. (Heidelberg, Germany: Springer-Verlag), pp. 209-299.

Nolfi, S. 1997, Using emergent modularity to develop control systems for mobile robots. **Adaptive Behavior**, 5, 343-363.

Oka, T., Tashiro, J. and Takase, K. 1998, Data-focused parallel modular software design for a communicating autonomous mobile robot. *Proc. of 6th International Symposium on Intelligent Robotic Systems (SIRS98)*, pp. 37-46.

Owen, C. and Nehmzow. U. 1996, Route learning in mobile robots through self-organising. *Proceedings of EuroBot 96, (Kaiserslautern, Germany), October 9-11*.

Prodanov, P., Drygajlo, A., Ramel, G., Meisser, M. and Siegwart, R. 2002, Voice enabled interface for interactive tour-guide robots. *Proceeding of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems (Lausanne, Switzerland)*, pp. 1332-1337.

Pulvermüller, F. 1999, Words in the Brain's Language. **Cognitive Neuroscience**, 22(2), 253-336.

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

Pulvermüller, F. 2001, Brain reflections of words and their meaning. **Trends in Cognitive Science**, 5(12), 517-524.

Pulvermüller, F., Härle, M. and Hummel, F. 2001, Walking or talking? Behavioral and neurophysical correlates of action verbs processing. **Brain and Language**, 78, 143-168.

Pulvermüller, F., Härle, M. and Hummel, F. 2000, Neurophysiological distinction of verb categories. **Cognitive Neuroscience**, 11(12), 2789-2793.

Pulvermüller, F., Mohr, B. and Schleichert, H. 1999, Semantic or lexico-syntactic factors: what determines word class specific activity in the human brain? **Neuroscience Letters**, 275, 81-84.

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

Reggia, J., Shkuro, Y. and Shevtsova, N. 2001, Computational investigation of hemispheric specialization and interactions. *In Emergent Neural Computational Architectures based on Neuroscience*, edited by Wermter, S., Austin, J. and Willshaw, D., (Heidelberg, Germany: Springer-Verlag), pp. 68-82.

Reilly, R. 2001, Collaborative cell assemblies: Building blocks of cortical computation. *In Emergent Neural Computational Architectures based on Neuroscience*, edited by Wermter, S., Austin, J. and Willshaw, D., (Heidelberg, Germany: Springer-Verlag), pp. 161-173.

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

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

Roy, D. 2002, Learning words and syntax for a visual description task. **Computer Speech and Language,** 16(3).

Roy, D. and Pentland, A. 2002, Learning words from sights and sounds: A computational model. **Cognitive Science**, 26(1), 113-146.

Sakagami, Y., Watanabe, R., Aoyama, C., Matsunaga, S., Higaki, N. and Fujimura, K. 2002, The intelligent ASIMO: System overview and integration. *Proceeding of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems (Lausanne, Switzerland)*, pp. 2478-2483.

Spitzer, M. 1999, *The Mind within the Net: Models of Learning, Thinking and Acting*, (Cambridge, MA: MIT Press).

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

Tani, J. and Fukumura, N. 1997, Self-organizing internal representation in learning of navigation: A physical experiment by the mobile robot YAMABICO. **Neural Networks**, 10(1), 153-159.

Taylor, J. 2001, Images of the mind: Brain images and neural networks. In *Emergent Neural Computational Architectures based on Neuroscience*, edited by Wermter, S., Austin, J. and Willshaw, D., (Heidelberg, Germany: Springer-Verlag), pp. 20-38.

Treves, A. and Roll, E. 1994, Computational analysis of the role of the hippocampus in memory. **Hippocampus**, 4(3), 374-391.

Voegtlin, T, and Verschure, P. 1999, What can robots tell us about brains? A synthetic approach towards the study of learning and problem solving. **Reviews in Neuroscience**, 10(3-4), 291-310.

Wermter, S., Austin, J., Willshaw, D. and Elshaw, M. 2001a, Towards novel neurally-inspired computing. *In Emergent Neural Computational Architectures based on Neuroscience*, edited by Wermter, S., Austin, J. and Willshaw, D., (Heidelberg, Germany: Springer-Verlag), pp. 1-19.

Wermter, S., Austin, J. and Willshaw, D. 2001b, *Emergent Neural Computational Architectures based on Neuroscience*, (Heidelberg, Germany: Springer-Verlag).

Wermter, S. and Panchev, C. 2002, Hybrid preference machines based on inspiration from neuroscience. **Cognitive Systems Research**, 3(2), 255-270.