A Neurocognitive Approach to Self-organisation of Verb Actions

Mark Elshaw and Stefan Wermter
Hybrid Intelligent Systems Group
School of Computing, Engineering and Technology
University of Sunderland
St Peter's Way, Sunderland, SR6 0DD, United Kingdom
http://www.his.sunderland.ac.uk
[Mark.Elshaw, Stefan.Wermter]@Sunderland.ac.uk

Abstract - Our approach integrates some neuroscience evidence for regional distributed modularity with cognitive evidence for action verb processing. The long-term goal is to provide a computational model for a neurally inspired self-organising robot action control system. Regional modularity in the brain involves distributed neural networks in diverse regions processing in a parallel manner. According to neurocognitive findings, action verbs are processed by associating specific cell assemblies with the appropriate body part. We present a model of the neurocognitive findings using self-organising networks that cluster action verbs into different locations of the network dependent on the body part they are associated with.

I. INTRODUCTION

Present approaches for computing have been able to perform many data processing activities but have been less successful when used to perform higher cognitive functions such as language processing. In language processing there is a need for a significant break from previous approaches that use passive, monolithic, static architectures [1, 8].

Regional distributed modularity in the brain is based on various distributed neural networks in diverse regions which carry out processing in a parallel distributed manner to perform specific cognitive functions [11]. The brain performs as a group of collaborating networks, none of which can deal with a complex task alone. In the brain it is possible to deal with complex processing using parallel distributed modularity [13, 14].

Brain imaging techniques have provided a great deal of information on regional modularity. The four main brain imaging techniques are positron emission tomography (PET), functional magnetic resonance imaging (fMRI), magnetoencephalogram (MEG) and electroencephalogram (EEG). PET and fMRI identify the regions with the greatest activation by considering blood flow and blood oxygen levels, respectively [5]. 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 establish brain activity using sophisticated superconducting quantum devices [5, 12].

These techniques outline the significant extent of distributed regional modularity organisation of language processing in the brain. Traditionally, there is an early model of language processing based on two modules linked via the arculate fasciculus [2]. The first module identified by Broca is used to create lucid speech and is found in the frontal left lobe of the cerebral cortex. The second module discovered by Wernicke is on the superior surface of the temporal lobe and is responsible for the grammatical correctness of language.

Since the development of this early model the number of additional brain regions that are held to be involved in language has increased considerably. Binder et al., 1997 [3] when requiring individuals to do a task that involved speech comprehension and recollection of information identified four principal regions in the left hemisphere of the cerebral cortex: i) a lateral and ventral temporal lobe area that incorporated the superior temporal sulcus, the middle temporal gyrus and parts of the inferior temporal gyrus; ii) sections of the inferior and superior frontal gyri, the middle frontal gyrus and the anterior cingulate; iii) angular gyrus; and iv) a section incorporating the posterior cingulate and gyrus areas.

Our proposed approach is based on neurocognitive evidence of Pulvermüller et al. on the processing of action verbs as its inspiration. Out of the various factors that are identified as fundamental in the brain we will concentrate here on regional modularity organization [13, 14]. By combining these two areas a new self-organising architecture for language processing for robot action is outlined.

II. NEUROCOGNITIVE EVIDENCE

First, we consider the neurocognitive evidence of Pulvermüller, 1999 [9] and Pulvermüller et al., 2000 [10] with particular reference to the processing of action verbs using Hebbian learning, synfire chains and meaning/semantics. This evidence offers a basis for our approach.

According to Pulvermüller et al. Hebbian learning is a model of how the brain stores memories through a simple synaptic approach based on cell assemblies for cortical processing. What is critical for Hebbian learning is (i) coactivated neurons are associated; (ii) associations can be between neurons that are close or far a part; (iii) if neurons become associated they become known as a cell assembly.

Spatiotemporal firing patterns are observed in cortical neurons which are known as synfire chains. This involves a cell assembly activating the next one to keep the chain going and hence representing the word. The assemblies are not only found to connect to the next cell assembly in the chain but skip to other cell assemblies. There is currently agreement that neurons in distant cortical regions perform together as functional elements and that there is stimulus specific cell assemblies.

When looking at the cell assemblies which process particular word types, Pulvermüller [9] noted that activation was found in both hemispheres of the brain for content words such as verbs and nouns. Action words that involve moving one's own body are represented by the perisylvian cell assembly and assemblies in 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. From the position of Hebbian learning, it is important to relate the neurons representing the word form with those neurons associated with its perception and actions reflecting elements of the meaning/semantic information. For content words the semantic factors that influence the cell assemblies come from various modalities and include the complexity of the activity performed, the type and number of muscles used, the colour of the stimulus, the tool used, the smell or taste of the object, and whether the person can see themselves doing this activity.

In order to support his findings on Hebbian learning and synfire chains Pulvermüller et al., 2000 examined the processing of action verbs. They hypothesized that actions verbs which are associated with for instance the leg are processed by cell assemblies that are distinct from some of those associated with the face. This involved an experiment using 20 students who were provided with leg-, arm- and face-related action verbs and pseudo-words and required to indicate by pressing a button whether the stimulus was a word or a pseudo-word. When performing this operation brain activity was recorded by using EEG electrodes positioned near the scalp.

This study found that the average response times for lexical decisions was faster for face-associated words than for arm-associated words and the arm-associated were faster than leg ones. There was a significant difference for the prefrontal and occipital regions and above the motor and premotor cortex. The prefrontal area was found to be associated mainly with arm-related verbs and the occipital visual areas for face-related verbs. Hence the brain processes action verbs by activating cell assemblies that are associated through meaning/semantic information with the appropriate body part.

III. SELF-ORGANISING NETWORK

Our approach uses self-organising networks that offer an unsupervised associative memory approach related to that of the cerebral cortex. Self-organising networks consist of an input and an output layer, with every input neuron linked to all the neurons in the output layer [4, 6]. The output layer creates a topographical representation of the critical characteristics of the different inputs by creating a pattern of active and inactive units (see Figure I) and cluster similar inputs together.

A typical self-organising network algorithm has 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

$$w_j = [w_{j1}, w_{j2}, ..., w_{jn}]$$

$$(1)$$

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

$$o_j = \sum w_{jk} i_k = w_j \bullet i$$

$$(2)$$

The weights are initalised 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:

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

$$(3)$$

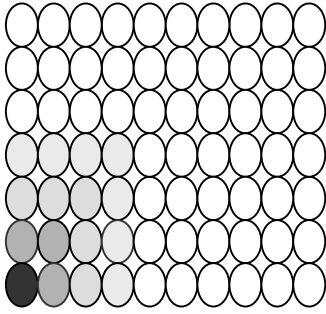where α is the learning rate parameter that is usually set between 0.2 and 0.5.

Fig I   A representation of the activity output layer of a self-organising network - The darker the neuron the greater the activation.

IV. SELF-ORGANISING EXPERIMENT

The first step in developing our model for verb actions was to ensure that self-organising networks are suitable for modeling the findings of Pulvermüller et al. with regards to the association of action verbs with the appropriate body part. On the other hand we wanted to use simple but real language material.

A. Experimental Method

We intended to use actual spoken language that involved a degree of training to perform actions. Hence, a data set was extracted from the transcripts of the dialog between a child (Peter) and adults from the Childes Corpus. This corpus is an on-line database of utterances by children of different ages (http://childes.psy.cmu.edu/).

The actual data set that was used consisted of questions/requests and responses between a child and adults. Interactions were extracted if they included action verbs that were associated with the hand (Get, Put, Pull, Push, Write, Draw), head (See, Eat, Drink, Sing, Ask) or leg (Stand, Walk, Come, Ride).

From these interactions the utterances containing the action verb were taken and a six word section that contained the action verb and the main focus of the phrase extracted. Then the action verb and the first two nouns were extracted to produce a three word phrase.

A representation was devised based on semantic questions and responses about the action verbs and nouns so that the three word phrase could be input for the self-organising neural networks [7]. The full set of questions for the actions verbs and nouns and responses are shown in Tables I and II.

As the input to the neural networks was in numeric form, the responses for the action verbs and nouns were represented with numeric values. For instance, for action verbs and nouns Small was represented by 01, Medium 10,

Large 11, Yes is 1, No 0. Based on this approach "Walk" was represented as 1101110001001, "Eat" – 0101010001011 and "Put" – 1011101001010. The noun "Peter" for example was represented as 101011 (human, non-edible, living, not a toy and solid). When there were not two nouns the missing ones were represented as 000000.

TABLE I
SEMANTIC REPRESENTATION OF ACTION VERBS

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 II
SEMANTIC REPRESENTATION OF NOUNS

Semantic Information	Possible 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

B. Unsupervised Learning

In the experiment the input layer to the self-organising networks was set at 25 units, one for each of the digits for the three action verb phrases, the output layers had various sizes (7 by 7 units, 10 by 10 and 12 by 12) and they were trained for 25, 50, 75 and 100 epochs. The location of each of the training phrases on the self-organising output layers was identified based on the units that had the highest activation. The trained network was then tested for its ability to generalise by identifying the location on the output layer of the unseen action verb phrases from the test set.

As the number of training and test samples were different between the action verbs (see Table III) and for each body part (see Table IV) percentage values rather than numbers were examined. These were the percentages for a specific action verb and the action verbs associated with a specific body part that produce the highest activation for a network unit. To remove the impact of outliers only those units that contained 10% of specific action verb and 5% of the specific body part action verbs were considered.

TABLE III
SAMPLE NUMBERS FOR ACTION VERBS

Verbs	Training	Test	Verbs	Training	Test
Get	50	27	Drink	6	5
Pull	152	74	Sing	18	11
Push	11	5	Ask	22	13
Write	58	27	Walk	27	14
Draw	13	10	Stand	27	16
See	56	32	Come	42	23
Eat	32	18	Ride	11	9

TABLE IV
SAMPLE NUMBERS BY BODY PART

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

V. RESULTS AND DISCUSSION

Initially when considering a 7 by 7 unit self-organising network there were no clear clusters for the three body parts when considering percentages for the individual action verbs and for the body parts. This indicated that the network output layer was too small to allow a clear split between the action verb phrases into the three body parts.

However, for the network with 10 by 10 units and so less restricted memory at a training time of 75 epochs there was clear clustering into the body parts (See Figures II to V). For instance, when looking at Figures II there were action verbs such as "Push", "Pull" and "Draw" etc. in the top left hand corner and clustered below this were action verbs such as "Drink", "Eat" and "See" that relate to "Hand" and "Head" areas respectively in Figure IV.

There was a split in the head action verbs with action verbs such as "Eat", "Drink" and "See" grouped together in one location on the output layers and communication-based head action verbs such as "Sing" and "Ask" in another. Given the diversity in the function performed between these action verbs it is likely that there is the same split to a certain degree between the cell assemblies that process these action verbs in the brain. It was also found that for one of the units in the "leg" region on the 10 by 10 units network for both test and training data that it had the highest activation for over 5% of both of the "leg" and "head" action verbs. Nevertheless, "leg" had the highest percentage value for both sets of data, with it being 20% greater on the training data.

Hence a network of this size can in principle realises the partitioning of the processing of action verbs with different clusters representing the specific body parts. For such an architecture on both training and test data the clusters were in very similar position on the output layer, which points to the ability of the network to generalise on data it has not seen before. Therefore, if the self-organising network was used in the control of a robot it may perform successfully in an on-line manner associating language inputs with the appropriate region of the output layer.

Turning to 12 by 12 units output layer network architecture there were not the clear clusters for the different body parts that were observed for the 10 by 10 units architecture. Two clusters were typically found on the 12 by 12 output layer. The top one contained action verbs for all three body part verbs and the lower one the "head" and "hand" action verbs. This indicated in this case that the output layer memory was too large.

		Pull Push Get							
								Sing	
Get			Push Pull Get		Draw				
			Pull	Draw Write	Write				
		Eat	Eat	Come Walk					
		Drink Eat See	Drink See	See				Ride	
				Sing					
						Come Walk	Ride Come Walk See	Ask Sing	
				Sing				Ride Stand	
								Stand	

Fig II The units on the training data for a 10 by 10 network with training time of 75 epochs that have 10% or more for specific action verb phrases whose activation is the greatest.

	Ride	Push Pull Get						Ride	
		Push						Sing	
		Pull		Push Pull Get	Draw				
				Pull		Draw			
			Draw	Draw Walk	Write				
		Drink Eat	Drink	Drink Eat				Ride	
				See Ask					
					Walk Come Stand	Ride Come Stand	Walk Ask Sing		
						See Ride Stand			Drink
							Sing		

Fig III The units on the test data for a 10 by 10 network with training time of 75 epochs that have 10% or more for specific action verb phrases whose activation is the greatest.
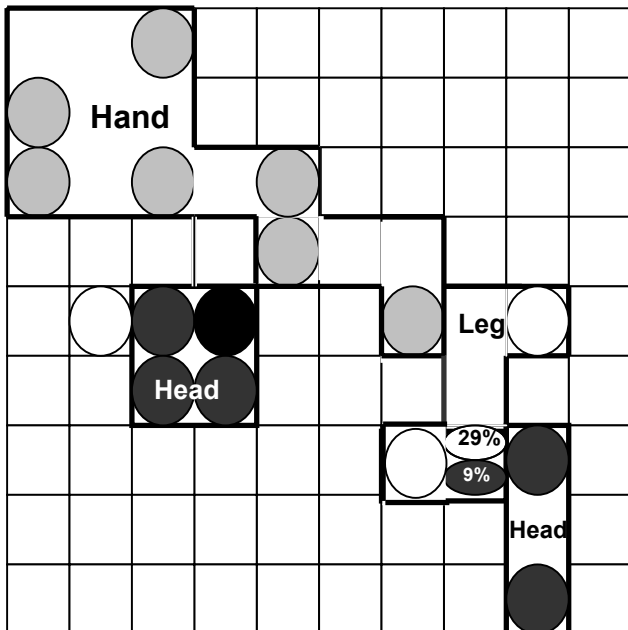

Fig IV The units on the training data for a 10 by 10 units network with learning over 75 epochs that have 5% or more for specific body parts whose activation value is the greatest.
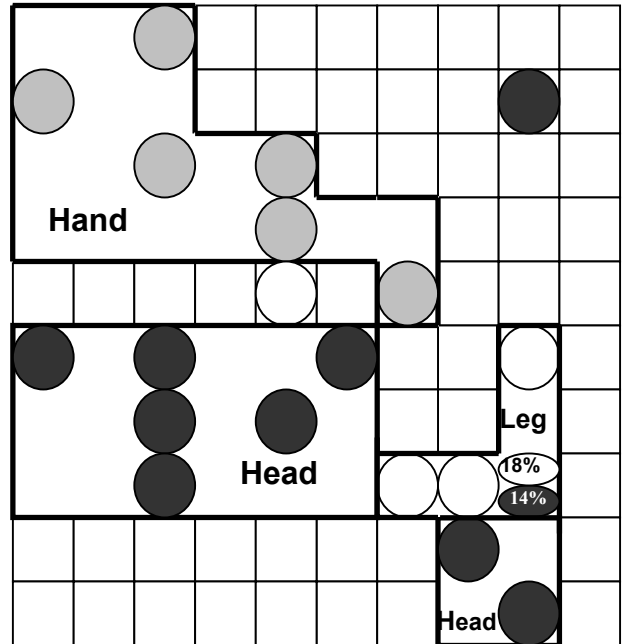

Fig V The units on the test data for a 10 by 10 units network with learning over 75 epochs that have 5% or more for specific body parts whose activation value is the greatest.

V1. FUTURE WORK

The experiment performed supports that self-organising networks cluster action words with the appropriate body part. Self-organising networks seem suitable for incorporation into a language processing system to control robot action that combines brain-inspired modularity with the neurocognitive evidence of Pulvermüller on action verb processing. This is to be achieved by a system's hierarchy structure of self-organising networks that perform an increasingly finer classification for an appropriate robot action.

An upper level self-organising network maybe used to associate the action verbs with the appropriate body part by clustering the verbs in different regions of the output layer. In the next processing stage the hierarchy involves a self-organising network for each of the body parts and different meaning/semantic input vector that enable the clustering of action verbs such as for hand movement words. The final level of the system is to have a network for each of the clusters in the previous layer and to have a new meaning/semantic input that can classify the action verb.

This approach offers some brain-inspired regional modularity and takes into account some neurocognitive evidence. The former is achieved by having multiple self-organising networks each performing a subtask of the overall task. However, they are all connected in a distributed overall memory organisation. Neurocognitive

evidence of Pulvermüller et al. is also considered in that cell assemblies in different regions are associated with specific action verbs as functional unit, with the association being best on the action verb's relationship with the appropriate body part.

VII. CONCLUSION

We have described an initial model of distributed regional modularity of the brain based on some neurocognitive evidence. This evidence suggests that cell assemblies in different regions of the brain are used to process action verbs based on their association with appropriate body parts. The association between the action verb and body part is due to the meaning/semantic information used to describe the verb in the learning process. Through a systematic self-organisation network approach it was possible to confirm this clustering on action verbs processing. In the future we will devise a hierarchical self-organising structure to control a robot based on this distributed regional modularity and neurocognitive evidence.

References

[1]   Abney, S., Measures and Models for Phrase Recognition, *Proceedings Workshop on Human Language Technology*, pp. 233-236, 1993.

[2]   Bear, M., Connors, B. and Paradiso, M., *Neuroscience: Exploring the Brain*, 1996.

[3]   Binder, J., Frost, J., Hammeke, T., Cox, R., Rao, S., and Prieto, T., Human Brain Language Areas Identified by Functional Magnetic Resonance Images, *The Journal of Neuroscience*, Vol. 17, No. 1, pp. 280-288, 1997.

[4]   Doya, K., What are the Computations of the Cerebellum, the Basal Ganglia and the Cerebral Cortex? *Neural Networks*, Vol. 12, No. 7-8, pp. 961-974, 1999.

[5]   Gazzaniga, M., Ivry, R. and Mangun, G., *Cognitive Neuroscience: The Biology of the Mind*, W.W. Norton & Company Ltd, 1998.

[6]   Kohonen, T., *Self-Organizing Maps*, Springer-Verlag, Heidelberg, Germany, 1997.

[7]   McCelland, J. and Kawamoto, A., Mechanisms of Sentence Processing: Assigning Roles to Constituents of Sentences, In McCelland, J. and Rumbelhart, D. editors, *Parallel Distributed Processing Vol. 2*, MIT Press, Cambridge, USA, pp. 272-331, 1986.

[8]   Magerman, D. M., Statistical Language Learning: Review, *Computational Linguistic*, 1995.

[9]   Pulvermüller, F., Words in the Brain's Language, *Behavioral and Brain Sciences*, Vol. 22, No. 2, pp. 253-336, 1999.

[10]   Pulvermüller, F., Hare, M. and Hummel, F., Neurophysiological Distinction of Verb Categories, *Cognitive Neuroscience*, Vol. 11, No. 12, pp. 2789-2793, 2000.

[11]   Reggia, J., Shkuro, Y., and Shevtsova, N., Computational Investigation of Hemispheric Specialization and Interactions, In Wermter, S., Austin, J. and Willshaw, D. editors, Emergent *Neural Computational Architectures based on Neuroscience*, Springer-Verlag, Heidelberg, Germany, pp. 68-82, 2001.

[12]   Taylor, J., Images of the Mind: Brain Images and Neural Networks, In Wermter, S., Austin, J. and Willshaw, D. editors, Emergent *Neural Computational Architectures based on Neuroscience*, Springer-Verlag, Heidelberg, Germany, pp. 20-38, 2001.

[13]   Wermter, S., Austin, J. and Willshaw, D., *Emergent Neural Computational Architectures based on Neuroscience*, Springer-Verlag, Heidelberg, Germany 2001.

[14]   Wermter, S., Austin, J., Willshaw, D., and Elshaw, M., Towards Novel Neuroscience-inspired Computing, In Wermter, S., Austin, J. and Willshaw, D. editors, *Emergent Neural Computational Architectures based on Neuroscience*, Springer-Verlag, Heidelberg, Germany, pp. 1-19, 2001.