A Mirror Neuron Inspired Hierarchical Network for Action Selection

Mark Elshaw, Cornelius Weber, Alex Zochios, Stefan Wermter
Hybrid Intelligent Systems
School of Computing and Technology
University of Sunderland, UK
[Mark.Elshaw,Cornelius.Weber,Stefan.Wermter]@sunderland.ac.uk

Abstract

In this paper we propose an approach to robot learning by imitation that uses the multi-modal inputs of language instruction, vision and motor. In our approach a student robot learns from a teacher robot how to perform three separate behaviours, 'pick', 'lift' and 'go' based on these inputs. We considered two neural architectures for performing this robot learning. First, a one-step architecture trained with two different learning approaches either based on Kohonen's self-organising map or based on the Helmholtz machine turns out to be inefficient or not capable of performing differentiated behaviour. In response we produced a hierarchical architecture that combines both learning approaches to overcome these problems. In doing so the proposed robot system models specific aspects of learning using concepts of the mirror neuron system [8] and the hierarchical organisation of the motor system with regards to demonstration learning.

1 Introduction

One learning approach that can make intelligent robots easy to use is imitation learning. Such learning allows the observer to gain skills by creating an abstract representation of the teacher's behaviour, and an understanding of the teacher's aims to produce the required solution [5]. There is growing interest in imitation learning as it offers a flexible way to programme robots by having the robot observe and imitate either another robot or a human.

Multimodal inputs are used in our robot learning model as it is only through the combination of language, vision and motor actions, that robots will be able to become service robots to benefit humans. By combining multimodal inputs service robots should adapt to changes in their environment and improve their decision-making. For instance, a robot performs grasping operations based on language, gestures and vision [12]. Language can be acquired by pairing words with raw multimodal sensory data [11]. A mirror neuron approach using multimodal inputs was applied to predictive behaviour perception and imitation [2]. Our approach incorporates a language element as input to the mirror neuron system to achieve learning by imitation.

Mirror neurons are a class of neurons in the F5 motor area of the monkey cortex which not only fire when performing an action but also when just seeing or hearing the action performed. They represent actions in an abstract sense so they are understood or can be imitated [9]. Mirror neurons in humans [4] have been associated with Broca's area which indicates their role in language development [8]. Their sensory property justifies our use of models designed for sensory systems such as the Helmholtz machine and the Kohonen algorithm.

The involvement of a hierarchical architecture as found in sensory or motor cortices is a necessary step to enhance robot learning to produce composed and selectable skills. Multi-modal inputs can be extended to the inclusion of reward values that are also represented in cortical structures [10]. This can allow in the future for goal driven, teleological behaviour based on models originally designed to account for cortical sensory systems.

2 Methods

A robot simulator was produced with a teacher robot performing 'go', 'pick' and 'lift' actions one after another in a loop in an environment (Fig. 1). The student robot observed the teacher robot performing the behaviours and was trained by receiving multimodal inputs. These multimodal inputs were *(i)* high-level visual inputs which were the x and y coordinates and the rotation angle φ of the teacher robot relative to the front wall, *(ii)* the motor directions of the robot ('forward', 'backward', 'turn left' and 'turn right') and *(iii)* a language description stating the behaviour the teacher is performing ('go', 'pick' or 'lift').

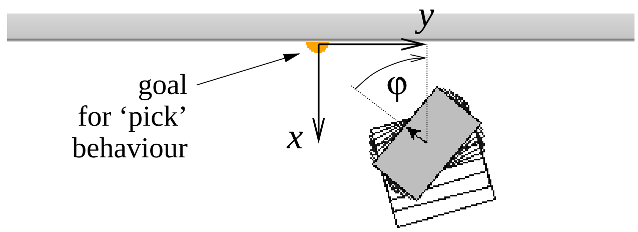

Figure 1: The simulated environment containing the robot at coordinates x, y and rotation angle φ. The robot has performed ten movement steps and currently turns away from the wall in the learnt 'go' behaviour.

The first behaviour, 'go', involves the robot moving forward in the environment until it reaches a wall and then turns away from it. The coordinates x and φ ensure that the robot avoids the wall, irrespective of y. The second behaviour, 'pick', involves the robot moving toward the target object depicted in Fig. 1 at the top of the arena. This "docking" procedure is produced by a reinforcement approach as described in [14] and uses all, x, y and φ coordinates. The final behaviour, 'lift', involves moving backward to leave the table and then turning around to face toward the middle of the arena. Coordinates x and φ determine how far to move backward and in which direction to turn around. These coordinates which are shared by teacher and learner are chosen such that they could be retrieved once the imitation system is implemented on a real robot.

When receiving the multimodal inputs corresponding to the teacher's actions the student robot was required to learn these behaviours so that it could recognise them in the future or perform them from a language instruction. Two neural architectures were considered.

2.1 Choice of Architecture

The first architecture depicted in Fig. 2 a) was run with two different learning algorithms. In [3] we have used a winner-take-all mechanism on the hidden area. This self-organising model is easy to handle, however, any hidden unit must "explain" all input modalities at once, i.e. if there are differences in the input in only one modality, then additional hidden units are needed. Thus different behaviours are represented on non-overlapping regions on the hidden area. If now, for example, one behaviour can be described by several different words, then it would have to be represented several times on the hidden area.

For a more efficient hidden representation, we considered a distributed code for the hidden area of Fig. 2 a) by using a Helmholtz machine learning algorithm rather than the winner-based self-organising map. Then a behaviour that is described by different words need not be entirely represented by separate units, but a small number of units might account for the word while others account for the visual-motor representation. As a result some units specialised to account for just one input modality. I.e. one unit's receptive field (RF) might be only in the language area while no connections are received from other input modalities, while another unit might make connections only with the motor area.

Our goal, however, is to account for motor output that differs with differing language input even if other sensory input is constant. Then the language input shall deliver the necessary bias to cause the different activation pattern. This bias needs to be situation-dependent, since behaviours differ in different situations, by activating different motor units. But we found that any language unit projects a certain input pattern onto the hidden units, i.e. it biases the hidden unit's activations dependent of the static connection pattern toward them. This bias was thus not situation-dependent, since the language area does not receive sensory input.

In response to these identified problems we will in the following concentrate on the architecture represented in Fig. 2 b). It associates the motor and high-level vision inputs using the first hidden layer, denoted HM area. The activations of the first hidden layer are then associated with the language region input at the second hidden layer, denoted SOM area. The first hidden layer uses a Helmholtz machine learning algorithm and the second hidden layer uses Kohonen's self-organising map algorithm. Such an architecture allows the features created on the Helmholtz machine hidden layer to relate a specific action for one of the three behaviours given the particular high-level visual information to "flexible" associations of pairs/patterns of activations on the hidden area.

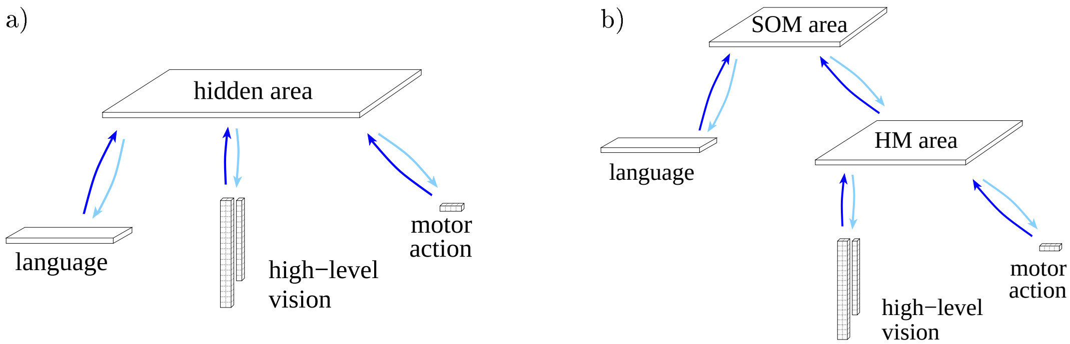

Figure 2: a) A single-step (3-to-1) architecture. b) A two-layer hierarchical architecture.

2.2 Learning Algorithm for the HM Area

The Helmholtz machine [1] generates representations of data using unsupervised learning. Bottom-up weights W^{bu}, depicted dark in Fig. 2 b), generate a hidden representation $\vec{r}$ of some input data $\vec{z}$. In the figure, $\vec{r}$ is the activation vector of the HM area units, $\vec{z}$ are the high-level vision and motor inputs, here formally concatenated to one vector. Conversely, top-down weights W^{td}, depicted light in Fig. 2 b), reconstruct an approximation $\tilde{z}$ of the data from the hidden representation. Both sets of weights are trained by the unsupervised wake-sleep algorithm which uses the local delta rule and which has two phases.

In the wake phase, a data point $\vec{z}$ is presented consisting of the motor and high-level visual components. The hidden representation is obtained as

$$\vec{r} = \vec{f}(W^{bu}\vec{z}), \quad \text{with } f_j(x_j) = \frac{e^{\beta x_j}}{e^{\beta x_j} + n}$$

where $\beta = 2$ and $n = 64$. These values are unchanged from a similar, visual network [13], however, variations would be possible. The reconstruction $\tilde{z} = W^{td}\vec{r}$ of the data is linear. The top-down weights w^{td}_{ij} from units j to units i are trained based on the difference between the original input data point and the reconstructed input data point from its internal representation:

$$\Delta w^{td}_{ij} = \eta\, r_j \cdot (z_i - \tilde{z}_i) \tag{1}$$

The learning rate was set to $\eta = 0.001$ and it was increased 5-fold whenever the teacher robot changed direction. This can be justified by networks of neurons that identify novel or significant behaviour to aid learning [6].

In the sleep phase an original binary topographic random internal representation $\vec{r}^s$ is projected onto the input as $\vec{z}^s = W^{td}\vec{r}^s$ and the reconstructed hidden representation $\tilde{r}^s$ is obtained from the linear input representation as $\tilde{r}^s = \vec{f}(W^{bu}\vec{z}^s)$, with the same activation function f as above. The bottom-up weights w^{bu}_{ji} from units i to units j are modified based on the difference between the original and the reconstructed random internal representation:

$$\Delta w^{bu}_{ji} = \varepsilon\,(r^s_j - \tilde{r}^s_j) \cdot z^s_i \tag{2}$$

with learning rate $\varepsilon = 0.01$. All weights W^{td} and W^{bu} were rectified to be non-negative. To ensure that weights did not grow too large a weight decay term of $-0.015 \cdot w^{td}_{ij}$ is added to Eq. 1 and $-0.015 \cdot w^{bu}_{ji}$ to Eq. 2.

Only the wake phases of training involve inputs from the motor and higher visual regions $\vec{z}$ based on observing the actions of the teacher robot performing the three behaviours.

2.3 Learning Algorithm for the SOM Area

The self-organising map algorithm [7] also learns internal representations of its training data. Its characteristic is that each single data point is represented by a single, winning unit on the hidden area. The network approximates a data point by this unit's weights.

The net input of unit k is established by determining the Euclidean distance of the weight vector to its inputs, given by $o_k = \|\vec{w}_k - \vec{i}\|$. The weights are originally randomised and hence one unit of the network will react more strongly than others to a specific input representation. This winning unit is the unit $k^{'}$ where $o_{k^{'}}$ is smallest. The activation of unit k is a Gaussian of the distance $d_{k,k'}$ to the winning unit $k^{'}$ on the SOM area grid:

$$T_{k,k'} = e^{-(d^2_{k,k'}/2\sigma^2)}$$

The weights are trained according to:

$$\Delta w_{kj} = \alpha \cdot T_{kk'} \cdot (i_j - w_{kj})$$

where we used a learning rate of $\alpha = 0.01$. Because of the activation function $T_{kk'}$ primarily the winning unit $k^{'}$ is trained, and depending on the variance σ^2 also its neighbours.

The hidden representation $\vec{o}$ is in our model the activation vector on the SOM area while its input data $\vec{i}$ is the concatenated vector from the language input together with the HM area activation $\vec{r}$. Only the bottom-up weights, depicted dark in Fig. 2 b), are trained. Top-down weights can formally be obtained from the bottom-up weights. Training of the SOM area weights was done after the HM area weights learning was completed.

At the beginning of training, a larger neighbourhood ($\sigma = 12$) achieved broad topologic learning. Following a fast linear reduction during training to $\sigma = 0.1$, finer training occurred with neighbourhood interaction widths σ reduced from 1 to 0.01.

2.4 Training and Testing Specifics

The multimodal inputs include first high-level vision representing the x and y coordinates and rotation angle φ of the robot. They are represented by two arrays of 36 units and one of 24 units, respectively. Fig. 3, left, shows an example. The centre of a broad Gaussian hill of activation denotes the corresponding coordinate as a neural population code.

The language region input is based on 46 English phonemes from the CELEX lexical databases (http://www.kun.nl/celex/). Each phoneme is represented by 20 features, which gives a different binary pattern of activation in the language input region for each phoneme with similar phonemes having similar structure. A region of 4 rows by 20 columns represents the words with each row representing one phoneme (Fig. 3, right top).

The robot motor directives are presented on the 4 motor units ('forward', 'backward', 'turn right' and 'turn left') with only one active at a time (Fig. 3, right bottom).

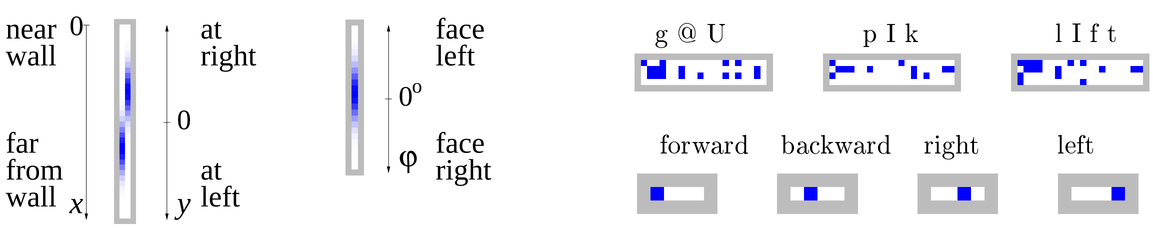

Figure 3: The network inputs. Left, example x, y and φ coordinate representations. Right top, the phonemes representing 'go', 'pick' and 'lift'. Right bottom, the motor activations.

To ensure that no input had a disproportional impact on the network as there was a significantly different size of the input regions, the motor regions input values were scaled to a value of 4 for an active unit. High-level vision inputs had a maximum value of 2. As the different word representations have different numbers of inputs, these vectors were scaled to the same length which produced ON-unit values between 2 and 2.3.

The size of the HM hidden layer is 32 by 32 units and the SOM layer has 24 by 24 units. The number of training examples was around 500000. The duration of a single behaviour depended on the initial conditions and may average at around 25 consecutive steps before the end condition (robot far from wall or target object reached) was met.

During training the student robot received all the inputs. When testing recognition of the behaviour that was performed the language input was omitted. Recognition was verified by comparing the units which are activated on the language area via W^{td} (depicted light in Fig. 2 b)) with the activation pattern belonging to the verbal description of the corresponding behaviour. When testing performance the motor input was omitted. The robot then continuously received its own current x, y and φ coordinates and the language instruction of the behaviour to be performed. It then had to produce the appropriate motor activations via W^{td}.

3 Results

First, we have trained a HM area to perform a single behaviour, 'pick', without the use of a higher-level SOM area. The robot thereby self-imitates a behaviour it has previously learnt by reinforcement [14]. Example videos of its movements can be seen on-line at:
`www.his.sunderland.ac.uk/supplements/NN04/`

Sample weights of the full trained network are shown in Fig. 4. The motor units in a) receive input from small regions in the HM area. SOM units in b) are connected to larger regions which comprise a part devoted to one of the motor units and additional parts devoted to x,y,φ input from high-level vision. The SOM units thus perform feature binding, or association of visually perceived input to a motor command.

Different SOM units establish different bindings and thereby mediate behaviour-modulated state-action pairings. Circles in Fig. 4 show that as one progresses along the SOM area from left to right, the neurons' RFs change from connecting to the 'turn left' motor unit to connecting to the 'forward' unit. Their differential activation reflects different phases within the 'go' behaviour.

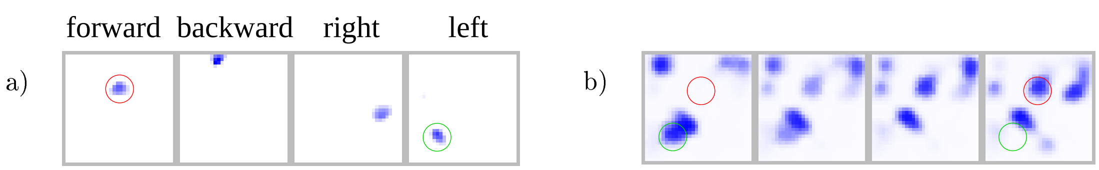

Figure 4: a) The four motor units' receptive fields (RF) in the HM area. Strong weights are depicted dark. Each unit receives input from a small region in the HM area. b) Four neighbouring SOM units' RFs in the HM area. These selected units are active during the 'go' behaviour. Circles indicate that the leftmost units' RFs overlap with those of the 'left' motor unit while the rightmost unit's RF overlaps with the RF of the 'forward' motor unit.

The action patterns during recognition of the 'go' behaviour action sequence and during its performance shown in Fig. 1 are shown in Figs. 5 and 6, respectively. At first glance, the activation patterns on the HM- and SOM areas are very similar between recognition and performance which suggests that most neurons display mirror neuron properties.

The largest difference can be seen within performance between the two activation steps of the HM area: in the first step it is activated from vision alone (top row of Fig. 6) in order to perceive the robot state and in the second step it is activated from the SOM area (third row of Fig. 6) in order to relay activation to the associated motor unit. The difference between these two steps comes from the lack of motor input in the first step and the completion of the pattern to include the motor induced activation as would come during full observation in the second step. Naturally, the second step's activation pattern resembles the pattern during recognition in the top row of Fig. 5, since patterns reconstructed from SOM units resemble the training data.

The differences in HM area unit activation patterns during recognition and performance are thus localised at the RF site of the active motor unit. If during training, the input differs only by the motor input (which happens if in the same situation a different action is performed according to a different behaviour) then the difference must be large enough to activate a different SOM unit, so that it can differentiate between behaviours. During performance, however, the absence of the motor input is not desired to have a too strong effect on the HM area representation, because the winner in the SOM area would become unpredictable and the performed action a random one.

The last row in Fig. 6 shows the activations of the language area as a result of the top-down influence from the winning SOM area unit during recognition. An error is made at the last time step which as far as the input is concerned (HM area activation in top row) is barely distinguishable from the second last time step. Note that the recognition error is in general difficult to quantify since large parts of some behaviours are ambiguous: for example, during 'go' and 'pick', a forward movement toward the front wall is made in large areas of the arena at certain orientation angles φ, or a 'turn' movement near the wall toward the centre might also be a result of either behaviour. Inclusion of additional information like the presence of a goal object or the action history could disambiguate many situations if a more complex model is used.

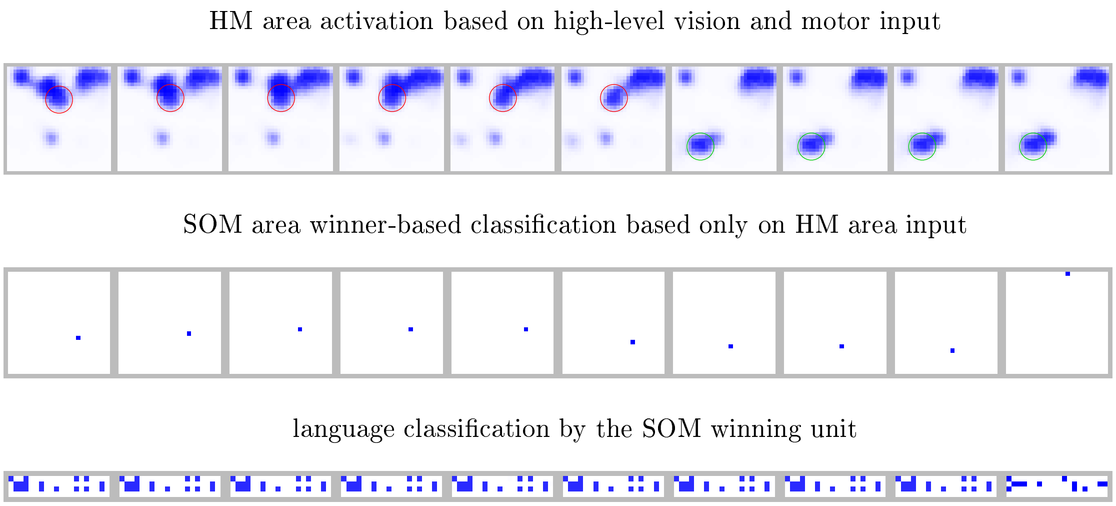

Figure 5: Activation sequences during observation of a 'go' behaviour, without language input. Strong activations are depicted dark, and shown at ten time steps from left to right. Circles mark the bottom-up input of the active motor unit of the teacher which changes from 'forward' in the first 6 steps to 'turn left' during the last 4 steps (cf. Fig. 4 b)). Language classification is correct except for the last time step which is classified as 'pick' (cf. Fig. 3 top)).

In the examples depicted in Figs. 5 and 6, the teacher and learner robots are initialised at the same position. Both then act similar during the first 4 time steps after which the learner decides to turn, while the teacher turns only after 6 time steps (see the circled areas in these figures).

Table 1 shows that the 'go' behaviour is mis-performed in about a quarter of 1000 cases where the robot has each time been placed in random start positions. Not-accounting for 'forward' movements was the primary error made in these cases. The 'pick' behaviour had the largest error rate as it is the most complex behaviour, depending very delicately on the particular y and φ coordinates, while the 'lift' behaviour is performed best.

4 Discussion

It is suggestive to identify the HM area of the model with area F5 of the primate cortex and the SOM area with F6. F5 represents motor primitives where the stimulation of neurons leads to involuntary limb movements. F6 rather acts as a switch, facilitating or suppressing the effects of F5 unit activations but it is itself unable to evoke reliable and fast motor responses. In our model, the

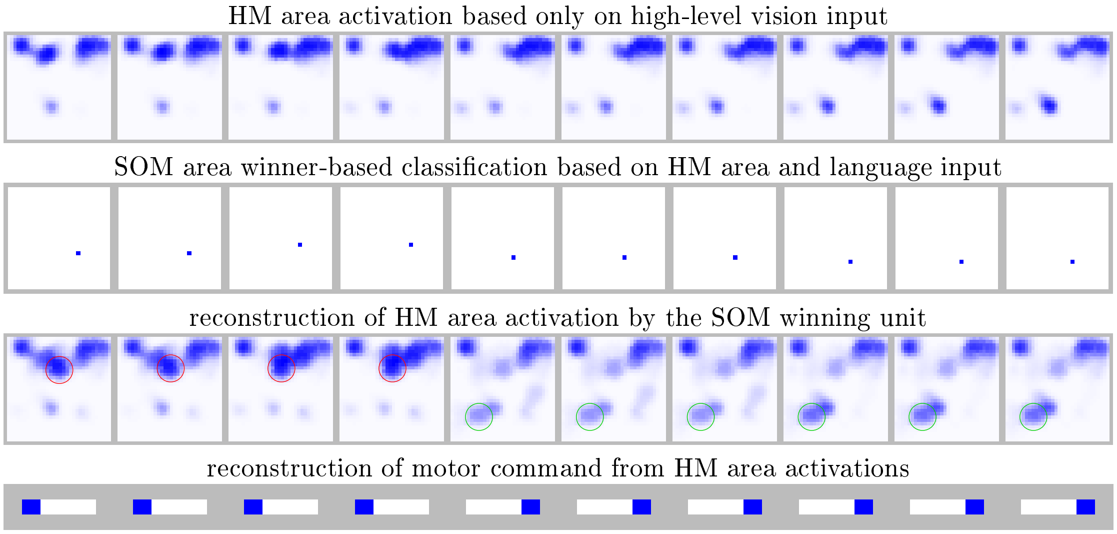

Figure 6: Activation sequences during performance of a 'go' behaviour, i.e. without motor input. The performed sequence is visualised in Fig. 1. Circles mark the region at each time step which has the decisive influence on the action being performed (cf. Fig. 4).

Table 1: Errors in 1000 learner's performance steps of each behaviour. In braces is the most failing motor unit and how often it was wrong ('-' wrongly inactive; '+' wrongly active).

Performing:	'go'	'pick'	'lift'
Errors	242 (forward, -205)	538 (right, -218)	62 (backward, +48)

HM area is directly linked to the motor output and identifiable groups of neurons activate specific motor units while the SOM area represents the channel through which a verbal command must pass in order to reach the motor related HM units.

Mirror neurons have so far been reported in F5. By design, our model uses the HM area for both, recognition and production, so an overlap in the activation patterns as observed in mirror neurons is expected. This overlap is mainly due to those neurons which receive high-level vision input. This perceptual input is tightly related to the motor action as it is necessarily present during the performance of an action and contributes to the "motor affordances" [4]. The decisive influence on the motor action, however, is localised in our model on smaller regions on the HM area, as defined by the motor units' receptive fields (Fig. 4 a)). The units in these regions would correspond to the canonical motor neurons which do not have mirror neuron properties and which are also found in F5.

A prediction of our model would then be that if the visually related mirror neurons alone are activated, e.g. by electrode stimulation, then neurons downstream would not be directly excited and no motor action would take place. It is, however, difficult to activate such a distinguished group of neurons since horizontal, lateral connections in the cortex are likely to link them to the canonical motor neurons.

5 Summary

We have developed a hierarchical approach to robot learning by imitation that combines Helmholtz machine and self-organising map learning algorithms in a hierarchical model. The model offers multimodal input processing of vision, language and action, and suggests analogies to the organisation of motor cortical areas F5 and F6 and to the properties of mirror neurons found in these areas. It provides insight to the organisation and activation of sensory-motor schemata from a computational modelling perspective. Considering functional processing logics it explains the position of mirror neurons connecting multiple modalities in the brain. In doing so a simulated student robot learns by repetitive observation to recognise and perform three behaviours in a real-world like environment.

Acknowledgments This work is part of the MirrorBot project supported by the EU in the FET-IST programme under grant IST- 2001-35282. We thank Prof. Friedemann Pulvermüller and Fermín Moscoso del Prado Martín at the Cognition and Brain Science Unit in Cambridge for their assistance with developing the language phoneme representation.

References

[1] P. Dayan. Helmholtz machines and wake-sleep learning. In M. Arbib, editor, *Handbook of Brain Theory and Neural Network*. MIT Press, Cambridge, MA, 2000.

[2] Y. Demiris and M. Johnson. Distributed, predictive perception of actions: a biologically inspired robotics architecture for imitation and learning. *Connection Science*, 15(4):231–43, 2004.

[3] M. Elshaw, C. Weber, A. Zochios, and S. Wermter. An associator network approach to robot learning by imitation through vision, motor control and language. In *Proceedings of International Joint Conference on Neural Networks*, 2004.

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

[5] I. Infantino, A. Chella, H. Dzindo, and I. Macaluso. A posture sequence learning system for an anthropomorphic robotic hand. In *Proceedings of the IROS-2003 Workshop on Robot Programming by Demonstration*, 2003.

[6] R. Knight. Contribution of human hippocampal region to novelty detection. *Computer Speech and Language*, 383(6597):256–259, 1996.

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

[8] G. Rizzolatti and M. Arbib. Language within our grasp. *Trends in Neuroscience*, 21(5):188–194, 1998.

[9] G. Rizzolatti, L. Fogassi, and V. Gallese. Motor and cognitive functions of the ventral premotor cortex. *Current Opinion in Neurobiology*, 12:149–154, 2002.

[10] E.T. Rolls. The orbitofrontal cortex and reward. *Cereb. Cortex*, 10(3):284–94, 2000.

[11] D. Roy and A. Pentland. Learning words from sights and sounds: A computational model. *Cognitive Science*, 26:113–146, 2002.

[12] J. Steil, F. Röthling, R. Haschke, and H. Ritter. Situated robot learning for multi-modal instruction and imitation of grasping. *Robotics and Autonomous Systems*, 2004. (Special Issue on Imitation Learning), in press.

[13] C. Weber. Self-organization of orientation maps, lateral connections, and dynamic receptive fields in the primary visual cortex. In *Proc. ICANN*, pages 1147–52. Springer Berlin, 2001.

[14] C. Weber, S. Wermter, and A. Zochios. Robot docking with neural vision and reinforcement. *Knowledge-Based Systems*, 17(2-4):165–72, 2004.