AN ASSOCIATIVE MODEL OF CORTICAL LANGUAGE AND ACTION PROCESSING

ANDREAS KNOBLAUCH[BA] HEINER MARKERT[A] GUENTHER PALM[A]

[A] *Department of Neural Information Processing, University of Ulm*
Oberer Eselsberg O27, 89069 Ulm, Germany
{knoblauch, markert, palm}@neuro.informatik.uni-ulm.de

[B] *Speech and Language Group, MRC Cognition and Brain Sciences Unit*
15 Chaucer Road, Cambridge CB2 2EF, England
andreas.knoblauch@mrc-cbu.cam.ac.uk

The brain correlates of words and their referent actions and objects appear to be strongly coupled neuron ensembles or assemblies distributed over defined cortical areas. In this work we describe the implementation of a cell assembly-based model of several visual, language, planning, and motor areas to enable a robot to understand and react to simple spoken commands. The essential idea is that different cortical areas represent different aspects of the same entity, and that the long-range cortico-cortical projections represent hetero-associative memories that translate between these aspects or representations.

1. Introduction

The brain correlates of words and their referent actions and objects appear to be strongly coupled neuron ensembles in defined cortical areas (Pulvermuller, 1999). Being one of the most promising theoretical frameworks for modelling and understanding the brain, the theory of cell assemblies (Hebb, 1949; Palm, 1990) suggests that entities of the outside world (and also internal states) are coded in overlapping neuron assemblies rather than in single ("grandmother") cells, and that such cell assemblies are generated by Hebbian coincidence or correlation learning. One of our long-term goals is to build a multimodal internal representation using several cortical areas or neuronal maps, which will serve as a basis for the emergence of action semantics, and to compare simulations of these areas to physiological activation of real cortical areas.

In this work we describe a cell assembly-based associative model of several visual, language, planning, and motor areas implemented on a robot to understand and react to simple spoken commands (cf. Fay et al., 2004). The task is to find certain fruits in a complex visual scene according to spoken or typed commands. This involves parsing and understanding of simple sentences, relating the nouns to concrete objects sensed by the camera, and coordinating motor output with planning and sensory processing.

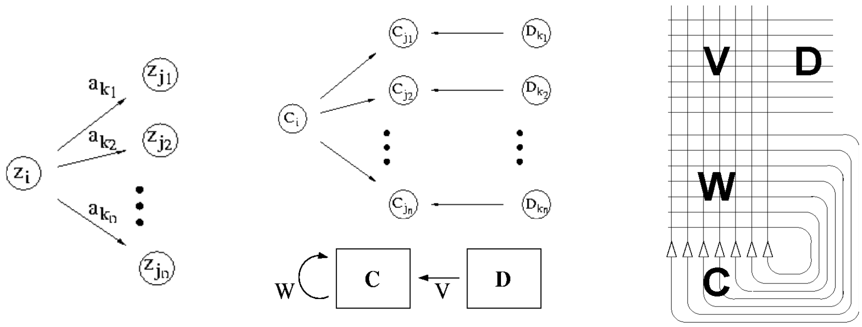

Figure 1. DFA (left), neural network (mid), and cell assemblies (right).

2. Language, finite automata, neural networks and cell assemblies

In this section we briefly review the relation between regular grammars, finite automata and neural networks. Regular grammars can be expressed by generative rules $A \rightarrow a$ or $B \rightarrow bC$ where upper case letters are variables and lower case letters are terminal symbols from an alphabet Σ.

Regular grammars are equivalent to deterministic finite automata (DFA). A DFA can be specified by $M=(Z,\Sigma,\delta,z_0,E)$ where Z is the set of states, Σ is the alphabet, $z_0 \in Z$ is the starting state, $E \subseteq Z$ contains the terminal states, and the function $\delta:(Z,E) \rightarrow Z$ defines the state transitions. A sentence $s=a_1a_2...a_n \in \Sigma^*$ is well formed with respect to the grammar if $\delta(...\delta(\delta(z_0,a_1),a_2),...,a_n) \in E$.

DFAs can be simulated by neural networks: For this it is sufficient to specify a simple model of recurrent binary neurons by $N=(C,D,W,V,c_0)$, where C contains the local cells of the network, D is the set of external input cells, W and V are binary matrices specifying the local recurrent and the input connections (Fig.2). The network evolves in discrete steps, where a unit is activated ($c_i(t)=1$) if its potential $x_i(t)=(Wc(t-1) + Vd(t-1))_i$ exceeds threshold Θ_i, and deactivated ($c_i(t)=0$) otherwise. A simple emulation of the DFA requires one neuron c_i for each state z_i, one neuron d_k for each input symbol a_k, synaptic connections $w_{ij}=d_{kj}=1$ for each state transition $(z_i,a_k) \rightarrow z_j$ and thresholds $\Theta_i=1.5$. If at the beginning only neuron c_0 is active the network obviously simulates the DFA. A biologically more realistic model would interpret the nodes in Fig.1 not as *single* neurons but as groups of nearby strongly interconnected neurons, i.e., local cell assemblies. This architecture has two additional advantages: First, it enables *fault tolerance* since incomplete input can be completed to the whole assembly. Second, overlaps between different assemblies can be used to express hierarchical relations between represented entities.

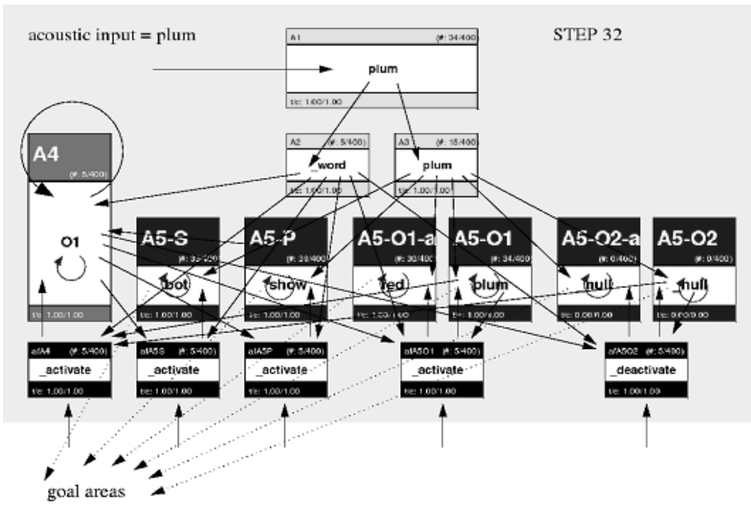

Figure 2. Architecture of cortical language areas.

3. Cortical language model

The language system consists of a standard HMM-based speech recognition system on the word level, and a cortical language system which can analyse streams of words detected with respect to simple regular grammars (Knoblauch et al., 2004). Fig. 2 shows the 15 areas of the language system. Each area is modelled as a spiking associative memory of 400 neurons (Knoblauch & Palm, 2001). Binary patterns constituting the neural assemblies are stored auto-associatively in the local synaptic connections by Hebbian learning.

The model can roughly be divided into three parts. (1) Auditory cortical areas A1, A2 and A3. (2) Grammatical areas A4 and A5-X. (3) Simple “activation fields” af-X that coordinate the activation or deactivation of the grammar areas.

When processing language, first auditory input is represented in area A1 by primary linguistic features (such as phonemes), and subsequently classified with respect to function in A2 and content in A3. The main purpose of area A4 is to emulate a DFA in a similar way as the neural network in Fig. 1. Each node corresponds to an assembly representing a grammatical state, and each edge corresponds to a state transition stored in delayed recurrent hetero-associative connections of area A4. E.g., processing of a sentence "Bot show red plum" would activate the state sequence S→Pp→OA1→O1→ok_SPO corresponding to expectation of processing of a subject, a predicate, an object or attribute, and finally an object. If the sentence was not well formed with respect to the grammar, then the sequence terminates in an error state.

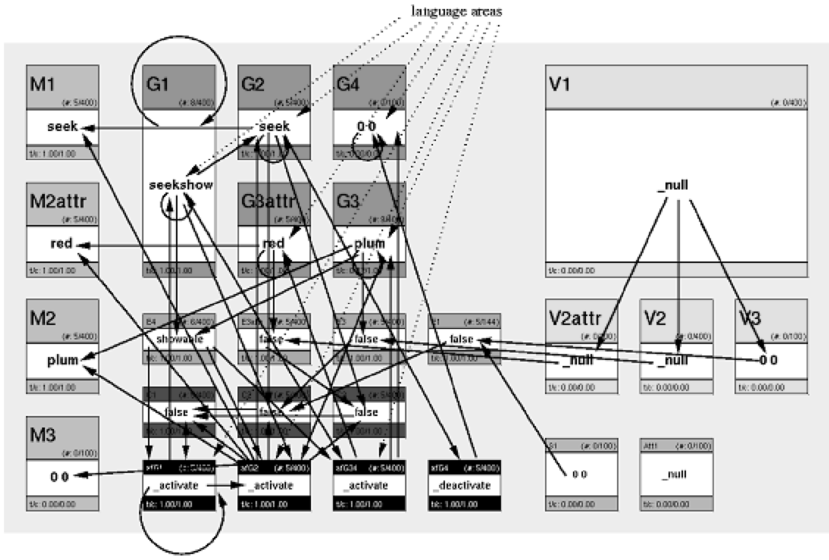

Figure 3. Architecture of the cortical action model.

Finally, the robot has to "understand" the sentence by transforming the word stream into an action representation. This is the purpose of areas A5-X which correspond to different grammatical roles. A5-S represents the subject "bot", A5-P the predicate "show" and A5-O1a/O1 the object "red plum" (Fig. 2).

4. Action processing

Our system for planning, action, and motor processing can be divided into three parts (Fig 3). (1) The planning/goal areas represent the robot's goal after processing a spoken command. Sequence assemblies in area G1 represent lists of actions necessary to complete a task, e.g., seek→point for the command "show object". Area G2 represents the current subgoal, and the remaining GX areas properties of involved objects. (2) The "motor" areas MX represent the motor command necessary to perform the current subgoal, and also control the low level attentional system. (3) Again there are "activation" fields (afX) and additionally "evaluation fields" that can compare representations of different areas. For illustration we describe the task „bot show red plum", where the robot has to point onto a red plum in the vicinity. First the robot has to understand the command as described in section 3 which activates the A5-representations. This information is routed to the goal areas where the first part of the sequence (seek→point) gets activated in G1. Similarly, object information is routed to areas G2, G3, G3attr. Since the plum's location is unknown, there is no activity in area G4. After checking semantics, corresponding assemblies in the motor areas are activated. This activates the attentional system and initiates the robot to seek the plum. Finding the plum activates corresponding visual assemblies in areas VX. The evaluation fields detect this and initiate areas G1, G2 to switch to the next state "point" of the action sequence. The robot will then adjust its “finger position” represented in area S1 in order to point to the plum. The matching of the positions will be detected by the evaluation fields and this eventually activates the final state in G1.

5. Conclusion

We have described the implementation of a cell assembly-based model of cortical language and action processing on a robot (cf. Knoblauch et al., 2004; Fay et al., 2004). The model consists of about 40 neuron populations each modelled as a spiking associative memory containing many "local" cell assemblies stored in local auto-associative connections (Knoblauch & Palm, 2001). The neuron populations can be interpreted as different cortical and subcortical areas, where it is a long term goal of this project to establish a mapping of our “areas” into real cortex (Pulvermuller, 1999).

Acknowledgments

This work has been supported by the MirrorBot project of the European Union.

References

Hebb D.O. (1949). *The organization of behavior. A neuopsychological theory.* New York: Wiley.

Fay R., Kaufmann U., Knoblauch A., Markert H., Palm G. (2004). Integrating object recognition, visual attention, language and action processing on a robot using a neurobiologically plausible associative architecture. In: *Proceedings of the AI-Workshop on NeuroBotics 9/04*, University of Ulm

Knoblauch A., Palm G. (2001). Pattern separation and synchronization in spiking associative memories and visual cortical areas. *Neural Networks, 14, 763-780*

Knoblauch A., Fay R., Markert H., Kaufmann U., Palm G. (2004) Associating words to visually recognized objects. In Coradeschi S. and Saffiotti A. (Eds.), *Anchoring symbols to sensor data. Papers from the AAAI workshop, Technical Report WS-04/03*. Menlo Park, CA: AAAI Press. (pp. 10-16)

Palm G. (1990). Cell assemblies as a guideline for brain research. *Concepts in Neuroscience, 1, 133-148*

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