Perception through Anticipation[*]

Ralf Möller   Horst–Michael Groß

Technical University of Ilmenau
Department of Neuroinformatics
98684 Ilmenau, FR Germany
E-mail: Ralf.Moeller@Informatik.TU-Ilmenau.DE

Abstract

A model of shape perception is presented. It is based on the anticipation of sensory consequences of motoric actions. With this model, we try to avoid the artificial separation between recognition and generation of behaviour. HEBBian cell assembly sequences are assumed to be the neural basis of the anticipation process. An unsupervised on–line learning rule system is introduced that enables the detection of complex, multimodal coincidences as a prerequisite for anticipation.

Keywords

VISUAL PERCEPTION – ASSEMBLY SEQUENCES – DETECTION OF COINCIDENCES

1   Behaviour-based Approach

Dealing with approaches to visual perception we faced a number of problems that, in our opinion, result from one major reason, namely the artificial separation between perception and generation of behaviour. Representational approaches usually own a level of purely sensory representation, as shown in figure 1 (left). Behaviour is generated using information provided by this level. Approaches of this type entail a number of problems, the most intricate being the 'reencoding problem' and the 'homunculus problem', which are closely related. Perception by transformation of visual information into a descriptive code is nothing else but a *'reencoding'* from one code into another. First, the descriptive code can be more compact than the visual information itself, but its interpretation is not necessarily simpler. Unintentionally in most cases an internal observer (an *'homunculus'*) is assumed, who evaluates the resulting code and derives appropriate behaviour. Second, the conversion into a descriptive code could be a detour. If an object partially covered by another shall be grasped, knowledge about covering has to be applied, covering has to be expressed in a desriptive code, and by use of knowledge about appropriate movements for grasping in the case of covering, motoric commands have to be derived from the code.

It seems to be much simpler to characterize the visual scenery *immediately in categories of behaviour*, e.g. by a sum of actions which describe possible methods of access to the covered and the covering object — the detour of reencoding could be avoided. This is the basic idea of our approach, that avoids the separation of perception and generation of behaviour and fuses both parts into one neural process (see figure 1, right). Perception (especially perception of shape as a fundamental ability of the brain) is considered to be the *internal simulation of a number of actions of the system and the anticipation of their consequences*. On one hand,

---

hypothetical actions and their anticipated consequences *characterize* the sensory situation; on the other hand, from this set of 'descriptive' actions *those* can be selected for execution, which result in a positive effect concerning the 'goal' of the system. An object in a certain distance may entail visual impressions that enable a prediction, what kind of tactile impressions can be expected in case of a movement. A number of such actions is 'simulated' (not consciously); actions having a *positive* consequence — e.g. allowing the avoidance of collision with the object — are preferred for execution; all actions that entail *negative* effects (collision, pain) will not be executed, but, nevertheless, they contribute to the *description* of the scenery, too.
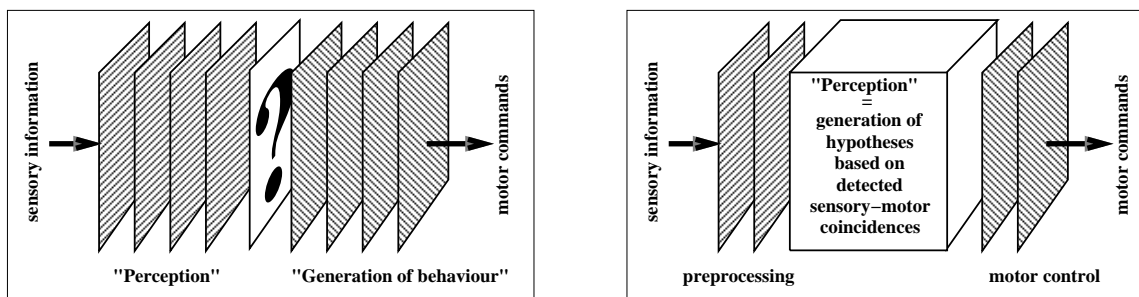

Figure1. Different approaches for systems in a closed perception–action loop. *Left:* 'Representational approaches', where sensory information is reencoded into an purely sensory representation. Behaviour is derived from this representation. Perception and generation of behaviour are stricly separated. *Right:* Our approach without explicit sensory representation. Perception and generation of behaviour are merged and originate from the same neural process. There are only relativly few levels of purely sensory or purely motoric processing.

2   Neural Code

A basic question for any neural architecture is the question of the neural code. Representational approaches in neural architecture often embody aspects of BARLOWs doctrine of 'Single Unit Perception' [BARLOW,H.B.72]. In BARLOWs approach the activity of a single cell encodes one usually complex aspect of a visual scene. It is close to classification or assignment of symbols to entities of the real world (mostly objects), and it entails the problems stated above as well as the problem of combinatorial explosion. In contrast, the theory of 'Cell Assemblies' introduced by HEBB [HEBB,D.O.49] promises a way avoiding combinatorial explosion by the assumption, that any entity is encoded by the simultaneous activity of a large number of neurons. Due to possible overlap of those cell assemblies, the coding space is enlarged drastically [PALM,G.90]. As a side effect of this theory, the notion of 'perception as classification' vanishes: it is replaced by an idea of perception at a lower, but fundamental processing level. This is perception of shape, guided by more general aspects of the physical world, which can be applied to every physical entity.

In our model, sensory situations (real or anticipated) are expressed by the activity of cell assemblies. Processes like completion or correction of data or generation of consistent hypotheses are performed inside these assemblies trough excitatory or inhibitory interaction. The anticipation process is based on assembly sequences as proposed by HEBB. Sensory assemblies (describing aspects of a sensory situation) and motor assemblies (hypothetical motor commands) fuse and entail the activation of other sensory assemblies coding a subsequent situation.

3 Neural Architecture

For simplicity, in the following description of the perception process (and in the corresponding figure 2, left), assemblies are reduced to the size of one neuron. Supposing the following coincidences have been detected formerly: if in a sensory situation **S1** a motor command **M1** is executed, a new situation **S2** is set up, the same being valid for **S2-M2** and **S3** as well as **S2-M3** and **S4**. **S3** and **S4** at the end of different sequences of hypotheses are assumed to be coincident with sensory situations of a genetically determined negative (pain) or positive (pleasure) meaning for the living being. Starting now from a given situation or event **S1**, different sequences are induced (in parallel or successively) by random activation of the motor command units below a level necessary for execution. In the hypothetical situation **S2** the generation process can take two different ways in dependence of the **random excitation** of **M2** or **M3**. If the chain of predicted events ends at a *negative impression* (**S1-M1→S2-M2→S3**), all motor command neurons which took part in the generation of this special sequence are suppressed by negative feedback, that is, actions supporting the real course predicted by the sequence of hypotheses cannot be executed. If the event is coincident with a *positive impression* (**S1-M1→S2-M3→S4**), the sequence is preferred for execution due to accumulation of excitatory feedback at the motor command neurons. The selection of appropriate behaviour in a given situation could be realized this way.
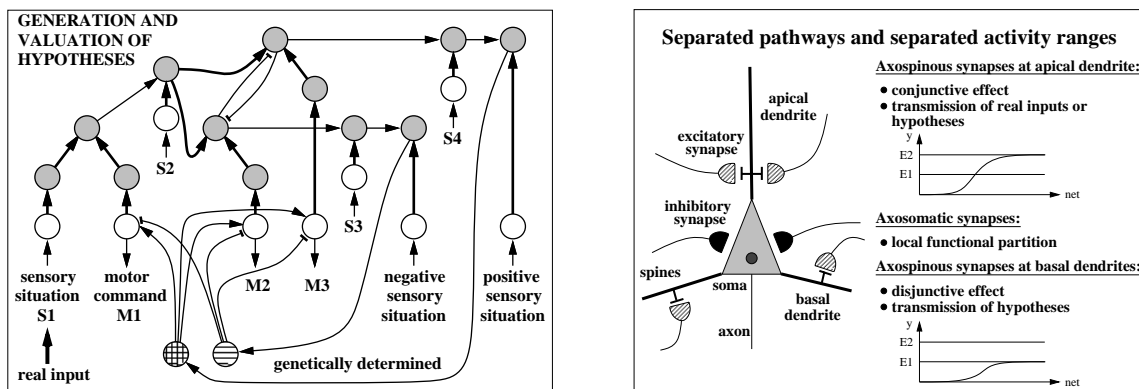

Figure2. *Left:* Perception and generation of behaviour in the neural architecture of our model. The 'conjunction path' is represented by thick arrows, the 'coincidence path' by thin arrows. For explanation see text. *Right:* Neuron model ('pyramidal cell'). The conjunction path is assumed to run via the apical dendrite, the coincidence path via the basal dendrites.

The neurons of our model — we use a spiking neuron model — transmit 'real' input signals as well as hypotheses. In a HEBBian learning rule these two cases have to be distinguished: modification of synaptic efficacies should only be induced by input activity, not by hypotheses. For this reason, we postulate *separated activity ranges* for neurons, low activity for hypotheses and high activity for input. The character of a signal should be preserved through all processing levels, which is obtained by introduction of *separated pathways*. Input information is transmitted via synapses at the apical dendrites of pyramidal cells, hypotheses via synapses at the basal dendrites (figure 2, right). Beside that, the two pathways perform different 'logical' functions. Synapses at the apical dendrite do only affect the neuron, if they are activated in *conjunction*, whereas synapses at the basal dendrites are supposed to have a more *disjunctive* effect. So together with inhibitory synapses an AND–OR–NOT system arises, which is able to perform *any* logical function.

4 Learning

Weights of the two pathways have to be trained in a way, such that complex coincidences can be integrated into this AND–OR–NOT system. We use on–line learning and coupled learning rules. The weight of a *basal* synapse is trained by a HEBBian pre–not-post rule, it reflects the statistics of coactivation of the pre- and the postsynaptic neuron by 'input activity' (hence the name 'coincidence path'). Using this mechanism, only coincidences between two single channels can be detected. More general, *complex coincidences*, i.e. coactivation of *conjunctions* of signals should be detected, too. A conjunction of signals has to be expressed in weights of the *apical* synapses, but only in the case when the conjunction of signals takes part in a coincidence, i.e. when the conjunction is *required* for the generation of an hypothesis. So modification of basal synapses influences the efficacy of apical synapses.

The effect of the learning rule system is demonstrated with a simple network (see figure 3). Four input channels in the lower part of the picture feed neurons A, B, C, and D immediately by apical synapses. A, B and C converge to E by apical synapses. A, B, C, and D are mutually connected by basal synapses as well as D and E. The input neurons of A, B, and C are activated randomly for $40ms$ approximately every $250ms$ with the 'input' firing frequency. This independent activation is superimposed by simultaneous activation of A, B, and D for $40ms$ with a mean interval of $1000ms$. After 30000 simulation steps, the symmetrical coincidence between the conjunction of channels A and B and channel D results in the specialization of neuron E to the conjunction of A and B (apical synapses) and the formation of symmetrical connections between E (which now represents the conjunction of A and B) and D (basal synapses).
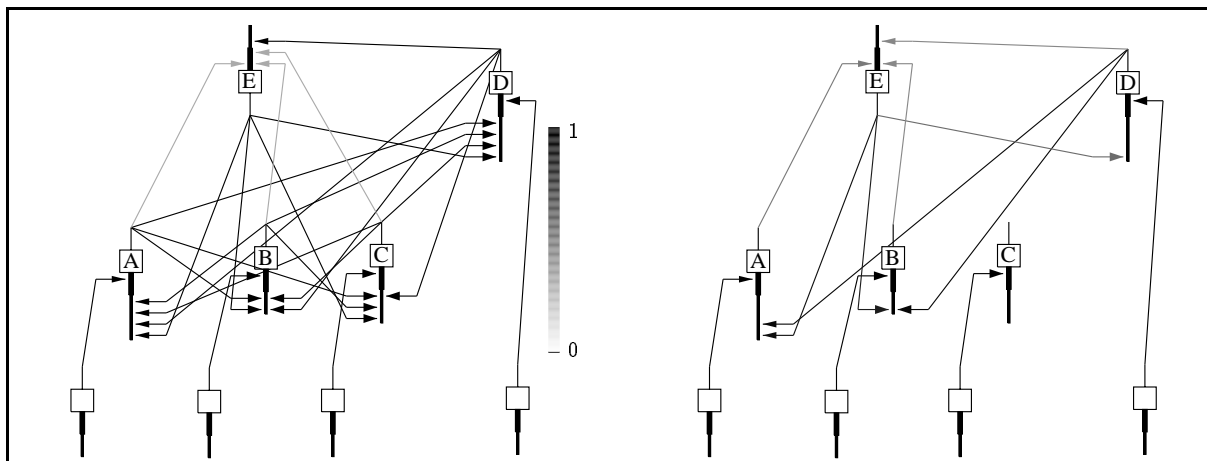

Figure3. Initial state of a small test network *(left)* and state after 30000 simulation steps (30 sec) *(right)*. Arrows symbolize axons, unfilled boxes somata, thick filled bars dendrites of the conjunction path and thin filled bars dendrites of the coincidence path. Weights are depicted in a linear gray scale *(middle)*; all weights below 0.2 have been skipped in the diagram.

References

[BARLOW,H.B.72] Barlow,H.B. Single unit and sensation: A neuron doctrine for perceptual psychology. *Perception*, 1:371–394, 1972.

[HEBB,D.O.49] Hebb,D.O. *The Organization of Behaviour*. Wiley, New York, 1949.

[PALM,G.90] Palm,G. Cell assemblies as a guideline for brain research. *Concepts in Neuroscience*, 1(1):133–147, 1990.