Integrating object recognition, visual attention, language and action processing on a robot in a neurobiologically plausible associative architecture

Rebecca Fay, Ulrich Kaufmann, Andreas Knoblauch, Heiner Markert, Günther Palm
Department of Neural Information Processing, University of Ulm, D-89069 Ulm
{*fay,kaufmann,knoblauch,markert,palm*}*@neuro.informatik.uni-ulm.de*

We have implemented a neurobiologically plausible system on a robot that integrates object recognition, visual attention, language and action processing using a coherent cortex-like architecture based on neural associative memories. This system enables the robot to respond to spoken commands like "bot show plum" or "bot put apple to yellow cup". The scenario for this is a robot close to one or two tables carrying certain kinds of fruit and/or other simple objects. Tasks such as finding and pointing to certain fruits in a complex visual scene according to spoken or typed commands can be demonstrated. 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.

The underlying cortical architecture is motivated by the idea of distributed cell assemblies in the brain [1, 2]. For visual preprocessing we use hierarchically organized radial-basis-function networks to classify objects selected by attention, where hidden states in this hierarchical network are used to generate sparse distributed cortical representations. Similarly, auditory input pre-processed by standard Hidden-Markov-Model architectures can be transformed into a sparse binary code for cortical word representations. In further cortical areas for language and action the sensory input is syntactically and semantically interpreted and finally translated into motor programs. The essential idea behind the cortical architecture is that different cortical areas represent different aspects (and correspondingly different notions of similarity) of the same entity (e.g., visual, auditory language, semantical, syntactical, grasp-related aspects of an apple) and that the (mostly bidirectional) long-range cortico-cortical projections represent hetero-associative memories that translate between these aspects or representations. These different notions of similarity can synergistically be used, for example, to resolve ambiguities within or across sensory modalities.

References

[1] D.O. Hebb. *The organization of behavior. A neuropsychological theory.* Wiley, New York, 1949.

[2] G. Palm. Cell assemblies as a guideline for brain research. *Concepts in Neuroscience*, 1:133–148, 1990.