Book Review **Robot Brains. Circuits and Systems for Conscious Machines.** Pentti O. Haikonen (Wiley, 2007, 214 pages hardcover, ISBN 978-0-470-06204-3). *Reviewed by Włodzisław Duch*[1]

The revival of research on consciousness, after almost 80 years of labeling the subject as unobservable and thus not amenable to rigorous investigation, has initially brought a lot of discussions among the philosophers of mind. Soon other cognitive scientists followed, and now consciousness research is in full swing on many fronts, with its own specialized journals (Consciousness and Cognition, Journal of Consciousness Studies) and series of conferences. In the past few years various attempts to define the problem of consciousness from engineering or computational perspective have emerged. Although detailed brain mechanisms responsible for consciousness are not yet known chances are that it might be possible to build conscious machines using some brain-like architectures. Indeed it has been argued [**?**] that a system with proper relations between its internals states that control its behavior, able to comment on these states, should behave as conscious.

The first book on the subject has been written by Stan Franklin [**?**], who created a "Conscious" Software Research Group at the University of Memphis. His "Intelligent Distribution Agent" architecture is based on the Global Workspace Theory [**?**]. The goal is to create a flexible system that may be "functionally conscious", but there is no claim that such system may show "phenomenal consciousness", or feel different qualities of conscious states. Owen Holland collected a number of interesting articles on consciousness in a book published in 2003 [**?**]. At the same time Pentti Haikonen, principal scientist at Nokia Research Center in Helsinki, proposed in his book [**?**] a cognitive architecture with a flow of inner speech, imagery and mental content. Adding the ability of introspection should endow such system with the awareness of both external environment and internal mental content. The inner interpretation of this mental content should seem to be immaterial. This down-to-earth approach to machine consciousness can in principle be tested if a sufficiently complex system could be build. It goes in the same direction as many other brain-inspired cognitive architectures (BICA) that have been formulated in recent years. Now Haikonen has written a second book "Robot Brains. Circuits and Systems for Conscious Machines", that explores the topic further, proposing more precise solutions.

So what, according to Haikonen, will make a robot conscious? In short it should be more human-like, endowed with some kind of mind, self-motivated, understand emotions and language and use it for natural communication, be able to react in emotional way, be self-aware and perceive its mental content as immaterial. While philosophers may still object that this is not sufficient building such robots may go a long way helping to understanding the problems of human consciousness. Therefore the book outlines an engineering approach that should lead towards cognitive and conscious machines.

The second chapter "Information, meaning and representation" argues for a vector representation of information as a substitute for non-numerical information processing by brains. Distributed vector representations are used a bit *ad-hoc*, without the attempt to present them as an approximation to probability distribution of activations of various brain areas at the microcolumn level (see for example [**?**]. The next chapter introduces neuron models and associative neural networks. The basic unit of the network is called here "Heikonen associative neuron", operating on binary signal vectors divided into the main signal, associative input, output and three signals that signify match, mismatch and novelty detection. These neurons are grouped together into networks of non-linear associators and trained using Hebbian principles. Such groups work as a "self-learning look-up tables". This is an interesting approach, in some ways similar to many other neural models, but its advantages have not been explicitly demonstrated. It does not seem to matter much, because the main purpose of the author is to quickly reach much higher level of complexity. This is done in Chapter 4 discussing circuit assemblies, built from associative neuron groups. These circuits are able to convert back and forth continuous signals to binary vectors, implement the "winner-takes-all" and "accept-and-hold" functions, convert serial signals to parallel, and be used in autoassociative predictors and sequencers of signals. Timing circuits are used for real-time learning and sequencing of signals. Defining such complex circuits allows for building higher-level modules. The author notes at the end of this chapter "the effect of noise and imperfect signal forms must be considered". Thus the approach proposed here is quite specific, but it seems that it has not really been tested.

Chapter 5 is about machine perception. After general discussion of the need for feedback loops in sensory recognition and perception-response loops acting as predictors, various circuits for forms of perceptions are introduced: kinesthetic, haptic, visual and auditory. While this is a long chapter, discussing many aspects of perception and outlining detailed solutions, it still remains on the theoretical level and thus may be easily criticized by experts. For example, a lot of work has been done in visual or auditory system modeling, with working models that can simulate various psychophysical effects, yet computer vision is still a very difficult subject. The same goes for motor control discussed in Chapter 6.

Chapter 7 discusses various aspects of machine cognition. It contains quite brief description of short and long-term memories, general remarks on perception of time, imagination and planning, and short section on deduction and reasoning. No specific solutions have been proposed here, just remarks that such functions can be implemented in memory-based architectures with match/mismatch conditions. The next chapter, on machine emotions, points out to the relations between emotions, significance, control and attention processes, and reduces the problem of emotions to combination of system reactions. A single page is devoted to machine motivation and willed action. In Chapter 9 the approach to natural language understanding and use in artificial systems is out-

[1]Department of Informatics, Nicolaus Copernicus University, Toruń, Poland; Google: W. Duch

lined. Multimodal approach advocated here treats language in conjunction with perception, imagination and motor actions. Embodied approach to language is of course quite popular but here specific circuits for grounding the meaning of words in multimodal information are proposed. This is the only chapter where results of some simulations that parses a simple sentence are presented.

In a short (only 5 pages) Chapter 10 a cognitive architecture for embodied, perceptive and interactive robot brains is presented. It is represented by block diagrams with modules for survival (pain and pleasure sensors), basic needs (self-sensors), self-image and environment sensing, perception, mental maps for orientation, learning and motor actions. “Machine consciousness”, the last chapter, is addressed to non-technical readers. An interesting observation here is that consciousness is determined by the focus of global attention and results from the way that operation of different modules “appears to the transparent cognitive system”.

The book should be interesting for people working in computational intelligence and cognitive robotics, and without doubt it is worth a careful reading. Also researchers in computational neuroscience, who usually focus on details, may benefit from the integrative approach presented in the “Robot brains”. But does it solve the problem? On the back cover of the books we can read: “The methods presented in this book have important implications for computer vision, signal processing, speech recognition and other information technology fields.” Certainly the book introduces a number of interesting points of view, but the proof of the pudding is in the eating. We already knew that memory, control, language and emotions are important, and that neural circuits are the way to implement them. No breakthrough technologies have been proposed, no software or hardware demonstration provided. Vision, language and cognition at the human competence level still remain a great challenge and it is not clear how far can one go without more detailed inspirations from the brain. For example the role of the brain stem or rather complex reward and motivation system needs to be addressed. In the preface the author writes that ”it is rather easy to implement these principles in the way of computer programs”. It still remains to be seen how easy, and how far will the methods described in this book be able to carry us towards autonomous, conscious robots.

REFERENCES

[1] W. Duch, “Brain-inspired conscious computing architecture”. *Journal of Mind and Behavior* 26(1-2), 1-22, 2005.

[2] S. Franklin, *Artificial Minds*. Bradford Books, MIT Press, 1997.

[3] B.J. Baars, *A cognitive theory of consciousness*. Cambridge University Press 1988.

[4] O. Holland, *Machine Consciousness*, Imprint Academic, 2003.

[5] P.O. Haikonen, *The Cognitive Approach to Conscious Machines*, Imprint Academic, 2003.

[6] W. Duch, P. Matykiewicz, J. Pestian, “Towards Understanding of Natural Language: Neurocognitive Inspirations”. *Lecture Notes in Computer Science* 4668, 953-962, 2007, and *Neural Networks*, forthcoming.