

# Reverse-engineering Mammalian Brains for building Complex Integrated Controllers

Ricardo Sanz, Ignacio López, Adolfo Hernando and Julia Bermejo

Autonomous Systems Laboratory  
Universidad Politécnica de Madrid, Spain  
Phone: +34-91-3363061, Fax: +34-91-3363010  
email: ricardo.sanz@upm.es

**ABSTRACT:** The ICEA project ([www2.his.se/icea/](http://www2.his.se/icea/)) is a four-year project on bio-inspired integrated cognitive control, bringing together cognitive scientists, neuroscientists, psychologists, roboticists and control engineers. The primary objective of the whole project is to develop a new cognitive system architecture to be used in technical systems and that integrates cognitive, emotional and autonomic control processes. The ICEA generic architecture will be based on the extraction of control design patterns from bioregulatory, emotional and cognitive control loops based on the architecture and physiology of the rat brain. The work of the ASLab team is focused in i) the development of a unified theory of integrated intelligent autonomous control, ii) the development of a cognitive architecture with self-awareness mechanisms and iii) the analysis of the applicability of such a technology in a broad domain of embedded, real-time systems. This paper describes the ICEA projects and the ongoing ASLab work.

**KEYWORDS:** Autonomous systems, integrated control, layered control, control design patterns, mammalian brain, emotion, cognition, autonomy.

## THE ICEA PROJECT

The ICEA project ([www2.his.se/icea/](http://www2.his.se/icea/)) is a four-year project funded by the European Commission Cognitive Systems Unit. The project is focused on bio-inspired integrated cognitive control, bringing together cognitive scientists, neuroscientists, psychologists, roboticists and control engineers. The primary objective of the whole project is to develop a *novel* cognitive systems architecture inspired by rat brains.

The ICEA project will develop the first cognitive systems architecture integrating cognition, emotion and autonomy (bioregulation and self-maintenance), based on the architecture and physiology of the mammalian brain. A key hypothesis underlying this four-year collaboration between cognitive scientists, neuroscientists, psychologists, computational modellers, roboticists and control engineers, is that emotional and autonomic mechanisms play a critical role in structuring the high-level thought processes of living cognitive systems.

The robots and autonomous systems developed will perceive and act in the real world, learn from that interaction developing situated knowledge (representations of their environments in spatial, emotional and behavioural terms), and use this knowledge in anticipation, planning and decision-making. The brain and behaviour of the rat will be an important starting point because of the large scientific literature available for this species. Rat cognition will be studied and emulated both through an ambitious program of empirical studies in real animals and through computational modelling, at different levels of abstraction, on several, real and simulated, robot platforms.

The project will develop two central, integrated platforms, rat-like in appearance, perceptual, and behavioural capacities. First, the ICEAbot robot platform, equipped with multimodal sensory systems will serve as a real-world testbed and demonstrator of the behavioural and cognitive capacities derived from models of rat biology. Second, a 3-D robot simulator, ICEAsim, based on the physical ICEAbot, but also offering richer opportunities for experimentation, will demonstrate the potential of the ICEA architecture to go beyond the rat model and support cognitive capacities such as abstraction, feelings, imagination, and planning. ICEAsim will serve as the main platform for exchange and technical integration of models developed in different parts of the project, but it will also be made freely available to the research community as a potential standard research tool. Other, more specialized platforms will be developed to investigate issues of energy autonomy, to model active whisker touch, and to evaluate the ICEA architecture's applicability to non-biomimetic robots.

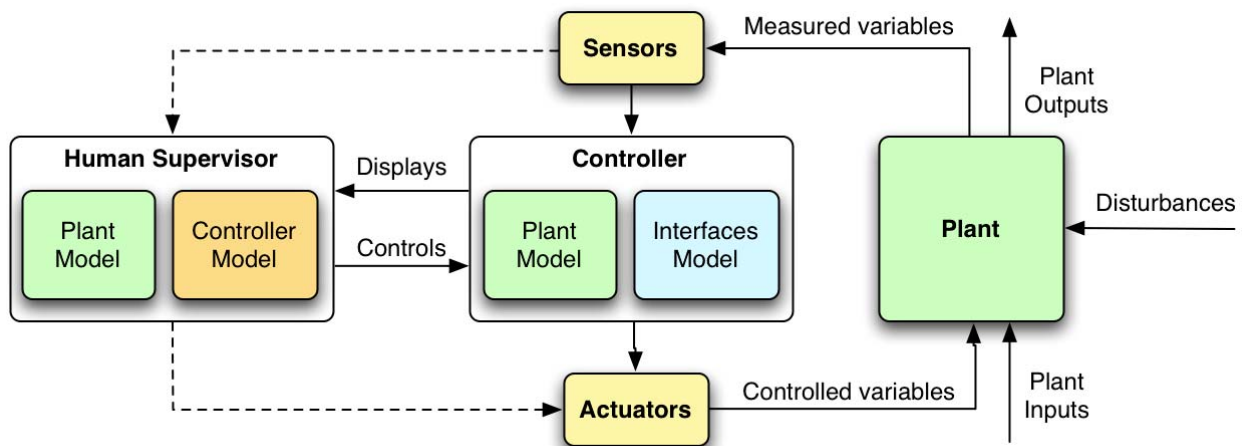
## CONTROL TECHNOLOGY IN CONTEXT

There are many reasons –well beyond the pure hubris of feeling like god– for pursuing the objective of fully autonomous machines. *Cost reduction* and *improved performance* were the main factors behind the drive for improved automation in the past. During the last years, however, a new force is gaining momentum: the need for *augmented dependability* of complex systems.

### AUTONOMY FOR PERFORMANCE

Cost reduction is usually achieved by means of reducing human labour, but another very important aspect is the increase in product quality derived from the improvement of the operational conditions of the controlled plants. Automated system performance is usually higher than manual system performance and there are many technical systems that cannot be operated manually in any way due to the limitations of human operators –for example high-speed machining– or practical or legal issues typically associated with worker health –for example space robotics.

In many cases, however, the problem of building a controller for a well known production process is mostly solved; only minor details persist. The problem appears when the plant is not so well known of when the operational and/or environmental conditions are of high uncertainty. This is so because having a good knowledge of the plant is the first necessary step to building a controller for it [1].



**Figure** Fehler! Unbekanntes Schalterargument.: A standard hierarchical (supervisory) control scheme.

There are many textbooks on controller design in general terms, centered in particular kinds of controller designs or centered in concrete application domains. In the last case, the domain typically constrains the kind of implementation that a controller may have (e.g. table-driven controllers in resource constrained electronic control units in vehicles or software-less controllers in some safety-critical applications).

### THE ICEA PROJECT FROM A CONTROL PERSPECTIVE

One of the biggest challenges –in a sense the only remaining challenge– of any control design paradigm is being able to handle complex systems under unforeseen uncertainties. This is the old-age problem of dirty continuous process plants or the problem of mobile robotics (or of any technical system that must perform in an uncontrolled environment). This problem is not restricted to embedded control systems but appears in many large system situations; e.g. web server scalability or security problems in open tele-command systems are examples of this.

The observation of the structure of the mammalian brain leads to the conclusion that the triple layering –autonomic, emotional and cognitive– is a form of objective/controller structuring and, at the same time, increase overall robustness of the animal.

### AUTONOMY FOR DEPENDABILITY

Dependability considerations about many systems –transportation, infrastructure, medical, *etc.*– has evolved from a necessary issue in some safety-critical systems to become an urgent priority in many systems that constitute the very infrastructure of our technified world: utilities, telecoms, vetronics, distribution networks, *etc.*

These large-scale, usually networked, systems improve the efficiency of human individuals and organizations through new levels of integration, control and communication. However, the increased distribution, integration and pervasiveness is accompanied by increased risks of malfunction, intrusion, compromise, and cascade failure effects. Improving autonomy into these systems can mitigate these risks by means of its impact in survivability.

Survivability [2] is the aspect of dependability that focuses on preserving essential services, even when systems are faulty or compromised. As an emerging discipline, survivability builds on related fields of study (*e.g.* security, fault tolerance, safety, reliability, reuse, verification, and testing) and introduces new concepts and principles.

A key observation in survivability engineering –or in dependability in general– is that no amount of technology –clean process, replication, security, *etc.*–can guarantee that systems will survive (not fail, not be penetrated and not be compromised).

The complexities of software-based system services, issues of function and quality in custom and/or commercial off-the-shelf (COTS) usage, and the proliferation of integrable and interoperable devices, combined with the growing sophistication of functionalities, present formidable engineering challenges in survivable system analysis and development.

In the following sections, we shall try to explore how the concept of autonomy is understood in artificial systems from an ICEA-wide point of view, considering how different aspects of autonomy emerge from different designs.

## OPERATIONAL ASPECTS OF SYSTEM AUTONOMY

The general principle for autonomy in artificial systems is *adaptivity*. This enables systems to change their own configuration and way of operating in order to compensate for perturbances and the effects of the uncertainty of the environment, while preserving convergence to their objectives. A series of aspects are studied in artificial systems in order to enhance adaptivity: cognition, modularity, fault-tolerance, *etc.*

### *Cognition*

In general, systems which are tightly grounded to the physical substrate have reduced adaptivity, due to mechanical constraints, than systems with cognitive capacities (reflections on cognition and autonomy in systems can be found in [3] [4]). It is understood that cognitive capacities in a system result from a combination of lower level aspects which have been studied in attempts to apply biological principles to artificial systems (*e.g.* [6][7][8]):

- Knowledge: Representation, retrieval, ontologies, types (procedural/declarative ...)
- Perception: Sensation, interpretation.
- Learning: Automation of tasks, chunking, self-reflection, inference.
- Intelligence: Inference, generalization, particularization, association of concepts.

### *Modularity*

Large systems may result in high levels of complexity and interdependence among parts. In order to structure interaction among system parts, systems may be designed as a combination of *modules*. A module is a conceptual, physically separable part of a system which usually performs a specific function, interacting with the rest of the system through a well-defined interface of inputs and outputs. Substituting a module for another with the same function and interface should result in an equivalent system. While module separability is not a remarkable characteristic of brains – maybe due to their organic nature– modularity has been well established as a core structuring mechanism of them.

Modular systems consist of a structure of parts that interact through their interfaces, presenting an explicit structure and functional decomposition. Interfaces make that dependencies between one module and the rest of the system are determined, allowing interchangeability of modules, as mentioned earlier.

Having an explicit structure and defined dependencies are critical factors for adaptivity. Uncertainty, perturbances and planning may eventually require reconfiguration of system parts, or in the way they interact with each other. Several examples can illustrate this point; some hybrid control architectures are based on a deliberative layer reconfiguring behaviour modules of the reactive layer in order to react to an unknown situation; implementing fault-tolerance mechanisms in systems involves identifying sources of error, faulty parts and eventually their isolation or reconfiguration.

### *Fault-tolerance*

System adaptivity depends on its capacity to achieve its objectives under perturbances and uncertainty. Eventually, parts of the system may be damaged or malfunction during operation, compromising system cohesion and therefore its capacity to achieve objectives. Fault tolerance techniques [9] have been developed to provide the system with mechanisms to react to these circumstances by adapting itself. Fault tolerant systems must evaluate self-performance in terms of their own *dependability*.

Three concepts distinguished in relation with reliability: a *failure* is a deviation of the system behaviour from the specifications. An *error* is the part of the system which leads to that failure. Finally, a *fault* is the cause of an error.

Fault-tolerance in artificial systems is usually implemented in four phases (error detection, damage confinement and assessment, error recovery and fault treatment and continued service). Faulty parts of the system are deactivated or reconfigured and the system continues operation. Fault tolerance in artificial systems usually distinguishes between hardware and software. Hardware fault tolerance is based on fault and error models which permit identifying faults by the appearance of their effects at higher layers in the system (software layers.) Hardware fault tolerance can be implemented by several techniques, the most known are: TMR-Triple Modular Redundancy (three hardware clones operate in parallel and vote for a solution,) dynamic redundancy (spare, redundant components to be used if the normal one fails,) and coding (including check-bits to test correct operation.)

Software fault tolerance can be based on a *physical model* of the system, which describes the actual subsystems and their connections, or on a *logical model*, which describes the system from the point of view of processing.

### *Soft computing*

In relation with artificial intelligence, a series of techniques have been developed in order to make systems capable of operating with uncertain, imprecise or partially representative measurements (neural networks, fuzzy logic, expert systems, genetic algorithms, *etc.*).

In general, following Checkland [10] we can summarise considering that *control* is always associated with the imposition of constraints between systems levels. Any account of a control process necessarily requires our taking into account at least two hierarchical levels. At a given level, it is often possible for simple systems to describe the level by writing dynamical equations in the Rouse sense, on the assumption that one element is representative of the subsystem and that the forces at other levels do not interfere. In a sense, the intra-level behaviour may be considered autonomous.

But any description of a technical control process entails an upper level –that may be a human operator (see previous figure)– imposing constraints upon the lower. The upper level is holder of lower level knowledge, *i.e.* it is a source of an alternative (simpler) description of the lower level in terms of specific functions that are emergent as a result of the imposition of constraints.

We can conclude this contribution to the NiSIS symposium saying that the two central questions that remain for engineering complex autonomous systems are: 1) how to decompose the system into a meaningful integrated hierarchy? and 2) what is the necessary knowledge that any particular layer must have?

Inspiration to answer these two questions is sought through the research in the ICEA project.

## ASLAB OBJECTIVES IN ICEA

ASLab objectives in ICEA are focused on three activities: i) the formulation of a theoretical framework, ii) the construction of a novel intelligent control architecture and iii) the study of the application of this technology to real-world systems.

ICEA activity A1 –Theoretical framework – will integrate the findings of the several project's tracks into a coherent integrated theory of the interaction of cognitive, emotional and (bio-) regulatory processes in natural and artificial cognitive systems. This will include the analysis and evaluation of potential impacts of ICEA results and technologies in the field of commercial embedded systems, which in many domains are still lacking an architecture-centric approach that specifically tackles the problem of robust system autonomy and survivability.

ICEA activity A7 – Representation, abstraction and planning– will be concerned with building a novel architecture to combine reverse-engineered emotion and feeling control structures with abstract controllers and representations of the world. This will let the technical system to reason about the environment and possible courses of action based on strongly embodied, emotionally biased underlying control structures up to the level of self-consciousness.

## REFERENCES

- [1] Conant, R. and Ashby, W. (1970). Every good regulator of a system must be a model of that system. *International Journal of System Science*, 1:89–97.
- [2] Ellison, R. J., Fisher, D. A., Linger, R. C., Lipson, H. F., Longstaff, T., and Mead, N. R. (1997). Survivable network systems: An emerging discipline. Technical Report CMU/SEI-97-TR-013, Software Engineering Institute, Carnegie Mellon University.
- [3] Heylighen, F. (1990). Self-Steering and Cognition in Complex Systems, chapter Autonomy and Cognition as the Maintenance and Processing of Distinctions, pages 89–106. Gordon and Breach. Editors: F. Heylighen, E. Rosseel and F. Demeyere.
- [4] Christensen, W. D. and Hooker, C. A. (2000). Autonomy and the emergence of intelligence: Organised interactive construction. *Communication and Cognition, Artificial Intelligence*, 17(3-4):133–157.
- [5] Meystel, A. (2000). Measuring performance of systems with autonomy: Metrics for intelligence of constructed systems. White Paper for the Workshop on Performance Metrics for Intelligent Systems. NIST, Gaithersburg, Maryland, August 14-16, 2000.
- [6] Newell, A. (1990). *Unified Theories of Cognition*. Harvard University Press.
- [7] Hayes-Roth, B. (1995). An architecture for adaptive intelligent systems. *Artificial Intelligence*, 72(1-2):329–365.
- [8] Albus, J. and Meystel, A. (2001). *Engineering of Mind: An Introduction to the Science of Intelligent Systems*. Wiley Series on Intelligent Systems. Wiley, New York.
- [9] Jalote, P. (1994). *Fault Tolerance in Distributed Systems*. P T R Prentice Hall.
- [10] Checkland, P. (1981). *Systems Thinking, Systems Practice*. John Wiley & Sons, New York.
- [11] Sanz, R., and López, I. (2006) What's going on in the mind of the machine? Insights from embedded systems. Computational Neuroscience & Model Integration workshop, Derbyshire June 2006.