Modeling Consciousness for Autonomous Robot Exploration

Raúl Arrabales Moreno, Agapito Ledezma Espino, and Araceli Sanchis de Miguel

Departamento de Informática,
Universidad Carlos III de Madrid,
Avda. de la Universidad 30, 28911 Leganés, Spain
{rarrabal,ledezma,masm}@inf.uc3m.es

Abstract. This work aims to describe the application of a novel machine consciousness model to a particular problem of unknown environment exploration. This relatively simple problem is analyzed from the point of view of the possible benefits that cognitive capabilities like attention, environment awareness and emotional learning can offer. The model we have developed integrates these concepts into a situated agent control framework, whose first version is being tested in an advanced robotics simulator. The implementation of the relationships and synergies between the different cognitive functionalities of consciousness in the domain of autonomous robotics is also discussed.

Keywords: Cognitive Modeling, Consciousness, Attention, Emotions, Exploration.

1 Introduction

Machine Consciousness could be considered as the field of Artificial Intelligence specifically related to the production of conscious processes in engineered devices (hardware and software). Undoubtedly, a multidisciplinary approach is necessary in order to approach such an intricate paradigm. Latest advances and contributions from psychology and philosophy in the scientific study of consciousness have lead computer scientist community to reconsider the possibility of engineering machine consciousness [1].

Although the phenomenal aspects of consciousness are still especially controversial [2][3], we argue that a purely functional approach can be successfully applied in the domain of autonomous robot control. In this work we present a machine consciousness model designed to command an autonomous robot, and the functionality of this model as a solution of the exploration problem. The phenomenal dimension, represented by the question '*Is the robot conscious of the exploration task he is doing?*' is deliberately neglected at this stage of our research.

In section two we introduce our model and the theories of consciousness in which it is based upon. Section three covers the software architecture where we have integrated the machine consciousness model. In section four we discuss the detailed design and interaction between model components. Finally, we conclude describing salient preliminary results.

2 Evading the Cartesian Theater

Materialist theories of consciousness are not supposed to rely on any link to the soul like the one located by Descartes in the pineal gland [4]. However, the so-called Cartesian materialism associates conscious experience with a concrete place in the brain. The Cartesian theater[1] refers to this materialistic *homunculus*, which would play the role of the director of the brain. In contrast to the Cartesian theater metaphor, there exist other accounts for consciousness based on the idea of interim coalitions of specialized processors running concurrently in our brains. These processors or agents are continuously collaborating and competing for the light of consciousness.

Our model is mainly based on two theories of consciousness: the Global Workspace Theory (GWT) [6] and the Multiple Draft Model (MDM) [2]. GWT depicts a theater where the processors compete for appearing in the scene spotlight, which is the attention focus. Aggregation of processors is produced by the application of contexts. Behind the scenes, context criteria are defined and coordinated (unconsciously) by the director. Context formation mechanisms select the event in the stage that will be illuminated by the spotlight. The MDM adopts the editorial review process metaphor, where coalitions of processors suffer reiterative edition and review until they are presented as the official published conscious content of the mind.

Taking the main ideas from the described metaphors of the mind, we have built a cognitive model of consciousness called CERA (Conscious and Emotional Reasoning Architecture) [7]. Key functionalities of the model can be directly mapped to functional aspects of both GWT and MDM. A layered and modular scheme has been defined, where layers represent levels of control and modules represent cognitive specialized functions. Modules are situated within layers, CERA core layer encloses the key functional modules identified in the mentioned theories of consciousness. This set of functional modules is designed to support the workflows described by consciousness metaphors. Initial version of CERA core layer comprises eight modules: attention, status assessment, global search, preconscious management, contextualization, sensory prediction, memory management, and self-coordination. In this framework, there is no central module representing consciousness. Consciousness is supposed to emerge from the interaction between modules and their management of specialized processors.

[1] The term Cartesian theater was coined by Dennett to define (and reject) the idea of a central point of the brain where all sensory data is projected and conscious experience is produced [5].

3 Software Architecture

CERA has been originally designed to be applied to the domain of autonomous robotics. Therefore, its three layers correspond to different levels of autonomous control. The external layer manages physical robot machinery, and has to be adapted to the particular robot and onboard sensors and actuators being used. Middle layer is called instantiation layer as it encloses the problem-specific components. In the case of unknown environment exploration, instantiation layer contains the map production primitives and robot basic 'innate' behaviors for exploring. Finally, the inner layer contains the mentioned general purpose cognitive functions of consciousness (Fig. 1).

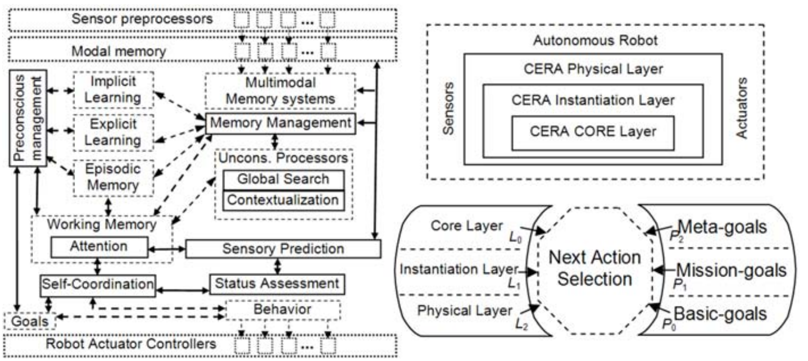

Fig. 1. In the left diagram solid lines represent CERA Core modules. Dashed lines represent CERA instantiation layer (domain-specific modules). Dotted lines represent CERA physical layer. Right diagrams illustrates CERA layered design and next action selection contributions.

Robot behavior is determined by a combination of the three level goals. At the physical level, the integrity of robot hardware is the highest priority (e.g. avoid collisions). Mission goals, unknown environment exploration in our case, are managed at the middle level. The *meta-goals* applied at the core level are related to the emotional dimension of the model of consciousness as explained in the next section.

In order to develop a flexible framework for experimentation with both simulated and real robots, we have integrated CERA into the Microsoft Robotics Studio (MSRS) platform [8]. A key component of MSRS is the Concurrency and Coordination Runtime (CCR) [9], which we use for asynchronous programming and unconscious processors concurrency management. A managed high-performance thread pool dispatches specialized processors tasks. Thread dispatching and asynchronous I/O operations follow diverse coordination patterns as required by CERA core modules.

MSRS is based on a light-weight distributed services-oriented architecture. A MSRS node run a set of services, and nodes can be installed in different machines. Communication and coordination between services is performed using the Decentralized Software Services Protocol (DSSP) [10]. The adaptation of CERA to this environment is the role of CRANIUM (Cognitive Robotics Architecture Neurologically Inspired Underlying Manager). CRANIUM is a wrapper for CERA that provides DSSP services creation and CCR parallel coordination patterns. Basically, CRANIUM is the interface for the creation of unconscious specialized processors and the management of their interactions. Like in a human brain, specialized regions of the brain perform concrete tasks concurrently, and emerging coordination is given by the neural connections between these areas (global access hypothesis) [11]. While CRANIUM provides the underlying neural-like mechanisms, CERA uses these services to produce the integrative function of consciousness, where only one (conscious) content can prevail at any given time.

4 Designing Robot Consciousness

A robotic application developed using MSRS is basically an orchestration of input and output between a set of services. CRANIUM provides a model to create the kind of services required by a cognitive robotics architecture like CERA. CRANIUM services represent the interface to unconscious processors like sensor preprocessors and actuator controllers. CRANIUM also defines the communication primitives between the processes that perform robot functions.

For our preliminary experiments we are using both simulated and real Pioneer P3 DX robots equipped with front and rear bumper arrays and a ring of eight forward ultrasonic transducer sensors (range-finding sonar) (Fig. 2). CRANIUM defines services for acquiring data from bumpers and sonar as well as commanding the differential drive motor system. Equivalent services are available for both real and simulated sensors and actuators.

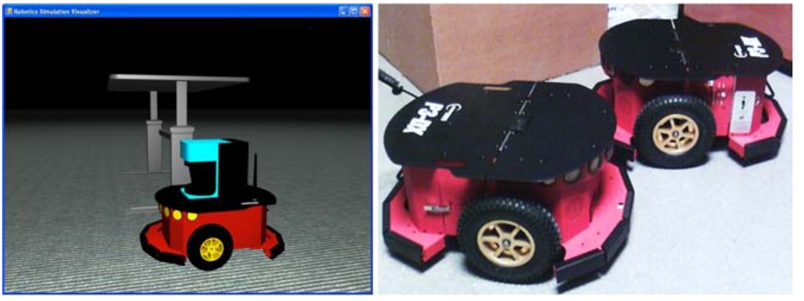

Fig. 2. Simulated and real Pioneer P3 DX robots

4.1 Physical Layer

CERA Physical Layer subscribes to sensors notifications using CRANIUM. Every time a sensor changes its state, the asynchronous operation is managed by a CERA handler. The process of acquiring a sensor state change corresponds to a minimal perceivable event for the robot. Following Aleksander and Dunmall notation for axioms of neuroconsciousness [12], where A is the agent (the P3 DX robot in our case) and S the sensory-accessible world, these minimal percepts δS_j are the atomic information acquired by sensor handlers. Therefore, these CERA handlers build an internal representation of the percept, called $N(\delta S_j)$. This perception process is twofold, as two differentiable pieces of information are obtained: sensed object or event and its relative position in the world. In a two dimensional world, j has two spatial dimensions, and $(x, y) = (0, 0)$ represents the robot reference system (his subjective point of view).

Measurement of j is provided by each sensor differently. For instance, P3 DX bumper arrays consist of five points of sensing. Bump panels are at angles around the robot (Fig. 3). In this case, j is calculated depending on the bumper panel being pressed.

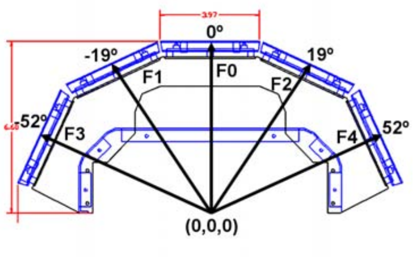

Fig. 3. P3 DX front bumper array consists of five bump panels at angles -52°, -19°, 0°, 19°and 52°to front of the robot. CERA bumper handler detects which bump panels are pressed and assigns values for every j accordingly. The resulting $N(\delta S_j)$ represent a physical obstacle at the relative location j.

As bump panels are a fixed part of the robot body and their activation is on contact, the j value is always the same for each bump panel. However, other sort of sensors would have to calculate relative position of the percept based on its own position or orientation. Like in natural nervous systems, all CERA handlers have to provide the ability to locate the source of the object or event being perceived.

The outputs from sensor handlers are combined into more complex percepts by sensor preprocessors. These preprocessors play the role of specialized group of neurons in charge of unconsciously detecting concrete features or patterns in perceived data. For instance, mammals visual system has specialized neural circuitry for recognizing vertical symmetry, motion, depth, color or shape [13,14]. Analogously, CERA sensor preprocessors provide the robot with feature extraction and recognition mechanisms appropriate for its environment. Some of the

CERA preprocessors that have been already implemented include wall detection and sonar invisible object detection (objects detected by bumper collisions but not detected by sonar or laser range finder).

In addition to sensor handlers and sensor preprocessors, CERA physical layer also contains unconscious processors related to behavior. Robot actuator controllers are defined as per CRANIUM interface to physical P3 DX robot. P3 DX is equipped with two motors (each wheel is connected to its own motor) that contain 500-tick encoders, forming a differential drive platform, where a third unpowered castor wheel provides balance. Initially, three basic actions have been implemented in CERA: stop, move forward, and turn. Move forward operation takes a motor power level for both wheels, and turn operation uses two power levels to apply to each motor in different directions. Thanks to CRANIUM, all the operations triggered by actuator controllers are executed in the context of the CCR dispatcher.

Basic actions are defined as δB_i, where i is the referent indicating the direction of the movement. Following the same notation as used for percepts, the robot representation for basic actions is $N(\delta B_i)$, and $N(B)$ corresponds to robot behavior. The composition of higher level behaviors in terms of physical basic actions is done at the instantiation layer under the coordination of CERA core layer.

Goals at this level can be seen as instincts, and more specifically as survival instincts. Basic goals are defined as a relation between perceptions and actions in the physical layer. In terms of our P3 DX survival a small set of basic goals, like avoiding collisions, have been defined. However, as explained below, higher layers can send inhibition messages that prevent physical layer goals to be accomplished.

4.2 Instantiation Layer

CERA instantiation layer makes use of sensor preprocessors in order to build a mission-specific representation of the world. This layer contains unconscious mission preprocessors, which are designed to recognize mission related objects and events using the perception information obtained in the physical layer. Wall segments and obstacles perceived by sensor preprocessors are internally combined in order to detect corridors or rooms. As percepts coming from the physical layer are j indexed, mission related percepts are built as *M(S)*, a partial description of the sensory accessible world S, where:

$$N(S) = \cup_j N(\delta S_j)$$
$$M(S) \subset N(S)$$

N(S) is the entire representation of the world built by the robot, while *M(S)* could be any subset of *N(S)*. In this case, the $N(\delta S_j)$ components of a concrete set *M(S)* are related due to their source location *j*.

In order to achieve the primary mission goal defined for the present work, unknown environment exploration, a simple two dimensional map representation

Table 1. Goal definition for a single mission (exploration). Layer 0, 1, and 2 refer to physical, instantiation, and core layers respectively. Execution time is discretized in *steps*, *updates* refer to $N(S)$ representation updates, and *mismatches* refer to failures to confirm a past percept, i.e. finding an obstacle where nothing was detected the last time the area was explored. E represents Emotion and n is the number of emotions being considered. Function *Energy* calculates the strength of a given emotion.

Goal	Layer	Description	Evaluation
G_{00}	0	Wander safely	$Eval(G_{00}) = (steps - collisions)/steps$
G_{10}	1	Map the environment	$Eval(G_{10}) = updates/steps$
G_{11}	1	Confirm created map	$Eval(G_{11}) = (updates - mismatches)/steps$
G_{20}	2	Positive emotional state	$Eval(G_{20}) = \sum_n Energy(E_n)$

has been chosen initially. As, for the time being, this is the only aspect of the world that we want the robot to be aware of, this map is actually *N(S)*. The robot keeps this map updated as he explores the world, mapping current perception of walls and obstacles into its two dimensional *N(S)*.

Similarly to percept aggregation, instantiation layer behaviors (called mission behaviors) are composed of the $N(\delta B_i)$ defined in the physical layer. Mission behaviors are the $M(B_i)$ (being $M(B) \subset N(B)$) that better fit mission goals needs. In terms of exploration, different wandering behaviors have been defined in the form of unconscious processors. These $M(B_i)$ compete for selection according to CERA Core layer cognitive rules.

4.3 Core Layer

CERA Core layer can be seen as a control center orchestrating the unconscious processor resources available in lower layers. The cognitive model implemented in this layer is intended to be domain independent, as all problem-specific representations are allocated in the instantiation layer. The general purpose functionality modules available in the core layer operate based on the basic percepts and actions from lower layers. The functionality of CERA Core modules is illustrated below applying the exploration problem.

Attention module is in charge of directing both perception and action. In order to be successful, the robot has to direct its attention to the fulfillment of mission goals, which can be recognized as full or partial solution of the specific problem being tackled. However, CERA design does not follow this strategy directly. Instead of taking mission goals as the drivers for the attentional focus, the meta-goals are considered. Meta-goals are related to the emotional state of the robot, and provide the means to have a general attention mechanism able to deal with multiple missions or different goals of the same mission. The definition of meta-goals characterizes the robot 'personality'. Initially, we have just considered one broad meta-goal: keeping a positive emotional state (Table 1).

Attention module calculates i referents for possible next $M(B_i)$ behaviors. In order to determine which $M(B_i)$ are applicable, contextualization mechanisms are used. The contextualization module provides possible associations between

$N(\delta B_i)$ based on available contextualization criteria. The primary criterion for building wander behaviors is based on the relation between j referent of perceived objects and action i referent. Basically, contextualization criteria for exploring will result in a set of promising directions to continue exploration, e.g. not to pay attention toward directions where an obstacle has been previously detected. Using this technique, attention focus is kept on the $M(S_j)$ perceived in the surrounding of the robot, and a set of possible actions $M(B_i)$ is calculated in that context with the aim of directing sensing.

The initial set of $M(B_i)$ behaviors calculated by CERA Core are considered gaze shifts, and are inspired in eye foveating saccades [15]. The robot is intended to direct its sensors to where relevant perception is predicted to take place. Even though our robot is not equipped with a motorized camera, he can rotate in place operating the differential drive, thus orienting the sonar coverage. Sensory prediction module is always active and listening sensor preprocessors output. Percepts $N(\delta S_j)$ are arranged into sequences, where $N(\delta S_j(t+1))$ is predicted based on past experience. As a first simplistic approach, sensory prediction is based on invariability. Therefore, the sensory prediction module will tell the attention module to direct the i referent of sensing to j locations where $N(\delta S_j(t))$ is different from predicted (or remembered).

As attention is serial, the Attention module has to select a concrete $M(B_i)$ at any given time (which could be composed of one or more $N(\delta B_i)$ and could take several time units to complete). The selection of winning attention focus and its associated behavior is not only based upon the factors explained above. The initial search on $N(S)$ in terms of contextualization criteria and sensory prediction, is extended further on $N(I)$ by the Self-Coordination module. $N(I)$ as defined in [12], is the representation of a imagined world. As both $N(S)$ and $N(I)$ are j-indexed, contextualization mechanisms can apply between perceived and imagined world. Self-Coordination module provides planning capability by searching trajectories in $N(I)$. Search on $N(I)$ is limited in depth and the sensory prediction function is also used to generate imagined perceptions $N(\delta I_j)$. The initial direction of the i referent of the most promising imagined behavior is used to finally select the next behavior to apply.

Evaluation of imagined behaviors is performed taking into account Status Assessment module output. This module implements a model of emotions, where basic emotions are defined and assigned an energy value. Emotions influence cognition, activating or inhibiting perception and action [16]. Additionally, in the context of CERA, emotions are the means to summarize the performance of the robot in terms of goal accomplishment. Consequently, CERA goals are assigned one or more emotional operators, which evaluate the progress being made in the goal achievement (see [7] for a detailed description of CERA emotional operators and associated emotional learning mechanism). Table 1 shows the evaluation functions used for some goals. Making good progress in goal achievement increases the energy of positive emotions like curiosity or joy. On the contrary, failure leads to increases in the energy of negative emotions like fear or anger. Emotional operators establish the relations between goals and specific emotions.

As described by Baars [11], global access is the capacity of accessing any piece of knowledge. The Global Search module is required to index and retrieve any unconscious processor, being a performance aid for the contextualizing function. Analogously, Preconscious management module is designed to be the interface between conscious and unconscious processes. It provides the required environment where different coalitions of unconscious processors can be built in the form of $M(B_i)$ and $M(S_j)$. Also, any 'editorial' review of these draft coalitions is managed in this domain, in order to have a consistent ('conscious') final version. Finally, the Memory Management module serves as an associative database manager, offering an interface to retrieve subsets of $N(S)$ and $N(I)$ related by any contextualization criteria.

5 Conclusion and Future Work

The described CERA architecture presents a novel approach to cognitive robotics where attention can be directed even without information from the real world. $N(I)$ provides a representation that permits the robot to plan possible behaviors. These imagined behaviors are emotionally evaluated the same way that actual performed behaviors. The emotional learning loop is closed when imagined behavior is physically performed, and the real and imagined outcomes are compared. An additional degree of flexibility beneficial to deal with real world is provided by CERA layered design, where Core layer can send inhibition messages that prevent physical layer goals to be accomplished when the threshold of energy of a particular emotion is reached.

There is still countless work to do in order to explore and compare the pros and cons of this kind of cognitive architectures. As of this writing we are working in the improvement of several components of CERA and CRANIUM. The application of forward models is being considered to improve sensory prediction functionality [17]. Additionally, real world experiments require much more effort: three dimensional representation and dealing with imperfect robot odometry.

Multiple mission accomplishment is other area where we believe that CERA can provide a good solution. The approach for this problem would be the creation of multiple instantiation layers. This design permits that the same architecture can be used for other domains, and facilitates the integration of different AI techniques into the unconscious processors.

Other challenges are robot vision and multi-robot collaboration. A pan-tilt onboard camera with foveating capability would increase the perception richness. Coordinated multi-robot exploration is also a challenging problem and very related to the field of autonomous exploration [18]. We believe that the application of inter-subjectivity models might be beneficial in this area.

Acknowledgments. We would like to thank Trevor Taylor from Queensland University of Technology (Brisbane, Australia) and Ben Axelrod from Georgia Institute of Technology (Atlanta, GA) for sharing their Microsoft Robotics Studio *Maze Simulator* code and supporting its application into this research work.

References

1. Holland, O.: Machine Consciousness. Imprint Academic (2003)
2. Dennett, D.: Quining Qualia. In: Consciousness in Modern Science. Oxford Universit Press (1988)
3. Nikolinakos, A.: Dennett on qualia: the case of pain, smell and taste. Philosophical Psychology **13** (2000) 505–522
4. Finger, S.: Descartes and the pineal gland in animals: A frequent misinterpretation. Journal of the History of the Neurosciences **4** (1995) 166–182
5. Dennett, D.C.: Consciousness Explained. Little, Brown and Co., Boston (1991)
6. Baars, B.: A Cognitive Theory of Consciousness. Cambridge University Press, New York (1988)
7. Arrabales, R., Sanchis, A.: A machine consciousness approach to autonomous mobile robotics. In: 5th International Cognitive Robotics Workshop. AAAI-06. (2006)
8. Microsoft, Corp.: Microsoft robotics studio. http://msdn.microsoft.com/robotics/ (2006)
9. Richter, J.: Concurrent affairs: Concurrency and coordination runtime. MSDN Magazine (2006)
10. Nielsen, H., Chrysanthakopoulos, G.: Decentralized software services protocol - dssp. Technical report, Microsoft Corporation (2006)
11. Baars, B.J.: The conscious access hypothesis: Origins and recent evidence. Trends in Cognitive Science (2002) 47–52
12. Aleksander, I., Dunmall, B.: Necessary first-person axioms of neuroconsciousness. In: IWANN. Lecture Notes in Computer Science (2003) 630–637
13. Braitenberg, V.: Vehicles: Experiments in Synthetic Psychology. MIT Press, Cambridge, Mass (1984)
14. Crick, F., Koch, C.: Towards a neurobiological theory of consciousness. Seminars in The Neurosciences **2** (1990)
15. Terzopolous, D.: Vision and action in artificial animals. In: Vision and Action. Cambridge University Press (1998)
16. Ciompi, L.: Reflections on the role of emotions in consciousness and subjectivity, from the perspective of affect-logic. Consciousness & Emotion **2** (2003) 181–196
17. Webb, B.: Neural mechanisms for prediction: do insects have forward models? Trends in Neurosciences **27** (2004) 278–282
18. Burgard, W., Moors, M., Stachniss, C., Schneider, F.E.: Coordinated multi-robot exploration. In: IEEE Transactions on Robotics. Volume 21., IEEE (2005) 376–385