XIM-Engine: a software framework to support the development of interactive applications that uses conscious and unconscious reactions in immersive mixed reality

Pedro Omedas
Alberto Betella
Riccardo Zucca
Xerxes D. Arsiwalla
Daniel Pacheco
SPECS, NRAS
Universitat Pompeu Fabra

Johannes Wagner
Florian Lingenfelser
Elisabeth Andre
Human Centered Multimedia
University of Augsburg, Germany

Daniele Mazzei
Antonio Lanatá
Alessandro Tognetti
Danilo de Rossi
Research Centre "E. Piaggio"
University of Pisa, Italy

Antoni Grau
Automatic Control Dept, Technical University of Catalonia, UPC, Spain

Alex Goldhoorn
Edmundo Guerra
Rene Alquezar
Alberto Sanfeliu
Institut de Robotica i Informatica Industrial CSIC-UPC, Spain

Paul F.M.J. Verschure
SPECS, NRAS, ICREA
Universitat Pompeu Fabra, Spain
paul.verschure@upf.edu

ABSTRACT

The development of systems that allow multimodal interpretation of human-machine interaction is crucial to advance our understanding and validation of theoretical models of user behavior. In particular, a system capable of collecting, perceiving and interpreting unconscious behavior can provide rich contextual information for an interactive system. One possible application for such a system is in the exploration of complex data through immersion, where massive amounts of data are generated every day both by humans and computer processes that digitize information at different scales and resolutions thus exceeding our processing capacity. We need tools that accelerate our understanding and generation of hypotheses over the datasets, guide our searches and prevent data overload. We describe XIM-engine, a bio-inspired software framework designed to capture and analyze multi-modal human behavior in an immersive environment. The framework allows performing studies that can advance our understanding on the use of conscious and unconscious reactions in interactive systems.

Categories and Subject Descriptors

H.1.2 [User /Machine systems]: Human factors, Software psychology. H.5.1 [Multimedia Information Systems]: *Artificial, augmented, and virtual realities.*

General Terms

Algorithms, Measurement, Design, Human Factors.

Laval Virtual VRIC'14, April 9-11, 2014, Laval, France.

Keywords

Cognitive systems, Bio-inspired, Human Computer Interaction, Biomimetics, Affective Computing, XIM.

1. INTRODUCTION

The analysis of complex datasets is a task that is becoming very common in the digital era where extremely large datasets are collected that overpass our capacity to extract useful meaning [2], [4]. There is a need of new tools in science and industries to analyze the datasets and support the process of understanding. This should be achieved by creating systems that can cooperate with humans and interact ubiquitously in ecologically valid conditions. One important aspect for such systems is how to understand the process of discovery and creativity where the unconscious reactions play a decisive role [5] and how we can capture this information from users. In the recent years, more and more research started to take into account the use of unconscious reactions from users (those which are not under explicit control by the user) but the challenge is still how we can interpret and use these signals [6] to feed an interactive system. Each of these signals provides valuable information of the user state but needs to be correctly balanced in order to determine the expected reaction. An optimal solution to approach this problem is from a human perspective, grounding on biological principles and designing a human-machine interaction in an ecologically valid environment which incorporates the social aspect of the interaction. Building a system to capture all this information in conditions that are suitable to perform psychological research poses technological challenges in sensors, signal synchronization, algorithms, audiovisual systems and the integration of all in a coherent interaction for the user. To advance on solving this problem we propose XIM-engine, a system that provides the technical framework to perform these studies, embedded in the, so called, eXperience Induction Machine (XIM) [7], an immersive mixed-reality room equipped with sensors, visualization and sonification systems. We have designed and implemented software to augment this infrastructure with capabilities to collect unconscious signals and support the exploration of large datasets[8][9][10]. Our assumption is that the spatial representation of information - through immersion - plays a key role in its understanding [5]. In this paper, we describe the software design and the grounding of the system in our understanding of mind and brain. In what follows, we detail the implementation of the system.

2. XIM-Engine bio-inspired design

The XIM-engine logical design is inspired on the functional organization of the brain proposed by the Distributed Adaptive Control (DAC) cognitive architecture. DAC is a theory of the design principles underlying the Mind, Brain, Body nexus (MBBn) that has been developed since 1992 [10][11]. DAC can act as a control system for embodied systems to act in real life situations with neuro-physiological constraints and support the deployment in robots and machines. It defines the inputs and outputs of the system in a hierarchical structure to regulate the decision making process. Indeed, DAC has already served as the integration framework for the large-scale synthetic organism ADA, a 180m^2 exhibition space that interacted and communicated with its more than 500K visitors during the Swiss Expo in 2002 [13]. Here we describe the engine inspired by the organization of DAC to control XIM, an embodied non-anthropomorphic machine which has a synthetic brain: the XIM-engine. The use of DAC in this framework is a novel perspective to develop applications with human-machine interaction, where we can treat XIM as a machine with human-like capabilities based on biological principles. In this sense, XIM interacts with humans in a cooperative way, by inferring their unconscious process from the signatures of implicit signals and adapting the information to the optimal level of understanding. This establishes a loop of communication where XIM tries to maximize its own goal: to inject new ideas to accelerate understanding. The approach we took, was to build incrementally from the lowest layer of the DAC architecture (the Soma), which defines the interfaces from the body (XIM) to the world (users interacting with XIM). In the XIM-engine, this bottom layer corresponds to the sensing architecture and to the effectors of the XIM infrastructure controlled by the Composition Engine component. The second layer of DAC organization (Reactive layer) is a direct mapping between the user’s action and the XIM effectors. When the user, for instance, performs a particular gesture, this is captured by the sensors architecture, processed and reflected as an action in the audiovisual content. In the next layer of organization (Adaptive layer) the output of the engine (e.g. audiovisual content) is adapted through a cognitive module, the so called Sentient Agent. This agent can represent different models of human behavior, setting the needs and the goals that drive the rest of the modules. In DAC, a goal is defined as a parameter that has to fulfill a need of the system. In the following sections, we describe the implementation of the Soma, Reactive and Adaptive layers of the DAC architecture.

2.1 Sensing Platform Architecture

A sensing platform [12] has been developed to collect human reactions through unobtrusive measurements of user reactions, deploying the Soma layer of the DAC architecture. It is responsible for capturing and processing raw sensor data delivered by newly developed wearable sensor devices in real time, extract high level features and forward the extracted information to the cognition module and the effectors controller of XIM-engine.

Table 1: Sensors along with derived signals/cues and their meanings

Sensor	Signal	Cues	Type	Meaning
Eye tracker	Gaze	Fixations	Implicit	Interest detection
	Pupillary response	Pupil size		Cognitive Load
Sensing Shirt	Acceleration	Energy		Activity
	Respiration	Breathing rate		Arousal
	Electrocardiogram	Heart rate		
Sensing Glove	Electrodermal activity	Peaks/Slopes		
	Forearm orientation		Explicit	Pointing, grabbing and translation of virtual objects
	Finger position	Grab gestures		
Depth Sensor	Body tracking			

Sensors included in the platform have been selected in order to allow an unobtrusive acquisition of subject's implicit and explicit cues through analysis of physiological and gestural signals. As reported in Table 1, five acquisition units -Eye Tracker [15], Sensing Shirt [16], Sensing Glove [17], a Depth sensor (Kinect sensor from Microsoft Corp. USA) and XIM-Tracking [18] - are included in the platform allowing the acquisition of eye gaze, head tracking, pupil size, ElectroCardioGram (ECG), Breath Rate (BR), Electro Dermal Response (EDA), Hand Gesture and Orientation, Body joints' Positions (described in detail in [14]) and the position of the user in the XIM space. Data fusion and integration is implemented using the Social Signal Interpretation framework[1] [19]. The SSI plug-in based architecture allows the integration of multi-modal sensor devices which acquire different signal typologies at various sampling frequencies guaranteeing synchronization and data storing. SSI also receives input events from other modules of XIM-engine, allowing, for example, to synchronize visual events with acquired signals. After synchronization, each type of signal is given a tailored treatment including pre-processing and cue-extraction steps. The architecture allows flexibility to add more sensors or to upgrade the sensors that are already integrated without affecting the rest of the system. Other components in the architecture can ask for an update on the signal state whenever required. A communication channel allows any engine component to send an update request to the sensing platform, which then replies by sending a message, which includes information on the current state in terms of cognitive load (from pupil) and arousal (from EDA, ECG). For each feature, the values are reported in terms of relative changes thus expressing the general tendency in time, e.g. whether a feature has been increased (value > 0) or decreased (value < 0). These features constitute the inputs to the upper layers of the system that are in charge of contextualizing the signals and events to the application. At the same time, the data streams are also steadily sent to the user response database (URDB), where all the information is stored.

[1] http://openssi.net

2.2 Effectors control: the Composition Engine

The Composition Engine (CE) is a module we developed to control the effectors of the XIM and drive the interaction: visual content projection, sound, floor lights and feedback sensors (e.g. haptic feedback). The CE implements the reactive layer of the DAC architecture. It provides interaction to the user (e.g. navigation controls or selection through gestures) and communicates with the sensing platform to send and receive information. It needs to be adapted to the specific interaction and content of the application, although it shares basic principles. As an example, all contents based on Kinect navigation, hand pointing and grabbing, are common to all scenarios that require natural interactions, and therefore these parts of control in the CE remain the same. The CE also defines how the XIM communicates with users defining a language for the human-machine interaction with different level of abstraction defined by the application. The Composition Engine is developed as a component of the audiovisual system of XIM, controlled by the real time engine Unity (Unity Technologies, CA, USA[2]) in charge of displaying the content and interaction. We adapted the 3D engine to work in a cluster to produce 360^{o} projections in the immersive environment, and we have created the interfaces to the sensing platform and the effectors.

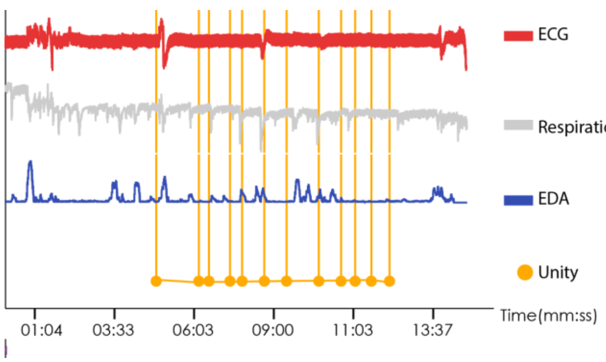

Figure 1: Plot of normalized raw data obtained from different sensors. The bottom part of the graph shows how the signals are aligned with events in the audiovisual content generated by the 3D engine (Unity). The data was collected by sensors and stored in the URDB in real-time during an experiment in XIM.

2.3 User Response Database (URDB): a database for collecting human behavior

In the contextual layer of the DAC architecture the autobiographical memory stores the experience with the interactions with the world, and labels these experiences. We implemented the User Response Database (URDB) designed to store and analyze the input/output sensor information that constructs the technological base for the autobiographical memory module of the DAC architecture. The URDB can be used in real-time and to perform off-line analysis of user experiences in XIM. By storing the signals that interprets the state of the user (cognitive load and arousal), and the sequence of stimuli presented, the system can thus build a synchronized map of conscious and unconscious reactions to the data. Later the experience can be recalled to bring the unconscious actions to the conscious scene. Figure 1 shows a plot of normalized signals captured by three different sensors worn by a user in XIM. Two types of signals can be seen: continuous numerical values and events. The events in the plot are categories with semantic meaning and can be used to store relevant events in the experience, such as the delivery of stimulus to the user. Furthermore, meta-information about an experiment can be stored (e.g. subject ID, date of the experiment and duration).

2.4 Sentient Agent

We have described a functional implementation of the sensing platform, the User Responses Database (URBD) and the Composition Engine (effectors control of the XIM-engine). Together they implement the Soma and Reactive layer of the DAC architecture, providing our system with the capability of perceiving and reacting to users. The next step is to implement the adaptive layer of the DAC architecture through the implementation of a module named as the Sentient Agent (SA). The SA provides adaptive capabilities to user interaction by regulating its actions on defined drives and needs, supporting goal oriented behavior. The SA is developed as an allostatic control system [21]. The inputs of the Sentient Agent are the signals interpreted by the sensing system: arousal (from EDA, HR) and cognitive load (from pupillometer). These inputs are balanced to support three main objectives: regain focus, maintain concentration and support exploration. This behavior of the system is supported by the Validation Gate (VG) or Anticipatory Field (HF) hypothesis on counter current attention modulation and the conscious scene [18]. Through this theory we claim that information of the presented data is regulated by user's cognitive load and arousal, and therefore the goal of the SA is to maintain an optimal level (allostasis) that will maximize the information processed.

2.5 Dataset visualization and analysis

XIM-engine supports the analysis and visualization of big datasets in the form of graphs. A wide range of data that are interesting to science can be expressed in the form of a graph. Fields of application can be found virtually in all domains varying from neuroscience, economy, life and applied sciences, natural language processing or social networks. The use of graphs allows researchers to infer relationships between nodes or regions, evaluate them quantitatively by computing relevant statistical coefficients and use them to characterize and compare different networks or to look for novel and unknown patterns. It also permits to perform analysis using complexity measures useful for moderating data presentation, using graph-theoretic methods, as well as information-theoretic measures [1][2].

2.6 Technical integration

Given the distributed architecture of the modules of the XIM-engine, we use Yet Another Robot Platform (YARP) [20], a framework that makes communication between different processes on possibly different machines easier. YARP has been designed for robots, where usually different processes run concurrently to control the robot and at the same time need to intercommunicate. Another advantage of YARP is that it supports different architectures, operating systems and programming languages. To communicate, a process can open one (or more) ports which has a unique name on the local network. A global name server keeps track of those names and their processes' addresses such that other processes are able to communicate with them. The XIM-engine is composed of several components, sensors and effectors which require intercommunication. For instance, each sensor module has one YARP port that allows either reading or sending information. For all the communication

[2] http://www.unity3d.com

lines, the data sent is pre-defined, thus allowing to easily changing a module provided that it uses the same communication name.

3. Future work: Towards the Adaptive and Contextual behavior in XIM-engine

Building on the histories of several users who have used the system for active exploration of the dataset, the SA, can learn collective user preferences and action sequences. Moreover, it can make associations of users' physiological states to subsets of the data. These trained sequences can be tried out as suggestions to new users, who can choose to either accept or reject them, thereby reinforcing those associations or waning them out. As these associations are learnt, not only on conscious user actions but also on unconscious or implicit behavioral cues, the system can potentially tap into the human subconscious and bring to the forefront an association or insight that would otherwise not have emerged into conscious realization of the user. We plan to further extend our implementation with a Narrative Generator (NG) module capable of structuring the information and the interaction delivered to the user. NG can be implemented as a real-time interactive event sequencer that generates actions plans to drive the user from the actual state (e.g. unfocused) to a goal state previously defined by the SA (e.g. engagement). Essentially, the Narrative Generator will guide the users through the dataset in the XIM environment by scripting the actions they will follow in time.

4. Conclusions

In this paper, we described the technological framework that we designed to support the collection of conscious and unconscious reactions to create an interactive system. The framework described is the XIM-engine, based on the DAC cognitive architecture and designed to control an immersive Mixed-Reality installation (the XIM) where datasets can be projected and analyzed. The novelty of our approach is that we centered the design of the whole system on the human experience and the processes of the mind and the brain that are behind the exploration of complex datasets in space using immersive environment. XIM-engine is a software tool designed to serve this purpose where complex psychological states are measured and evaluated while, at the same time, unconscious user actions are taken into account to drive the system interaction. This makes it unique and not comparable to any other existing system. In this direction, our work in the future is to perform experiments that validate the system proposed, prove the effectiveness of it to speed up the exploration processes and ultimately become a tool for the research community.

5. ACKNOWLEDGMENTS

The work described in this paper has been supported by the European Commission, Seventh Framework Programme under the research grant CEEDs (FP7-ICT- 258749).

6. REFERENCES

[1] Barabási, A. L. (2009). Scale-free networks: a decade and beyond. Science, 325(5939), 412-413.

[2] G. Bell, T. Hey, and A. Szalay, “Beyond the data deluge,” Science (80)., vol. 323, no. 5919, pp. 1297–1298, 2009.

[3] Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience,10(3), 186-198.

[4] J. Lessiter, A. Miotto, J. Freeman, P. Verschure, and U. Bernardet, “CEEDs: Unleashing the Power of the Subconscious,” *Procedia Comput. Sci.*, vol. 7, pp. 214–215, 2011.

[5] P. F. M. J. Verschure, “The complexity of reality and human computer confluence: stemming the data deluge by empowering human creativity,” in *Proceedings of the 9th ACM SIGCHI Italian Chapter International Conference on Computer-Human Interaction: Facing Complexity*, 2011, pp. 3–6.

[6] E. Andre, “Exploiting Unconscious User Signals in Multimodal Human-Computer Interaction,” *ACM Trans. Multimed. Comput.*, vol. 9, pp. 1–5, 2013.

[7] U. Bernardet, S. B. i Badia, A. Duff, M. Inderbitzin, S. Le Groux, J. Manzolli, Z. Mathews, A. Mura, A. Väljamäe, and P. F. M. J. Verschure, “The eXperience Induction Machine: A New Paradigm for Mixed-Reality Interaction Design and Psychological Experimentation,” in in *The Engineering of Mixed Reality Systems*, E. Dubois, P. Gray, and L. Nigay, Eds. London: Springer London, 2010, pp. 357–379.

[8] X. D. Arsiwalla, A. Betella, E. Martinez, P. Omedas, R. Zucca, and P. F. M. J. Verschure, “The dynamic connectome: a tool for large-scale 3D reconstruction of brain activity in real-time,” in *Proceedings of the 27th European Conference on Modelling and Simulation*, 2013.

[9] A. Betella, R. Carvalho, J. Sanchez-Palencia, U. Bernardet, and P. F. M. J. Verschure, “Embodied interaction with complex neuronal data in mixed-reality,” in *Proceedings of the 2012 Virtual Reality International Conference*, 2012, p. 3.

[10] A. Betella, E. Martínez, R. Zucca, X. D. Arsiwalla, P. Omedas, S. Wierenga, A. Mura, J. Wagner, F. Lingenfelser, E. André, D. Mazzei, A. Tognetti, A. Lanatà, D. De Rossi, and P. F. M. J. Verschure, “Advanced Interfaces to Stem the Data Deluge in Mixed Reality: Placing Human (un)Consciousness in the Loop,” in *ACM SIGGRAPH 2013 Posters*, pp. 115:1–115:1.

[11] P. F. M. J. Verschure, B. J. A. Kröse, and R. Pfeifer, “Distributed adaptive control: The self-organization of structured behavior,” *Rob. Auton. Syst.*, vol. 9, no. 3, pp. 181–196, 1992.

[12] P. F. M. J. Verschure, T. Voegtlin, and R. J. Douglas, “Environmentally mediated synergy between perception and behaviour in mobile robots,” *Nature*, vol. 425, no. 6958, pp. 620–624, 2003.

[13] K. Eng, A. Babler, U. Bernardet, M. Blanchard, A. Briska, J. Conradt, M. Costa, T. Delbruck, R. J. Douglas, K. Hepp, and others, “Ada: Constructing a synthetic organism,” in *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on*, 2002, vol. 2, pp. 1808–1813.

[14] J. Wagner, F. Lingenfelser, E. André, D. Mazzei, A. Tognetti, A. Lanatà, D. De Rossi, A. Betella, R. Zucca, P. Omedas, and P. F. M. J. Verschure, “A sensing architecture for empathetic data systems,” in *Proceedings of the 4th Augmented Human International Conference*, 2013, pp. 96–99.

[15] A. Lanatà, A. Armato, G. Valenza, and E. P. Scilingo, “Eye tracking and pupil size variation as response to affective stimuli: A preliminary study,” in *2011 5th International Conference on Pervasive Computing Technologies for Healthcare and Workshops, PervasiveHealth 2011*, 2011, pp. 78–84.

[16] G. Loriga, N. Taccini, D. De Rossi, and R. Paradiso, “Textile Sensing Interfaces for Cardiopulmonary Signs Monitoring,” in *EMBC 2005. The 27rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2005, pp. 7349–7352.

[17] N. Carbonaro, A. Greco, G. Anania, G. Dalle Mura, A. Tognetti, E. P. Scilingo, D. De Rossi, and A. Lanatà, “Unobtrusive Physiological and Gesture Wearable Acquisition System: A Preliminary Study on Behavioral and Emotional Correlations,” in *GLOBAL HEALTH 2012, The First International Conference on Global Health Challenges*, 2012, pp. 88–92.

[18] Z. Mathews, P. F. M. J. Verschure, and others, “PASAR: An integrated model of prediction, anticipation, sensation, attention and response for artificial sensorimotor systems,” *Inf. Sci. (Ny).*, vol. 186, no. 1, pp. 1–19, 2012.

[19] J. Wagner, F. Lingenfelser, and E. André, “The Social Signal Interpretation Framework (SSI) for Real Time Signal Processing and Recognition,” in *Proceedings of Interspeech 2011*, 2011.

[20] P. Fitzpatrick, G. Metta, and L. Natale, “Towards long-lived robot genes,” *Robot. Auton. Syst.*, vol. 56, no. 1, pp. 29–45, 2008.

[21] M. Sanchez-Fibla, U. Bernardet, E. Wasserman, T. Pelc, M. Mintz, J. C. Jackson, C. Lansink, C. Pennartz, and P. F. M. J. Verschure, “Allostatic control for robot behavior regulation: a comparative rodent-robot study,” *Adv. Complex Syst.*, vol. 13, no. 03, pp. 377–403, 2010.