Robot Dream paradigm for Anthropomorphic Social Agent

Vlada Kugurakova[1] and Denis Ivanov[2]

[1] Kazan Federal University vlada.kugurakova@gmail.com
[2] Kazan Federal University denis.999.ivanov@gmail.com

Abstract
In this paper, we discuss the possibilities of integration of the Robot Dream paradigm with Anthropomorphic Social Agent (ASA). Anthropomorphic Social Agent is essentially a user interface specifically designed for simulating the realistic human-to-human interaction. The main idea behind the Robot Dream paradigm is to provide highly realistic social behavior when resources of the machine are limited by outsourcing resource-intensive computation to the cloud. We believe that these two concepts go together nicely.

Keywords: virtual humans, behavioral modeling, cognitive science, computational models, artificial emotions, affective computing

1 Introduction

In our previous work [5], we have proposed a mathematical model of emotional decision making, which will be integrated into intelligent agent as a part of his inner architecture. The model will be firmly based on ideas proposed by Marvin Minski in his *"Model of Six"* [7], Tomkins psychological model and neurobiological studies of Lövheim [6], and also based on Integration of neuromodulators effects on human perception and reasoning and brain rhythms [13]. With interest observing the development of neurobiologically inspired cognitive architecture *NeuCogAr* [16] realizing the need of neurobiological plausibility as part of criteria for highly realistic cognitive architectures [4], we use new paradigm of neuromorphic robot dream [14, 15] most suitable for the implementation of a anthropomorphic social agent, don't forget to raise important points for overcoming the phenomenon of an *"uncanny valley"* [21, 8].

2 Human-machine interactions and the role of anthropomorphism

As bonds between computer and human become more personal, the demand grows for means of communication more closely resembling these between humans. The way humans convey information, however, is much different from how the computers are accustomed to receiving it – humans largely rely on context, undertones, and emotions to augment their voice-transmitted message with more meaning, while computers – and most of the software they are running – are best suited to work with a complete, structured and easy to parse information. How do we build a bridge between these different mechanisms of data processing?

This is where the Anthropomorphic Social Agent comes in. As a user interface, ASA is specifically designed to close the gap between human and computer – it would both decode the language of human for computer and do the opposite.

What does "Anthropomorphic" in ASA mean? A range of things, basically. It is not so simple to define how exactly "human-likeness" can manifest – appearance first comes to mind, but there is more. The way the agent talks, the way he mimics, even the structure and the content of its sentences – all contribute to some kind of "average anthropomorphic score", so it's in our best interest to try to maximize it.

As we have mentioned before in our previous paper [5], as characters become more human-like in their behavior and appearance, discrepancies in appearance and behavior tend to confuse and put the audience in unease, hence, falling into the so-called *"uncanny valley"* [21, 8]. This effect is commonly attributed to people viewing uncanny valley victims as ill and disfigured humans, which provokes a natural negative reaction.

Even though the valley had since become a much debated topic, the exact psychological phenomena behind it remain unclear. The basic premise is also raising some serious questions, such as how exactly should the *"anthropomorphism factor"* should be measured, what qualities of an entity contribute to it, and how do we order the test agents on the human-likeness sliding scale. Besides, it only deals with appearance and does not cover any other possible sides to the definition of anthropomorphism. Hence, it is necessary to expand the definition of the Uncanny valley to include them and to design the consistent criteria that would help to measure and prove the existence (or absence) of this effect. This is a work we are currently undertaking.

3 ASA as an extensible framework

It is up for ASA to respond to multiple challenges involved with anthropomorphism. Since every one of these challenges is a complex problem in it's own right, the mountain of work and research successful prototype of ASA requires seems insurmountable, and the date when we will finally be able to converse with with PC in a natural way drifts away into the ocean of uncertainty.

Fortunately, there is a way to simultaneously develop and test various parts of ASA – a principle of modularity that we suggest it to be put in the basis of ASA's design. What we propose is a framework-like solution, which allows to easily plug-in whichever modules are currently necessary.

At the very core, ASA has two jobs: receiving the input from the user and passing into the machine, translating and reformatting into computer-friendly data language (think of JSON and XML data formats); and giving back the resulting output it the similar process. This raises some questions on what the structured and reformatted input should contain, such as what emotions and mental states could be affected and how to properly annotate them, noting their intensity.

The emotions have a significant influence on rational judgment, social interaction, perception, memory, creativity, learning and many other cognitive functions in humans [9]. Hence, for behavior of ASA to be recognized as affective, the system needs routines that provide information about the emotional processes involved in the decision making which this human-recipient does while interacting with ASA. Following are some examples on technologies that can be used to detect and recognize emotions:

- emotional speech recognition [3];
- facial affect detection [1];
- motion detection from body gesture and physiological monitoring [17].

Since ASA's job is not only to read and interpret human input, but also to show output resulting from machine's computations, it's necessary to correctly display all the same information we previously tried to structurise and understand. There are certain advances in this field already. For example, New Zealand researchers led by Mark Sagar [10, 11, 12] came close to solving the *"uncanny valley"* problem. They use biologically accurate models to animate interactive life-like renditions of human faces.

They are developing a visual modeling methodology for the construction, visualization and animation of neural systems, a novel simulation environment for neural models.

This allows users to create animations and real-time visualizations from biologically based neural network models, allowing simulation effects to be viewed in an interactive context. Such a visual environment is not only suitable for visualizing a simulation; it is also ideal for model development.

A highly detailed biomechanical face model has been constructed from MRI scans and anatomic reference, akin to Wu [20, 19, 18].

Having the ability to accurately display human facial expressions through the interface of the anthropomorphic agent, laying the basis for the biological mechanisms of facial expression, it is necessary that they be read adequately to the context.

4 Common Sense Based Reasoning

Implementing human-like logical reasoning mechanism requires an approach based on common sense. In works Tchitchigin et al. [14, 15] are presented an approach to neurobiologically plausible implementation of emotional reactions and behaviors for real-time autonomous robotic systems. They use metaphors *"day"* and the *"night"* phases of mammalian life. During the *"day phase"* a robotic system stores the inbound information and is controlled by a light-weight rule-based system in real time. In contrast to that, during the *"night phase"* information that has been stored is transferred to a supercomputing system to update the realistic neural network: emotional and behavioral strategies.

To enable (soft) real-time behavior from a robotic system it should run some kind of traditional rule-based control system. We do not constrain type or architecture of this system, it might be some sort of Boolean expert system, fuzzy-logic based or Bayesian-logic based system, possibly augmented with Deep-learning preprocessing cascades for visual and/or audio input channels. In our approach robots control system should not only produce appropriate reaction for input stimulus but also record these stimuli for post-processing during *"night phase"*.

So we propose that robotic system records and stores all input signals for further post-processing. It might store either raw inputs (video and audio streams for example) or some higher-level aggregated form like Deep-learning cascade output (if control system uses such preprocessing mechanisms). And when robot "goes to sleep", it transmits all new recorded inputs to the supercomputer over some network connection. Actually robotic system might not "sleep all night long", it can stay active while supercomputer processes new information, we assume only that the robot regularly "takes a break" to transmit recorded information and later to receive results and update rules of its control system.

Finally, the supercomputer runs realistic spiking neural network that simulates brain regions involved into neuromodulation pathways. This network receives as an input actual signals recorded by robotic system in a raw or pre-processed (aggregated) form. Then the network processes inputs in the simulated time (which might be faster or slower than clock time depending on available computational power).

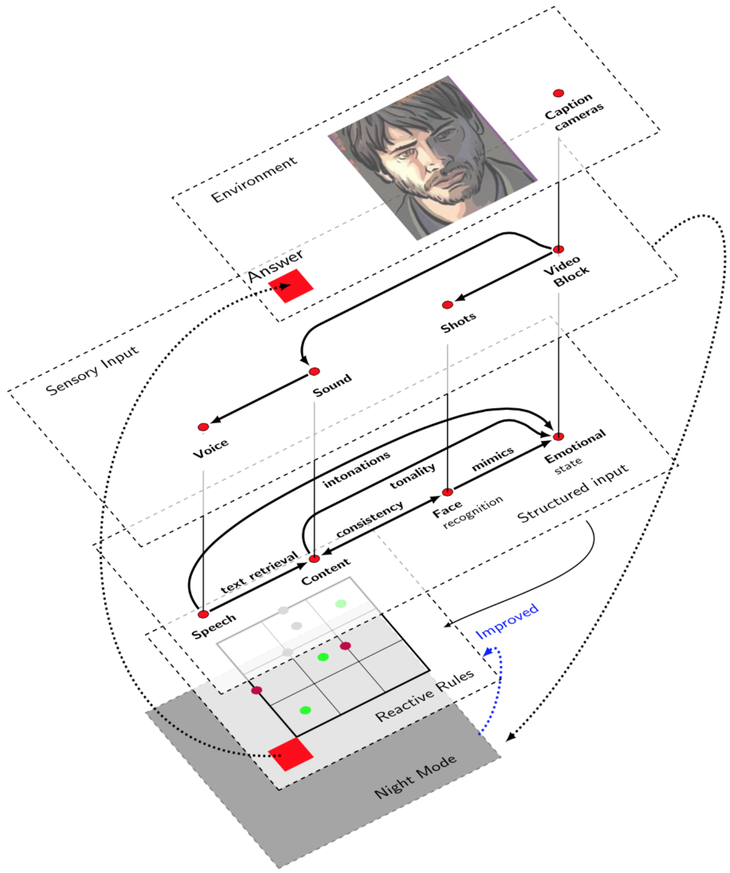

Figure 1: “Day and night” phases from ASA solving.

5 Robot Dream and ASA: possible synergy

In opinion of Dehaene and Naccache [2], a problem of cognitive neuroscience, though empirically challenging, is conceptually simple. “...Human subjects routinely refer to a variety of conscious states. In various daily life and psychophysical testing situations, they use phrases such as *I was not conscious of X*, *I suddenly realized that Y*, or *I knew that Z, therefore I decided to do X*. In other words, they use a vocabulary of psychological attitudes such as believing, pretending, knowing, etc., that all involve to various extents the concept of ‘being conscious’. In any given situation, such conscious phenomenological reports can be very consistent both within and across subjects...” The task of cognitive neuroscience is to identify which mental representations and, ultimately, which brain states are associated with such reports. Within a materialistic framework, each instance of mental activity is also a physical brain state. The cognitive neuroscience of consciousness aims at determining whether there is a systematic form of information processing and a reproducible class of neuronal activation patterns that systematically distinguish mental states that subjects label as "conscious" from other states. These systematic form of information processing can called *"reactive rules"*.

Such *"reactive rules"* should allow to quickly generate or even choose from ready-made index files a suitable answer, just as a average person does not sway to react to the usual situation.

Despite the fact that it is important for us what is behind the appearance of such rules, but it is important to note that for the operation of ASA it does not matter what leads to the emergence of these fast rules. This approach makes it possible to separate the interface and logical parts, which is common for the design of information systems.

If we are to introduce the concept of *"reactive rules"*, which allow quick generation of an appropriate answer or it's selection from ready-made index files, just as a person reacts to a typical situation without thinking, then for ASA functionality it is irrelevant what exactly initiates these rules. Such approach allows us to split interface and logical components, which is typical for system design. Reactive rules creation is under the competence of assigned logical component. Input data processing, interpretation and creation of reactive rules and further associations is handled by assigned logical engine. One of these engines may be developed via *"robot dream paradigm"*, utilizing *NeuCogAr* [16] or congruent cognitive architecture. However, early we stated our adherence to biologically inspired approach.

"Reactive rules" are implemented in *"day phase"*. *"Night phase"* contains the selected cognitive architecture and a separate unit, the module forming, on the basis of its work, an improvement / modification of the *"reactive rules"*. The interface part (ASA) works only with *"reactive rules"*, but transfers all the sensory input received from raw data, and the structured input associated with it to the night phase, for long processing. Since the agent is inherently virtual human and can receive data from the real world only by means of accessible sensors (see. Fig. 1), Raw data is only using video capture from available cameras, in particular from dialogue with a real person communicating with the virtual agent, but when connecting other available sensors (infrared sensors, thermal imagers, ultrasound, vibrations, etc.), the range of raw data it receives can be expanded and somehow interpreted. All subsequent ASA experience, except for the start-up phase of the kindergarten, should be formed under the influence of information from available sensors. The *"reactive rules"* suggest a best match response of the system to a set of structured input, which is a complex of the corresponding ASA reaction – a smooth transition from the initial emotional state expressed by facial expression to a new, with the right intonated sound, corresponding to the emotional reaction to the previous context of the dialogue.

6 Conclusion and future works

We considered a new approach of phase separation – the use of ready-made rules (*"reactive rules"*) for the rapid response of an anthropomorphic social agent and a parallel, but longer and more resource-intensive phase for processing information with the resulting improvement after this "astral thinking".

With lack of positronic brains at the current stage of computer systems' evolution, it is reasonable to use robot dream paradigm for ASA reactive rules calculations and building expected behavioral responses. Furthermore, an unlimited amount of avatars with ASA interfaces, distinct behavioral masks and shared experience are able to connect to a single *"association* *processor"*. At this point, in *"night phase"* it is necessary to develop *"reactive rules"* , used in *"day phase"*.

7 Acknowledgment

This work was funded by the subsidy of the Russian Government to support the Program of competitive growth of Kazan Federal University among world class academic centers and universities.

References

[1] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan. Analysis of emotion recognition using facial expressions, speech and multimodal information. In *ICMI'04 - Sixth International Conference on Multimodal Interfaces*, pages 205–211, 2004.

[2] S. Dehaene and L. Naccache. Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79(1-2):1–37, 2001.

[3] M. El Ayadi, M.S. Kamel, and F. Karray. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, 44(3):572–587, 2011.

[4] V. Kugurakova, M. Talanov, and D. Ivanov. Neurobiological plausibility as part of criteria for highly realistic cognitive architectures. In *Procedia Computer Science*, volume 88, pages 217–223, 2016.

[5] V. Kugurakova, M. Talanov, N. Manakhov, and D. Ivanov. Anthropomorphic artificial social agent with simulated emotions and its implementation. In *Procedia Computer Science*, volume 71, pages 112–118, 2015.

[6] H. Lövheim. A new three-dimensional model for emotions and monoamine neurotransmitters. *Medical hypotheses*, 78(2):341–348, 2012.

[7] M. Minsky. *The Emotion Machine and The Society of Mind,NY*. Simon & Schuster, 2006. ISBN: 0743276647.

[8] M. Mori, K. F. MacDorman, and N. Kageki. The uncanny valley. *IEEE Robotics and Automation Magazine*, 19(2):98–100, 2012.

[9] R. W. Picard. *Affective computing for future agents*, volume 1860 of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2000.

[10] M. Sagar. *Creating Models for Simulating the Face*, pages 394–394. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[11] M. Sagar, P. Robertson, D. Bullivant, O. Efimov, K. Jawed, R. Kalarot, and T. Wu. *BL: A Visual Computing Framework for Interactive Neural System Models of Embodied Cognition and Face to Face Social Learning*, pages 71–88. Springer International Publishing, Cham, 2015.

[12] M. Sagar, M. Seymour, and A. Henderson. Creating connection with autonomous facial animation. *Commun. ACM*, 59(12):82–91, December 2016.

[13] M. Talanov, J. Vallverdu, S. Distefan, M. Mazzara, and D Radhakrishnan. Neuromodulating cognitive architecture: Towards biomimetic emotional ai. *AINA*, pages 587–592, 2015.

[14] A. Tchitchigin, M. Talanov, L. Safina, and M. Mazzara. *Robot dream*, volume 58 of *Smart Innovation, Systems and Technologies*. 2016.

[15] A. Tchitchigin, M. Talanov, L. Safina, and M. Mazzara. Neuromorphic robot dream. *BioNanoScience*, 7(1):199–200, 2017.

[16] J. Vallverdu, M. Talanov, S. Distefano, M. Mazzara, M. Manca, and A. Tchitchigin. Neucogar: A neuromodulating cognitive architecture for biomimetic emotional al. *International Journal of Artificial Intelligence*, 14(1):27–40, 2016.

[17] S. Walter, S. Scherer, M. Schels, M. Glodek, D. Hrabal, M. Schmidt, R. Bck, K. Limbrecht, H. C. Traue, and F. Schwenker. *Multimodal emotion classification in naturalistic user behavior*, volume 6763 LNCS of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2011.

[18] T. Wu. *A Computational Framework for Modeling the Biomechanics of Human Facial Expressions*. The University of Auckland, Auckland, New Zealand, 2014. Ph.D. thesis.

[19] T. Wu, A. Hung, and K. Mithraratne. Generating facial expressions using an anatomically accurate biomechanical model. *IEEE Transactions on Visualization and Computer Graphics*, 20(11):1519–1529, 2014.

[20] T. Wu, A.P. . Hung, P. Hunter, and K. Mithraratne. Modelling facial expressions: A framework for simulating nonlinear soft tissue deformations using embedded 3d muscles. *Finite Elements in Analysis and Design*, 76:63–70, 2013.

[21] Y. Yamada, T. Kawabe, and K. Ihaya. Categorization difficulty is associated with negative evaluation in the uncanny valley phenomenon. *Japanese Psychological Research*, pages 20–32, 2013.