Building Machines in the Biological Way: Reactivity and Emotions First, Reasoning Later

Horia-Nicolai L. TEODORESCU[1,2], Silviu Ioan BEJINARIU[2]

[1]Gheorghe Asachi Technical University of Iasi, [1,2] hteodor@etti.tuiasi.ro, [2]Institute of Computer Science from Romanian Academy, Iasi Branch, Carol I nr. 8, Iasi, Romania

Abstract. We expose a vision for affective computing and a deployment of firmware to support the vision. In interactions, humans and animals perceive and convey expressions of emotions and intentions, in the first place. Computers have no means to express either of them. The current approach to endow machines with interactivity is to add software applications that mimic humans' abilities. We suggest a different approach, based on embedding essential elements of the user's state recognition deep at the UEFI level.

Keywords. Affective computing, emotions, firmware interface, perceptive I/O

1 Introduction

Users enjoy sensors and interfaces when these improve communication and enhance experience. However, sensors and interfaces are largely seen today as separated elements. We propose sensor-interface combinations that determine and adapt to the general state and current mood of the user. The affective and state characteristics are used to better control interfaces, to help the subjects to apprehend their state, and to make known their mood and general state to fellows during dialogue and activities.

State and mood controlled interfaces and dialog for adaptation of the interface to the current state of the user could tremendously enhance the communication and interaction experience, and for increasing self-awareness. This paper proposes to synergize the sensors and the interfaces. The proposal is in line with the (loose) definition due to [1], namely "Cyber-physical systems (CPSs) are the synergy of the physical world with the cyber world," as well as with the aim of CPSs, "The … objective is to attain a better quality of life by monitoring, affecting, and controlling the physical world within which we live using the capabilities of the cyber world. These capabilities encompass sensing, computation, communication, and actuation tasks" [1].

It is well known that the quality and level of interaction impacts work efficiency and the human experience in general. We propose, extending some general ideas presented in [2], in line with the proposals in [3], the concept of a HUman-MAchine INter-SYStem (HUMAINSYS), where the machine is able to recognize the pattern of the user's behavior, to determine the human state and emotions and to react accordingly by suitably changing the operation of the interfaces. By increasing the capabilities of collecting information on the user and of determining the users' state, mood and emotions, we aim to enable the machine to produce an apprehensive dialog and adaptive interactions. The proposal goes to the core of the man-machine interaction addressing the root elements of the interfacing with the user. A syntax of the user-machine interrelationship should be built to endow the machine with the required interactivity in image, speech and overall reactivity. The concept of the proposal refers to a way of integrating extant and new interfaces and of adding intelligence to them. The input interfaces are endowed with functionalities for detecting the human physical state and personal behavior, in order to improve the interaction, in view of enhancing the work efficiency and users' experience by self-adapting both the input and output interfaces. The integration of interfaces, according to the proposal concept, refers to the influence that the operation of the input interfaces have on the output ones. A key idea is the entrenching of the new functionalities of the interfaces starting with driver and UEFI level, such that the machine itself inherently performs the adaptation and interaction, instead of an application-level program.

The organization of this position paper is as follows. In the first part, we revisit the broad framework of human-machine interaction. The next two parts are devoted to building a vision on the requirements and essential tools needed by future machines for they are able to interact meaningfully with humans. The fourth part shows how to address the development of machines in the nature's way, first developing perceptive I/O and a computational equivalent of the amygdala at the hardware and software level. This part puts into the central place of the computer the interfaces and a generalization of the current Unified Extensible Firmware Interface (UEFI, which is replacing BIOS) and shows how to build the other sections of the system around the generalized UEFI. The final part is a glance on future potentialities and developments.

2 Research directions

Starting with Babbage, computers were imagined as "computing machines". Yet, the computation ability is the last addition to the power of the brain in the evolution of species, and a non-existing feature in most species. In many respects, computers constitute prostheses for helping us in our poorer skill – making fast computations. Life has not evolved by first endowing animals with computation capacity, and then with senses and emotivity; it was developing in the first place basic reactivity, perceptions, sensations, and basic emotional states, as fear and stress. Computers are deprived of all of these, because of the opposite direction they were developed and progressed. To appreciate the difference, imagine that the first cells and structure developed in the womb would be the neo-cortex cells and the neo-cortex structure, from which bones and limbs are created.

Research has been made to detect the human emotional states using various input interfaces, such as video cameras for monitoring facial expressions and gestures [4-8], microphones for acquiring speech signals and for assessing (partly language-independent) the emotional state [7-12], and sensors for monitoring the bio-signals [13-15]. Fusion methods for multimodal emotion recognition are used in most researches today, for combining various sources of information, including audio, video, and physiological signals [16-19]. Yet, the sensors, input interfaces and output interfaces are still largely seen as separated elements. We aim to make the computer interactions more humanly-like by endowing them with an understanding of the user.

There are three directions of recent progresses that invite research in improved man-machine interactivity: large progresses in state and emotion recognition with computer applications; significant progresses in sensor technology, and evolution at the level of the hardware and software.

3 Expected progress

The main directions that should be followed in the development of the proposal are the emotion detection and the interfaces functionality redesign. In order to improve the results in the field of affective interaction between the computer and the human user, we aim to add new functionalities to the existing computers at the core level that will make the computer more responsive to the user state. Therefore, we want to change the current behavior of the computers and of the computing systems by making the current interfaces adaptive to the way that they are manipulated and by adding new functionalities that are able to respond to the user emotional state. By integration of the new features at the hardware level, the application developers would gain access to the system resources that monitors the user state.

Drawing a comparison between humans' nervous system and computing machines, several similarities appear, as reasoning power, subsystems for gathering information from the interlocutor (e.g., keyboard and microphone, for the machine, vs. sensing organs), output subsystems for communicating information. Yet, while computers seem to have several basic capabilities of the humans, they still lack the ability to grasp the human general / affective state, which are essential in establishing an efficient, higher level communication with a human– an ability that even a toddler has. Consequently, they are unable to produce meaningful emotive expressions, an ability that is essential in an interaction with humans. At the level of system architecture, there are more obvious differences, while the functional differences above mentioned find some explanations. Computers do not have a separate unit, as the brain has the amygdala, to produce and react to emotions and moods. Moreover, there are no 'shortcut connections' between output devices and sensors, as there is between the effectors and the sensing organs – where a shortcut means faster connections through lower level parts of the nervous systems. With user's emotions detection in computers today at the application level, and with modest panoply of sensing devices, emotion recognition and response are inherently slow, difficult to perform, and primitive, even for basic emotions, while too sophisticated in implementation. No emotion detection and no responses are produced at the kernel level or at the interfaces proper level today. We see five main limits of the current machine interaction capabilities [2, 3]:

i) Provide, as a standard feature of machines, state, mood and emotion detection of the users and adapting its outputs accordingly.

ii) To create inherently sensing and responsive machines by rethinking the user's state sensing at the level of its input devices and of the basic software.

iii) Solve the issue of shortcut connections between interfaces in a way similar to the one in living organisms.
iv) Increment the capability of the machine to grasp the users' state and emotions.
v) Produce a richer reactivity of the machine to users' state and emotions.
vi) An enrichment of the sensing capabilities should be addressed.

We propose to look at the human-machine system as a whole by making the machines responsive to the user state and affect, also taking into account the surrounding environment, whenever relevant.

4 Standard interfaces

Several basic features that process the data coming from the devices and sensors that are designed to monitor the human user parameters should be included at the core level of the system that for personal computer is UEFI. A novelty of the proposal is that the user's state evaluation and the output devices parameters adjustment are partly performed at UEFI and devices drivers level. The human is seen by the machine as a first rank interface. This means that the "standard interfaces" would work according to how the "main interface" – the human – is characterized. For this, the UEFI should be re-thought to constantly take into account the user. The modified UEFI should take in charge the adaptation of the input and output interfaces and the whole machine behavior to the human-interface.

The new UEFI could be upgraded to include camera or voice operation. This functionality is used for user behavior recognition as an early stage of verification and adaptation before loading the operating system. The integration at the driver level of the software component that is performing emotion recognition and controls the output interfaces (ERCI) is feasible and may bring benefits to the proposed tasks.

The proposed responsive UEFI, which will be named User UEFI (U2EFI), together with improved drivers, should identify the users and adapt to their state. By integrating the sensors in the input and output interfaces, the human general state and mood should be determined from biological signals, posture, movements, voice, and gestures. Depending on the resources required for the computer ability to respond to user physical states, the complex features, such as the intelligence that recognize the user emotional state may be integrated at the drivers' level. Among the tasks involving the keyboard/mouse drivers for improving the collection of information on users' working patterns during standard tasks are:

- keyboard activity monitoring by evaluation of:
  - the time between two successive keys pressing – for example, if it is higher than a specified time (TBD), the user may be tired, otherwise she/he is in normal state;
  - the spreading of the above time gives information about the user state – for example higher values for the spreading may mean abnormal user states.
  - the keys hold time – for example, higher values may mean tiredness.

Because the user behavior analysis is supposed to focus at the first stage mainly of words writing, the keys that are not letters are excluded from the analysis. Word writing is detected if time for three successive keys strokes is lower than a specified time.

• mouse activity monitoring by evaluating:
  - the precision of positioning in a clicking point – for example if the precision is low may mean that the user is tired;
  - the position changing during the button is pressed – for example if the mouse is moved on short distances during clicks the user may be nervous;
  - the movement speed – e.g., if the speed is low, the user may be bored or tired;
  - the trajectory between two points gives information about the user identity.

We intend to make functional interconnections between input and output interfaces to obtain the influence of one operation on the operation of the others. The system would detect the user state by evaluating the way the input interfaces are manipulated and by using non-invasive sensors embedded in the interfaces. As a matter of feasibility example of state recognition, we designed and developed a database to store information that describe the work sessions on a computer. The aim is the analysis of the relationship between a person's general state and the way he interacts with the computer. For every work session, the data base contains (see Fig. 1):

- information about the operator's general status (relaxed, tired, nervous, calm, mood, moodiness, melancholy) and health status,
- sequences of mouse and keyboard events arising from the computers usage.

The events recording starts after the operator gives his consent and is performed using a background application that allows a normal usage of the computer. The database contains sequences of input events as:

- mouse events: event code, time, mouse position, name of the active application (AA) that actually process the event;
- keyboard: event code, time, key code, functional key properties, and AA name.

The application allows determining when the events are grouped in sequences.

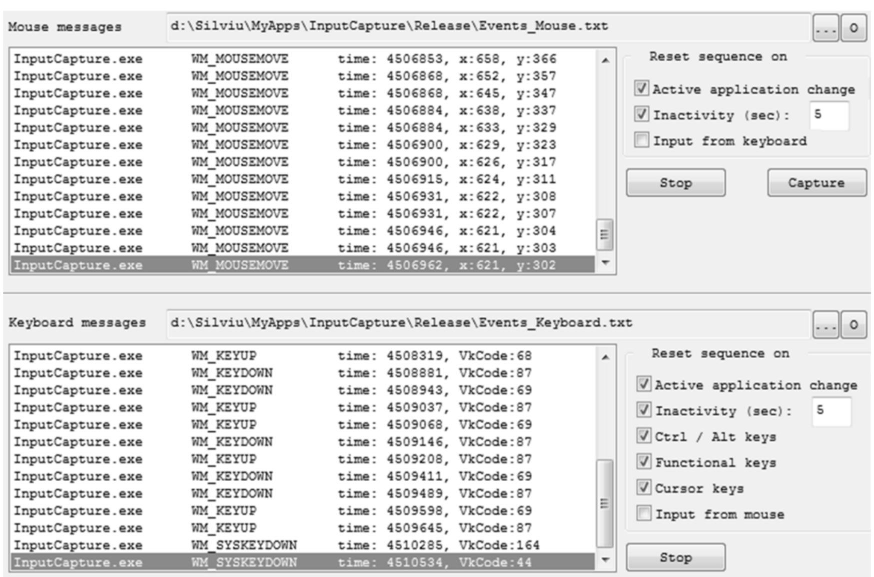

Fig. 1. Main window of the recording application

After determining the state and personality traits of the user, the next task is the design of drivers for a screen that changes in response to user's state, mood, emotions. We suggest controlling the output peripherals devices by:
- contrast and brightness adjustment and screen color management;
- suitably changing the desktop gadgets and desktop background image;
- starting the narrator automatically – for example when the user is tired - and adjust narrator "mood" (prosody, loudness and other parameters);
- manipulate the facility "Make it easier to focus on task";
- automatically add/remove/turn on-off assistive technologies (e.g., sound assistance);
- automatically adjust the repeat delay for keystrokes, the double-click delay for the mouse, and the sound amplitude and tone balance;
- automatically set the preferences of the user account.

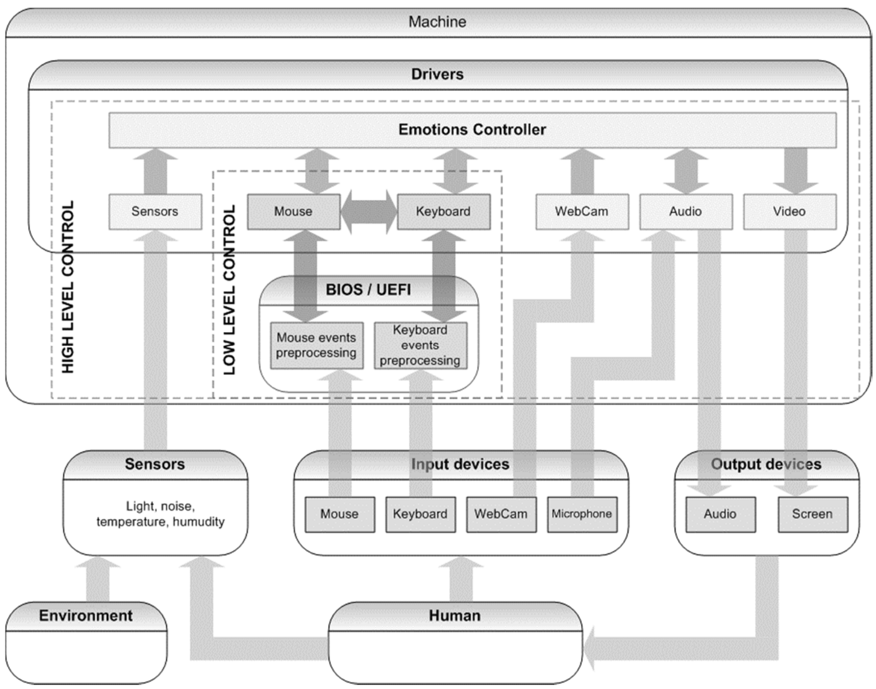

Fig. 2. Integration of the human-interface features on a personal computer architecture

The above mentioned features should be integrated at the UEFI/BIOS level in a computationally efficient manner. The user should have the possibility to switch off the computer responsiveness to his affective state and to customize the computer response when the emotional states are detected. Also, because processing of the video image may raise privacy problems, this feature will be controlled independently. The system will operate on two different levels, see Figure 2, namely:

Low level. In this level, only the basic input devices, mouse / touchpad and keyboard should be considered. Their functional parameters, such as mouse double-clicks delay, mouse speed, keyboard auto-repeat delay should be automatically adapted to the user's state. On this level, the user's state evaluation and parameters adjustment is performed at UEFI / drivers level, without involving other devices or applications.

High level. At this level, a more accurate evaluation is performed using external sensors to obtain information about the users' biological state and input devices, as microphone or webcam for user recognition and gesture analysis.

The "Emotions controller" is the main component on this level. It evaluates the input information, makes decisions, and changes the output devices functional parameters by sending commands to the drivers. The core of the Emotions Controller is an expert system based on rules, which makes decisions using the received information. In case of voice recognition and gesture analysis, the input information may be processed by external applications specialized for sound and image processing, analysis and classification. There are several issues to solve:

- The low level component must be enough fast, while low resources consuming, to be implementable at BIOS / U2EFI and driver level;
- To avoid slowing down the system, the high level component should record the received information in real time, but the effective processing should start with low priority only when the processor is in idle state;
- The designed components must be enough secure for not inserting vulnerabilities in the operating system;
- The external sensors should be connected through the Universal Serial Bus (USB), and specific sensor drivers have to be implemented.
- The use of external sensors, microphone and webcam must be made optional.

Further features of the proposed system could include: emotion recognition from information fusion of speech, spoken interaction, face mimic, gestures, and biological signals; social signal interpretation; human eye-gaze detection and interpretation; complex emotions representation for choosing the suitable behavior for interfaces; monitoring environmental parameters, such as temperature, humidity, and light level.

5 Conclusions

We proposed a radically new approach to the embedding of the "communication intelligence" in computing systems (computers, tablets, phones and the like), deep in the root of the OS. A systematic analysis of the requirements for developing an UEFI that is responsive to the user has been laid down and several steps of development have been sketched. We exemplified the proposal for Unified Extensible Firmware Interface (UEFI) in connection with Windows, but the same principles apply to whatever other systems. The proposal for Unified Extensible Firmware Interface (UEFI) was designed in connection with Windows, but the same principles may be applied to other platforms like Android™ and Windows RT / Mobile™.

Acknowledgements. We acknowledge the participation of Dr. M. Hulea in several discussions on the topic; however, he has not contributed to the ideas in this paper, or to its writing.

The research is connected to the priority research Cognitive Systems of the Romanian Academy. We thank colleagues R. Luca and A. Ciobanu for corrections and for suggestions.

References

1. Ahmad T. Al-Hammouri, A comprehensive co-simulation platform for cyber-physical systems. Computer Communications, Vol. 36, Issue 1, 2012, pp. 8-19.
2. H.N. Teodorescu, Interrelationships, communication, semiotics, and artificial consciousness, in: what should be computed to understand and model brain function, Ch. 6, Editor: T. Kitamura, World Scientific, 2001, pp. 115-148.
3. H.N. Teodorescu, Affect, state, & mood interface for apprehensive dialog and adaptive interactions and control (abstract), CHIST-ERA Conf., Edinburgh 5-6 Jul 2012, p. 13.
4. Y. Sun, N. Sebe, M. S. Lew, T. Gevers, Authentic emotion detection in real-time video, computer vision in human-computer interaction, LNCS, 2004, Vol. 3058/2004, pp. 94-104.
5. J.F. Cohn, Foundations of human computing: facial expression and emotion, Proc. Eighth ACM Int'l Conf. Multimodal Interfaces (ICMI '06), 2006, pp. 233-238.
6. J. C. Martin et al., Multimodal complex emotions: gesture expressivity and blended facial expressions, Int. J. Human. Robot. Vol. 03, Issue 03, 2006, pp. 269-291.
7. Z. Zeng, G. Roisman, and T. Huang, A survey of affect recognition methods: audio, visual, and spontaneous expressions, IEEE Trans. PAMI, Vol. 31, no. 1, 2009, pp. 39-42.
8. Z. Zeng et al., Audio-visual affect recognition in Activation-Evaluation Space. Proc. ICME 2005, IEEE Int. Conf., 6-8 Jul, Amsterdam, Netherlands, 2005.
9. L. Devillers, N. Campbell, Special issue "Affective speech in real-life interactions", Computer Speech & Language, Vol. 25, No. 1, 2011, pp. 1–3.
10. H.N. Teodorescu, M.S. Feraru, A study on speech with manifest emotions. In Text, Speech and Dialogue, Matousek, V; Mautner, P. (Eds), Proc. 10th Int Conf Text, Speech, and Dialogue, Pilsen, Czech, Sep 3-7, 2007. Lecture Notes in AI, Vol. 4629, pp. 254-261.
11. H.N. Teodorescu, M. Feraru, Analyzing emotions in spoken Romanian. Proceedings of the Romanian Academy Series A-Mathematics Physics Technical Sciences Information Science, Vol. 8, Issue 2, pp. 161-168, 2007.
12. W. Rodriguez, H.N. Teodorescu, F. Grigoras, et al., A fuzzy information space approach to speech signal non-linear analysis. International Journal of Intelligent Systems, Vol. 15, Issue 4, pp. 343-363, Apr 2000.
13. R. Picard, E. Vyzas, and J. Healy, Toward machine emotional intelligence: Analysis of affective physiological state, IEEE Trans. PAMI, vol. 23, no. 10, 2001, pp. 1175-1191.
14. K. H. Kim, S. W. Bang, S. R. Kim, Emotion recognition system using short-term monitoring of physiological signals, Med Biol Eng Comput, Vol. 42, no. 3, 2004, pp. 419-427.
15. P. Abhang, S. Rao, B. W. Gawali and P. Rokade, Emotion recognition using speech and EEG Signal – A Review, Int. J. Computer Applications, Vol. 15, No. 3, Feb 2011.
16. J. Wagner, F. Lingenfelser, E. André, J. Kim, T. Vogt, Exploring fusion methods for multimodal emotion recognition with missing data, IEEE Trans. Affective Comp, vol. 2, No. 4, 2011, pp. 206-218.
17. T.W. Bickmore, R.W. Picard, Establishing and maintaining long-term human-computer relationships, ACM Trans. Comput-Hum Inter, Vol. 12, no. 2, 2005, pp. 293-327.
18. R. W Picard, J. Klein, Computers that recognise and respond to user emotion: theoretical and practical implications, Interacting with Computers, Vol. 14, No. 2, 2002, pp. 141-169.
19. N. Fragopanagos, J.G. Taylor, Emotion recognition in human-computer interaction, Neural Networks, Vol. 18, Issue 4, May 2005, pp. 389-405.