2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

Emotional Architecture for the Humanoid Robot Head ROMAN

Jochen Hirth
Robotics Research Lab
Department of Computer Science
University of Kaiserslautern Germany
Email: j_hirth@informatik.uni-kl.de

Norbert Schmitz
Robotics Research Lab
Department of Computer Science
University of Kaiserslautern Germany
Email: nschmitz@informatik.uni-kl.de

Karsten Berns
Robotics Research Lab
Department of Computer Science
University of Kaiserslautern Germany
Email: berns@informatik.uni-kl.de

***Abstract*— Humanoid robots as assistance or educational robots is an important research topic in the field of robotics. Especially the communication of those robots with a human operator is a complex task since more than 60% of human communication is conducted non-verbally by using facial expressions and gestures. Although several humanoid robots have been designed it is unclear how a control architecture can be developed to realize a robot with the ability to interact with humans in a natural way. This paper therefore presents a behavior-based emotional control architecture for the humanoid robot head ROMAN. The architecture is based on 3 main parts: emotions, drives and actions which interact with each other to realize the human-like behavior of the robot. The communication with the environment is realized with the help of different sensors and actuators which will also be introduced in this paper.**

***Index Terms*— humanoid robot head, emotional architecture, behavior based control**

I. INTRODUCTION

The combination of intelligent machines and emotions is a topic of research for several decades. M. Minsky [1] told in his book "Society of Mind":

> *The question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without any emotions.*

The realization of emotions should be a central aspect of any intelligent machine which wants to communicate with humans. A major part of human communication is expressed non-verbally via gestures and facial expressions. Those expressions are mainly influenced by the current mood generally spoken the emotional state of the human. Therefore it is necessary to include the presentation and even the recognition of emotions into humanoid robots.

Worldwide, several research projects focus on the development of emotional architectures for robot human interaction (e.g. see [2] or [3]). Certainly Kismet [4], [5] and WE4 [6], [7] are some of the best known projects in this area. The base of Kismet's architecture are 3 drives. Depending on sensor data these drives calculate their content. The most discontent drive determines the group of possible behaviors. The main disadvantage of the selected approach is its inflexibility. That means if a new drive should be integrated the complete system has to be changed. The same problem exists at WE4. A better approach is presented in [8], [9]. This architecture is more flexible because the drives determine the action of the robot. They use a "maximum fusion" so that the most discontent drive controls the robot. A new drive only had to be integrated into the fusion. The problem with this architecture is that the drives have no priority that means more important drives have no chance to inhibit less important ones. This is a big problem in the usage of the robot for a certain assistance assignment.

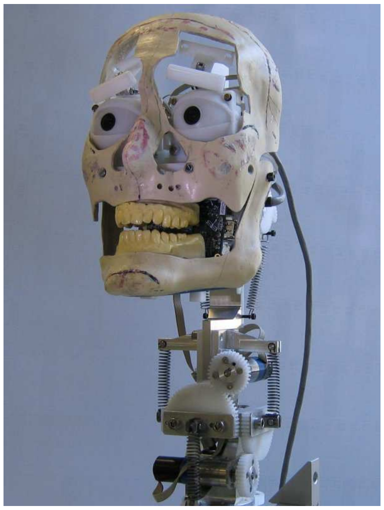

Fig. 1. The humanoid robot head "ROMAN" (ROMAN = RObot huMan interAction machiNe) of the University of Kaiserslautern.

II. SYSTEM DESIGN OF ROMAN

The following section will present the mechanical design of ROMAN including the basic head, eyes and neck construction as well as the computer architecture and sensor system.

A. Mechanics

The mechanics of the humanoid head (see Fig. 1) consists of a basic unit of mounting plates which is fixed to the 4 DOF neck. These plates are the mounting points for the eyes, the servo motors, and the cranial bone consisting of lower jaw, forehead and the back of the head. The artificial skin of the robot is glued onto the cranial bone and can be moved with 8 metal plates, which are connected to 10 servos via wires. The positions of theses movable metal plates are selected according to Ekman‘s action units. The plate areas as well

Fig. 2. Artificial eyes with Dragonfly cameras and Faulhaber stepper motors. The eyes are build of the lightweight material POM (PolyOxyMethylene) with a weight of 150g per eye.

as its fixing positions on the skin and the direction of its movement are optimized in a simulation system according to the basic emotions which should be expressed. Additionally, a single servo motor is used to raise and lower the lower jaw.

The eye has a compact and lightweight design, so that it could be included in the restricted space of the head frame. The eyeballs can be moved independently up/down and left/right. The upper eyelid can also be moved. This is necessary for the expression of specific emotions. In comparison to the human eyeball which has a diameter of $23mm$ to $29mm$ the eyeball of the robot has a diameter of $46mm$. This extension was necessary because of the size of the used Dragonfly cameras, which are included in the eyeball. The complete construction of one eye has a weight of $150g$ including the motors and gears. The minimal distance between the two eyes is $65mm$. The eyes are able to move $\pm 40^\circ$ in horizontal and $\pm 30^\circ$ in vertical direction and can be moved from left to right or top to bottom in about $0.5s$. The upper eyelid can be moved 70° down- and 10° upwards from the initial horizontal position. A blink of an eye can be realized in about $0.4s$.

The neck has 4 active DOF (degree of freedom). For the design of the neck basic characteristics of the geometry, kinematics and dynamics of a human neck are considered. From the analysis of the human neck a ball joint could be selected for the construction of the 3DOF for basic motion. Because of the high construction effort for the realization of such a joint a serial chain similar to Cardan joint solution was applied. The first degree of freedom is the rotation over vertical axis. The range of this rotation for artificial neck was assumed as $\pm 60^\circ$. The second degree of freedom is the inclination of the neck over horizontal axis in the side plane (range of motion $\pm 30^\circ$). The third degree of freedom is the inclination of the neck in frontal plane. It is rotating around the axis which is moving accordingly to the second degree of freedom (range of motion $\pm 30^\circ$). In addition there is a 4th joint used for nodding ones head (range of motion $\pm 40^\circ$). The axis of the fourth joint is located next to the center of the head to realize a rotation along the heads pitch-axis. The rotation is assisted by two springs with a maximum force of $20N$ each since the center of gravity is placed in the front part of the head.

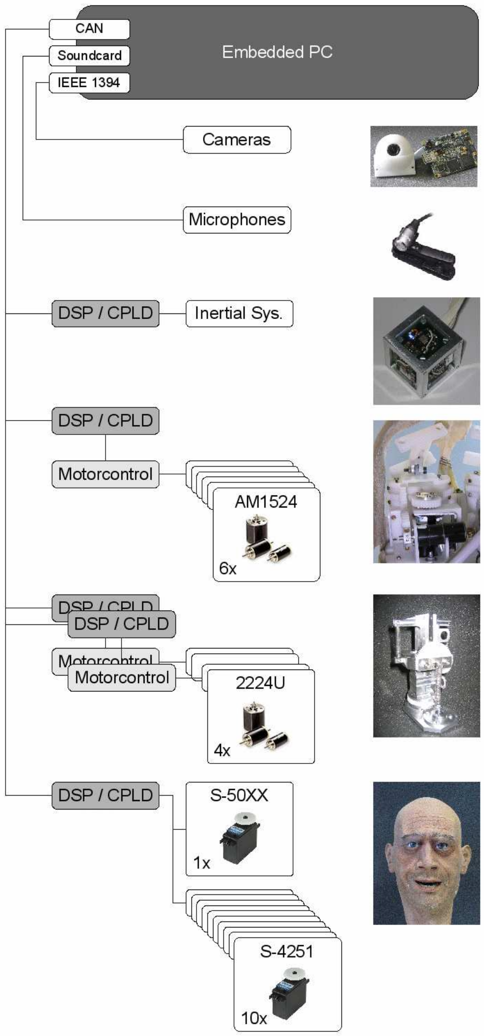

Fig. 3. Hardware architecture including sensor systems and actuators as well as all necessary connections to the embedded computer.

B. Computer architecture and sensor system

A design goal of the ROMAN project is the integration all mechanical and most electronic components inside the robot head. The sensor system of ROMAN consists of stereo vision cameras, microphones and inertial system. In Fig. 2 the eye construction is shown including the Dragonfly cameras and the motor units for the control of the 3 DOF. The inertial system and a smell sensor will be integrated in the near future. Microphones which will be used for sound localization will be directly connected to the sound card of an embedded PC. Fig. 3 shows an overview of the hardware system including all necessary connections to sensors and actuators.

The actuator system of the robot consists of 21 different motors including electric, stepping and servo motors. All motors are connected to DSPs. The 6 stepping motors for the eyes (Faulhaber AM 1524) are controlled by a single DSP. 2 additional DSP's are needed to control the 4 electric motors (Faulhaber 2224 gear ratio 246:1) of the neck. In

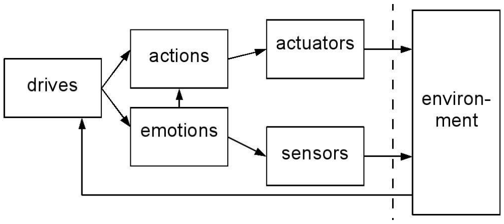

Fig. 4. The design concept of the emotional architecture. The information flow of the main module groups (drives, emotions, actions, sensors, and actuators) is presented.

combination with the precise encoder a accurate positioning system can be realized. A final DSP controls the 11 servo motors which move the artificial skin.

III. THE EMOTIONAL ARCHITECTURE

The idea of our project is the design and implementation of an emotional architecture which allows an easy integration of new drives, sensors, actors and behaviors. Due to these requirements 5 basic modules are specified: sensors, drives, emotions, actions and actuators (see Fig. 4). As pointed out in the last section the sensor system of ROMAN detects the environment based on a stereo camera system and microphones. Based on this information the drives calculate their satiation. Depending on the satiation the emotional space is modified and the actions are selected. The emotion module calculates the emotional state of the robot. This state influences the actions. That means that in our architecture the drives determine "what to do" and the emotions determine "how to do". Complex actions are movements like: looking at a certain point, moving the head at a certain point, moving the eyes at a certain point, eye blink, head nod and head shake. These actions consist of several simple movements like: head up, head down, left eye up, left eye down, etc. These actions select and influence system behaviors like: man-machine-interaction or the exploration of the environment.

The drives shown in Fig. 4 represent a group that consists of several drive modules. These drives are implemented based on our behavior node concept which is used for the control of all our robots see [10]. In Fig. 5 the drive module is presented. It has 2 inputs one for sensor data and one for inhibition which could be determined by other drives and 3 output parameters one for the inhibition of other drives one for the emotional state and one for the action activation. It also has two internal functions. The first function $t()$ calculates the discontent of the drive depending on sensor data. The other function is the activity $a(t(), i)$ (Eq. 1), where $i \in [0,1]$ means the inhibition input. The activity is a piecewise defined function in which the interval$[0, t_0]$ means the inactive area of the drive, $[t_0, t_1]$ means the area in which the activity is calculated based on the $sigmoid$ function Eq. 2 and $[t_1, 1]$ means the satisfaction area. The codomain for the discontent function just as for the activity function is $[0, 1]$. The correlation of the activity and the discontent of a drive is shown in Fig. 5.

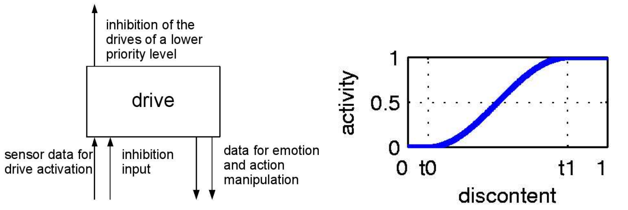

Fig. 5. Left: The single drive node. Right: The general correlation of the activity of a drive and the discontent of a drive (inhibition $i = 0$).

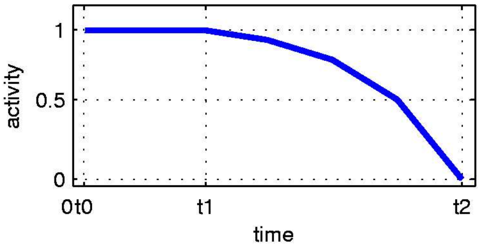

Fig. 6. Behavior of the activity of a drive. t_0 = occurrence of a discontent causing stimulus, t_1 = the drive is more saturated, t_2 = the drive is saturated and its activity is 0.

$$a(t(), i) = \tilde{a}(t()) \cdot (1 - i)$$

$$\tilde{a}(t(s)) = \begin{cases} 0 & \text{if } t() < t_0 \\ 1 & \text{if } t() > t_1 \\ sigmoid(t()) & \text{else} \end{cases} \quad (1)$$

$$sigmoid(x) = \frac{1}{2} + \frac{1}{2} \cdot sin\left[\pi \cdot \left(\frac{x - t_0}{t_1 - t_0} - \frac{1}{2}\right)\right] \quad (2)$$

As described the drive gets active if the discontent value is over t_0. The drive than calculates parameters which change the emotional state and which select the actions of the robot. The aim of the drive is to reach a saturated state by the selection of the actions. If the saturation of the drive is getting higher the activity of the drive is getting lower. When after a certain time the drive is saturated. The activity will turn to 0 again (see also Fig. 6).

To expend the number of drives easily a hierarchical drive system is implemented. That means every drive has a certain priority level. The drives of a higher level inhibit the drives of a lower level. This is realized with the inhibition input of our drive nodes. Because of a fusion of the drives output only the drive with the highest activity is able to determine the actions of the robot (see Fig. 7). This means that if a new drive should be added to the architecture only the priority level has to be determined. Then the connections of the inhibition output of the drives of the next higher priority had to be connected to the inhibition input of the new drive. Its inhibition output has to be connected to all drives of the next lower priority level.

The actual emotional state of the robot is defined according to the emotion cube shown in Fig. 8. In this cube

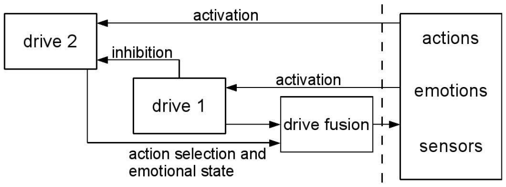

Fig. 7. The interaction of two drives, where ”drive 1” has a higher priority level than ”drive 2”.

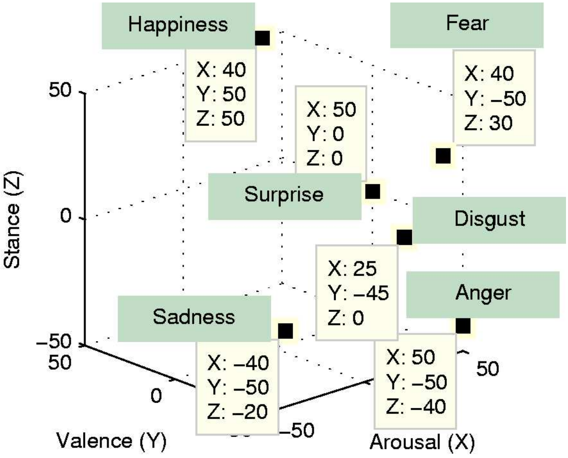

Fig. 8. The cube for the calculation of ROMANs emotional state.

the 6 basic emotions (anger, disgust, fear, happiness, sadness and surprise) [11] are represented. The calculation of the emotional state is done similar to the Kismet-Project [5], by a 3-dimensional input vector (A, V, S) (A = Arousal, V = Valence, S = Stance). This vector selects a specific emotional state which could be a mixture of the basic emotions. The activation of emotion-i, ι_i is calculated in Eq. 3, where $diag$ stands for the diagonal of the cube. P_i means the point that represents the emotion i and I the input vector (A, V, S).

$$\iota_i = \frac{diag - |P_i - I|}{diag} \tag{3}$$

The selected emotional state has two tasks. The first task is to reach a situation adapted behavior according to the emotional state. That means if a drive activates a certain action this action is adapted according to a certain emotional offset. For example: If an object causes fear, the robot will fulfill the moves decided by the drives but it will keep distance to this object.
The second task is to realize manlike facial expressions and gestures. This is very important for man-machine-interaction to support nonverbal communication. In [12] our concept for the realization of facial expressions is described. In addition to this system new action units for the movement of the neck and the eyes are implemented. These action units will also be used to express emotions, as described in [13]. In table I the action units of the robot head ROMAN are presented.

TABLE I

TABLE OF THE ROBOT HEAD ROMANS ACTION UNITS.

AU Nb.	Description	AU Nb.	Description
1	raise and lower inner eyebrow	51	head turn left
2	raise and lower outer eyebrow	52	head turn right
5	lid tightener	53	head up
7	lid raiser	54	head down
9	nose wrinkle	55	head tilt left
12	raise mouth corner	56	head tilt right
15	lower mouth corner	57	head backward
20	stretch lips	58	head forward
24	press lips	63	eyes up
26	lower chin	64	eyes down

TABLE II

THE REALIZED DRIVES WITH THEIR PRIORITY LEVELS. THE HIGHER THE LEVEL, THE LOWER THE PRIORITY, THAT MEANS: THE HIGHEST PRIORITY LEVEL IS 1.

Name	Priority	Name	Priority
Law 1	1	EnergyConsumption	5
Law 2	2	Communication	6
Law 3	3	Exploration	7
Fatigue	4		

IV. IMPLEMENTATION OF DRIVES

In our emotional architecture 7 different drives are realized. 3 drives should make sure that the robot laws presented by Isaac Asimov in his short story ”Runaround” (1942) will be considered. *Law 1*: ”A robot may not harm a human being, or, through inaction, allow a human being to come to harm.”, *Law 2*: ”A robot must obey the orders given to it by human beings except where such orders would conflict with the First Law.”, *Law 3*: ”A robot must protect its own existence, as long as such protection does not conflict with the First or Second Law.”. In addition 2 survival drives are implemented. The first one is ”EnergyConsumption” the second is ”Fatigue”. The last two drives are ”Communication” and ”Exploration”. The drives and the corresponding priority levels are shown in Table II.

In the following the different drives are explained in more detail:
Law 1 - takes care that the robot doesn't harm any human. Because the robot exists only of a neck and a head the only way it could harm a human is to bite. To avoid this a sensor that is placed in the mouth gives information weather there is something in the robots mouth or not. If there is something in the robots mouth this drive is getting active and makes sure that the mouth won't be closed until the object is removed. The other function of this drive is to protect humans. Therefore, the robot informs its environment by a warning if it detects a danger. At the moment a certain color is defined to detect something dangerous (color detection is done with CMVision [14]). In the simulation the usage of an artificial nose is realized which will be implemented on ROMAN in the next months. With this sensor the robot could detect fire or a dangerous gas and give humans warning of it. Therefore, it uses spoken language. This is done with help of

the speech synthesis program Multivox 5 (see [15]). Because the main maxim of the robot is not to harm humans this drive has the highest priority level.

Law 2 - makes sure that the robot follows every order it receives from a human. The restriction Asimov gave to this law is, that it may not be in conflict with the first law. This is easily achieved by our priority system. To realize this drive the robot is able to understand several keywords with help of a speech recognition software based on the HTK Toolkit[1]. These keywords are ”stop”: the robot stops all its actions, ”ok”: after a ”stop” the robot only continues its work if ”ok” is detected, ”roman”: if someone is calling for the robot it turns toward this person. Therefore the position of the speaker is determined. Because this couldn’t be done exactly the robot turns to this position and looks for a human in the neighborhood of this position. Then it turns towards this person. For human detection we use ”HaarClassifierCascade” from the ”open cv library[2]”.

Law 3 - realizes self protection of the robot. At the moment this drive is activated with the help of a distance sensor. If something is getting to close to the robot this drive becomes active and causes an evasion maneuver. The robot also asks a person that is to close to it to step back. Another function of this drive is to cause fear of dangerous objects.

Fatigue - is the first one of the survival drives. That means these drives take care that the robot works adequately. ”Fatigue” means in the sense of our robot, that a system cycle needs to much time. The time that elapses, before a certain sensor date is processed, is too long. If the drive detect this, it gets active. Then no new processes can be started and some for the robots survival non-relevant processes are ended. If the cycle time is small enough this drive gets inactive again.

EnergyConsumption - this drive is getting important if the robot uses batteries as energy source. The input of this drive is the actual energy level. If this is lower than a certain threshold, this drive is getting active. It forces the robot to tell a human about its energy problems. If the problem is not solved and the energy level falls under a second threshold the robot will not start any new process if it is not important for the robots survival. As result the exploration and the communication will be inhibited.

Communication - is most important for the interaction with humans. The communication drive gets active if the robot identifies a human. If this is done this drive realizes two things: The first is, the robot follows the human with its cameras. If the human walks around the robot follows him with its head and eyes. The second function is that the robot starts a conversation with the human. Because the communication is on a higher priority level than the exploration, the exploration is inhibited by the communication drive if the robot detects a human being. Our tests show, that the robot could easily follow an once detected face.

Exploration - the drive on the lowest priority level. It takes care that the robot is always looking for something new. Therefore a certain time, called ”boring time” is defined. If this time is elapsed and nothing interesting (no other drive is getting active and no interesting object is detected) happened, the robot is getting bored. Than the robot looks randomly to a new location under consideration that this differs from the last 10 explored locations. This approach allows the robot to explore different situations. If the robot detects an interesting object (defined by color) the robot focuses it. If the object moves, the robot follows this object with its head and eyes. For testing an interesting object was defined as a red object. The tests show that the detection of the red color and the following of these red objects is no problem for the robot.

V. EXPERIMENTS AND RESULTS

To test and verify our work we realized several experiments. In the following we present the results of two representative runs. The setup of the first experiment was as follows: The exploration-drive and the communication-drive are both activated. A person is placed beside ROMAN. That means the robot needs to turn its head to see this person. We record the activity (a) and the discontent (r) of the different drives. In addition we record the activity (a) and the target rating (r) of the ”follow object” behavior.

The result is shown in Fig. 9. The exploration drive gets discontent and active. The robot detects the person. At approx. cycle 10 the communication-drive gets active and inhibits the exploration. Additionally the Follow-Object behavior is activated. At approx. cycle 15 the communication-drive is getting inactive and the exploration is getting active again.

The setup of experiment 2: We tested the complete system for a longer time. All seven drives are activated. The robot is placed in an office environment. That means all sensors are activated more or less randomly. The activation (ι), the activity (a) and the discontent (r) of the drives are recorded. The results are shown in Fig. 10. The results show the correct work of the system.

As conclusion of the experiments is to say that all drives behave in the expected way.

VI. SUMMARY AND OUTLOOK

Based on the mechanical realization of the humanoid robot head ROMAN we realized a control architecture with an integrated emotional system. This emotional architecture consists of 3 main modules: the emotions, the drives and the actions. Drives specify the needs and global goals, emotions describe the internal state of the robot and actions realize the interaction with the robots environment.

The actuator system which is directly controlled by the actions realizes all necessary movements like neck, eye and skin motions. Besides this actuator system there exists the sensor system consisting of cameras and microphones as well as inertial system and artificial nose in near future. These sensor measurements have a direct influence on the drives which are activated depending on the current sensor inputs.

All modules of the robot are realized as behavior based modules which results in a very flexible control architecture.

---

[1] http://htk.eng.cam.ac.uk/

[2] http://opencvlibrary.sourceforge.net/

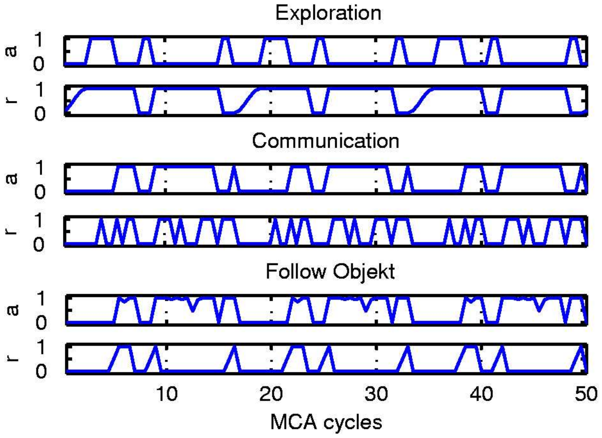

Fig. 9. The results of experiment 1: The activity and discontent of the "exploration drive", "communication-drive" and "follow object behavior" during a communication are messured.

Final experiments with the robot showed a natural behavior of communication and exploration drives.

Our future work will include the extension of drives and actions as well as the integration of all sensors and actors. Especially the interaction between different drives will be of interest for future developments.

ACKNOWLEDGMENT

Supported by the German Federal State of Rhineland-Palatinate as part of the excellence cluster "Dependable Adaptive Systems and Mathematical Modeling".

REFERENCES

[1] M. Minsky, *Society of Mind.* Simon and Schuster, 1988.

[2] N. Esau, B. Kleinjohann, L. Kleinjohann, and D. Stichling, "Mexi - machine with emotionally extended intelligence: A software architecture for behavior based handling of emotions and drives," *Proceedings of the 3rd International Conference on Hybrid and Intelligent Systems (HIS'03)*, December 2003.

[3] G. A. Hollinger, Y. Georgiey, A. Manfredi, B. A. Maxwell, Z. A. Pezzementi, and B. Mitchell, "Design of a social mobile robot using emotion-based decision mechanisms," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Beijing, China, October 9 - 15 2006.

[4] C. L. Breazeal, "Sociable machines: Expressive social exchange between humans and robots," Ph.D. dissertation, Massachusetts Institute Of Technoligy, May 2000.

[5] ——, "Emotion and sociable humanoid robots," *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 119–155, 2003.

[6] H. Miwa, K. Ioh, D. Ito, H. Takanobu, and A. Takanishi, "Introduction of the need model for humanoid robots to generate active behaviour," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, Nevada, USA, October 27 - 31 2003, pp. 1400–1406.

[7] A. Takanishi, H. Miwa, and H. Takanobu, "Development of human-like head robots for modeling human mind and emotional human-robot interaction," *IARP International workshop on Humanoid and human Friendly Robotics*, pp. 104–109, December 2002.

[8] M. Malfaz and M. A. Salichs, "Design of an architecture based on emotions for an autonomous robots," in *2004 AAAI Spring Symposium*, Stanford, California, USA, March 2004.

[9] ——, "A new architecture for autonomous robots based on emotions," in *5th IFAC Symposium on Intelligent Autonomous Vehicles*, Lisbon. Portugal, July 2004.

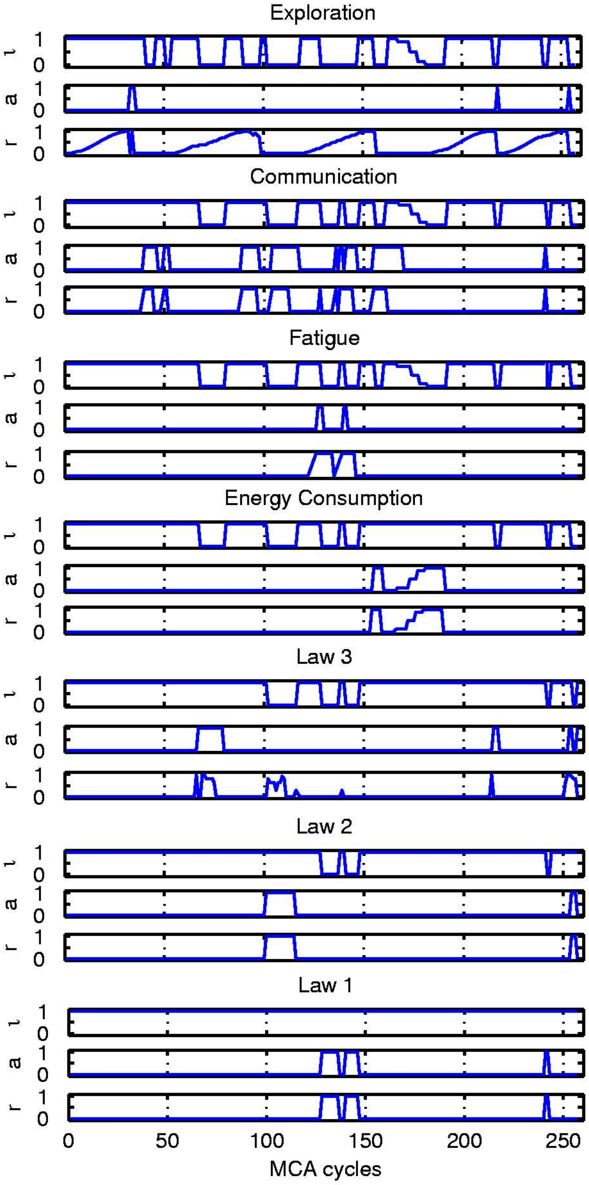

Fig. 10. The results of experiment 2: The activation, activity and discontent of all realized drives when the robot is placed in an office environment are messured.

[10] J. Albiez, T. Luksch, K. Berns, and R. Dillmann, "An activation-based behavior control architecture for walking machines," *The International Journal on Robotics Research, Sage Publications*, vol. vol. 22, pp. pp. 203–211, 2003.

[11] P. Ekman and W. Friesen, *Facial Action Coding System.* Consulting psychologist Press, Inc, 1978.

[12] K. Berns and J. Hirth, "Control of facial expressions of the humanoid robot head roman," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Beijing, China, October 9 - 15 2006.

[13] P. Ekman, W. Friesen, and J. Hager, *Facial Action Coding System.* A Human Face, 2002.

[14] J. Bruce, T. Balch, and M. Veloso, "Fast and inexpensive color image segmentation for interactive robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Takamatsu, Japan, October 30 - November 5 2000.

[15] G. Olaszy, G. Nemeth, P. Olaszi, G. Kiss, C. Zainko, and G. Gordos, "Profivox - a hungarian text-to-speech system for telecommunications applications," in *International Journal of Speech Technology.* Springer Netherlands, 2000, pp. 201–215.