A NEW ARCHITECTURE FOR AUTONOMOUS ROBOTS BASED ON EMOTIONS

María Malfaz, Miguel A. Salichs

*Carlos III University of Madrid
28911 Leganés, Madrid, Spain
E-mail: mmalfaz@ing.uc3m.es, salichs@ing.uc3m.es*

Abstract: With the emergence of personal robots the need to create autonomous robots with an easier form of human-robot interaction has become apparent. Emotions will play an important role in the autonomy issue as well as in this interaction. In the proposed architecture emotions are generated from the evaluation of the wellbeing of the robot. Behaviour selection mainly depends on the dominant drive and emotions have a supervisory role. *Copyright © 2004 IFAC*

Keywords: Autonomous mobile robots, Decision making, Drives, Emotions, Supervisory Control, System Architectures, Human-machine Interaction.

1. INTRODUCTION

1.1 Motivation

This work belongs to a new research project focused on personal robotic assistants. The goal of this project is to develop autonomous personal robots that can act as companions and as domestic assistants. Obviously in this context, human-robot interaction is very important. In relation to interaction, emotions have a fundamental role. In this paper we will give a brief review of the work done in this area.

Several authors such as Cañamero (2003), Picard (2003) and Bellman (2003), have expounded different reasons to explain why the introduction of emotions in robotics would provide a giant step forward in the robotic “evolution”. From these reasons we will highlight the most pertinent, by order of importance, that justify the use of emotions in robots. Most important is that emotions help facilitate believable human-robot interaction. Secondly, emotions can provide feedback to the user, such as indicating the robot’s internal state, goals and intentions. Finally, emotions can act as a control and learning mechanism, driving behaviour and reflecting how the robot is affected by, and adapts to, different factors over time (Fong, *et al.*, 2002).

1.2 Personal Robots

For personal robots, such as pet-robots, emotions are essential for the life-like appearance. They are autonomous robots and therefore, they must establish their own goals and the right behaviours in order to fulfil them. In recent years, an emergence of this type of robots has been occurred on a large scale, mainly

in Japan. The most representative, and the highest selling product, has been AIBO from SONY a dog-like robot (Sony, 2003). The latest model was released last September, with voice and face recognition software. We can also find a cat-like robot, NeCoRo from Omron (Omron, 2003), which is very popular in Japan. On the other hand many personal robots have been developed, but not commercialized. This is the case of PaPeRo from NEC (NEC, 2003) and Qrio (Sony, 2003) the new humanoid from SONY, next to be released.

The main objective of these pet-robots is entertainment and its fundamental characteristic is the emotional behaviour (Tashima *et al.*, 1999). A Pet-robot must behave in a similar way to a real pet. This implies responding differently according to its emotional state. It is expected that a pet behaves in an unpredictable way. In the case of AIBO, only a few behaviours are documented, the rest must be discovered by the owner. Moreover, these behaviours could differ among robots depending on their “personality”.

The potential applications of personal robots are numerous and could vary along with the owner’s characteristics. For a child it could be a toy as well as a teacher. For an adult it could be an assistant, in the work place and/or at home. On the other hand, a personal robot could help elderly or disabled people to overcome some of their physical limitations. Obviously, the human-robot interaction system must be user friendly and make the robot easy for people to use, without special training. In this interaction system, the implementation of an emotional model is quite important. For example, in the case of elderly people, for whom most do not like machines, the fact that the robot has emotions and behaves as if it is

“alive”, would help them to think about robots not as simple machines but as real companions. Evidently, for children and many adults, a robot that “feels” and has his own “personality” is much more attractive than one that simply executes the orders that he is programmed to do.

2. EMOTIONS AND ROBOTICS

2.1 Introduction

One of the main objectives in robotics and artificial intelligence research is to imitate the human mind and behaviour. For this purpose the studies of psychologists on the working mind and the factors involved in the decision making are used. In fact, it has been proved that two highly cognitive actions are dependant not only on rules and laws, but on emotions: Decision making and perception (Picard, 1998). In fact, some authors affirm that emotions are generated through cognitive processes. Therefore emotions depend on ones interpretation, i.e. the same situation can produce different emotions on each agent, such as in a football match (Ortony, 1988). Moreover, emotions can be considered as part of a provision for ensuring and satisfaction of the system’s major goals (Frijda, 1987).

2.2 Emotional models in robotics

Emotions play a very important role in human behaviour, communication and social interaction. Emotions also influence cognitive processes, particularly problem solving and decision making (Damasio, 1994). In recent years, emotion has increasingly been used in interface and robot design, primarily in recognition that people tend to treat computers as they treat other people.

There are several theories about emotions (Frijda 1987; Ortony, 1988; Sloman, 2003; Rolls, 2003), but the results of Damasio (1994) can be considered the basis, for many A.I. researchers, to justify the use of emotions in robotics and their computation.

Rosalind Picard in her book *Affective Computing* (1998), writes a complete dissertation about this subject based on several psychologists, including Damasio. Picard (1998) proposed a design criterion in order to create a computer that could express emotions. Moreover, she established that a computer has emotions if it has certain components that are present on the emotional systems of healthy people. Picard (2003) expounded four motives for giving certain emotional abilities to machines: The first goal is to build robots and synthetic characters that can emulate living humans and animals, such as a humanoid robot. The second is to make machines that are intelligent. A third objective is to try to understand human emotions by modelling them. Although these three goals are important, the main one is to make machines less frustrating to interact with, i.e. to facilitate the human-machine interface.

Cañamero (2003) considers however, that emotions, or at least a sub-group of them, are one of the mechanisms founded in biological agents to confront their environment. This creates ease of autonomy and adaptation. For this reason she considers that it could be useful to exploit this role of emotions to design mechanisms for an autonomous robot. Emotions are used as mechanisms that allow the agent (robot) to:

1. Have fast reactions.
2. Contribute to resolve the selection among multiple objectives.
3. Signal important events to others.

Bellman (2003) agrees, to some degree, with Cañamero and her reasons for considering emotions in robotics. The author states that emotions allow animals with emotions to survive better than the others without emotions. Therefore, we can presume that some type of analogy to emotional abilities is required within robots, if we want an intelligent and independent behaviour within a real environment.

Changing subject, Picard (2003) gives an advice about the implementation in machines of functions implemented by the human emotional system. Computers do not have emotions as human beings in any natural experimentation sense. Science methodology is to try to reduce complex phenomena, such as emotions, to a functional requirements list. The challenge of many computing science researchers, is to try to duplicate these in computers at different levels depending on the motives of the investigation. But we must be careful when presenting this challenge to the general public, who may perceive that emotions are the frontier that separates man and machine.

3. SYSTEM ARCHITECTURE

3.1 Introduction

The emotional model that is proposed in this work has been implemented on a real robot. This robot is supported by the AD architecture. The AD architecture (Automatic-Deliberative Architecture), was developed by Barber (2000; 2001), and although it was inspired by the human mind, it does not consider emotions in its original design. Therefore, our emotional model has been included in it, rounding off its functions taking into account the role of emotions in human behaviours.

According to the modern psychology theories (Shiffrin and Scheider, 1997), there exists two mechanisms through which the human being processes the perceived information from the environment: Automatic and controlled processes. Therefore, we can distinguish between two levels of mental activity: Deliberative level and Automatic level. Based on these ideas, only two levels are established in this architecture. The Deliberative Level is associated with reflective processes and the Automatic Level with the automatic ones. Both

levels are formed by skills. These skills are activated by execution orders generated by others skills or by a sequencer, returning data and events to the skills or sequencers that activated them. These skills are the basis of the AD architecture.

In the Automatic Level exists low-level modules which act directly with the actuators, and modules that get data from the different sensors of the system. These skills are activated and deactivated by the Deliberative Skills. An Automatic Skill can be carried out simultaneously with other automatic skills. In this level there also exists Reflex Actions, which are involuntary and priority responses to the stimuli.

In the Deliberative level we can find the modules that require reasoning or a decision making capability (such as trajectories planning). These modules do not produce an immediate response as the Automatic Skills, since they need more time to process the information that they work with to make decisions. These modules form the Deliberative Skills, and they are carried out sequentially, the same as in human beings.

In addition, there exist two kinds of memories, the short term and the long term memory. The short term Memory appears in this architecture as an interchange element of information between both levels of the architecture. This memory is different from the long term Memory because the stored information is the state robot information, while the stored information in the long term Memory is the data that can be considered as stable over time.

3.2 Proposed Architecture

The goal of our research is to construct autonomous personal robots. These robots are being built on mobile platforms (iRobot Magellan Pro), on which various sensors (laser, sonar and camera) and devices are installed. These sensors and devices will make possible the interaction of the robot with its environment and users.

Fig. 1. AD Architecture with an emotional control System (ECS).

Figure 1 illustrates how an emotional control system (ECS) has been added to the AD architecture (Barber

and Salichs 2001). Inside the ECS we can observe three different modules: Drives, Activity Selection and Emotional Supervisory System (ESS). The Drives module is the one that controls the basic drives of the robot. The Activity Selection module on the other hand, determines goals and action tendencies of the robot. Finally, the ESS module generates the emotional state of the robot.

Drives. Several authors, such as Frijda (1987), propose that an independent system should not have to wait for someone to maintain, succour, and help it. Therefore, an autonomous agent should be capable of determining its goals, and it must be capable of selecting the most suitable behaviour in order to reach its goals. Similarly to the Velasquez's proposal (1998), the robot's autonomy relies on a motivational model based on drives and emotions. Within the model developed by Velasquez, drives are the robot's internal needs. These drives are related to the feedback of internal variables such as its battery and temperature. By contrast, within our control architecture the drives are defined in a different way. They are related to the feedback of internal variables (e.g. battery level) as well as external variables (e.g. social interaction).

In this architecture the following drives are initially used:

- Energy: With the battery level as its input.
- Entertainment: Depending on the activity and human-robot interaction.
- Affect: Depending on the level of affection received through the human-robot interaction.
- Health: Related to physical damage.
- Agenda: This drive reflects the motivation of the robot to carry on with activities which have been previously planned.

Since our objective is to construct an autonomous personal robot, it must show social behaviours. Therefore, as it is shown, social motivations are included as robot's needs. The Entertainment and Affect drives are similar to the Stimulation and Social drives defined by Breazeal (2002).

Inside the Drives module we define a new concept in this architecture: the Agenda. This implies that the robot will select the current behaviour not following its internal and social needs, but because it has to complete a plan.

In order to determine the dominant drive, each drive has a specific weight depending on the importance given to survival (energy and health), social interaction (entertainment and affect) or the "must do" (agenda). Therefore, these drives will compete to be the dominant one and determine the goal of the robot.

Emotions. Most of authors working with emotion-based control architectures, such as Velasquez (1998), Gadanho (2002), Breazeal (2002), etc., give a

basic role to emotions i.e. behaviour selection is conditioned by emotions. Nevertheless, according to several authors such as Ortony (1988) and Frijda (1987), not all behaviours are emotional, that there exist non emotional behaviours, i.e. we can avoid a chair without being afraid of it. Following this way of thinking, in our control architecture behaviour selection mainly depends on a specific drive, the dominant drive, and emotions have a supervisory role.

When there is not any active emotion, the dominant drive will decide the goal of the robot, and therefore it will determine which behaviour will become active in order to reach the goal, i.e. if the Energy drive is low, then the goal of the robot will be to recharge itself, so the behaviour “find the power station” will become active. Only when an emotion becomes active, will the emotion influences the Activity Selection module.

The emotions used in this architecture are: Happiness, Anger, Fear and Sadness. These emotions have persistence in time. This classification is the same used by other authors (Gadanho and Hallam 2001).

Emotion generation. In this architecture emotions are generated from the evaluation of the wellbeing of the robot. Happiness is produced because something good has happened, i.e. an increment of the wellbeing is produced. On the contrary, Sadness is produced because something bad has happened, so its wellbeing decreases. Fear appears when the possibility of something bad is about to happen. In this case, we expect that the wellbeing drops off. Finally, Anger is produced when a decrement of the wellbeing of the robot happened due to an other-initiated act. These definitions of emotions are based on Ortony's works (Ortony, Clore and Collins 1988).

Fig. 2. Emotional Supervisory System

As it is shown in Figure 2, in order to generate these emotions we evaluate two signals: the current wellbeing and the prediction of the future wellbeing. The evaluation of the past wellbeing signal would also produce emotions such as shame, pride, gratitude, etc. Nevertheless, these kinds of emotions will not be considered by now in our architecture.

The intensity of each emotion is related to the derivate of the current or future wellbeing signal. Depending on the absolute value of the slope of the wellbeing signal, the intensity of the correspondent emotion will vary.

The current wellbeing signal is formed by two contributions. The first one is a general appraisal of the drives, and the second one is the emotion influence. The future wellbeing signal can be obtained from the analysis of the consequences of the current situation.

The functions of the ESS in our control architecture are as follows:

- To establish general goals: i.e. happiness. These global goals allow the architecture to develop behaviour selection mechanisms and a global appraisal of the situation. This global appraisal can be used also as reinforcement in learning processes.
- Supervision: When the changes of the wellbeing signal (current or future) are greater than a certain threshold, one emotion will become active. In this sense, the ESS will act as an alarm system where emotions are the alarm signals.

4. CONCLUSIONS

In this paper an emotion based architecture for mobile autonomous robots has been presented. In order to be autonomous, the robot must decide its goals and how to get them. Its decision will be mainly based on its drives and not exclusively on its emotional state. Only in certain circumstances the selection of the robot's goal and behaviour will be based on the current emotion. These drives are the robot's internal and social needs, and a set of activities which has been previously planned. Emotions in this architecture are generated as an evaluation of the wellbeing of the robot. In this architecture emotions have a supervisory role, and reflect the general appraisal of the system. Emotions will not create behaviours but action tendencies. This general appraisal is used as reinforcement in making decision processes.

5. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the funds provided by the Spanish Government through the DPI2002-00188 project, named Personal Robotic Assistant (PRA), of MCyT (Ministerio de Ciencia y Tecnología).

REFERENCES

- Aaron Sloman (2003). How many separately evolved emotional beasts live within us. In: *Emotions in Humans and Artifacts*. (Robert Trappl, Paolo Petta and Sabine Payr), pp 35-115. The MIT Press. Cambridge, Massachusetts
- Antonio Damasio.(1994). *Descartes' Error – Emotion, reason and human brain*. Picador, London.

- Barber, R. (2000). Desarrollo de una Arquitectura para Robots Móviles Autónomos. Aplicación a un Sistema de Navegación Topológica. PhD Thesis. Universidad Carlos III de Madrid.
- Barber, R., Salichs, M.A. (2001). A new human based architecture for intelligent autonomous robots. The Fourth IFAC Symposium on Intelligent Autonomous Vehicles, pp85-90. Sapporo, Japón.
- Breazeal C. (2002) Designing Sociable Robots. The MIT Press.
- Edmund T.Rolls (2003). A Theory of emotion, its functions, and its adaptive value. In: Emotions in Humans and Artifacts. (Robert Trappl, Paolo Petta and Sabine Payr), pp 11-35. The MIT Press. Cambridge, Massachusetts.
- Fong, T., Nourbakhsh, I., Dautenhahn K (2002). A survey of socially interactive robots: Concepts, design, and applications. Technical Report CMU-RI-TR-02-29.
- Frijda, N. and Swagerman, J. (1987) Can computers feel? Theory and design of an emotional model. *Cognition and Emotion*. 1 (3). 235-357.
- Kirstie L. Bellman (2003). Emotions: Meaningful mappings between the individual and its world. In: Emotions in Humans and Artifacts. (Robert Trappl, Paolo Petta and Sabine Payr), pp 149-188. The MIT Press. Cambridge, Massachusetts.
- Lola D. Cañamero. (2003) Designing emotions for activity selection in autonomous agents. In: Emotions in Humans and Artifacts. (Robert Trappl, Paolo Petta and Sabine Payr), pp 115-148. The MIT Press. Cambridge, Massachusetts.
- NEC.PaPeRo.
<http://www.incx.nec.co.jp/robot/2003papero/english/gijyutu3.html>
- OMRON.NeCoRo.
<http://www.necoro.com/newsrelease/>
- Ortony, A., Clore, G. L., and Collins, A. (1988). The Cognitive Structure of Emotions. Cambridge University Press. Cambridge, UK:
- Rosalind W. Picard. (1998). Los ordenadores emocionales. Ed. Ariel S.A.
- Rosalind W.Picard.(2003) What does it mean for a computer to have emotions?. In. Emotions in Humans and Artifacts. (Robert Trappl, Paolo Petta and Sabine Payr), pp 213-235. The MIT Press. Cambridge, Massachusetts
- Sandra Clara Gadanho and John Hallam (2001). Emotion-triggered learning in Autonomous robot control. In: *Cybernetics and Systems: An International Journal*, 32, pp 531-559.
- Sandra Clara Gadanho (2002). Emotional and Cognitive Adaptation in Real Environments. In: Symposium ACE'2002 of the 16th European Meeting on Cybernetics and Systems Research, Vienna, Austria.
- Shiffrin,R.M., Schneider, W. (1997) Controlled and automatic human information processing: II Perceptual learning, automatic attending and a general theory. *Psychological Review* 84, p 127-190.
- Tashima, T.; Saito, S.; Kudo, T.; Osumi, M. (1999). Interactive pet robot with an emotion model. *Advanced Robotics*, vol.13, no.3, pp. 225-6.
- Sony. AIBO. <http://www.aibo.com/>
- Sony, Qrio. <http://www.sony.net/SonyInfo/QRIO/>
- Velásquez, J (1998). When Robots Weep: Emotional Memories and Decision Making. In: *Proceedings of AAAI-98*.