Int J Soc Robot (2011) 3:273–290
DOI 10.1007/s12369-010-0087-2

Towards Social Robots: Designing an Emotion-Based Architecture

Jochen Hirth · Norbert Schmitz · Karsten Berns

Accepted: 22 December 2010 / Published online: 7 January 2011

Abstract Service and assistant robots operating in "real world" become more and more important for everyday life. Requirements like social intelligence, social interaction, and social behavior are of enormous importance. Researchers agree that the realization of emotions for a robot is one of the most important steps on the way towards social robots. Still, an open question in robotics is how to model emotion in a way that all its functions are realized. In this paper an emotion-based architecture is presented. The architecture supports the implementation of the five functions of emotion, namely: regulative function, selective function, expressive function, motivational function, and rating function. Scalability, perceptibility, and parameterability are also considered as design features. As basis for the proposed architecture, psychological emotion and motivation theories are used. The derived control architecture is implemented using a behavior-based approach, in order to realize a modular extensible system design. The developed architecture is tested on the humanoid robot ROMAN.

Keywords Emotion-based control · Humanoid robots

1 Introduction

Application areas for socially intelligent robots reach from the care of elderly people, over nursing robots or household robots, to tour guiding robots or museum guide robots as mentioned in [1]. Thus completely new requirements for the control systems arise. Compared to traditional service or industrial robots, these robots need to have the abilities to interact and behave socially. They need some kind of understanding of human behavior and they also require some kind of empathy in order to behave in an appropriate way. Psychologists call this kind of ability "social intelligence" [2]. Social intelligence mainly depends on emotion. Besides this characteristic, which enables a robot to react to unexpected changes in its environment, the emotion also provides a second benefit. As described in [3], emotions are crucial for the motivation of humans' behaviors, since they realize some kind of internal goal generation. Therefore, it is obvious that autonomous robots operating in real world need some kind of emotion-based control. One important step towards social robots is the realization of a well grounded emotional state, see [4]. The generation of such an emotional state still is a hard problem in robotics since it is not clear how to determine the necessary parameters depending on internal and external sensor information. The Robotics Research Lab at the University of Kaiserslautern is developing an emotion-based architecture for autonomous robots based on theories mentioned in [5] and [6]. This paper summarizes the work done in the concept of this project till now. Compared to previous publications like [7] the newly developed appraisal system will be introduced and explained in this article. Furthermore, this article allows presenting the emotion-based architecture as a whole and therefore provides a much better idea of it.

This article is arranged in the following way: At first, the psychological background of emotions and motives are highlighted. Section 3 provides a brief summary and discussion of the state of the art in realizing emotion-based control

J. Hirth (✉) · N. Schmitz · K. Berns
Robotics Research Lab, Department of Computer Sciences,
University of Kaiserslautern, 67663 Kaiserslautern, Germany
e-mail: j_hirth@informatik.uni-kl.de

N. Schmitz
e-mail: nschmitz@informatik.uni-kl.de

K. Berns
e-mail: berns@informatik.uni-kl.de

architectures. Afterwards, the UKL[1] Emotion-based Architecture is introduced and exemplarily explained by describing its implementation on the humanoid robot ROMAN.[2] In Sect. 6, the results of the experiments are discussed to prove the realization of the functions of emotion. The quality of the emotional expressions has been rated by "non-expert" subjects in a post-interaction evaluation. Furthermore, the sequence of the changes of the robot's emotional state during an interaction has been recorded and investigated. In the conclusion, an outlook to a tangram scenario is given that should be used as complex testing scenario to evaluate the interaction abilities realized by the proposed architecture.

2 Psychological Insights

In order to have a socially intelligent robot, an emotion system is needed [3]. Therefore, the first step is to figure out what emotion is. In psychology there exist several definitions and theories about the generation and the dimensions of emotion. An overview of emotion theories can be found in [8]. According to [9] it can be argued, that the function of emotion is most important and should be investigated.

Five functions of emotions are pointed out in [6]:

Regulative function: Emotions signal whether there are any abnormal external or internal values perceived. That way, they can protect the organism from injuries.

Selective function: Emotions influence the perception of the environment as well as the perception of internal stimuli.

Expressive function: Facial expressions, gesture, body posture, and the tone of the voice are strongly influenced by emotions. These expressions are used to conduct non-verbal information.

Motivational function: Emotions activate and control the behavior of humans. Humans try to experience comfortable rated emotions again and to avoid uncomfortable emotions.

Rating function: Emotions can be used to evaluate situations. Depending on the emotions, experiences can be classified e.g. comfortable or uncomfortable situations.

To describe motivation, psychologists invented the construct of motives. Motives are used to explain the reasons for human behavior and actions [10]. Active motives stimulate goal directed behaviors, a process which is called motivation. Therefore, motivation is a state of activity that is controlled by motives. This motivation lasts until the specific goal is reached or a motive of a higher priority gets active.

According to [5], a motive is defined by four criteria:

Activation: Specific behaviors are stimulated by an active motive.

Direction: The activity is directed towards a specific goal and lasts until this goal is reached or a motive of a higher priority level gets active.

Intensity: The level of the activity of a motive can vary. This level of activity is described by the intensity.

Duration: In most cases the activity is maintained until the specific goal is reached.

As mentioned in [5], the activity of motives can be explained using the control-loop-theory. The difference between the actual state and the target state leads to an activity of the motives. Regarding [11] the motivation of humans can be defined as the sum of all active motives controlling the humans' actions and behavior.

Most psychologists nowadays agree that the generation of emotions can be described using appraisal theories [12]. These theories claim that emotion is the result of an evaluation—an appraisal—of the current situation. There exist different appraisal theories which use different categories to appraise an event, namely: Criteria, attributions, themes, and meanings. The different theories also vary in the number of dimensions used—some scientists use up to 15 criteria, others just use three.

In summary, emotions as well as emotional expressions are necessary for the interaction with and the adaptation to a complex environment. These abilities are called social intelligence. Emotions include changes in physical and psychical state. These changes are perceived by humans and they are rated as comfortable or uncomfortable. Therefore, emotion mainly influences the behavior of humans.

As pointed out in [13], the realization of a social intelligence is only possible based on an emotional state. Therefore an emotion-based control architecture seems to be an important pre-condition for socially intelligent robots. As mentioned in [4], robots need a well grounded emotional state. Therefore, important implementation aspects for emotion-based architectures are the five functions of emotion mentioned above (regulative, selective, expressive, motivational, and rating). Because of the motivating function of emotion, also the characteristics of motives must be realized. For the determination of an emotional state, an appraisal system seems to be an appropriate solution.

3 Emotion-Based Approaches

This section will give a brief overview of the state of the art in the development of emotion-based control architectures. More detailed overviews of that kind of control architectures can be found in [14, 15]. The different approaches are classified and evaluated using three categories: Emotion displays, virtual agents, and social robots.

---

[1] University of Kaiserslautern.

[2] http://agrosy.cs.uni-kl.de/en/robots/roman/

Emotion displays are robots with the ability to express emotions. Therefore, these robots utilize facial expressions, gestures, or other non-verbal signals. Some of them also use speech to express emotions. One of these systems is the artificial head Eddie [16]. Eddie is able to generate different facial expressions corresponding to the six basic emotions—anger, disgust, fear, happiness, sadness, and surprise. Another system for the generation of emotional expressions is the so-called animation engine [17]. This engine is used to enable the robot iCat to generate emotional expressions. iCat is developed as user interface and should interact with humans using emotional expressions. Two more robots in this area are the expressive robots Probo [18] and Keepon [19]. These systems have been developed for usage in hospital environments. Especially the interaction with children is in the focus of these projects. Another expressive robot is Barthoc, which has the capabilities to use gestures, body posture, and spoken words besides the facial expression to display its emotions [20, 21]. Other projects in this area are the gesture generating steward robot [22] and the facial expression generating robot described in [23].

Virtual agents are fully or partly simulated creatures. The agents have the advantage that they can easily perceive all the information they need from their simulated environment. Furthermore, they have no mechanical limitations for their expressions since they can move their simulated body parts in any direction with any velocity. These agents have the drawback that the simulated parts are not present in the real world and therefore they can not physically interact. One of the partly simulated systems is the Roboreceptionist of the Carnegie Mellon University. The underlying emotion architecture is explained in [24]. The robot is equipped with a virtual face, able to express emotions. A very sophisticated architecture is used for the fully simulated virtual agent Max developed at the University of Bielefeld [25]. Max is able to realize social interactions with humans and can therefore behave in an emotional way. More projects in this area are described in: [26]—a gesture generating embodied agent, [27]—a robot equipped with a screen showing a virtual face, or [28]—an expressive virtual agent.

Social embodied robots, that are physically present and provide interaction abilities are describe in the following:

According to [29], a social robot can be defined as:

> “An autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioral norms expected by the people with whom the robot is intended to interact.”

One of the first projects world-wide that tried to develop a socially interactive robot based on emotions was the Cog project [30] at the Massachusetts Institute of Technology. Depending on their studies, the most popular project in the area of social robots was developed, Kismet [31, 32]. This architecture consists of emotions, drives and a behavior system. The drives represent the robot’s internal goals; depending on their satisfaction, behavior classes are selected. The single behaviors within these classes are stimulated depending on the current emotional state. Another emotion-based architecture was developed for WE-4RII at the Waseda University [33]. The so-called mental model consists of emotions, moods, and needs. A very advanced architecture in this area is described in [34, 35]. This architecture contains all important components of the psychological theories and also realizes all the functions of emotions. The proposed framework regards a wide range of time-varying affect-related phenomena. Further systems in this area are: The small humanoid robot Qrio [36, 37] or the interactive robot Maggie described in [38] and [39]. Most of these architectures are specially designed to work on exactly one robot and are optimized for handling a few scenarios. Furthermore, it seems hard to adapt these architectures to different applications or robots. Therefore, a general description how to design an emotion-based architecture for robots and how to determine the necessary parameters seems to be necessary.

This section gave a brief overview of related efforts in realizing emotions for a robot system. There exist a variety of approaches, some focusing on one function of emotion, others realizing a complete emotion system very closely related to psychological theories. The new approach presented in this paper provides a concept how to model an emotion-based architecture in general. Furthermore, the architecture remains a scalable, perceptible, and describable system. In addition, rules how to derive the parameters necessary to determine the robot’s emotional state are provided.

4 The UKL Emotion-Based Architecture

Since emotion-based architectures are crucial for the realization of intelligent social robot behavior, a new emotion-based architecture that regards all the conditions mentioned in Sect. 2 is developed at the Robotics Research Lab of the University of Kaiserslautern. This section provides design guidelines how to develop and structure an emotionbased system. Besides the conditions mentioned before the following three secondary conditions need to be fulfilled: The system should be scalable, so that the system can be extended and adapted and still remain manageable (scalability). All the necessary information need to be perceptible by a robot’s sensor system (perceptibility). In addition, there need to be rules how to derive the parameters necessary to determine the emotional state of the system (parameterability). Otherwise it is not possible to handle complex scenarios, like e.g. human-robot interaction situations.

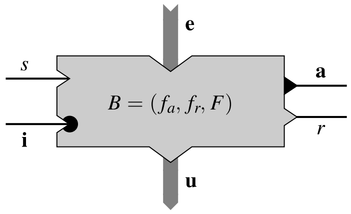


Fig. 1 Basic iB2C behavior module: The standardized interface consists of the inputs *stimulation* s and the *inhibition vector* $\mathbf{i}$ as well as the outputs *activity vector* $\mathbf{a}$ and *target rating* r. Arbitrary ports are provided for the *input vector* $\mathbf{e}$ and *output vector* $\mathbf{u}$. The transfer function $F(\mathbf{e}, \iota)$ determines the output vector with regard to the input vector and the internal *activation* ι

For the realization of this architecture, the **i**ntegrated **B**ehavior-**b**ased **C**ontrol (iB2C)[3] [40] is used. The basic component of iB2C is the so called behavior module, see Fig. 1.

A behavior module can be described as a three-tuples of the form

$$B = (f_a, f_r, F) \tag{1}$$

where f_a is the *activity function*, f_r is the *target rating function*, and F is the *transfer function* of the behavior. These functions generate *activity* information $\mathbf{a}$, a *target rating* r, and an *output vector* $\mathbf{u}$, respectively. Additionally, each behavior receives an *input vector* $\mathbf{e}$, a *stimulation* s, and an *inhibition vector* $\mathbf{i}$. In the following, these characteristics are explained in more detail.

Input Vector $\mathbf{e}$: Behaviors receive data required for fulfilling their work via the *input vector* $\mathbf{e} \in \Re^m$ which can be composed of sensory data (e.g., detected expressions of an interaction partner) or information from other behaviors (e.g., their target rating).

Output Vector $\mathbf{u}$: The *output vector* $\mathbf{u} \in \Re^n$ transmits data generated by the behavior (e.g., intended expressions). This output describes the data which are used for actuator control or as input for other behaviors.

Transfer Function $F(\mathbf{e}, \iota)$: The transfer function determines the output vector of a behavior depending on the input vector, the stimulation of the behavior, and the behavior's inhibition. Furthermore, the transfer function can also depend on internal variables representing a certain state of the behavior. That way, both reactive sensor responses and deliberative behaviors can be implemented.

Stimulation s: All behaviors receive their stimulation from other parts of the control system. The co-domain of the stimulation is $[0, 1]$, where $s = 0$ means no stimulation and $s = 1$ full stimulation, values in between refer to a partially stimulated behavior. The stimulation of a behavior can be seen as the intended relevance of this behavior in the current situation.

Inhibition $\mathbf{i}$: A behavior can be inhibited by several other behaviors. The inhibition i of a behavior is defined as: $i = \max_{j=0,\dots,n-1} i_j$, where n denotes the number of inhibiting behaviors. The inhibition has the inverse effect of stimulation, $i = 0$ refers to no inhibition and $i = 1$ full inhibition.

Activation ι: The activation of a behavior defines the effective relevance of a behavior in the behavior network. It is composed of the stimulation s and the inhibition i, with $\iota = s \cdot (1 - i)$.

Activity a: The behavior signal activity $a \in [0, 1]$ represents the amount of influence of a behavior in the current system state. $a = 1$ refers to a state where all output values are intended to have highest impact, whereas $a = 0$ indicates an inactive behavior. Values between 0 and 1 refer to a partially active behavior. If the behavior realizes different internal states it might be usefull to derive different activities. These derived activities can be used to stimulate other activities with different intensities.

The activity a and the derived activities $\underline{\mathbf{a}}$ are defined by the activity function f_a with

$$\begin{aligned} &f_a : \Re^m \times [0, 1] \to [0, 1] \times [0, 1]^q \\ &f_a(\mathbf{e}, \iota) = \mathbf{a} = (a, \underline{\mathbf{a}})^T \end{aligned} \tag{2}$$

where

$$\underline{\mathbf{a}} = (\underline{a}_0, \underline{a}_1, \dots, \underline{a}_{q-1})^T \tag{3}$$

with

$$\underline{a}_i \le a \quad \forall i \in \{0, 1, \dots, q - 1\} \tag{4}$$

The derived activities $\underline{\mathbf{a}}$ allow a behavior to transfer only a part of its activity to other behaviors.

Target Rating r: The behavior signal target rating $r \in [0, 1]$ is an indicator for the contentment of a behavior. A value of $r = 0$ indicates that the behavior is content with the actual state, while $r = 1$ shows maximal dissatisfaction. Values between 0 and 1 refer to a partially content behavior.

The target rating is defined by the target rating function f_r with

$$f_r : \Re^m \to [0, 1], \qquad f_r(\mathbf{e}) = r \tag{5}$$

If multiple behavior outputs try to influence the same control system parameters, these outputs must be coordinated. Therefore, iB2C provides so called fusion modules. A fusion module is a special behavior module. It receives the output vectors as well as the behavior signals of the behavior modules to be coordinated and generates a combined output vector depending on the fusion function. There are three

---

[3] http://rrlib.cs.uni-kl.de/

fusion functions implemented: maximum fusion, weighted fusion, and weighted sum fusion.

In case of the maximum fusion the output vector of the most active behavior is forwarded and all other behaviors are disregarded. The transfer function of the maximum fusion is shown in (6). The activity and target rating of the fusion module are set to the activity and target rating of the most active behavior, see (7).

$$\mathbf{u} = \mathbf{u}_s \quad \text{where } s = \underset{c}{\operatorname{argmax}}(a_c) \tag{6}$$

$$a = \max_c(a_c), \qquad r = r_s \quad \text{where } s = \underset{c}{\operatorname{argmax}}(a_c) \tag{7}$$

For a weighted fusion the output of the fusion module is calculated depending of the activity of the involved modules. The control output of the fusion module is calculated as described in (8), where N means the number of involved behavior modules. The activity and the target rating of the fusion module are calculated in the same manner.

$$\begin{aligned} \mathbf{u} &= \frac{\sum_{i=0}^{N-1} a_i \cdot \mathbf{u}_i}{\sum_{j=0}^{N-1} a_j}, \qquad a = \frac{\sum_{i=0}^{N-1} a_i^2}{\sum_{j=0}^{N-1} a_j} \cdot \iota \\ r &= \frac{\sum_{i=0}^{N-1} a_i \cdot r_i}{\sum_{j=0}^{N-1} a_j} \end{aligned} \tag{8}$$

The weighted sum fusion is used for summing up the control values of the involved behaviors according to their activity. The transfer function of the weighted sum fusion is described in (9), the activity and the target rating are defined as shown in (10).

$$\mathbf{u} = \sum_{j=0}^{p-1} \frac{a_j \cdot \mathbf{u}_j}{\max_c(a_c)} \tag{9}$$

$$a = \min\left(1, \sum_{i=0}^{p-1} \frac{a_j^2}{\max_c(a_c)}\right) \cdot \iota, \qquad r = \frac{\sum_{i=0}^{p-1} a_i \cdot r_i}{\sum_{j=0}^{p-1} a_j} \tag{10}$$

Coming back to the introduction of the concept of the UKL Emo tion-based Architecture: According to Sect. 2, the proposed architecture consists of an emotion center divided into a motivational system and an appraisal system, a perception system called percepts and the habits as an expressive system (see Fig. 2). The goal of the implementation is to realize the functions of emotions: regulative (emotion), selective (percepts), expressive (habits), motivational (motives), and rating (emotion), as well as the secondary functions. Compared to previous publications [7] where the rating and the regulative function where not included, all five functions of emotion mentioned in [6] are realized, since the rating and the regulative function are very important with a view to more cognitive or learning applications. The perception system perceives and interprets information of the environment. Depending on this information, direct responses, performed by the habits, are activated and the motives calculate their satisfaction. This satisfaction changes the current emotional state of the robot. Besides this, the motives activate several habits to change the robot's behavior in order to reach a satisfied state. They also determine which information of the percepts is needed in the current situation. That way, the selective function is realized. The current emotional

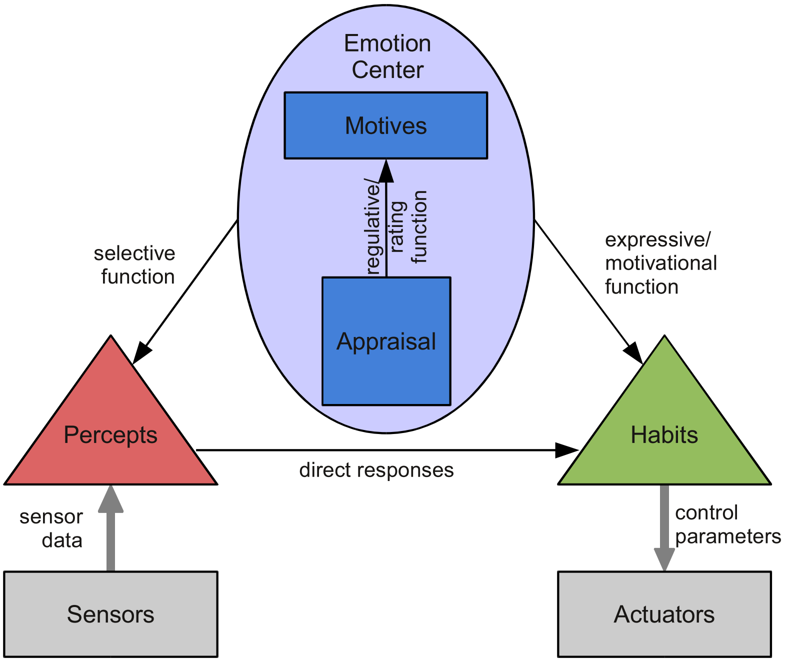


Fig. 2 The UKL Emotion-based Control Architecture, consisting of the four main groups, motives, emotional state, habits of interaction, and percepts of interaction

state, determined by the appraisal system, influences the percepts, the motives, and the habits.

The motives are realized considering the four characteristics: Activation, direction, intensity, and duration (cf. Sect. 2). Motives are extended behavior modules using an additional satisfaction function. Every motive owns a special goal. The motive's satisfaction is calculated depending on the distance to this goal. This distance is determined using the information provided by the percepts. Depending on this satisfaction, the motive calculates its target rating and activity—the more unsatisfied a motive, the higher its activity. If a motive is active, it stimulates different habits in order to reach a satisfied state. The intensity of this stimulation depends on the motive's activity. A motive remains active until the goal is reached or it is inhibited by another motive.

Depending on the remarks in Sect. 2 an appraisal system using three dimensions for evaluating the current situation has been selected. The robot system has to provide the capabilities to perceive all required information for the appraisal. Therefore the dimensions arousal (A), valence (V), and stance (S) seem to be appropriate. Projects like [31] or [25] have proven that three dimensions are sufficient to determine an emotional state for a social interactive agent. Arousal specifies how thrilling a stimulus is to the robot. Valence describes how favorable or unfavorable a situation is to the system. High valence means the robot is very content with the situation and low valence means the robot is discontent. Stance specifies the attitude of the robot to the environment in a certain situation. High stance means the currently perceived stimuli are welcome; low stance means the stimuli are unrequested.

The stimuli are perceived by the perception system. An overview of the structure of this system is depicted in Fig. 3. The perception system consists of three different module types: Source modules (*Image Source*, ...) provide raw input data captured by audio and video systems as well as the kinematic chain of the robot itself. Perception modules (*Flow Detector*, ...) use the raw information and optionally already existing percepts to generate new percepts. Here, a single percept represents an information that may be of relevance for the modeling of the user or the environment. All percepts are combined in a hierarchy of Fusion modules (*InView Fusion*, ...). First, the source specific percepts are fused and finally an overall fusion combines audio, video and kinematic information. The memory is classically divided into ultra-short-term-memory, short-term-memory and long-term-memory [41]. Each source and fusion module has access to its own memory area. The *Perception Fusion* finally generates the time varying user model and passes it to the *Habits*.

The habits are realized as a behavior network, starting with basic habits representing motor primitives up to complex habits considering whole tasks. Complex habits are generated by combining several basic habits. That way a complex robot behavior can be realized in an easy way. In Fig. 4 the structure of the habits is described. The lowest layer is represented by the *Basic Habits* (e.g. *Turn Head Left*, *Turn Head Right*)—two basic habits per active joint. At the next layer, *Actions* and *Responses* are located. Actions

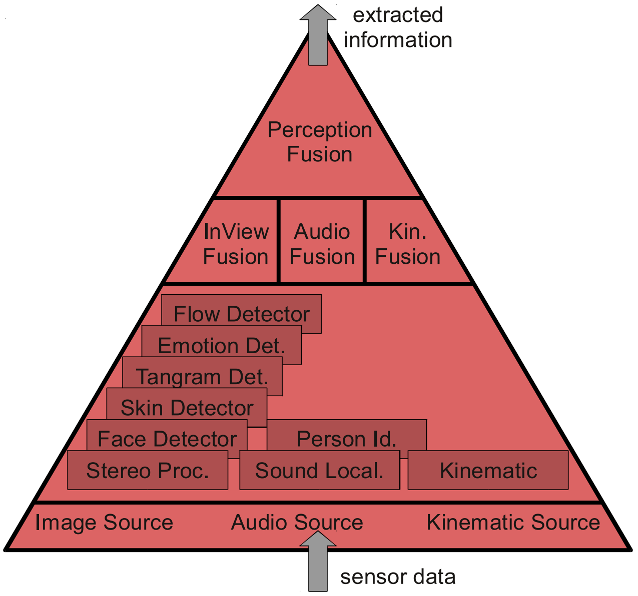


Fig. 3 The perception system of the robot is triggered by the data sources kinematic, audio, and video. Each source generates a set of basic perceptions using detectors like the *Face Detector* or *Flow Detector*. The output of each percept is fed into one of the three fusion modules for kinematic, audio, and video. All fusion modules are combined into a central *Perception Fusion* which combines all percepts into a central perception model

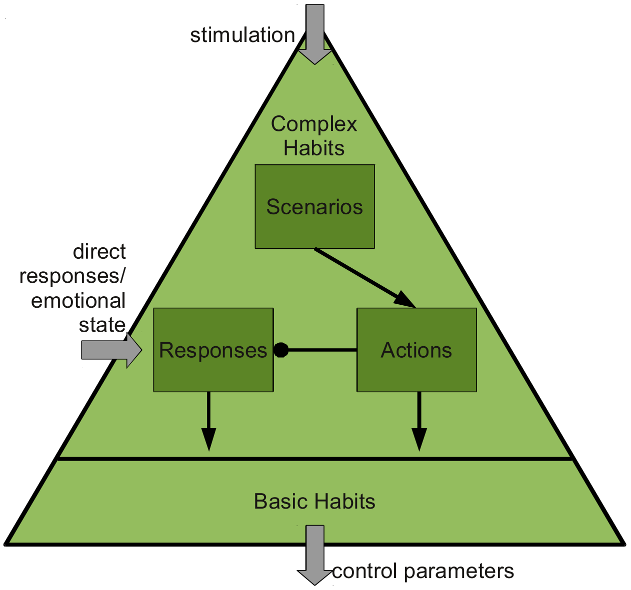


Fig. 4 Design structure of the behavior-network of the Habits

are consciously realized movement primitives like, e.g., nodding or focusing a certain object. Responses are unconsciously realized movements triggered by certain stimuli, e.g., an avoiding movement. Both, *Actions* and *Responses*, stimulate the *Basic Habits* to realize the desired movements. In addition, *Actions* can inhibit *Responses* in order to fulfill their tasks. *Actions* need to be stimulated by higher layers, whereas *Responses* get active depending on sensor information. At the highest layer, *Scenarios* (e.g. playing a game) are located. They provide information in which order different *Actions* need to be stimulated in a certain situation. In this structure, emotional expressions are realized either by the *Actions*, as consciously performed communication signals, or by the *Responses*, as a reaction to the information of the "internal sensor" perceiving the emotional state.

5 Exemplary Implementation of the Architecture

The emotion-based architecture described in Sect. 4 is implemented and tested using the humanoid robot ROMAN (see Fig. 5). This section explains the realization of the different parts of the emotion-based architecture in detail. The introduced implementation is used to test whether the proposed architecture fulfills all the functions of emotions.

5.1 Emotion

As depicted in Fig. 2, the emotion center is divided into two subparts, namely motives and the appraisal system. According to psychological theories, motives and emotions are strongly related, since the emotional state mainly depends on the satisfaction and the activity of the different motives.

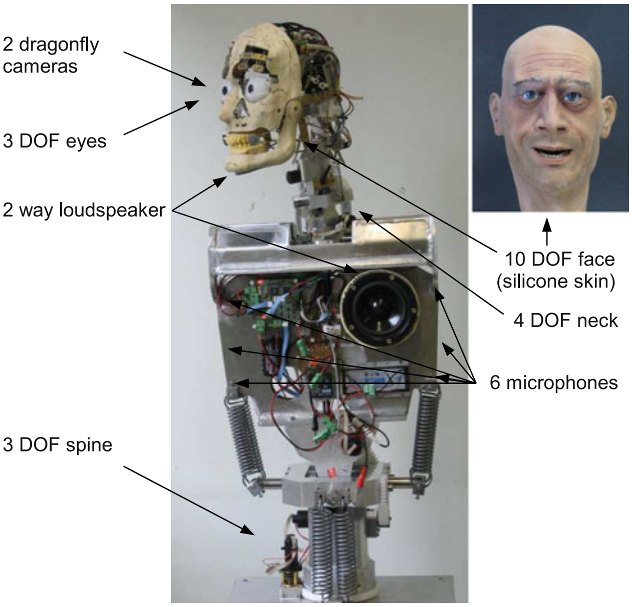


Fig. 5 The humanoid robot ROMAN of the University of Kaiserslautern. To describe the mechatronics the robot is depicted without its cover

5.1.1 Motives

For the realization of the motives, behavior modules as mentioned in Sect. 4 are used. In comparison to the standard behavior module, a motive owns an additional internal function, $\text{sat}(\mathbf{e})$, to calculate the motive's satisfaction. Although a motive inherits the target rating function of a behavior, the satisfaction function is necessary since a motive can be over- or under-satisfied, see [31]. That means the co-domain of a motive's satisfaction must be $[-1, 1]$, whereas the co-domain of the target rating is $[0, 1]$. A motive's goal is to keep its satisfaction in a medium state, that means in an epsilon-neighborhood around 0. If the motive is over-satisfied, it tries to avoid the corresponding stimuli, if the motive is under-satisfied it gets active in order to experience the requested stimuli. The motive gets active in both cases, but the corresponding output vector is different. The motive's target rating is calculated depending on the satisfaction, since a motive has reached its target if it is satisfied whereas the target is not reached if it is unsatisfied, no matter whether it is over- or under-satisfied.

The motives used for testing whether the proposed architecture fulfills the functions of emotion are:

Obey Humans: If a human gives the robot an order to do something, it will stop its actual work and obey the order.
Self Protection: Generate an evasive movement if a too close object is detected.
Communication: If a person is detected, this motive tries to start a conversation and takes care that the person is focused.
Exploration: If the robot is getting bored because of the absence of stimuli, this motive starts the exploration of the robot's surrounding.

Obey Humans and *Self Protection* are called *Basic Motives*, since they are responsible for the "survival" of the robot, whereas *Communication* and *Exploration* are named *Social Motives* since they correspond to ROMAN's "social" behavior.

In the following, the implementation of the motive system, target rating function, activation function, activity function, and transfer function, as well as the corresponding parameters, will be explained. As an example, the realization of the *Exploration Motive* and the *Communication Motive* are investigated. The function of the *Exploration Motive* is to look for new stimuli. Therefore it searches for interesting objects, e.g. human faces in the robot's surrounding. The *Communication Motive* starts an interaction if a human is detected and keeps the interaction partner in the focus of the robot.

Target Rating: A motive's target rating represents the distance to the motive's goal—reaching a satisfied state. A high target rating denotes a big distance to a satisfied state, a low target rating means the motive is almost satisfied. To calculate the target rating of a motive, at first the satisfaction has to be determined, see (11) and (12). Equation (11) is used to realize a cyclic motive characteristic. In this equation, $t_{\max}$ denotes the maximal time a situation has to last until the motive reaches either a fully satisfied or unsatisfied state. $t_{\max}$ is predefined during the creation of a motive since it characterizes this motive. A more relaxed motive can be designed by a rather big $t_{\max}$. In that case it will take a while until the motive wants to get active again. A more nervous motive, on the other hand, is described by a very small $t_{\max}$. The function $\mathrm{h}(\mathbf{e}, \mathbf{w})$ calculates whether the current situation is satisfying or not. Therefore the information of $\mathbf{w}$ is needed. It provides the information which value of the input vector $\mathbf{e}$ is of importance for the motive, $w_i = 1$ means the i-th input value is of importance $w_i = 0$ means unimportant. The vector $\mathbf{w}$ is also defined during the creation of the motive, since it represents the motive's view to the "world". The longer the desired stimulus is absent, the more unsatisfied the motive gets, which results in a negative satisfaction value. Furthermore, a motive can get over-satisfied if a certain stimulus is present for a too long time. It is also possible to realize a non-cyclic motive just reacting on perceived stimuli. Therefore, the calculation of the update function (11) can be adapted to make it time independent. Depending on the satisfaction, the target rating of a motive is calculated by (13). Assuming the satisfaction of a motive decreases and assuming that the motive was satisfied previously, the target rating increases since the distance to the motive's goal—a satisfied state—increases. On the other hand, if the satisfaction of a motive increases and again assuming that the motive was in a satisfied state, the distance to the motive's target also increases and because of this also the target rating will increase. In (12), the actual satisfaction of the motive is calculated. For this calculation, the cosine is used to ensure that the satisfaction value is in the range of $[-1, 1]$ and to realize a smooth behavior.

$$\mathrm{update}_t(\mathbf{e}, \mathbf{w}) = \begin{cases} \mathrm{update}_{t-1} + \frac{1}{t_{\max}}, & \text{if } \mathrm{h}(\mathbf{e}, \mathbf{w}) = 1 \\ \mathrm{update}_{t-1} - \frac{1}{t_{\max}}, & \text{else} \end{cases} \qquad (11)$$

$$\mathrm{sat}(\mathbf{e}, \mathbf{w}) = \cos\bigl(\pi \cdot (\mathrm{update}(\mathbf{e}, \mathbf{w}) + 1.5)\bigr) \qquad (12)$$

$$r = \bigl|\mathrm{sat}(\mathbf{e}, \mathbf{w})\bigr| \qquad (13)$$

For the *Exploration Motive*, the relevant input is *Interesting Object Detected*, therefore $\mathbf{w}$ contains the value 1 for *Interesting Object Detected* and all other elements of the vector are set to 0. The target rating of the *Exploration Motive* rises depending on the time elapsed since the environment was investigated. For testing purposes, motives that are inactive for a long time are not appropriate, therefore $t_{\max}$ is selected rather small, to 30 s. An interesting percept for the *Communication Motive* is *Communication Takes Place*. Because of this, the vector $\mathbf{w}$ of the *Communication Motive* has only one entry set to 1, the one correlated to *Communication Takes Place*. Therefore, the target rating of the *Communication Motive* rises over time until a human is detected and the communication takes place. For $t_{\max}$, again, a rather small value, 10 s, was selected in order to have no delays during the experiments—since exactly this motive should be active.

Activation: The activation ι of a motive is calculated as shown in (14). As already mentioned, the activation is an upper limit for the activity. The co-domain of the activation is $[0, 1]$, it depends on the stimulation (s) and on the inhibition ($\mathbf{i}$) of the motive; N denotes the number of inhibiting inputs. At the moment all motives are permanently stimulated, but imagining that there would be a cognitive layer on the top of the emotion-based architecture it would also be possible that this layer stimulates motives.

$$\iota(s, \mathbf{i}) = s \cdot (1 - \max(i_0, \dots, i_{N-1})) \qquad (14)$$

In absence of stimuli, the robot looks for humans, therefore the stimulation of the *Exploration Motive* is always set to 1. The *Exploration Motive* is inhibited by the *Basic Motives* and the *Communication Motive* as the robot's goal is to communicate with humans. Because of this the stimulation of the *Communication Motive* is also set to 1 and since the interaction should not be interrupted until it is finished the *Communication Motive* is only inhibited by the *Basic Motives*.

Activity: A motive's activity describes whether a motive is currently trying to influence the robot's behavior in order to reach a satisfied state. The higher the motive's activity the higher the motive's influence on the resulting robot behavior. The activity is calculated depending on its target rating, activation, input vector, and vector $\mathbf{w}$, see (15) to (17); the co-domain of the activity is $[0, \iota]$. The higher the activity of a motive, the higher is its influence on the robot's behavior. The activity is defined as the maximum of the functions $\mathrm{g}(r, \iota)$ and $\mathrm{h}(\mathbf{e}, \mathbf{w})$, see (15). In (16), two thresholds l_0 and l_1 are used. Like $t_{\max}$ in the target rating function the thresholds are predefined and represent the motive's character. For a more "good-natured" motive, l_0 and l_1 are selected rather big, that means the motive must be very discontent until it gets active and will take a long time until its activity reaches the maximum. For a more "short-tempered" motive, the thresholds are selected rather small, so that the motive already gets active if it is just a little discontent and it also reaches its maximum very fast. An active motive should remain active until its task is fulfilled, also if it is already satisfied. In this case, the target rating of the motive $r = 0$ and because of this $\mathrm{g}(0, \iota) = 0$. Therefore, the function $\mathrm{h}(\mathbf{e}, \mathbf{w})$, (17), has been introduced. If the motive was active in the previous state $t - 1$ and the corresponding stimulus is still present (in that case $\sum_{i=0}^{N-1} \frac{w_i}{\sum_{j=0}^{N-1} w_j} \cdot e_i = 1$). The motive remains active and its activity is set to the activity value of the previous state, a_{t-1}.

$$a = \max\big(\mathrm{g}(r, \iota), \mathrm{h}(\mathbf{e}, \mathbf{w})\big) \tag{15}$$

$$\mathrm{g}(\mathrm{r}, \iota) = \begin{cases} 0, & \text{if } r \cdot \iota < l_0 \\ 1, & \text{if } r \cdot \iota > l_1 \\ r \cdot \iota, & \text{else} \end{cases} \tag{16}$$

$$\mathrm{h}(\mathbf{e}, \mathbf{w}) = \left(\sum_{i=0}^{N-1} \frac{w_i}{\sum_{j=0}^{N-1} w_j} \cdot e_i \right) \cdot a_{t-1} \tag{17}$$

In the example of *Exploration* and *Communication*, the lower threshold for the *Exploration Motive* is chosen as 0 and the upper threshold $1 - V$, where V means the valence value of the robot's emotional state. That way the *Exploration Motive* reaches the maximum activity much faster when the robot is discontent in the actual situation and so it looks for new stimuli in order to reach a more content state. For the *Communication Motive*, the lower threshold is 0 and the upper threshold is 1. If the *Communication Motive* gets active, it inhibits the *Exploration Motive* by its activity. Again, the motives' parameters are defined to get the motives' responses quickly in order to reduce the latency for the experiments.

Transfer Function: To realize an appropriate behavior the transfer function is used to calculate the control values for the modulation of the different habits. As already mentioned in the description of the behavior module in Sect. 4, the data output is calculated depending on the data input. Besides this, the satisfaction, as a representation of the motive's internal state, is also used to calculate the data output. The output of a motive is used to change the emotional state of the robot and to stimulate different habits. This transfer function must be predefined for each motive. It provides the intelligence of the motive, calculating which habits to stimulate depending on the percepts and on the motive's satisfaction. Therefore, the transfer function is defined as:

$$\mathbf{u} = \mathrm{F}(\mathbf{e}, \mathrm{sat}(\mathbf{e}, \mathbf{w})) \tag{18}$$

The data output of the *Exploration Motive* used for the stimulation of the habits is calculated depending on the information whether an interesting object is in the focus of the robot or not. If for example a face is present, the habit to focus a recognized face is stimulated by the motive, otherwise the search habit is stimulated in order to find an interesting object. The data output of the *Communication Motive* also depends on the information of a human face. If a human face is detected, the *Communication Motive* stimulates the habit to focus a human face and it also stimulates the habit that starts the dialog. If no possible interaction partner is present or a possible partner does not communicate with the robot, the attraction habit will be stimulated.

To increase the clarity of the system, similar motives are arranged in groups. Within these groups, the different motives can be realized on different priority levels. Motives on higher priority levels inhibit motives on lower levels depending on their activity. For generating the output of a motive group, the outputs of the single motives are merged by a weighted fusion, see Sect. 4. The different motive groups can also be realized on different priority levels. Groups on higher priority levels inhibit groups on lower levels. The activity of a motive group is represented by the activity of the corresponding fusion module and the inhibition of a motive group is realized by inhibiting this fusion module. Because of this modular setup the motive system remains maintainable and it can easily be modified or extended.

In the investigated example, two motive groups are realized, see Fig. 6. Within the *Social Motives*, the *Communication Motive* is on a higher priority level than the *Exploration Motive*, therefore the *Exploration Motive* is inhibited by the *Communication Motive*. The control data output of both motives is merged by a weighted fusion depending on the activity of the different motives. In the *Basic Motives* group, the *Obey Humans Motive* has a higher priority than the *Self Protection Motive*. These groups are also considered as motives. In this example, the *Basic Motives* are on a higher priority level than the *Social Motives*, so the *Social Motives* are inhibited by the *Basic Motives*. The output vectors of the single motive groups are also merged by a weighted fusion depending on the activity of the different groups. On one hand the output of the motives is used to stimulate different habits and to generate goal directed behavior. On the other hand the output is used by the appraisal system to determine the current emotional state.

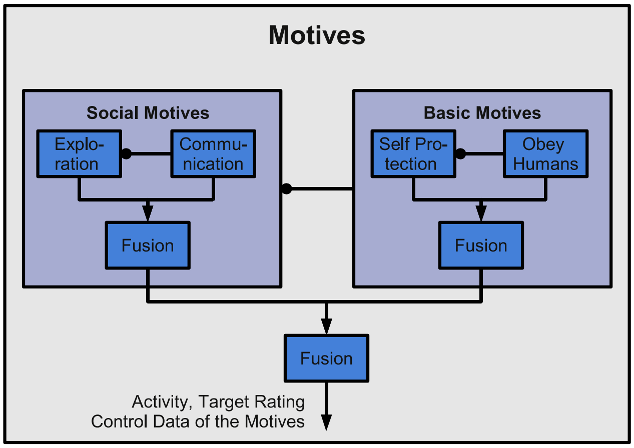


Fig. 6 An extract of the realized motive system; basic motives have a higher priority level than social motives, represented by inhibiting links

5.1.2 Appraisal System

As already mentioned, the appraisal system evaluates the situation using three dimensions (A, V, S) (arousal, valence, and stance). To implement this system the behavior-based modules are used, too. Each dimension is realized by one behavior module. The modules' output represents the result of the appraisal for the different dimensions. This (A, V, S)-value is propagated throughout the system. It represents the current emotional state. The different habits locally interpret the (A, V, S)-value and adapt their action to the current emotional state. E.g., the habits concerned with the facial expressions map the (A, V, S)-value to a certain emotion, see Sect. 5.3. Another example is the habits that are responsible for ROMAN's speech. Depending on the (A, V, S)-value, they change the frequency, speed, and volume of the speech.

The evaluation of the situation is realized depending on the activity and the target rating of the motives, and on the environmental information. In Algorithm 1, the calculation of arousal is described. The robot's arousal depends on the intensity of the perceived stimuli. Where, the intensity of a stimulus is increased by a predefined factor *unexp* if this stimulus is not expected in the current situation. The factor *unexp* can be seen as a representation of the robot's nervousness. The expected stimuli for a situation can be derived from the active motives, since every motive expects certain inputs. If a perceived stimulus is in the list of recommended inputs (represented by the vector $\mathbf{w}$, see Sect. 5.1.1), this stimulus is rated as expected, otherwise as unexpected. The robot's arousal is defined according to the intensity of the perceived stimuli and it is decreased in absence of any stimulus. *Weight_arousal* is used to define how fast the arousal value of the robot changes. A high value for *weight_arousal* means fast changes, whereas a low value means slow changes, since *weight_arousal* defines the influence of the currently calculated arousal to the determination of the new arousal value A_t (A_{t-1} denotes the previous arousal value).

In Algorithm 2, the robot's valence is calculated. Therefore, the target rating of the motives is used. The target rating represents the satisfaction of a motive. To derive the valence the weighted sum of all target ratings is generated, since some motives can be more important for the robot than others. The weights for the different motives must be predefined and they depend on the application of the robot. For an interactive robot, e.g. the weight for a communication motive will be set to 1, whereas for a security robot the weight of an exploration motive should be 1. For the realized experiments, the weights for the *Basic Motives* and for the *Communication Motive* are set to 1. The weight for the *Exploration Motive* is set to 0.75. The *Basic Motives* are important for the robot's safety and the *Communication Motive* represents the robots most important task, at least regarding the realized experiments. According to the previous algorithm, *weight_valence* defines the influence of the currently calculated valence value and the previous one (V_{t-1}) to the resulting valence value V.

Algorithm 1 Calculating arousal (A)

```
if $number\_of\_stimuli > 0$ then
    for $i = 0$ to $number\ of\ stimuli - 1$ do
        $\hat{A} = intensity\_of\_stimulus\_i \cdot unexp$
    end for
    $A_t = \frac{weight\_arousal \cdot \hat{A} + A_{t-1}}{weight\_arousal + 1}$
    limit $A$ to the domain of $[-1, 1]$
else
    $A_t = A_{t-1} - (\frac{A_{t-1}+1}{2})^2$
end if
```

To derive the stance value, Algorithm 3 is used. Stance describes the robot's attitude concerning environmental stimuli. High stance means the perceived stimuli are requested, low stance means they are unrequested. The robot's stance depends on the currently active motives. As already mentioned, the motives provide information which stimuli are requested and which are unrequested. Depending on the activity of the motives a list of currently requested stimuli can be generated. It the perceived stimuli are required by the active motives, these stimuli enable the motives to increase their satisfaction. This is indicated by a high stance value. If the currently perceived stimuli are unrequested they will lead to a decrease of the satisfaction of the motives. Therefore, the stance value will be rather low and a retreating be-

Algorithm 2 Calculating valence (V)

```
for $i = 0$ to $number\_of\_motives - 1$ do
    $\hat{V} = -2 \cdot (target\_rating\_motive\_i$
        $\cdot\ weight\_motive\_i - \frac{1}{2})$
    $V_t = \frac{V_{t-1} + weight\_valence \cdot \hat{V}^3}{weight\_valence + 1}$
end for
```

Algorithm 3 Calculating stance (S)

```
for $i = 0$ to $number\_of\_motives - 1$ do
    if motive is active then
        $req\_stim_t = req\_stim_{t-1} \cup req\_stim_m$
        $unreq\_stim_t = unreq\_stim_{t-1} \cup unreq\_stim_m$
    end if
end for
for $i = 0$ to $number\_of\_stimuli - 1$ do
    if stimulus $\in req\_stim$ then
        $S_t = S_{t-1} + pos\_step\_size$
    end if
    if stimulus $\in unreq\_stim$ then
        $S_t = S_{t-1}$ - $neg\_step\_size$
    end if
end for
```

Table 1 Exemplary definitions for the robot's parameters according to the traditional personality types

Parameter	Phlegmatic	Melancholic	Choleric	Sanguineous	Neutral
unexp	1	1	6	2	2
$weight_{arousal}$	2	1	4	2	2
$weight_{valence}$	2	1	4	0.5	2
pos_step_size	1	0.2	4	4	2
neg_step_size	0.2	0.2	4	0.5	2

havior will be initiated. Since the stance value should increase with a different gradient than it decreases, there are two different step sizes (*neg_step_size* and *pos_step_size*) defined. That way, the characteristic behavior of a robot's stance can be defined. The previous stance value is represented by S_{t-1}, the resulting stance value is by S_t. The list of currently requested stimuli is described by *req_stim* and the unrequested by *unreq_stim*; the requested and unrequested stimuli of a motive are describe by req_stim_m and $unreq\_stim_m$, resp.

By defining the values for the parameters used in these algorithms—the factor *unexp*, *weight_arousal*, *weight_motive_i*, *weight_valence*, *pos_step_size*, and *neg_step_size*—the personality of the robot can be set. As already mentioned, these parameters are predefined and remain constant throughout the lifetime of the system. By setting *unexp*, the robot's reaction to unexpected stimuli can be defined. For example, if the robot is communicating with a human, it expects the sound of the voice of a human. If the robot suddenly perceives a loud sound in its back, it will not have expected that. Because of this, the intensity of the perceived stimulus is increased by *unexp* and the robot will be very aroused because of this. By *weight_arousal*, *weight_valence*, *pos_step_size*, and *neg_step_size*, the degree of change in the different dimensions can be set. If e.g. *weight_arousal* is set to 2, the influence of the newly calculated arousal value is twice the influence of the previous arousal value. That means the higher *weight_arousal*, the faster the robot's arousal changes. *Weight_valence* has the same meaning to the robot's valence. *Pos_step_size* and *neg_step_size* define how fast the stance value changes, in positive and in negative direction, resp. If e.g. *pos_step_size* is set to 2 and *neg_step_size* is set to 1, stance will increase twice as fast as it decreases. In Table 1, some exemplary definitions of the robot's personality are presented. For the test implementation on ROMAN, the *Neutral* personality was chosen.

5.2 Percepts

A perception system for a humanoid robot fulfills the task of sensor data abstraction. This is required to extract relevant information from the constant stream of raw data provided by the cameras, microphones and kinematic system. This information extraction uses the prototype of natural human-interaction to select a set of important and useful interaction signals out of the infinitely large amount of perceivable entities. In addition to the vast amount of percepts also ambiguities and situation depending signals can be perceived.

The interaction process triggered by the user model itself should be "natural" which does not allow environmental changes like markers and includes verbal as well as nonverbal interaction signals. Fortunately, the interaction signals and their meaning is in the focus of scientific researchers in the areas of psychology and social sciences. Based on this information, seven major categories of signals with varying importance can be found.

Paralanguage: This term defines all information transmitted on top of speech signals but not directly related to the content of the spoken words. Further subdivision distinguishes perspective, organic, expressive and linguistic aspects.

Kinesics: The term kinesics describes any type of body motion including head, facials, trunk, hands, and so on.

Proxemics: The perception of distance zones, spatial arrangements and sensory capabilities is a key factor of any conversation.

Olfactory: The scent of interaction partners has an influence although it is limited to a close area.

Haptics: Haptic interaction signals are related to specific situations like hand shaking, beating, or grabbing. These communicative actions require an embodied agent and direct feedback of the robot.

Artifacts: Artifacts in the environment and even the environment itself influences a communication process. Depending on the situation, artifacts may even be necessary. Examples of relevant artifacts are gaming pieces during play or goods in sales conversations.

Language: Besides the previously mentioned non-verbal aspects, there is of course also the verbal aspect of interaction. This includes the spoken words and the content that is being transmitted. Although non-verbal aspects are considered important, most of the conversational content is transferred via the speech.

An example of kinesic interaction signals is the nodding and head shaking action which are commonly annotated with agreement and disagreement. These signals are commonly used and form a basic non-verbal signal in a broad range of ethnic groups. The following section introduces all basic percepts and fusion modules that are required for the experiments described in Sect. 6.

Basic Percepts The notion of "basic percepts" has been introduced to describe a module which analyzes the real sensor information with respect to a specific property. This concept allows a modular integration of additional sensors for a specific type of information. Relevant percepts for the detection of head nodding are

Stereo Processing: The stereo module uses a dense stereo algorithm to assign a depth value to each pixel in the observed area. This information is relevant to obtain distances and to reduce the number of false detections of the following percepts.

Skin Detector: The skin detector assigns a skin color probability to each pixel. This information is relevant to extract facial regions which are often visible in interaction scenarios.

Face Detector: The face detector uses a pattern recognition algorithm to generate a set of face candidates. Each candidate is verified by using distance and skin color information. The combination of these percepts dramatically reduces the number of false detections to a minimum.

Flow Detector: The flow detector assigns motion directions and distances to a set of well trackable points. All motion vectors within a previously detected facial region provide an estimation of head motion direction.

In View Fusion: The in view fusion realizes the time dependent tracking of faces and face regions and combines the flow-information with the tracking data. The output of the visual fusion is then passed to the overall Perception Fusion which may add audio and kinematic information.

The Perception Fusion provides a detailed description of all users as well as specific artifacts in the environment of the robot. Table 2 lists a subset of the properties provided by the Perception system. One object descriptor is provided for each object of interest.

5.2.1 *Nodding and Shaking*

The probabilities of nodding $p_{nodding}$ and shaking $p_{shaking}$ are calculated by

$$p_{nodding} = \|(ampl_y - ampl_x) \cdot 10\| \cdot \|ampl_y\| \qquad (19)$$

while $\| \bullet \|$ indicates a limitation to the interval $[0, 1]$ and $ampl_x, ampl_y$ are the average amplitudes of the optical flow in x and y direction. The activity of the module *nodding* is high when the amplitude of the humans face in y-direction of the perceived image is high and the difference between the amplitude in y and x direction is positive and high. These heuristics describe the typical observations when a person is nodding. The head shaking percept is implemented similarly to the nodding percept with inverted amplitudes in x-direction and y-direction.

Table 2 The output of the *Perception Fusion* module is a list of objects containing a large amount of assigned properties like position information. The selection listed here is limited to the information required to extract nodding and head shaking movement. Additional information like emotional state, gaze direction, and person identification are also available

Object	Type $t = \langle id \rangle$
	TimeStamp $t = \langle time \rangle$
	Pos $p = (x, y)$
	AvgDistance $avgdist = d$
	AvgOpticalFlow $avgflow = (f_x, f_y)$
	AmountFlowVectors $afv = \langle num \rangle$
	Nodding $p_{nodding} = \langle prob \rangle$
	Shaking $p_{shaking} = \langle prob \rangle$
	...

The habits can observe the probabilities of head shaking and nodding during operation at any time for any object in the visual focus of the robot. The perception system itself does not directly influence the motion or movements of the robot. All motions are controlled by the habits based on the information provided here.

5.3 Habits

As mentioned above, the habits are realized as a behavior network using the standard behavior module of iB2C, see Sect. 4. The description of the inputs and outputs as well as of the internal functions can be found there.

For the explanation of the motives the example of the *Communication Motive* was used. During the conversation the emotional state of the robot changes and because of this the robot's emotional expressions will change. Therefore the habits are explained using the example of generating facial expressions. As already mentioned, basic habits correspond to the motor primitives of the robot-system, like e.g. *Head Up/Head Down*. The basic habits that are involved in the process of generating facial expressions are: *Mouth Corner Up/Down*, *Mouth Corner Forward/Backward*, *Wrinkle Nose*, *Inner Eyebrow Up/Down*, *Open/Close Mouth*. Several of these basic habits can be combined to generate complex facial expressions, e.g. happiness. For the generation of facial expressions, the expressions corresponding to the six basic emotions are defined according to [42]. Every expression is represented by one behavior module.

For the calculation of the activity of an expression, the six basic emotions can be imagined as points within a three-dimensional space (according to the three dimensions of the appraisal system). The activity of a certain emotional expression is calculated using (20). In this equation, a_i is the activity of emotional expression i where $i \in \{$anger, disgust, fear, happiness, sadness, and surprise$\}$ and d denotes the largest possible distance between two points within this space. P_i is the point representing emotion i in the three-dimensional space and $\mathbf{e}$ represents the current emotional state of the robot, the outputs of the arousal, valence, and stance modules. That way, an activity value in the range of [0, 1] for every basic emotional expression is calculated.

$$a_i = \frac{d - |P_i - \mathbf{e}|}{d} \tag{20}$$

Every basic emotional expression calculates its data output to stimulate the basic habits that are necessary for the generation of facial expressions. Therefore, a basic emotional expression owns a vector, **bh**, containing the information how strong the different basic habits should be activated, $\mathbf{bh} = (bh_1, \ldots, bh_N)$. The strength for basic habit i is represented by the vector element bh_i. The output vector $\mathbf{u}_i$ of basic emotional expression i is defined as the product of the basic emotional expression's activity and its vector **bh**, see (21). The higher the activity of a basic emotional expression the stronger is the corresponding facial expression displayed.

$$\mathbf{u}_i = \mathbf{bh_i} \cdot a_i \tag{21}$$

The output vectors of all basic emotions are merged by a weighted fusion depending on the activity of the single basic emotions, see (22). In this equation $\mathbf{s}$ represents a vector containing the stimulation values for the basic habits, where u_i denotes the output vector of basic emotion i and a_i resp. a_j represents the activity of emotion i or j resp.

$$\mathbf{s} = \sum_{i=0}^{5} \frac{a_i}{\sum_{j=0, j \neq i}^{5} a_j} \cdot \mathbf{u}_i \tag{22}$$

That way, different emotional expressions can be generated and because of the weighted fusion any combination of basic emotions can be displayed. So, the expressive system of the robot is not limited to the expression of basic emotions. Depending on the stimulation calculated in (22), the different basic habits get active. The actuator layer maps the activity of the basic habits to specific motor positions and velocities. This information is send to the motor controllers. Using the abstract representation provided by simulation and activity, the *Habits* remain hardware independent. Just the actuator layer needs to be adapted or exchanged if the hardware changes. This also allows to easily replace the layer for the "real" actuators by a simulation layer [43]. The complete process of generating emotional expressions is depicted in Fig. 7.

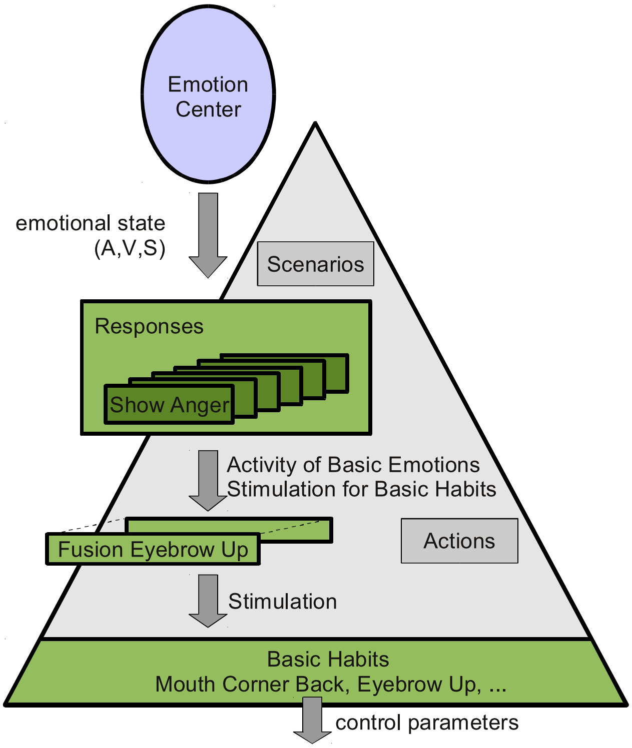


Fig. 7 The generation of emotional expressions using the concept of habits. The *Emotion Center* provides the emotional state. The expression modules located in the *Responses* get active and stimulate the *Basic Habits*. The *Basic Habits* produce motor control parameters

Besides the nonverbal behavior and expression also verbal communication is realized. As mentioned above, for verbal communication ROMAN is equipped with a dialog system. The dialog structure needs to be specified in an XML description file. Besides the speech output the dialog system also provides information about the meaning of the recognized text, e.g. friendly or unfriendly, and of the state of the current interaction. This information is treated by the control architecture in a similar way as the percepts and they are also used to calculate the satisfaction of the motives to change the emotional state, or to stimulate direct nonverbal responses. The sound of the robot's voice—volume, frequency, and velocity—is changed depending on the emotional state. If e.g. the robot is very aroused, its speech is much faster and the tone of the voice is much higher than in a more relaxed state.

6 Experiments

The experiments conducted so far, have been realized in order to prove whether the proposed architecture implements the five functions of emotion. Therefore, the different functions have been tested separately.

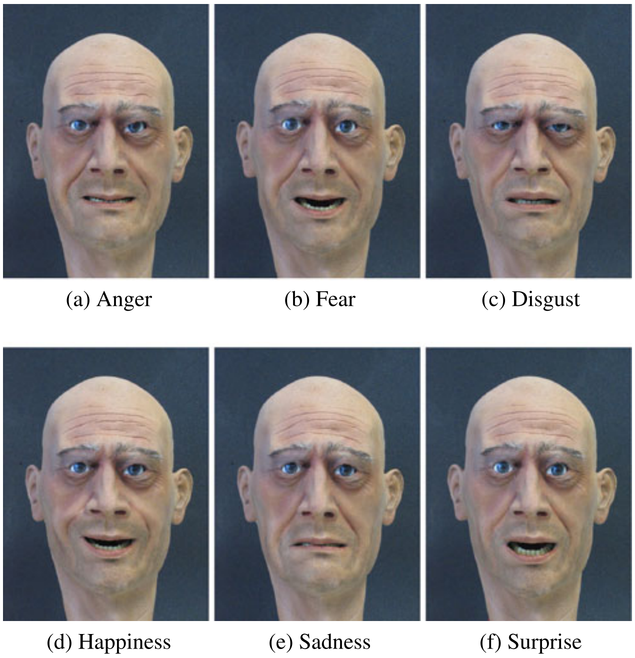
At first, experiments have been conducted in which non-expert subjects had to rate and classify ROMAN's expressions. The results of these experiments have been evaluated using multivariate variance analysis to determine significant differences between the ratings for the different expressions (significance $\alpha < 5\%$). To figure out whether there are differences between static or dynamic expressions the experiments was realized twice, using images and videos, resp. The hypothesis was that the rating for the expressed emotion is significantly higher than for all other, currently not expressed, basic emotions. The second hypothesis was that there is no significant difference between the staic and the dynamic expressions. Furthermore, using analysis of variance it had been analyzed whether significant differences between the currently displayed emotion and a single currently not displayed emotion exists. For this analysis the "Statistical Package for the Social Scientist" has been used.

Experimental setup: To test the quality of ROMAN's facial expressions, nine images showing different emotions were presented to 32 persons (13 women and 19 men) aged 21 to 61 years [44]. Every person had to rate the correlation of the presented expression to the six basic emotions, using a scale from 1 to 5 (1 means a weak and 5 a strong correlation). The facial expressions of ROMAN are displayed in Fig. 8.

Experimental results: First of all the hypothesis that there are no significant differences between the static and the dynamic presentation of the expressions was confirmed. The results of the evaluation are shown in Tables 3 and 4. In Tables 3 and 4, the left column contains the displayed expression. The right column contains the average values of the detected correlation to the six basic emotional expressions. Table 5 displays single significant differences between the displayed expression and the basic emotion expressions. The evaluation of the results showed that the correct recognition of the expressions anger, happiness, sadness, and surprise is significant. But subjects found it hard to distinguish between the expressions of fear, and disgust. Furthermore, two understated expressions (fear 50% activation and sadness 50% activation) have been presented. These expressions have been realized by stimulating the corresponding habits (expression modules) with only 50% of the intensity. The analysis of the results for these expressions showed: For fear, the subjects recognized that the presented emotion is not that intensive as in the case of a 100% stimulation. For sadness they were not able to distinguish between 50% and 100% stimulation. Finally, a mixture of several emotions have been presented and as expected, no specific emotion could be recognized. Nevertheless, the involved emotional expressions are identified in most cases.

However, it is still necessary to increase the quality of ROMAN's expressions. especially *disgust* and *fear* need to be improved. Therefore, ROMAN's expression system has been extended by gestures. On that score, arms and hands are currently under development. To implement and test the control software in parallel, a simulation of the robot's actuators as well as a visualization of the robot itself have been integrated [43]. Using this simulation, similar experiments have been conducted.

Experimental setup: In this experiment, twelve images showing the six basic emotions have been presented. Six of


(a) Anger (b) Fear (c) Disgust

(d) Happiness (e) Sadness (f) Surprise

Fig. 8 The facial expressions of the 6 basic emotions generated by ROMAN

Table 3 The results of the experimental evaluation of ROMAN's facial expressions (static presentation). On the left of each row the represented emotion is named, followed by the strength of the correlation to the different basic emotion

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	4.4	1.9	1.6	1.0	1.5	1.1
Disgust	3.7	2.6	1.8	1.0	2.6	1.1
Fear	1.3	1.8	3.6	1.4	1.8	3.8
Happiness	1.1	1.0	1.2	4.4	1.0	2.2
Sadness	2.1	2.1	2.8	1.0	3.9	1.3
Surprise	1.3	1.3	2.8	1.4	1.6	4.1
50% fear	1.7	1.7	2.9	1.6	1.8	2.5
Anger, Fear, Disgust	3.1	2.3	2.0	1.0	2.4	1.3
50% sadness	2.1	2.1	2.8	1.0	3.9	1.3

Table 4 The results of the experimental evaluation of ROMAN's facial expressions (dynamic presentation). On the left of each row the represented emotion is named, followed by the strength of the correlation to the different basic emotion

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	3.7	2.1	2.6	1.3	1.7	1.7
Disgust	3.1	2.7	1.9	1.0	2.4	1.5
Fear	1.7	1.6	3.3	2.4	1.5	4.0
Happiness	1.1	1.0	1.1	4.8	1.1	2.4
Sadness	1.5	1.4	1.9	1.1	3.5	1.4
Surprise	1.7	1.7	3.5	1.2	1.2	4.6
50% fear	1.8	1.8	3.2	1.2	1.8	3.4
Anger, Fear, Disgust	1.5	1.4	2.4	1.1	3.3	1.4
50% sadness	1.5	1.4	1.9	1.1	3.5	1.4

Table 5 The results of the experimental evaluation of ROMAN's facial expressions. On the left of each row the represented emotion is named, followed by the information whether a significant difference to the other basic emotions was detected or not (significance $\alpha < 5\%$). ● denotes a significant difference, ○ denotes no significant difference

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	–	●	●	●	●	●
Disgust	○	–	●	●	○	●
Fear	●	●	–	○	●	○
Happiness	●	●	●	–	●	●
Sadness	●	●	●	●	–	●
Surprise	●	●	●	●	●	–
50% fear	●	●	–	●	●	○
Anger, Fear, Disgust	○	○	○	○	○	○
50% sadness	●	●	●	●	–	●

these images depicted the facial expressions of the simulated robot, the other six a showed the combination of facial expressions and gestures, see Fig. 9.

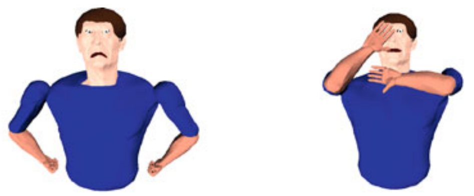
Experimental results: It turned out that in all cases the combination of facial expressions and gestures increased the number of correct rated results. Especially, it seemed to be easier for the subjects to identify the expression of disgust and not mix it up with the expressions of anger and sadness. Unfortunately, it is still hard to distinguish between fear and surprise. These results using the simulation can be seen as a good hint that the integration of arms and hands might also increase the quality of ROMAN's expressions.

All the subjects involved in these experiments were from Germany. This is an important fact, since the cultural background plays an important role for the interpretation of emotional expressions. If ROMAN should be used in a different cultural environment, its expression might need to be adapted. Because of the modular setup of the control architecture, in that case only the habits responsible for the different expressions would have to be changed.

In a second experiment for testing the functions of the emotion architecture, a very simple interaction situation was realized [7]. Since the focus of the experiment was not on handling interaction situations but on the ability of the robot's motivation system, a very simple scenario seems to be suitable. Compared to previous experiments, this time the functioning of the appraisal system was tested in addition.

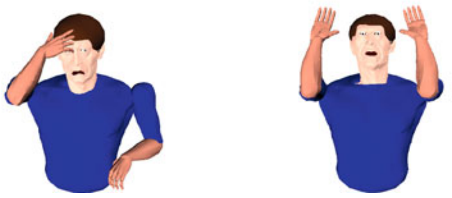
Experimental setup: ROMAN is placed in the hall of the Robotics Research Lab where it introduces the Lab to the people walking by. For that reason, the robot searches for humans, welcomes them and asks whether they are interested in information about the lab or not. If they are interested, ROMAN provides them with the recommended information. To achieve this goal, the following components of the emo tion-based architecture are mainly involved: The percepts to detect humans as well as to detect the non-verbal signals nodding and shaking the head are used. The mainly used habits are searching and focusing humans. To detect humans in the surrounding of the robot, the *Exploration Motive* is needed. Whenever a person is detected, the interaction is triggered by the *Communication Motive*.

Course of the experiment: To evaluate the performance of the architecture the standardized signals—activation (ι), activity (a), and target rating (r)—of the involved motives and habits were recorded and drawn against the time in seconds, see Fig. 10. At first, the target rating of the exploration motive rises, the motive gets active and stimulates the *Search* habit. The robot turns its head and the percepts detect the person. After approx. 5 s, the *Communication Motive* gets active and inhibits the *Exploration Motive*. During the activity of the *Communication Motive*, the *Focus Face* habit is activated and the percepts still have to deliver the position of the detected face. During the communication, ROMAN asks



(a) Anger (b) Fear

(c) Disgust (d) Happiness



(e) Sadness (f) Surprise

Fig. 9 The facial expressions of the 6 basic emotions in combination with gestures

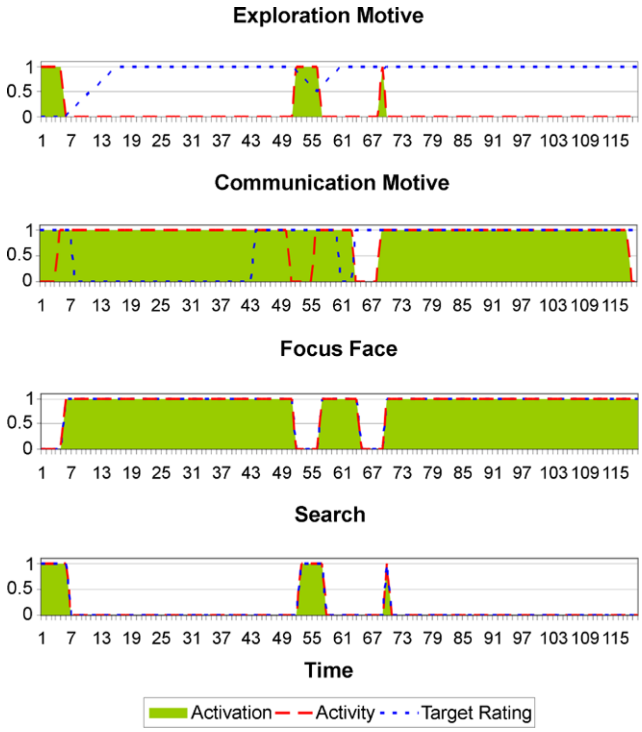


Fig. 10 The activations, activities, and target ratings of the *Exploration Motive* and the *Communication Motive*, as well as of the involved habits *Focus Face* and *Search* during a communication situation

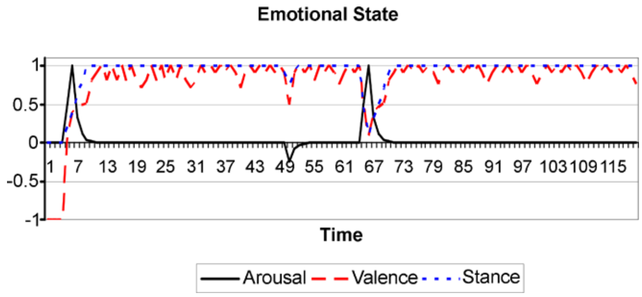


Fig. 11 The robot's emotional state over time during a communication situation

the human whether he or she is interested in some information about the lab, the human can either answer this question verbally by saying "Yes" or "No", or non-verbally by nodding or shaking his head. Therefore the percepts also have to provide the information whether nodding or shaking the head is detected. After approx. 45 s the target rating of the *Communication Motive* increases as the probability rate of the human detection decreases. Approx. at 50 s, the activity of the communication reaches 1 and the exploration gets active again. A few seconds later, the subject is detected again and the communication continues at the point where it was interrupted. It is quite interesting that the human interaction partner did not notice this. After approx. 58 s, the activation of both motives decrease. The reason for this is that they have been interrupted by the *Basic Motives* since the human interaction partner steps too close to ROMAN. ROMAN asks the person to step back and afterwards, after approx. 68 s, the communication is completed. During the whole communication process, ROMAN has been generating emotional expressions depending on its emotional state. This state was changed by the different motives and by the information derived from the dialog. In Fig. 11, the trace of the emotional state during the conversation is displayed. When recognizing the human standing next to ROMAN, its arousal increases since a new intensive stimulus is perceived. Since the robot's goal is to communicate, valence and stance increases after the communication has started. After approx. 50 s the arousal value decreases since the stimulus—the recognized human—is getting weaker—therefore, stance decreases—and since the target rating of the *Communication Motive* increases—the motive is discontent—the valence value decreases. When the robot is interrupted by the human after approx. 58 s, the arousal increases as the robot did not expect such an intensive stimulus, and valence because the target ratings of the motives increase; stance decreases because the perceived stimulus was unrequested. and the robot blames the human. Afterwards, the communication continues, the arousal decreases and since the target rating of the motives decreases valence increases. The perceived stimulus—the interaction partner—is requested and because of this stance increases.

Experimental results: The presented experiment demonstrated that it is possible to realize the five functions of emotion with the proposed architecture. The expressive function is realized by the emotion and the habits when generating emotional expressions. The motivational and the selective function are fulfilled by the motives by stimulating searching or focusing a face and by demanding the information whether a face is present or whether the interaction partner is nodding or shaking his or her head. The emotional state and the motives implement the regulative and the rating function. The *Basic Motives* cause a significant change in the emotional state when the interaction partner steps too close to ROMAN. Finally the rating and the regulative function are fulfilled by the appraisal system. The appraisal system generated a high level description of the actual state.

7 Conclusion

Although the necessity and the benefit of emotions for robots is already analyzed and proven by researchers, the transfer to robot control architectures is a hard problem.

In this paper the UKL Emotion-based Control Architecture was presented which fulfills all five functions of emotion—regulative, selective, expressive, motivational, and rating—and realizes additionally secondary conditions like scalability, perceptibility, and parameterability.

For testing purposes, the architecture was implemented on the humanoid robot ROMAN—a human-sized natural looking upper body system. Experimental results performed in public environments showed that the functions of emotion are in principle fulfilled.

Although the results are promising these studies are just the first step towards social interaction. In the future the proposed architecture needs to be tested in more complex situations where more interaction takes place. To handle these scenarios additional motives, habits, and percepts need to be realized. One of these complex situations planned for the future is a tangram game where the robot can influence its human interaction partner either in a motivating or demotivating way, depending on the robot's emotional state. The advantage of a tangram game situation, compared to other interaction scenarios, is that it provides still enough human-robot interaction and in addition the possible human actions are known and can be perceived. Every human move can be rated by the robot, whether it leads to the goal or not. First steps towards this scenario have already been taken. At the *Hannover Messe 2009*,[4] first experiments where ROMAN rated the tangram playing of fair attendees have been conducted. This scenario will be improved in the future in order to realize "real world" interaction experiments.

Acknowledgements This work is supported by the Carl-Zeiss-Stiftung. Additionally, this work is funded by the International Research Training Group of the University of Kaiserslautern (IRTG 1131) which is funded by the Deutsche Forschungs-Gemeinschaft (DFG).

References

1. Tapus A, Matarić M, Scassellati B (2007) The grand challenges in socially assistive robotics. IEEE Robot Autom Mag 14(1):35–42
2. Canamero D, van de Velde W (1997) Socially emotional: using emotions to ground social interaction. In: Proceedings of the AAAI fall symposium on socially intelligent agents, Cambridge, USA, pp 10–15
3. Canamero D (2000) Designing emotions for activity selection. Technical report, LEGO Lab, University of Aarhus
4. Adolphs R (2005) Could a robot have emotions? Theoretical perspective from social cognitive neuroscience. In: Fellous J, Arbib M (eds) Who needs emotions? The brain meet the robot. Oxford University Press, London, pp 9–28
5. Hobmair H, Altenhan S, Betscher-Ott S, Dirrigl W et al (2003) Psychologie. Bildungsverlag EINS, Troisdorf
6. Lückert HR, Lückert I (1994) Einführung in die Kognitive Verhaltenstherapie Allgemeine Grundlagen. Ernst Reinhardt Verlag, München
7. Hirth J, Berns K (2009) Emotion-based architecture for social interactive robots. In: Choi B (ed) Humanoid robots, in-tech, pp 97–116
8. Ulich D, Mayring P (1992) Psychologie der Emotionen, vol 5. Kohlhammer, Stuttgart
9. Fellous J (2004) From human emotions to robot emotions. In: Proceedings of the AAAI spring symposium, Menlo Park, USA, pp 37–47
10. Mietzel G (2000) Wege in die Psychologie. Klett-Cotta, Stuttgart
11. Heckhausen J, Heckhausen H (eds) (2006) Motivation und Handeln. Springer, Berlin
12. Scherer K (1999) Appraisal theory. In: Dalgeish T, Power M (eds) Handbook of cognition and emotion. Wiley, New York, pp 637–663
13. Kelley A (2005) Neurochemical networks encoding emotion and motivation—an evolutionary perspective. In: Fellous J, Arbib M (eds) Who needs emotions?—The brain meets the robot. Oxford University Press, London, pp 29–77
14. Fong T, Nourbakhsh I, Dautenhahn K (2003) A survey of socially interactive robots. Robot Auton Syst 42:143–166
15. Vernon D, Metta G, Sandini G (2007) A survey of artificial cognitive systems: implications for the autonomous development of mental capabilities in computational agents. IEEE Trans Evol Comput 11(2):151–180
16. Bittermann A, Kühnlenz K, Buss M (2007) On the evaluation of emotion expressing robots. In: 2007 IEEE international conference on robotics and automation (ICRA), Roma, Italy, pp 2138–2143
17. van Breemen A, Xue Y (2006) Advanced animation engine for user-interface robots. In: Proceedings of the 2006 IEEE/RSJ international conference on intelligent robots and systems (IROS), Beijing, China, pp 1824–1830
18. Saldien J, Goris K, Vanderborght B, Vanderfaeillie J et al (2010) Expressing emotions with the social robot probo. Int J Soc Robot 2(4):377–389
19. Kozima H, Michalowski MP, Nakagawa C (2009) Keepon—a playful robot for research, therapy, and entertainment. Int J Soc Robot 1(1):3–18

---

[4] http://www.hannovermesse.de/homepage_e

20. Spexard T, Hanheide M, Sagerer G (2007) Human-oriented interaction with an anthropomorphic robot. IEEE Trans Robot 23(5):852–862. Special issue on Human-Robot Interaction
21. Beuter N, Spexard T, Lütkebohle I, Peltason J et al (2008) Where is this?—Gesture based multimodal interaction with an anthropomorphic robot. In: Proceedings of the IEEE-RAS international conference on humanoid robots (Humanoids), Daejeon, Korea, pp 585–591
22. Kim HH, Lee HE, Kim Y, Park KH et al (2007) Automatic generation of conversational robot gestures for human-friendly steward robot. In: Proceedings of the international conference on robot and human interactive communication, Jeju, Korea, pp 1155–1160
23. Lee H, Park J, Chung M (2006) An affect-expression space model of the face in a mascot-type robot. In: Proceedings of the 2006 6th IEEE-RAS international conference on humanoid robots, Genoa, Italy, pp 412–417
24. Gockley R, Simmons R, Forlizzi J (2006) Modeling affect in socially interactive robots. In: Proceedings of the international symposium on robot and human interactive communication (RO-MAN), Pittsburgh, USA, pp 558–563
25. Becker-Asano C, Kopp S, Pfeiffer-Leßmann N, Wachsmuth I (2008) Virtual humans growing up: from primary toward secondary emotions. Künstl Intell 22(1):23–27
26. Hartmann B, Mancini M, Pelachaud C (2003) Implementing expressive gesture synthesis for embodied conversational agents. In: Proceedings of the international gesture workshop (GW), Genova, Italy, pp 189–199
27. Bruce A, Nourbakhsh I, Simmons R (2002) The role of expressiveness and attention in human-robot interaction. In: Proceedings of the IEEE international conference on robotics and automation (ICRA), Washington, DC, USA, pp 4138–4142
28. de Melo C, Gratch J (2009) Expression of emotions using wrinkles, blushing, sweating and tears. In: Proceedings of the international conference on intelligent virtual agents, Amsterdam, Netherlands, pp 188–200
29. Bartneck C, Forlizzi J (2004) A design-centred framework for social human- robot interaction. In: Proceedings of the IEEE international symposium on robot and human interactive communication (Ro-Man), pp 591–594
30. Brooks R, Breazeal C, Marjanocic M, Scassellati B et al (1998) The Cog project: building a humanoid robot. In: Nehaniv C (ed) Computation for metaphors, analogy and agents. Springer lecture notes in artificial intelligence, vol 1562. Springer, Berlin, pp 52–87
31. Breazeal C (2000) Sociable machines: expressive social exchange between humans and robots. PhD thesis, Massachusetts Institute of Technology
32. Breazeal C, Brooks R (2005) Robot emotion: A functional perspective. In: Fellous J, Arbib M (eds) Who needs emotions? The brain meets the robot. Oxford University Press, London, pp 271–310
33. Itoh K, Miwa H, Matsumoto M, Zecca M et al (2005) Behavior model of humanoid robots based on operant conditioning. In: Proceedings of 2005 5th IEEE-RAS international conference on humanoid robots (Humanoids 2005), Tsukuba, Japan, pp 220–225
34. Moshkina L, Arkin R, Lee J, Jung H (2009) Time-varying affective response for humanoid robots. Tech rep, Georgia Tech Mobile Robotic Lab
35. Moshkina L (2006) An integrative framework for time-varying affective agent behavior. In: Proceedings of the IASTED conference on intelligent systems and control (ISC), Honolulu, USA, pp 536–802
36. Fujita M, Hasegawa R, Costa G, Takagi T et al (2001) An autonomous robot that eats information via interaction with humans and environment. In: Proceedings of the 10th IEEE international workshop on robot-human interactive communication (RO-MAN), Bordeaux and Paris, France, pp 383–389
37. Fujita M, Hasegawa R, Costa G, Takagi T et al (2001) Physically and emotionally grounded symbol acquisition for autonomous robots. In: Proceedings of the AAAI fall symposium: emotional and intelligent II, North Falmouth, USA, pp 35–36
38. Salichs MA, Malfaz M (2006) Using emotions on autonomous agents. The role of happiness, sadness and fear. In: Integrative approaches to machine consciousness, part of AISB'06: adaption in artificial and biological systems, Bristol, England, pp 157–164
39. Salichs M, Barber R, Khamis A, Malfaz M et al (2006) Maggie: a robotic platform for human-robot social interaction. In: Proceedings of the IEEE international conference on robotics, automation and mechatronics (RAM), Bangkok, Thailand, pp 1–7
40. Proetzsch M (2010) Development process for complex behavior-based robot control systems. RRLab Dissertations. Verlag Dr. Hut, Munich, ISBN 978-3-86853-626-3
41. Baddeley A, Eyenck M, Anderson M (2009) Memory. Psychology Press, New York
42. Ekman P, Friesen W, Hager J (2002) Facial action coding system. A human face
43. Schmitz N, Hirth J, Berns K (2010) A simulation framework for human-robot interaction. In: Proceedings of the international conferences on advances in computer-human interactions (ACHI), St Maarten, Netherlands Antilles, pp 79–84
44. Berns K, Hirth J (2006) Control of facial expressions of the humanoid robot head ROMAN. In: Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS), Beijing, China, pp 3119–3124

Jochen Hirth studied computer sciences at the University of Kaiserslautern. During this time, he specialized in robotics. In 2006 he received his diploma degree in computer sciences. Since this time, he is a Ph.D. student at the Robotics Research Lab of the Department of Computer Sciences at the University of Kaiserslautern, funded by the Carl-Zeiss-Stiftung. His research interests include humanoid robotics, social human-robot interaction, as well as emotion-based and behavior-based control architectures.

Norbert Schmitz studied computer sciences at the University of Kaiserslautern with the focus on Embedded Systems. In 2005 he received his diploma degree in computer sciences in the research area of outdoor robot localization. Since 2006 he is a Ph.D. student in the Robots Research Lab of Prof. Berns focusing on the perception system of humanoid robots for natural human-robot interaction.

Karsten Berns studied computer sciences at the University of Kaiserslautern till 1988. In 1994 he received his Ph.D. from the University of Karlsruhe. From 1989 till April 2003 he was employed at the Research Center of Information Technologies (FZI) at the University of Karlsruhe where he led the group "Interactive Diagnosis and Service Systems (IDS)". Since April 2003 he is a full professor for robotic systems at the University of Kaiserslautern. Present research activities are in the area of autonomous mobile robots and humanoid robots with a strong focus on control system architecture and behavior-based control.