Automatic Interpretation of Affective Facial Expressions in the Context of Interpersonal Interaction

Emilia I. Barakova, *Member, IEEE*, Roman Gorbunov, and Matthias Rauterberg

***Abstract*—This paper proposes a method for interpretation of the emotions detected in facial expressions in the context of the events that cause them. The method was developed to analyze the video recordings of facial expressions depicted during a collaborative game played as a part of the Mars-500 experiment. In this experiment, six astronauts were isolated for 520 days in a space station to simulate a flight to Mars. Seven time-dependent components of facial expressions were extracted from the video recordings of the experiment. To interpret these dynamic components, we proposed a mathematical model of emotional events. Genetic programming was used to find the locations, types, and intensities of the emotional events as well as the way the recorded facial expressions represented reactions to them. By classification of different statistical properties of the data, we found that there are significant relations between the facial expressions of different crew members and a memory effect between the collective emotional states of the crew members. The model of emotional events was validated on previously unseen video recordings of the astronauts. We demonstrated that both genetic search and optimization of the parameters improve the accuracy of the proposed model. This method is a step toward automating the analysis of affective expressions in terms of the cognitive appraisal theory of emotion, which relies on the dependence of the expressed emotion on the causing event.**

***Index Terms*—Appraisal theory of emotion, interpretation of affective expressions, Mars-500, mathematical modeling of affective interactions, second-person perspective.**

I. Introduction

THE methods for automatic measurement and interpretation of spontaneous (as opposed to posed) human emotions are becoming an integral part of intelligent products and services that can lead to a breakthrough in domains such as healthcare, marketing, security, education, and entertainment [14], [27]. This paper features a very specific application of using facial affect recognition in combination with game technology for diagnostic purposes in space missions.

There is related research on long-term missions that are performed by small crews in isolation. Examples of these types of missions are those performed on the international space station, polar research stations, submarines, oil platforms, and meteorological stations. The success of such missions strongly depends on the emotional states of and the interpersonal relationships between the crew members. Past experiences in space have shown that the emotional state and quality of cooperation of a crew can have a great effect on the success or failure of a mission. In this context, it is very important to develop methods for monitoring and improvement of the psychosocial atmosphere in isolated, goal-oriented teams, i.e., their emotional state (estimated from facial expressions), interpersonal relations (inferred from measuring fairness and cooperation in game behavior), and dependences between these.

One important problem that needs to be resolved is monitoring of the affective state within the context of the causing events. Reviews on affect recognition by Zeng *et al.* [4] and Calvo and D'Mello [6] concluded that affective expressions can never be divorced from context, but almost none of the available systems for automatic affect processing integrate contextual cues with facial feature tracking. Some aspects of this problem were addressed by Conati (2002), who proposed a probabilistic framework for affective user modeling that integrated, in a dynamic decision network (Dean and Kanazawa, 1989), information on both the possible causes of the user's affective reaction and its observable effects. This work was further used to build a model of user affect during interaction with an educational computer game. We needed to monitor the emotions in evolving interpersonal relationships in groups of humans. Schilbach and colleagues [19, p. 449] argue that *"Social cognition is fundamentally different when an individual is actively and directly interacting with others. In such cases, an individual adopts a 'second-person perspective' in which interaction with the other can be thought of as essential or even constitutive for social cognition, rather than merely observing others and relying on a 'first- (or third-) person grasp' of their mental states."* This finding calls for a new experimental design and methods that are able to analyze the human social and emotional behavior in interaction.

We used a collaborative computer game to monitor the evolving interaction between groups of three astronauts and analyzed a temporal history of facial expressions of emotions during the collaborative game playing. We aimed to find a method for interpreting the observed temporal history of facial expressions in the context of the events that caused the expressed emotions. For this purpose, we developed a model of emotional events that assumes that the time dynamics of facial expressions are determined by events of different types and intensities and by

Manuscript received February 25, 2014; revised February 10, 2015; accepted March 20, 2015. Date of publication May 12, 2015; date of current version July 11, 2015. This work was supported by NWO-User Support Program Space Research, which funded AMHA project ALW-GO-MG/07-13. This paper was recommended by Associate Editor H. Liu.

The authors are with the Department of Industrial Design, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands (e-mail: e.i.barakova@tue.nl; R.Gorbunov@gmail.com; g.w.m.rauterberg@tue.nl).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/THMS.2015.2419259

facial expressions at the moments when these events take place. Within these data, we searched for dependences in the emotional reactions caused by socially significant events, i.e., caused by the game behavior of another person.

To find the way in which the facial expressions are determined by the underlying events, we used the genetic programming (GP) approach [21]–[23]. GP was chosen among other optimization methods since it searches for functions that fulfill certain criteria (and does not search for a set of parameters). Therefore, we did not need to predefine the structure of the functional relations between the facial expressions and emotional events.

The findings of this paper are as follows.

1) We showed that there are statistically significant relations of different types (positive and negative) between the facial expressions of different crew members.
2) A mathematical model that can explain the observed facial expressions in terms of the events of different intensities and types was developed.
3) It is demonstrated that both genetic search and optimization of the parameters improve the accuracy of the model.

This paper is organized as follows. Section II gives the background information of the experiment and the methods used. Section III presents the model of emotional events. In Section IV, the results of a validation experiment show the predictive power of the model. Section V provides a discussion.

II. Background and Data Collection

A. Frameworks for Measuring of Affect

Studies on affect recognition (or interpretation) systems are based on different theoretical frameworks. Most research on facial expression recognition searched for methods for accurate interpretation of the facial expressions in terms of basic emotions. This approach describes expression of affect in terms of discrete basic emotions or categories of emotion. One popular classification framework stems from the cross-cultural studies of Ekman [1], who defined six basic emotional categories that are found in most cultures. This categorical scheme provides an intuitive description in everyday-life terms and supports user studies in which explicit subjective evaluations of emotional expressions are needed. Questionnaire-based user studies is one example of applying this scheme. This scheme is suitable for identifying emotional states of items from discrete datasets, as well as from video data streams [2], [3], [18]. Automatic emotion recognition tools based on the assumption that the basic emotions correspond to facial models [26] have emerged to systematically categorize the physical expressions of emotions [29].

A major theoretical drawback to the categorical scheme of affect is that there is no agreement on the definition of a basic emotional state, which, according to Barrett, results in several different classifications of the so called basic emotions [7]. Moreover, Zeng *et al.* [4] concluded that the categorical approach fails to describe the range of emotions that occurs in natural communication settings. To alleviate this problem, the existing software solutions (such as the software proposed by Vicar Vision [25]) recognize the degree to which an emotion is present in the snapshot of a facial expression. In addition, the software can depict a combination of emotions in a single facial expression. A limitation of this framework is that by itself, it cannot relate the emotional expressions to the causing events.

The dimensional theories of affect provide a possibility for a continuous representation of the affective space. Another fundamental difference between the representations provided by the two frameworks is that the dimensional framework does not measure the perceived emotion from the observer's perspective, while the categorical representation does. The dimensional theories of affect propose a rather limited number of dimensions such as evaluation, activation, control, and power, and concern elicitation/control of the emotion but not its expression in the face or movement. Each dimension in the different theories [12], [13] ranges from positive to negative, and different control variables define affective states as points in the N-dimensional elicitation space. The dimensional frameworks are not intuitive to use for explicit evaluation of affect (like questionnaires), but are very suitable for methods that use implicit and objective measures, such as bodily signals (heart rate variability, sweating of the hands, tension of posture), or measurable qualities of movement [16], [17]. The shortcomings of this scheme from the point of view of automatic recognition of emotions are the information loss due to the limited choice of several (typically two) dimensions, necessary to represent such a complex notion as emotion. This could be the reason some well-known emotions from daily life are indistinguishable (e.g., fear and anger), or difficult to represent (e.g., surprise). The dimensional theories as well do not provide a framework for connecting the expressed emotions to the causing events.

The appraisal theory of emotion provides a suitable framework for the development of affective recognition (or interpretation) systems in which the affective expression must be related to the causing event. In addition, it can connect to the categorical framework of recognition of affect and respectively utilize the available systems for automatic recognition of facial affect that are developed according to the categorical scheme. The meaning of a stimulus, determined by a particular human in a particular context and at a particular moment of time, leads to elicitation and differentiation of emotions [7]. According to the formulation of Scherer [11], appraisal theory accounts for the contextual dependence of an emotional response from the environmental changes through stimulus evaluation checks, including the novelty, intrinsic pleasantness, goal-based significance, coping potential, and compatibility with standards. It explains the evaluation and the resulting emotional expression as continuous and changing in time process, wherein the variability is caused by changes in environment (i.e., changing environmental stimuli or events) and reappraisals of the changed situation [11]. Directions for creating a computational framework that could also serve the purposes of automatic emotion recognition have been proposed by Sander *et al.* [5]. However, these directions are too coarse and difficult to realize because of the lack of computationally efficient nonlinear dynamics algorithms.

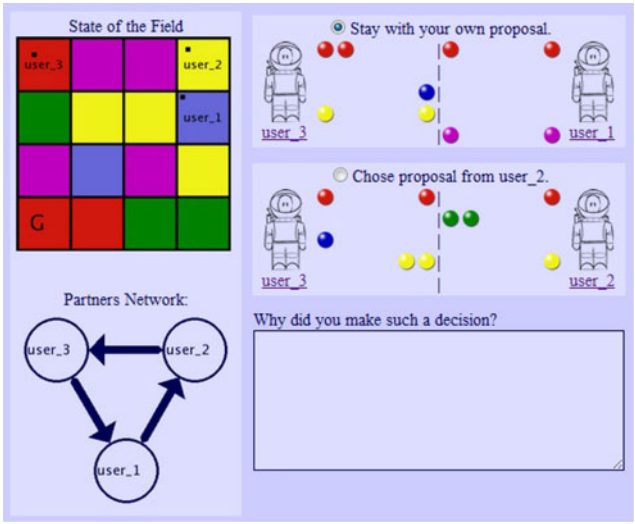

Fig. 1. Snapshot of the CT game. Three players form a partner network as shown in the bottom-left plot. If one has the chip with the same color, then he can move to the next colored field of the board as demonstrated in the top-left plot, aiming to get as close as possible to the goal square. If an exchange proposition is made, then each player can exchange their chips as shown in the second plot on the right or choose to stay with the set of chips he already has.

B. Experiment: Background Information

In 2004, the Institute for Biomedical Problems (IBMP) in Moscow and the European Space Agency started to plan a full-scale ground-based simulation of a manned mission to Mars. Such a full-scale mission requires from 520–700 days of isolation. Referring to the lower end of this time frame, the initiative was named Mars-500. The goal of the Mars-500 study was to gather the data, knowledge, and expertise required to prepare for a mission to Mars. Hence, all key features expected during future missions to Mars were reflected: ultralong duration, need for autonomy, affected communication due to signal delay, and limited stocks of expendables. This ensured that the psychological and physiological impacts of isolation through such an extended period were simulated with high fidelity. A crew of six candidates (four Russians and two from EU countries) were sealed inside the facilities of the IBMP in Moscow.

We created a modified version of the colored trails (CT) game [8], [9], [15], [24] developed at Harvard University. The CT is a multiplayer negotiation game that can be classified as a computer version of a board game involving strategic and logical reasoning. It resembles situations in which people have different goals and need some resources to reach these goals. In the initial state of the game, players can be in different situations, and as a consequence of that, they require different kinds of resources to reach their goals. A redistribution of the resources can be done through negotiations between players. The game supports analysis of developing social relations since it contains both competitive and collaborative components.

The game is played on a rectangular board composed of colored squares (see Fig. 1).

The game is played by three persons. In the beginning of the game, every player is placed on one of the colored squares of the board and one square is assigned to be the goal-square. Every player also receives a set of colored chips (resources), which are the same as the board squares. Players can move their chips horizontally or vertically to one of the neighboring squares if they have a chip of the same color as the new square. The player then surrenders a chip of that color. The goal of the player is to move as close as possible to the goal-square, spending a minimum number of chips. Before making their moves, players are allowed to exchange some of their chips with another player if both participants agree. Every player can take the role of a proposer and can choose to whom to make a proposition. This yields additional information about the social atmosphere in the group. Every player can potentially receive a proposition and play the role of a responder. The responders can have from one to three propositions, and in this way, they experience a broader range of situations in comparison with the original version of the game.

C. Data Collection

The data collected during the cooperative game were separated into two approximately equal parts: before and after the simulated landing on Mars. The data from the first part of the experiment were used to develop a model to explain the data from the second part. We collected three kinds of data: behavior in a cooperative computer game, self-assessment questionnaires, and video records of facial expressions during game play. The video records of facial expressions were collected during the Mars-500 isolation experiment from six participants over a period of 520 days. Every second week, the participants were required to interact with each other through a computer environment for approximately 30 min as a part of our experiment. During these sessions, the participants were seated in front of the computers, performing different learning tasks and playing the modified CT game with one another [9], [10]. The frontal video records of facial expressions were made by the cameras located on the personal computers of the participants.

To extract facial expressions from the videos, we used FaceReader [25]. FaceReader can recognize facial expressions by distinguishing six basic emotions (happy, sad, angry, surprised, scared, and disgusted) plus neutral with an accuracy of 89% [18], [25] and is based on the facial action coding system [26]. We used FaceReader to generate automatic classifications in terms of components of the facial expression for every third frame of the video. This provided 120-ms time separation between the two neighboring data points. In total, 33 experiments were conducted. Every experiment provided six videos for different participants. The average duration of the videos was about 30 min. If the astronauts' facial expressions could not be interpreted, the video segments were removed from the analyzed data.

III. Model of the Dependence of the Emotional Expression on the Underlying Events

A. Correlations in the Behavior of the Participants

We attempted to find a relation between the facial expressions of the participants of the Mars-500 experiment and external events. Such a relation could have been caused by two

different factors. First, every participant interacted with his/her instance of the computer program and different instances of the program were synchronized. As a result, some users may have simultaneously observed the same event and, therefore, had an emotional reaction to this event at the same time. Second, the participants could interact with each other through the provided computer program. In particular, within the CT game, participants made and accepted or rejected proposals made by others. Moreover, the participants could communicate with each other using chat. This interaction could be accomplished with some emotions from both sides; therefore, we could expect that these emotions were related in some way. By an analysis of the relations between the facial expressions of different participants, we can potentially determine the kind of the emotional bonding between two given participants. Moreover, we can expect different kinds of relations reflecting different kinds of emotional bonding between the participants. For example, we can determine if the second user is more likely to be happy if the first one is happy, or as an alternative, the users could tend to have opposite facial expressions at the same time (for example, the first user tends to be happy when the second user is sad).

To find a possible relation between the facial expressions of a given pair of users, we considered joint 2-D distributions of the different components of the facial expressions. In particular, to generate a joint 2-D distribution, we paired the values of a given component of the facial expression for a given pair of users if these values corresponded to the same experiment and the same frame index. These pairs of values were then used to calculate a joint 2-D distribution. If there is no relation between the facial expressions of the two given users, the obtained distributions should be independent.

To compare the calculated joint distribution with the joint distribution for conditionally independent variables, we generated 2-D distributions by a random pairing of the values corresponding to a given pair of users (two randomly paired values correspond to different experiments and/or different frame index). We calculated about 2600 joint distributions of randomly paired values. Each of these distributions were compared with the distribution obtained for the values that are paired according to their experiment and frame indices. For every grid on which the distributions have been calculated, we calculated the number of times when the real joint distribution was larger, smaller, or equal, respectively, to the corresponding value of the distributions of the randomly paired values. We denoted these numbers as n_l, n_s, and n_e, respectively, and used them to calculate the magnitude of the difference:

$$S(n_l, n_s, n_e) = \begin{cases} +1 \cdot (n_l + n_s + n_e + 1) / (n_s + 1) \\ \quad \text{if} \quad n_l > n_s \\ -1 \cdot (n_l + n_s + n_e + 1) / (n_l + 1) \\ \quad \text{if} \quad n_l < n_s \\ 0, \\ \quad \text{if} \quad n_l = n_s. \end{cases} \tag{1}$$

Let us assume that for a considered grid, the real joint distribution is larger than the distribution for conditionally independent variables in most of the cases. For example, in only one case, out of 500 the real distribution was smaller than the distribution of independent variables; therefore, the number 500 can be used as a measure of statistical significance.

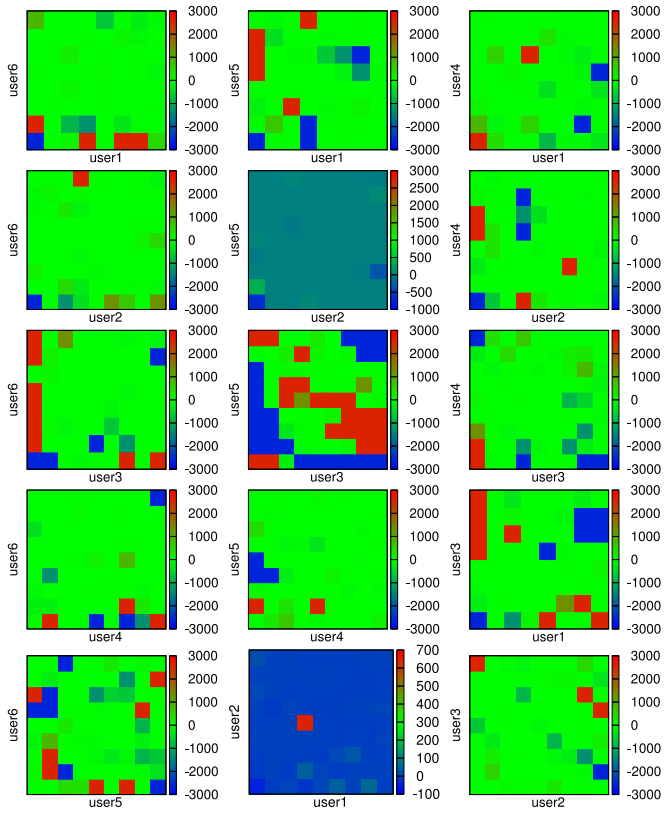

Fig. 2. Relations between the neutral components in the facial expressions of each pair of astronauts. The deviations of the joint distribution of the neutral components of the expressions are used to show these relations. The bar on the right side of each individual plot shows the correspondence between the colors and their values. The color corresponding to a high positive value (red) means that if one user has a neutral expression, then it is more likely that the second user will also have a neutral expression.

The measure calculated for the neutral component of the facial expressions is shown in Fig. 2 for all 15 possible pairs of users.

Please note that the pattern observed in the figure corresponding to users 3 and 5. The bar on the right side gives the correspondence between the colors and their value. In the left bottom corner of the plot, we see the red square, which means that if user 3 has the neutral expression, it is more likely that user 5 also has the neutral expression. This effect is also supported by the blue squares along the two axes. The blue squares (the situations in which one of the users has a neutral facial expression and another one does not) are less frequent than we could expect for unrelated facial expressions.

In the comparison made for the "surprised" component (see Fig. 3), similar pattern (as the pattern for neutral component between users 3 and 5 in Fig. 2) is clearly seen only for one pair of users: user3-user6. In a weak form, this pattern is present for another two pairs: user1-user3 and user4-user6. However, the "surprised" component of the facial expression is interesting because we can see another pattern that has an opposite meaning. This pattern can be seen with the pair user3-user5. We see the blue square in the left bottom of the plot and the red squares close

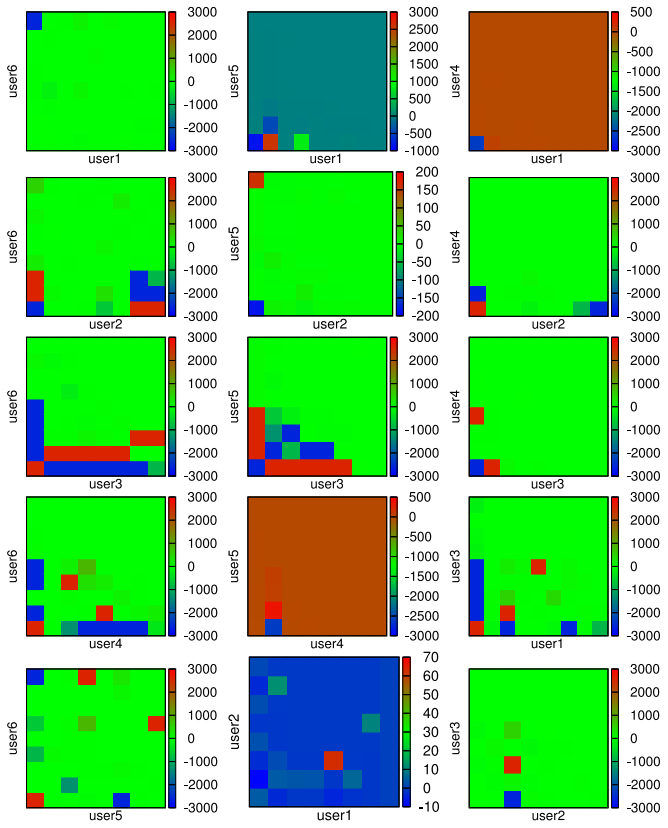

Fig. 3. Relations between the surprised components of the facial expressions of each pair of astronauts. The deviations of the joint distribution of the surprised expression components are used to show these relations.

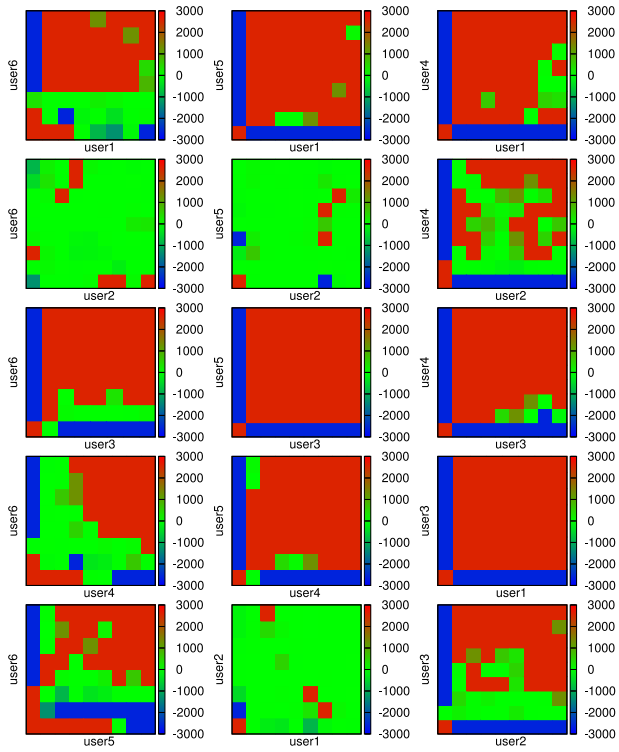

Fig. 4. Relations between the joint distributions of the facial expressions of each pair of astronauts. The joint distribution combines all the components of emotional expressions detected by the FaceReader. The color corresponding to a high positive value (red) is very pronounced, which means that at any given time, the facial expression between two astronauts is most likely the same.

to the two axes. This pattern means that if one user is surprised, it is more likely that another will not be. A possible explanation of this relation between the "surprised" components of the facial expressions could be that in some cases, one user surprises the other. As a consequence, when one user is surprised, another one is not.

We have also considered joint distributions that combine the different components of the facial expressions by pairing the values corresponding to the same experiment, frame index, and type of the component of the facial expression. For the combined consideration of different components of the facial expressions, we performed the same analysis as for the separate consideration described above. The measures of difference between the real distributions and distributions of unrelated variables are shown in Fig. 4. As can be seen in Fig. 4, the first pattern is even more pronounced in this case. This pattern means that the facial expressions of the considered pair of users tend to be similar at the same moment of time. In nine out of 15 cases, the pattern is observed clearly. Only in three cases, the pattern is seen in a weaker form.

Thus, this method can be used to calculate the strength of the dependence between the facial expressions of different users. Moreover, by analyzing the strength and the type of the relation for different components, we can potentially discover different types of the relations between the facial expressions of different users.

B. *Model of Emotional Events*

An emotional event is defined as any event that influences emotions and, therefore, the facial expressions of the participants. We propose a model that makes the assumption that the components of the facial expressions $\vec{f}_{k+i}$ are given by the last emotion-causing event $\vec{s}_k$ and the facial expression at the moment of the event $\vec{f}_k$:

$$\vec{f}_{k+i} = \vec{F}\left(\vec{f}_k, \vec{s}_k, i\right). \tag{2}$$

In other words, we assume that after an emotional event, the facial expression changes from the current state to the state corresponding to the emotional event. We call the $\vec{F}$-function a response function because it determines the response of the facial expressions to the emotional events. In (2), the subscripts are used for the time frames of the video: k gives the position of the last emotional event and i is the number of time steps between the given facial expression and the moment when the last event happened.

In general, the emotional events can be described as a set of parameters denoted as vectors $\vec{s}$. We consider emotional events as 2-D vectors in which the first component indicated the type of an event t (e.g., "sad," "funny"), and the second component indicates its intensity I (how "sad" or "funny" was it?). The type of the event is given by the type of the emotion that is provoked by this event. Hereafter, we will use a terminology in which a

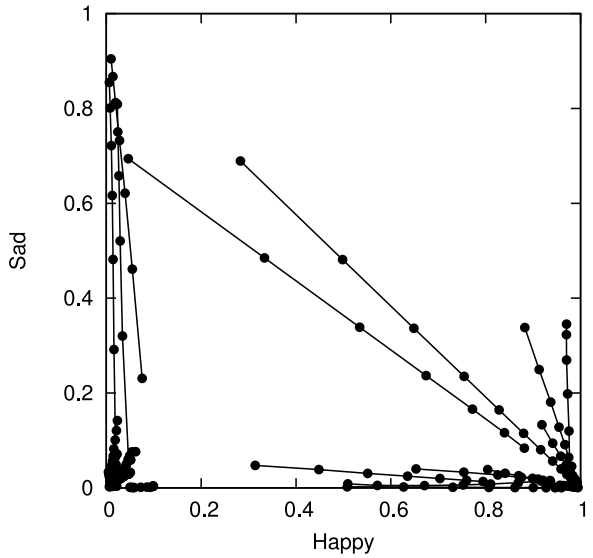

Fig. 5. Examples of the segments forming straight lines in the space of facial expressions. The linear segments are directed to a specific location in the 7-D space of emotional expressions. Two groups of lines in this plot point to the happy and sad facial expressions.

"happy" or "sad" event is understood as an event that provokes feelings of happiness or sadness, respectively.

Thus, the model can be written as

$$f^c_{k+i} = F^c\left(\vec{f}_k, t_k, I_k, i\right) = F^c_t\left(\vec{f}_k, I_k, i\right). \tag{3}$$

Using an index rather than a vector notation, the superscript c indicates different components of the facial expressions. It ranges from one to seven, reflecting the FaceReader's values for facial expressions. Initially, we found segments in which components of the facial expressions are smooth functions of time. By a further analysis of these patterns, we found that changes in different components of facial expressions are approximately linearly proportional to each other.

Fig. 5 and the following mathematical expression indicate these segments as lines in the 7-D space:

$$f^c_{k+i} = f^c_k + \mu(i)\left(s^c_k - f^c_k\right). \tag{4}$$

The function $\mu(i)$ should be equal to zero if $i = 0$ and to one if i is large enough. In this case, the facial expression starts to change from the expression at the moment of the event (f^c_k) and moves along a line to the final expression (s^c_k) corresponding to the given emotional event.

In Fig. 5, we can see that linear parts of the dependences form a pattern: All the lines are directed to specific locations in the space of the facial expressions. For example, in the figure, we can see two groups of lines that point to "happy" and "sad" facial expressions, respectively. Because of this property, the mathematical expression for the dynamics of the components of facial expressions can be rewritten as

$$f^c_{k+i} = f^c_k + \mu(i)\left(\delta_{tc} I_k - f^c_k\right) \tag{5}$$

where δ_{tc} is the Kronecker's delta. t can be considered as the type of the event, and I as its intensity. By fitting the observed

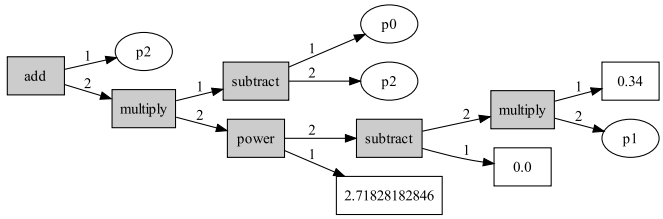

Fig. 6. Tree representation of the first (F_1) of two functions that determine a simplified response facial expression to the causing event. The tree represents the initial exponential guess for the first component (F_1) of the response function. The evolutionary optimisation takes place by introducing mutations of this tree until the response functions are found.

dependences, we found that $\mu(i)$ can be approximated by an exponential function $\mu(i) = \exp(-\alpha \cdot i)$. Equation (5) can be considered as a partial case of (2) and can be rewritten in the following form:

$$\begin{aligned} f^c_{k+i} = {} & \delta_{tc}\{f^c_k\left[1 - \mu(i)\right] + \mu(i) I_k\} \\ & + (1 - \delta_{tc})\{f^c_k\left[1 - \mu(i)\right]\}. \end{aligned} \tag{6}$$

In this study, we used GP to find the shape of the response function. To reduce the search space, we restricted the form of the response function. As the structure of the response function should be able to capture not only the considered linear segments, given by (6), but also more complex dependences, our compromise response function is as follows:

$$f^c_{k+i} = \delta_{tc} F_1\left(f^c_k, I_k, i\right) + (1 - \delta_{tc}) F_2\left(f^c_k, I_k, i\right). \tag{7}$$

Compare (7) with (3); (3) includes 49 functions corresponding to different values of c and t, as any of the components can follow any event. For instance, by expression (3), we specify how the "happy" component of the facial expression changes after a "sad" event, or how the "angry" component changes after a "happy" event, and so on. In contrast, (7) contains only two functions (F_1 and F_2). This restriction assumes that the way in which the ith component of the facial expression is influenced by the ith event is independent of i. The "happy" component of the facial expression is assumed to depend on a "happy" event in the same way as the "sad" component depends on a "sad" event (assuming that the initial values of the components of the facial expressions and the intensities of the events were the same in the two mentioned cases).

C. Genetic Programming Approach for Finding Response Functions

To find the functions F_1 and F_2 that determine the response of the facial expression on emotional events of different types and intensities, we used GP in a tree structure. For example, the exponential function for F_1 corresponding to (6) is shown in Fig. 6.

As nodes of the tree, we used either basic functions or real constants or arguments. The set of the basic functions consisted of: 1) addition, 2) subtraction and 3) multiplication functions of two arguments, 4) "if-function," 5) "greater-than-function," 6) power function, and 7) arctangents. The "if-function" is a

function of three arguments. It compares the first argument with zero, and if it is larger than zero, the function returns the second argument. Otherwise, the third argument is returned. The "greater-than-function" is a function of two arguments. It returns 1 if the first argument is larger than the second one. Otherwise, the second argument is returned.

The functions F_1 and F_2 have three arguments. For the function F_1, they are indicated in Fig. 6 as $p0$, $p2$, and $p3$. The first argument ($p0$) is the value of those components of the facial expression that correspond to the type of the emotional event [f_k^c in the (7)]. k indicates that this value is taken at the moment of the event. For example, in case of a "happy" event, $p0$ should be equal to the value of the "happy" component of the facial expression at the moment when the event was observed by the participant. The second argument ($p1$) is the number of frames between the current moment and the moment of the last emotional event [i in the (7)]. The third argument ($p2$) is the intensity of the event [I_k in (7)].

To find a response function, we used a simple evolutionary process. The evolution started from the earlier defined pair of the exponential functions given by (6) and shown in Fig. 6 for the function F_1. For the given pair of trees, we calculated a score that indicated how well the given pair of functions explains the observed dynamics of the facial expressions (more details about the calculation of the score will be given later). Then, we created a new pair of trees by mutation of every tree in the old pair. To mutate a tree, we randomly choose a node in the tree and replaced it by a random tree. A random tree is generated in the following way. First, we create a root node. We randomly decide if it should be a function or not. The probability for the node to be a function is set to 0.5. If a node is chosen to be a function, then a function is randomly chosen from the earlier given list of the basic functions. If the node is not chosen to be a function, then we decide if it should be a parameter (argument) or a constant. The probability for the node to be a parameter is set to 0.6. If a node is chosen to be a parameter, one of the parameters is randomly chosen ($p0$, or $p1$ or $p2$). If a node is chosen to be a constant, a random number is generated and associated with the node. A random number generator with the uniform distribution between zero and one is used. After a root node is created, we loop over its parameters (arguments) and generate nodes associated with them. The procedure is repeated recursively for every node in the tree whose child nodes are not yet specified. The procedure stops if there are no nodes that require child nodes (constants and parameter nodes). The maximal depth of the tree is set to four to prevent generation of large trees.

D. Training Set and Score Function

We searched for the response function that could model segments around the patterns described earlier and shown in Fig. 5. To find a generalization of the dependence (6), we extended the linear segments by preceding and subsequent steps of the data. The addition of nonlinear segments required the use of a more general function. To find this function, we used GP techniques. Specifically, we selected the parts of the trajectories in the space of the facial expression that lay on a line (if the angle between the line connecting the first and second points and the line connecting the second and third points was not larger than 3°). Three video records were considered. The number of segments with the above-described properties in these records was 26, 52, and 21, respectively. The minimal and maximal length of the segments was 6 and 12 steps, respectively. The average length of the segments was 7.3 steps. To capture patterns happening immediately before and after the considered segments, we added to them 20 preceding and 31 subsequent steps. For every extended segment, we searched for the best emotional event that could explain the dependences observed in the segment. Specifically, we looped over all possible locations, types, and intensities of the event. The loop over intensities of the events was run from 0.0 to 1.0, with the step equal to 0.01. For every considered event, we used the available response function to predict the dynamics of the facial expressions. First, we combined the intensity and type of the event with the facial expression at the moment of the event to estimate the facial expression on the next step. Then, the difference between the estimated and observed facial expression was calculated. In particular, the estimated and real (observed) facial expressions can be represented as points in the 7-D space of the facial expressions. As a measure of the difference between the estimated and observed facial expressions, we used the distance between the two points, representing the two kinds of the facial expressions, divided by the average length of the vectors connecting the origin of the coordinate system and the two points:

$$d = 2\frac{|\vec{o} - \vec{p}|}{|\vec{o} + \vec{p}|} \tag{8}$$

where $\vec{o}$ and $\vec{p}$ are the observed and predicted facial expressions, respectively. The predicted facial expression was accepted if its deviation from the observed expression was smaller than 0.03, according to (8). After the prediction for the given step was accepted, a prediction for the next step was generated and evaluated in the same way. The procedure was repeated until an unaccepted prediction was reached. Then, the total length of the prediction was calculated. This way, we obtained a location, type, and intensity of the event that maximized the length of the prediction for the considered segment. This procedure was performed for all segments with a given response function, and the total length of the predictions was used as a measure of the quality of the considered response function.

E. Optimization Procedure

We started the evolutionary process from the response function given by (6). Then, we generated new response functions and evaluated their scores until a function with a score larger than or equal to those of the initial function was found. The new response function then replaced the initial function and the whole procedure was repeated. The procedure was stopped if the score did not improve for a large enough number of generations. After the evolutionary search was stopped, we ran a hill-climbing optimization algorithm to find new values for the constants involved in the trees, to improve the predictive power

TABLE I
AVERAGE LENGTH OF THE PREDICTIONS FOR DIFFERENT DATASETS AND RESPONSE FUNCTIONS

Optimization method	Video 1	Video 2	Video 3
None (Initial Guess)	10.15	10.98	10.48
Evolutionary	12.50	12.85	12.29
Hill Cclimbing	12.65	13.46	12.38

of the response function. We iterated over all constants in the pair of trees. For every constant, we considered the two neighboring values, separated by 0.1 from their original values. Then, we chose the variable and the direction of the shift over this variable that maximized the predictive power of the response function. If no improvement was possible, we decreased the current step by 1.1.

We ran three independent optimization procedures for three different video records. After that, the response functions optimized on the three independent sets of data were tested on the data that were not used during the optimization.

IV. VALIDATION OF THE MODEL OF DEPENDENCE OF THE EMOTIONAL EXPRESSION ON THE UNDERLYING EVENTS

We ran the evolutionary optimization procedure for three video records. These optimization procedures were stopped after 2165, 156, and 672 steps of the evolution, respectively, based on the observation that the score did not improve for several hundred steps. Table I summarizes the improvement in the predictive power of the response function when compared to the initial assumption/starting point that this dependence will be exponential, which was given in (6). Specifically, the average length of the accepted prediction made with the initial assumption for exponential dependence was equal to 10.15, 10.98, and 10.48 steps for the three video records, respectively. After the evolutionary optimization, the predictive power increased to 12.50, 12.85, and 12.29 steps, respectively. The additional hill-climbing optimization of the response functions found in the evolutionary optimization also led to an increase of the predictive power of the response functions in all three cases; however, the improvement was very small. After the hill-climbing optimization, the predictive power in the three cases increased to 12.65, 13.46, and 12.38 steps.

Examples of the found response functions are shown in Figs. 7 and 8.

To address possible overfitting, the response functions were tested on data not used in the optimization procedures (see Table II).

The first response function that was obtained with the first video record performs well for the second and third video records. The average lengths of the prediction for the second and third video records are even larger than those for the first one. Moreover, the considered response function has a higher predictive power for the third video record than the third response function, which was obtained with this record. We can conclude that the first response function has not been overfitted. The second response function performs best for the second video record and has a low predictive power for the other two records. However, even for these two records, the predictive power of the considered response function is larger than those of the exponential response function used as the initial guess. The third response function, obtained with the third video, also performs well with the second record but not so well with the first one. As a general conclusion, we can say that response functions obtained just with one video record are meaningful and could perform well for other records. However, since a small effect of overfitting is present, for further optimizing the response functions, it is recommended to use a larger set of data.

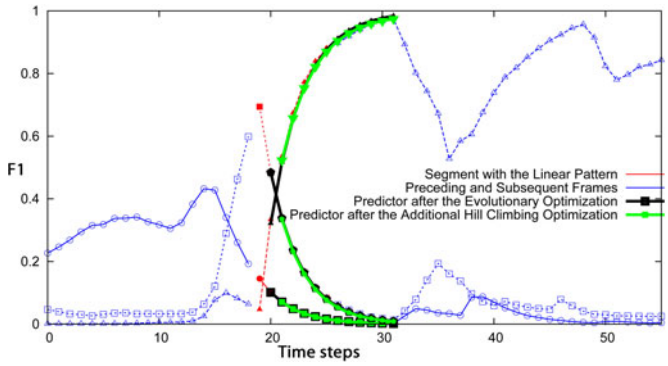

Fig. 7. Example of the first response function (F_1) representing the different components of facial expressions. The horizontal axis shows the number of frames between the current moment and the moment of the last emotional event.

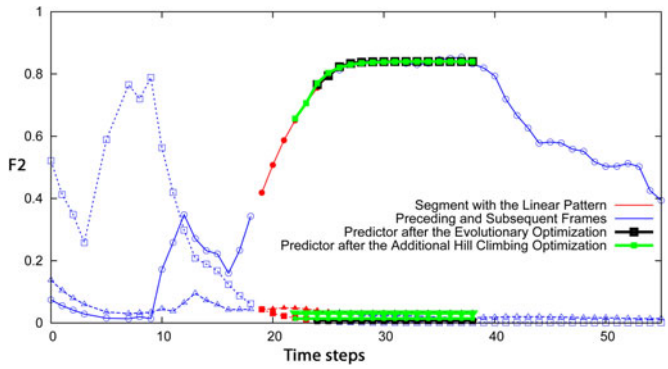

Fig. 8. Example of the second response function (F_2) representing the different components of facial expressions. The horizontal axis shows the number of frames between the current moment and the moment of the last emotional event.

TABLE II
CROSS VALIDATION OF THE RESPONSE FUNCTIONS

Response function (RF)	Video 1	Video 2	Video 3
No RF used	10.15	10.98	10.48
RF1	12.65	12.92	13.24
RF2	10.31	13.46	11.86
RF3	10.35	13.08	12.38

V. Discussion

The main contribution of this paper is the development of a model that, on the basis of the available technology, opens possibilities for interpreting affective facial expressions in the context of the events that caused these affects. The emotional events depicted during collaborative gaming could be of different intensities and types. Our model assumes that the current facial expression can be represented as a mathematical function of the following three arguments: 1) the intensity and type of the event; 2) time separation between the current moment and the moment when the event happened; and 3) the facial expression at the moment of the event. The approach is applicable in the context of the interpersonal relations and long-term effects of isolation.

We have also developed a computational procedure that can help identify the locations of the emotional events as well as their types and intensities based on the observed sequence of the facial expressions. This procedure is also used to determine the way in which the dynamics of the facial expressions are influenced by the emotional events. This is especially useful when multimodal information is available, and one could compare the moments in time when something (an event) occurs in a complementary data stream. This method, therefore, opens possibilities for applications to interpersonal communication.

Since we analyzed video records of facial expressions taken during the Mars-500 isolation experiment, we have studied how facial expressions depend on the interactions that happen between each two players during a collaboration game. The interactions of the players are mediated by a computer but reflect the events that cause emotion during this game play, and these events could be either a result of the game that all the players observe, or reaction of a player as a game act, or as a reaction to the posted chat item. Within this respect, the model was developed and used to analyze the real interaction and emotional engagements between people, rather than mere observation. Therefore, this model supports going beyond spectator accounts of social knowing, which have been central to the Western intellectual tradition [28].

To study the long-term effects of isolation on the emotional states of the crew members, we analyzed how different components of facial expressions change with time. We worked under the restriction of limited data from the very specific experimental setting. The video data were separated into two approximately equal parts: before and after the simulation of the landing on Mars. This way, we artificially created training and test datasets. The difference between the average components of the facial expression, corresponding to the two parts of the experiment was calculated.

To study the short-term effects, we found the correlations between the facial expressions from the neighboring experiments (i.e., experiments separated by two weeks). To derive information about the interpersonal relations in the crew, we analyzed the correlations between facial expressions of different crew members. To study the relation between the facial expressions of different crew members in more detail, we considered the joint distributions of the components of the facial expressions corresponding to two different crew members. Deviations of these distributions from the distributions of unrelated variables were calculated and found to be statistically very significant. The relations between the emotional states of different crew members can be used as a measure of how emotionally bound two persons are.

The method is an advancement in analyzing the affective expressions in the framework of appraisal theory, which is best supported by a multidisciplinary community of researchers. In addition, the events are depicted during the interpersonal interaction; therefore, the experimental design, as well as the model, supports analyzing social and emotional behavior in the context of interpersonal interaction. The developed method for analysis of social behavior during interaction can contribute to research in the multidisciplinary study of behavior from the so-called second-person perspective, i.e., the perspective of a person in interaction with another person, which differs from the well-established first-person perspective in science, where the subject reports on his experiences, and from the third-person perspective, when there is an observer who is not participating in the experiment.

The importance of studying the second-person perspective was emphasized by Schilbach [19], [20]. The experimental design that we proposed gives a good example of how a second person perspective can be empirically studied. In addition, the model for analysis was developed and tested to study the social relations of participants who are involved in a setting in which they need to collaborate.

Acknowledgment

The authors would like to thank the workgroup of the MECA project, particularly Mark Neerincx, for their support and collaboration. They are also grateful to their partners from the Moscow Institute of Biomedical Problems, particularly Vadim Guschin, for support with preparation and conducting of experiment within the MARS-500 study.

References

[1] P. Ekman, "An argument for basic emotions," *Cognition Emotion*, vol. 6, pp. 169–200, 1992.

[2] A. Kapoor, W. Burleson, and R. W. Picard, "Automatic prediction of frustration," *Int. J. Human-Comput. Stud.*, vol. 65, pp. 724–736, 2007.

[3] J. E. Resnicow, P. Salovey, and B. H. Repp, "Is recognition of emotion in music performance an aspect of emotional intelligence?" *Music Perception*, vol. 22, pp. 145–158, 2004.

[4] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009.

[5] D. Sander, D. Grandjean, and K. R. Scherer, "A systems approach to appraisal mechanisms in emotion," *Neural Netw.*, vol. 18, pp. 317–352, 2005.

[6] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affective Comput.*, vol. 1, no. 1, pp. 18–37, Jan. 2010.

[7] L. F. Barrett, "Are emotions natural kinds?" *Perspectives Psychol. Sci.*, vol. 1, pp. 28–58, 2006.

[8] Y. a. Gal, B. J. Grosz, S. Kraus, A. Pfeffer, and S. Shieber, "Colored trails: A formalism for investigating decision-making in strategic environments," in *Proc. IJCAI Workshop Reason., Representation, Learning Comput. Games*, pp. 25–30, 2005.

[9] R. Gorbunov, E. I. Barakova, R. M. Ahn, and M. Rauterberg, "Monitoring interpersonal relationships through games with social dilemma," in *Proc. Int. Conf. Evol. Comput. Theory Appl.*, 2011, pp. 5–12.

[10] N. Voynarovskaya, R. Gorbunov, E. Barakova, R. Ahn, and G. W. M. Rauterberg, "Nonverbal behavior observation: Collaborative gaming method for prediction of conflicts during long-term missions," in *Proc. 9th Int. Conf. Entertainment Comput.*, 2010, pp. 103–114.
[11] K. R. Scherer, "Appraisal theory," in *Handbook of Cognition and Emotion*, T. Dalgleish and M. J. Power, Eds. New York, NY, USA: Wiley, 1999, pp. 637–663.
[12] D. Watson, L. A. Clark, and A. Tellegen, "Development and validation of brief measures of positive and negative affect: The PANAS scales," *J. Personality Social Psychol.*, vol. 54, pp. 1063–1070, 1988.
[13] J. A. Russell, "A circumplex model of affect," *J. Personality Social Psychol.*, vol. 39, pp. 1161–1178, 1980.
[14] R. P. Hill and M. B. Mazis, "Measuring emotional responses to advertising," *Adv. Consum. Res.*, vol. 2, pp. 164–169, 1986.
[15] R. Gorbunov, "Monitoring emotions and cooperative behavior," Ph.D. dissertation, Dept. Ind. Design, Eindhoven Univ. Technol., Eindhoven, The Netherlands, 2013.
[16] E. I. Barakova and T. Lourens, "Expressing and interpreting emotional movements in social games with robots," *Personal Ubiquitous Comput.*, vol. 14, pp. 457–467, 2010.
[17] T. Lourens, R. van Berkel, and E. Barakova, "Communicating emotions and mental states to robots in a real time parallel framework using laban movement analysis," *Robot. Auton. Syst.*, vol. 58, pp. 1256–1265, 2010.
[18] V. Terzis, C. Moridis, and A. Economides, "Measuring instant emotions based on facial expressions during computer-based assessment," *Pers. Ubiquitous Comput.*, vol. 17, pp. 43–52, 2013.
[19] L. Schilbach, "A second-person approach to other minds," *Nature Rev. Neurosci.*, vol. 11, pp. 449–449, 2010.
[20] L. Schilbach, B. Timmermans, V. Reddy, A. Costall, G. Bente, T. Schlicht, and K. Vogeley, "A second-person neuroscience in interaction," *Behavioral Brain Sci.*, vol. 36, pp. 441–462, 2013.
[21] W. Banzhaf, P. Nordin, R. E. Kellner, and F. D. Francone, *Genetic Programming—An Introduction: On the Automatic Evolution of Computer Programs and Its Applications*. San Mateo, CA, USA: Morgan Kaufmann, 1998.
[22] J. R. Koza, "Human-competitive results produced by genetic programming," *Genetic Program. Evolvable Mach.*, vol. 11, pp. 251–284, 2010.
[23] T. Segaran, *Programming Collective Intelligence: Building Smart Web 2.0 Applications*. Sebastopol, CA, USA: O'Reilly Media, 2008.
[24] B. J. Grosz, S. Kraus, S. Talman, B. Stossel, and M. Havlin, "The influence of social dependencies on decision-making. initial investigations with a new game," in *Proc. 3rd Int. Joint Conf. Auton. Agents Multiagent Syst.*, vol. 2, 2004, pp. 782–789.
[25] M. J. D. Uyl and H. van Kuilenburg, "The facereader: Online facial expression recognition," in *Proc. 5th Int. Conf. Methods Techn. Behavioral Res.*, 2005, pp. 589–590.
[26] P. Ekman and W. V. Friesen, *Manual for the Facial Action Coding System*. Palo Alto, CA, USA: Consulting Psychol. Press, 1977.
[27] E. I. Barakova, A. S. Spink, B. de Ruyter, and L. P. Noldus, "Trends in measuring human behavior and interaction," *Personal Ubiquitous Comput.*, vol. 17, pp. 1–2, 2013.
[28] J. Dewey, *Reconstruction in Philosophy*. New York, NY, USA: New Amer. Library, 1950.
[29] A. A. Salah, N. Sebe, and T. Gevers, "Communication and automatic interpretation of affect from facial expressions," in *Affective Computing and Interaction*. Hershey, PA, USA: IGI Global, 2011, pp. 157–183.

Emilia I. Barakova (M'94) received the Master's degree in electronics and automation from the Technical University of Sofia, Sofia, Bulgaria, and the Ph.D. degree in mathematics and physics from Groningen University, Groningen, The Netherlands, in 1999.

She is currently with the Department of Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands. She has expertise in modeling social behavior, social robotics, functional brain modeling for applications in robotics, learning methods, and human-centered interaction design. Her recent research is on modeling social and emotional behavior for applications to social robotics and robots for social training of autistic children.

Roman Gorbunov received the M.S. degree (*cum laude*) in theoretical physics from the Dnipropetrovsk National University, Dnipropetrovsk, Ukraine, in 2000, and the Ph.D. degree in quantum chemistry from Goethe University, Frankfurt, Germany, in 2007. He also received the Ph.D. degree from the Department of Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands, in 2013.

From 2007 to 2009, he was a Postdoctoral Research Associate with Biophysics and Physiology Department, Albert Einstein College of Medicine, Yeshiva University, New York, NY, USA.

Matthias Rauterberg received the B.S. degree in psychology from the University of Marburg, Marburg, Germany, in 1978; the B.A. degree in philosophy in 1981, the B.S. degree in computer science in 1983, the M.S. degree in psychology in 1981, and the M.S. degree in computer science in 1986 from the University of Hamburg, Hamburg, Germany; and the Ph.D. degree in computer science/mathematics from the University of Zurich, Zurich, Switzerland, in 1995.

He is currently a Professor and Head of the Designed Intelligence Research Group, Department of Industrial Design, Eindhoven University of Technology, Eindhoven, The Netherlands. His recent research interests include entertainment computing, cognitive systems, human–computer interaction, and design science.