Long-Term Feedback Mechanisms for Robotic Assisted Indoor Cycling Training

Sebastian Schneider
Applied Informatics
Bielefeld University, Germany
sebschne@techfak.uni-bielefeld.de

Luise Süssenbach
Applied Informatics
Bielefeld University, Germany
lsuessen@techfak.uni-bielefeld.de

Ingmar Berger
Applied Informatics
Bielefeld University, Germany
iberger@techfak.uni-bielefeld.de

Franz Kummert
Applied Informatics
Bielefeld University, Germany
franz@techfak.uni-bielefeld.de

ABSTRACT
We present a concept for long-term feedback during robot assisted indoor cycling training. Our feedback model captures different aspects from sport motivation theory. Furthermore, we present our designed measurements to evaluate the robot's persuasiveness and user's compliance. We conducted an intensive 18-day isolation study in two campaigns (e.g. socially assistive robot vs. display instructed, n=16) in cooperation with the German Aerospace Center. The results show that users tend to comply to the robot's instructions and that there is a significant difference in compliance between the two conditions.

INTRODUCTION
When questioning which aspects from Human-Human Interaction (HHI) we can use for Human-Robot Interaction (HRI), we have to keep the focus on the domain and the role the robot plays. Clearly, in each domain we will find certain human-like skills a robot needs to have. And yet we are far beyond in building a unique role model description with formalized task descriptions, interfaces and responsibilities for every possible social role a robot could encounter in real world scenarios. Hence, we have to adjust the social skills for each scenario manually. For example, a robot for animal assisted therapy[1] needs to have different capabilities than a robot that acts as a teacher or trainer. For toys it is fully sufficient to be reactive. They do not need to engage pro-actively in long lasting interactions and do not need to remember what they have played yesterday and whether the user was kind. A trainer has a far different role than a toy. It needs to engage the user in long-term interaction in order to achieve a training success and sustain motivation to continue. Hence if we want to deploy a robot as an interactive trainer, we need to equip it with the necessary social and evaluative skills.

In this paper we will focus on robot assisted sport activities with an emphasis on long-term feedback mechanisms for indoor cycling. We report a long-term HRI study which lasted for 18 days with 50 minutes of interaction each day and 16 participants. The participants were assigned to one of two groups: socially assistive robot vs. display instructions. The goal of our work is to identify which mechanisms and informations are important for user's to comply to a robotic assisted training and how to implement concepts of intrinsic motivation to boost the user's motivation for adhering to long-term interventions. Furthermore, we want to develop metrics that help to evaluate such long-term assistive systems. The paper is structured as follows: The next section gives an introduction and motivation for the scenario. Following, we give an overview of related work and the theoretical background for sport assistance. Afterwards, we describe the concept of long-term feedback mechanisms for user engagement, which is followed by a description of our designed evaluation measurements. The last sections give an outlook and conclusion.

MOTIVATION AND SCENARIO
Engaging in sport activities is becoming an increasingly demanding task in our future society. Governmental departments start many campaigns to motivate children and adults to participate in daily sport activities or clubs[2]. Yet, many people have a lack of time or enthusiasm to exercise regularly. This opens the focus for new approaches to engage and motivate people to work out every day with more enthusiasm and goal orientation.

Research shows that motivation, performance and goal orientation can be boosted through the company of a trainer [7]. In our work we focus on people working in isolated environments (e.g. space, submarine or Arctic stations). These environments have further demanding challenges for the habitants. They will be faced with psychological and physiological problems therefore daily work out is even more important.

[1] http://paro.jp/english/index.html, visited on 5/22/2014

[2] http://www.in-form.de/, visited on 5/22/2015

HAI 2015, October 21 - 24, 2015, Daegu, Kyungpook, Republic of Korea.

978-1-4503-3527-0/15/10...$15.00.
DOI: http://dx.doi.org/10.1145/2814940.2814962

Specially on space missions work out is not only for relaxation, but it is needed to sustain vital functions and decrease bone degradation. The needed assistance is usually given by ground control. But during far reaching missions real time communication and thus motivating feedback is not possible anymore and communication protocols and schedules do not allow such assistance. Also sending personalized trainers is not an option, because they will face the same psychological and physiological problems. Hence, social robotics are a promising technology for such application areas and have already been widely used in different socially assistive contexts. The range goes from tutoring, dieting to rehabilitation scenarios [11, 9, 5] . The requirements concerning the interactive capabilities of such systems are highly demanding. Regarding our indoor cycling scenario the robot has an explicit role with certain attributes, responsibilities and expectations attached. It needs to implement standards of motivation theory and sport instructions. The later has been studied so far from an interactional point of view [16]. In this work we will just focus on the global feedback mechanism for multiple and single training sessions in this domain (i.e. indoor cycling). In the course of this work global feedback mechanisms are defined as the social assistance the robot gives in terms of evaluative information about training performance across different training sessions.

RELATED WORK AND THEORETICAL BACKGROUND

Motivation and Feedback Theory

During therapy or training it is important to use the concept of motivation for establishing commitment to the task and to induce behavioral change. Motivation is usually divided into extrinsic and intrinsic motivation [13]. To be intrinsically motivated means to do something because it is inherently interesting or enjoyable, and to be extrinsically motivated refers to doing something because it leads to a separable outcome. The sub-theory *cognitive evaluation theory* of *self-determination theory* explains the variablity of intrinsic motivation influenced by social and environmental factors [3]. The theory claims that events like social feedback or reward can enhance intrinsic motivation [2]. Therefore, intrinsic motivation can also be effected by an external entity (e.g. through an assistive robot). This can be established by *feedback* to the user regarding the performance, which can result in an increase of intrinsic motivation and change of behavior. Feedback in motivation theory shows that it is important for the user to have transparent information about her/his performance and progress. It can be motivational and informational, as well as elicit learning effects and induce behavior modification. Hence, feedback can be divided into **positive reinforcing feedback**, which is motivating and spurs to higher performance [8]. It also gives the person appreciation which boosts self esteem and encourages to continue a certain behavior. The other side is **negative critical feedback** which is less motivating but contains more informational content for the user to enhance error correction and decrease the target-actual gap [12]. Nevertheless, both feedback types can have informational content which can either be evaluative (e.g. good, bad, right, wrong) or quantitative (e.g. too fast, too slow, too less). So if the feedback also includes information that can modify or improve a behavior, feedback can have a learning function. Additionally, the feedback frequency can give a subtle message about training progression. It can be preplanned, randomly or systematically reduced. Another important aspect for enhancing motivation and increasing enjoyment in non-competitive task is to create a competition [17]. This can be done by competing against a ghost player, which could be recorded training data from previous sessions. In our work, we want to focus on the quantitative, qualitative and comparative aspects of reinforcing and critical global feedback to the user in course of an 18-day indoor cycling training. Moreover, we want to analyse how these motivational feedback mechanism improve the user's compliance to interact with the system and investigate the persuasiveness of the feedback.

Socially Assistive Robots

The effects of socially assistive robots giving feedback have already been studied in many applications [11, 9, 14]. They show that feedback has a distinct effect on user's performance and is able to change human behavior. Feedback to the user can be mainly distinguished between quantitative, qualitative and self-referential feedback. The influences of qualitative positive and negative social feedback together with quantitative feedback on a user's energy consumption has been studied in [11]. The results show that people are sensitive to social feedback and the system is able to persuade the user to use less energy. The effects of long-term interaction and feedback on a dieting task have been studied in [9]. The robot collects quantitative data in terms of the user's dieting behavior (i.e. how much the user's have eaten or exercised). Using this information the system can support the users on updating their goals for daily exercising and calorie intakes. The results show that the users had a stronger alliance and used the robot longer for dieting than in any other tested condition. Further types of quantitative and qualitative feedback of SARs supporting on cognitive tasks have been studied in [4]. They tested different conditions where a robot either gave A) instructive feedback (reporting only scores; quantitative feedback), B) praising feedback related to the task performance (i.e. reporting scores and praising the user for correct sequences or reassuring the user in case of failure, qualitative feedback) or C) implicit feedback in terms of changing the difficulty of the test (i.e. same verbal feedback as in the previous conditions). The enjoyment was the highest in the implicit feedback condition and the task was perceived less frustrating. Furthermore, the role of implicit feedback from a SAR during a cognitive test has been studied in [14]. The users received either generic motivational feedback or performance-based qualitative feedback. The qualitative feedback was given based on the information whether the user has changed their answer too often, was in general too fast or too slow. The system reported this information to the user. The results showed that people in the condition with performance-based feedback did better on the cognitive task than those who received generic motivational statements.

While there is more literature on the topic of the short-term effects of feedback from SAR, there exists limited research that investigates the long-term effects of feedback on exercising tasks from a SAR. In the current research we focus on the development of suitable feedback mechanisms for long-term sport assistance and report our findings from an extended 18 day long-term HRI study.

MODEL FOR LONG-TERM FEEDBACK

We have implemented a global feedback model for the user, which gives him/her qualitative and quantitative feedback about her/his current and past performance as well as feedback about her/his training progression. It is based on the motivation theory explained in the previous section. We want to enhance the robot's capabilities for long-term interaction so that users keep motivation to engage and to comply with the robotic assistance high. Thus, we will describe our design and approach for building an assistive system for long-term exercising in the following.

Design Requirements

To be capable to evaluate the presented system with upcoming design iterations or study setups, different requirements are necessary. The variety goes from issues regarding training plan conception, fitness level of participants to system adaption.

Training Plan Representation

The robot needs an internal representation of each workout plan for the user. This training plan has to be designed, tested and adjusted beforehand study realization. It is composed of a warmup phase, a training phase, a pause phase between two training phases, and a cooldown phase. These phases consist of 1 to n movements. These movements represent the target heartrate, speed in rounds per minute (rpm), power in watt and the training posture the athlete should cycle in. Moreover, each interval can be of the type *speed*, *normal*, *power*, *pause*, *cooldown* or *warmup* indicating that the focus is either on cycling with a high cadence or power, for relaxation, warming up or cooling down. The session characteristic is based on the distribution of these intervals. Sessions including many speed intervals have their focus on a high intense cadence training, whereas sessions with many power intervals focus on high power training. The session types are evenly distributed over the period of our study to offer the participants a balanced training plan.

Movement Representation

We categorized different workout exercises as movements. These movements capture the targets, posture, preparations, instructions, micro-exercises and reparations a user receives while cycling. For example: **Preparation**: "*Attention! It's getting faster.*"; **Instruction**: "*Increase the resistance, pedal with 120 rpms and 100 watts.*"; **Acknowledge**: "*That looks good. Continue!*". The targets for each movement are: a heartrate threshold $w(hr_m)$, a target cadence and target power in watt. The postures are *standing* or *sitting* . Micro-exercises are *pushups* or *jumps*.

Design Issues Regarding Study Comparisons

Research on motivation shows that not only the feedback given by the instructor can enhance intrinsic motivation but also the task itself. The task challenge should be optimally balanced for enjoyment. Therefore, the task difficulty should rise with the user's experience in performing the task [1]. However, for research purposes and a controlled study setup it is not desirable to have an adaptive training plan yet. In order to be able to compare different conditions between different study setups (e.g. robot-present versus no-robot-present vs. enhanced system), the training plan has to be fixed and can not adapt over time to the user's requirements. This would complicate the evaluation of the effectiveness of the system. Hence, we have designed a suitable training plan for the user for 18 days indoor cycling training.

Design Issues Regarding Different Fitness Level

However, the static training plan needs to be individual to varying fitness levels of participants. This enables the system to evaluate the user's performance based on the commitment and not on the fitness level. To obtain the different fitness levels, the athletes had to perform a standardized test procedure called "IPN" test [10]. People cycle a cardio trainer with stepwise increase and decrease of the resistance. In the meantime, a chest strap records the heartrate. Based on the age, height, weight and recorded data the test results are used to calculate a recommended training heart frequency and a Watt/Kg (body weight) measure, which gives us the training power.

Global Feedback Design

The global session feedback given by the robot bases on the different aspects of feedback theory sketched in the section about motivation theory. In our model we included the quantitative and qualitative informational aspects of global session feedback.

Quantitative Feedback

The quantitative feedback is parted into feedback that is *non self-referential* feedback and *self-referential* feedback referencing on past training episodes. The non self-referential feedback has a non-evaluating character and gives the participant only the quantitative information about his/her performance without further explanations. The participants can decide on their own how to interpret the feedback. The other part is the self-referential feedback which creates a kind of competition with themselves and increases intrinsic self-motivation.

Non Self-Referential Feedback

The aggregated average heart rate, cadence and power values for each session (e.g. "Your average heart rate was 96 bpm, speed 98 rpm and power 120 watt).

Self-Referential Feedback

The comparison between current and past performance gives the participants the opportunity to challenge against themselves. The feedback depends on the type of the session (e.g. *power*, *speed*, *normal*) and compares the related average training values (i.e. rpm or watt) to the previous values of the same session type. If the performance has increased compared to the last session, the system also checks whether the user has exceeded her/his maximal heart rate. In cases where users exceeded this value the system warns the user not to train above her/his limits. If the performance has increased while the user has not exceeded her/his maximum heart frequency the robot gives encouraging feedback.

Qualitative Informational Feedback

The user also gets specific qualitative feedback about the phases of the training. This feedback is based on the movement success (see Eq. 1) of each interval. If the user succeeds in every movement of an interval, the system gives her/him $positiveglobalfeedback$ (e.g. "you were very good during speed intervals"). If there are also intervals where the participant only succeeded in half or less of the movement, the system checks the possible reasons. This type of feedback is called $globalcriticalfeedback$ and can deliver the following information: too much/less $power/speed$ for $power/speed/normal/warmup/cooldown$ intervals. Depending on the results the robot gives appropriate advice to cycle faster, slower or with less or more power. Depending on the phase of the session this information is also backed by sports scientific knowledge (e.g. "it is important for your muscles to warmup/cooldown smoothly").

Movement Success

To evaluate a movement success, each athlete has an individual heart rate threshold, which she/he should either exceed or fall below during a movement, depending on the current phase of the training (e.g. *cooldown* or *warmup*). Based on this, the system can compute a movement success, as well as the success for each interval. Using this information, we are able to give each athlete individual feedback. The movement success for an interval is computed as follows:

$$success(I) = \frac{\sum_{\forall m \in I} \tau(\overrightarrow{hr_m})}{|M_i|} \tag{1}$$

where M is the set of movements in interval i, I is a specific interval consisting of movements m and $\overrightarrow{hr_m}$ is the recorded heart rate data for a movement m. τ is the success heart rate threshold for a specific movement and computed as follows:

$$\tau(\overrightarrow{hr_m}) = \begin{cases} 1 & \text{if } max(\overrightarrow{hr_m}) > w_{m\{training \vee warmup\}} \\ 0 & \text{if } max(\overrightarrow{hr_m}) < w_{m\{training \vee warmup\}} \\ 1 & \text{if } min(\overrightarrow{hr_m}) < w_{m\{cooldown \vee pause\}} \\ 0 & \text{if } min(\overrightarrow{hr_m}) > w_{m\{cooldown \vee pause\}} \end{cases} \tag{2}$$

w is the threshold parameter that exists for each movement in each interval, which are categorized in $training, cooldown, warmup, pause$. During the *training* and *warmup* phases the participant should exceed the heart rate threshold, which results in a positive score. During the *cooldown* and *pause* phase the participant's heart rate should fall below the boundary.

The positive qualitative feedback consists of the interval with highest compliance score (e.g. "your warm up was real smooth today"). The critical qualitative feedback is about the interval with compliance score below .5 (e.g. half of the movements of one interval were not successful). The system computes the reason for the low compliance score (e.g. "too fast during cool down" or "too slow during warm up") and gives the user this information as feedback. If two or more intervals have the same lowest success score one of them is chosen randomly.

Performance Prediction

If there are enough data points for a specific session type available, a trend analysis about the performance can be computed and communicated to the user. The trend analysis is computed by a linear regression of the different training variables: compliance, heartrate, power, rpm.

METRICS FOR HUMAN-ROBOT INTERACTION

In order to measure the effectiveness of a social robot several different metrics have been proposed [15]. These measurements can directly be feedback to the user to elicit behavioral change. The potential measurements of our scenario are the user's compliance towards the robots instructions, the robot's persuasiveness and the user's training engagement. In the following we will describe how these measurements can be modeled and implemented in an assistive robot scenario in order to evaluate the motivational aspects of an robotic training assistance.

Modeling User Compliance

The user compliance is defined as the user's commitment to follow the instructions of the robot. In our scenario we use a training plan with fixed instructions for each movement for each interval and session. Hence, the compliance for an interval can be modelled as the percentage of instructions from the robot the user puts into action.

The overall compliance score of a session is computed as:

$$comp(S) = \frac{\sum_{\forall i \in S} success(i))}{|S|} \tag{3}$$

where S is the session consisting of a set of intervals i and $|S|$ the amount of intervals in one session.

Modeling Robot's Persuasiveness

The persuasiveness of the system can be modelled as the effectiveness of the robot's instructions to the user which were effectively put into action over time of the instruction. If the user actually follows the instruction of the robot and corrects negative execution of training movements the persuasiveness of the robot can be considered as effective. The persuasive measurement is modelled as:

$$pers(U) = \frac{\sum_{\forall f_i \in F} (|f_i| - 1)}{|N|} \tag{4}$$

where U is the user, $N = \{\forall f_i \in F, |f_i| \geq 2\}$, F is the set of global critical session feedbacks, f_i is the global critical feedback given of type i. So the persuasiveness score is the ratio of the same critical feedback given several times divided by the amount of global feedback types given that occurred more than one time. It shows how often on average the robot had to give a certain feedback to the user before it results in a change of behavior.

SYSTEM AND INTERACTION OVERVIEW

Robot System

Our previous HHI analysis revealed that interaction during indoor cycling is fine-grained and in sequential manner to the trainee's actions [16]. From our observations we have build an autonomous system that guides users through indoor cycling traing and gives them feedback based on their execution over extended periods of training sessions. As robotic target platform we use the humanoid Nao[3].

The design of such a system comes with a variety of requirements. The system has to perceive the training execution, vital data and own position. Furthermore, it needs to be able to make decisions based on these parameters and to be reactive in order to put these decisions into multimodal feedback (e.g. speech, gesture, head orientation and colour changing of eye-LEDs) to give the user corrective instructions or positive advice.

The robot's behavior during scenario-specific workout situations is triggered by the *action-based motivation model* designed as a state chart [16, 6]. The usage of state charts allows us to build reconfigurable patterns that can be adapted to the different exercises we encounter in indoor cycling regime. States trigger situation-specific interaction patterns, which are designed in our dialog sytem [4]. These interaction patterns execute multimodal behaviors modelled as the Behavior Markup Language (BML)[5]. We used BML to synchronize the gestures, head movements and the speech of the robot (e.g. if the robot instructs the user to stand up it also emphasizes this statement with an according gesture). Moreover, we also used BML to implement a synchronizer that allows Nao to show a beat gesture in time with the current bpm of a song, like human trainers would do to help their trainees to cycle synchronously to the music.

Besides the embodied vision for face detection and marker detection for localization, the robot has not-embodied perception for detecting the user's physical state and posture. We have used a bike computer from the indoor bike from SRM[6] which provides values for current cadence, power and speed. This enables the system to detect deviating values and to react in an adequate way. Furthermore we have a posture and pedal detection to assess the participant's posture on the bike. For those components we used two 3D depth cameras, one in front of the bike and one beside the bike. The posture component was used to identify whether the user was sitting or standing. In order to evaluate the participant's performance and to detect the physical limits online the robotic system needs to know the participant's heart rate using a chest belt. This allowed us to record and analyse the heartrate.

Local Human-Robot Interaction Protocol

Besides the global feedback and instructions that are given to the user as explained in section about feedback design, the system also gives immediate feedback during training according to the user's training performance. The general designed local robot behavior during the indoor cycling training includes the following interactional aspects: A) Structuring the workout sessions; B) Giving instructions during the exercises; C) Providing athletes with multimodal positive and corrective feedback.

The system instructs each workout session by announcing the target cadence, power and posture. While the user is cycling, the system observes the vital and trainings parameter. If one of the targets is violated the robot gives feeedback in a hierarchical manner. After a repair instruction the user has time to put the feedback into action, which is either acknowledged or regretted by the system accordingly.

Global Human-Robot Interaction Protocol

After a session has finished the system evaluates the performance based on the recorded data online. It announces the user's basic parameters like average heart rate, cadence and power. The system gives a self-referential feedback based on the comparison with previous sessions. In case compliances for any interval were below the threshold, the systems evaluates what might be the reason for this and feeds this information back to the user. At last it also gives a positive feedback according to the interval with the highest compliance of this session. Afterwards the system instructs the user to clean the bike and says good bye.

STUDY PROCEDURE

We conducted an 18-day randomized, controlled isolation study, where 16 participants (average 23.63 years) were tested in two campaigns with 8 participants each. The aim of the study was to simulate conditions of manned long-term space missions with one group accompanied by a robot assistance system compared to a control group without robot assistance. The daily activities and schedules of the two groups were identical. Participants had to do a workout instructed by our system everyday for approximately 50 minutes.

In both conditions the participants had a similar workout plan, i.e. the physical workout was the same for both groups. They exercised instructed by the robot each day. The participants in the control group had a non-interactive computer display, which showed them relevant information as text. It only provided instructions and structural information. After each session the average heartrate, power and cadence values for the current and the last session were displayed. Participants in the robot condition were greeted by the robot, took part in an assisted and reactive workout that, besides giving guiding instructions, also produced repair hints as well as positive and/or corrective local feedback during workout. Figure 1 shows the study setup and an exemplary interaction with the robot.

Participants

After an extensive pretesting phase, potential participants with extreme values on personality characteristics or physical fitness as well as persons with prior experience in robotics were excluded from participation. All 16 participants were healthy and had a Body Mass Index (BMI) between 20-25

[3] http://www.aldebaran-robotics.com/, visited on 5/22/2015

[4] https://toolkit.cit-ec.uni-bielefeld.de/node/368, visited on 6/14/2015

[5] http://www.mindmakers.org/projects/bml-1-0/wiki, visited on 6/14/2015

[6] http://www.srm.de/, visited on 6/14/2015

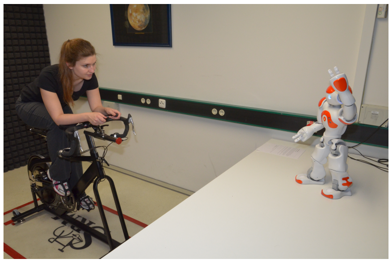

Figure 1. The participant is instructed by the assistive robot.

kg/m2. All participants were nonsmokers and successfully completed the medical as well as the psychological qualification. Each participant was matched with a corresponding partner from the other campaign based on personality and physiological parameters. They received monetary compensation for their participation. The study was approved by the Ethics Committee of the Deutsche Gesellschaft für Psychologie (DGPs; German Psychological Society).

RESULTS

We aim at evaluating the participants compliance towards the instructions of the robot in the course of an 18-days isolation study. Therefore we evaluate the compliance of the users during the study and how the global feedback is different between both study groups. We define the overall compliance as the percentage of successful intervals for a training session based on our described model. Figure 2 shows the average session compliance for the participants over the 18 days. However, due to missing data from the last day we excluded day 18 from our analysis. The session compliance is shown

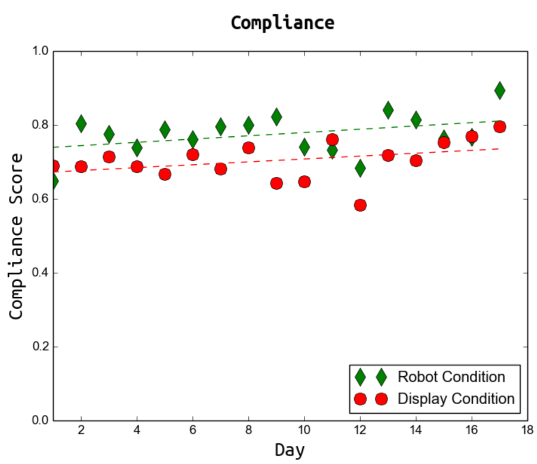

Figure 2. Global session compliance score over the time of the experiment.

in Figure 2. Due to some recording errors for day 3 (robot condition) and 14 (display condition), we replaced the values with the average values. We replaced missing individual values (due to sickness of participants) with the participants individual average value. A simple linear regression was calculated to predict compliance based on days of interaction with the system. The results show that the user's compliance is slightly increasing in both conditions. However, we found no significant regression equation. The linear regression error for the robot condition is $R^2 = 0.15$ and for the display condition $R^2 = 0.14$. Also the calculated slopes for both conditions (robot: $0.0044x + 0.73$, display: $0.0039x + 0.66$) show no significant difference. The average compliance over all days and all users (robot: M=.774 SD=.055, display: M=.703 SD=.05) shows a significant difference ($t = 3.766, p < .01$).

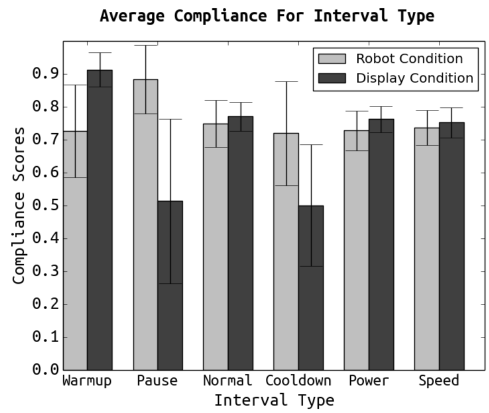

Figure 3. Average compliance for each interval type.

Figure 3 shows the average compliance for each interval. The two conditions (robot vs. display) differ significantly in the cooldown ($t = 4.131, p < 0.01$), warmup ($t = -4.565, p < 0.01$) and pause phase ($t = 6.896, p < 0.01$). The distribution of positive feedback is depicted in Fig. 4. In total there were 128 positive feedbacks in the robot condition. The users in the display condition did not receive any global positive or corrective feedback during the course of the study. In order to compare the two conditions we computed the feedback users in the display condition would have received based on their recorded performances. Hence, the users would have received 113 positive feedbacks in the display condition. The

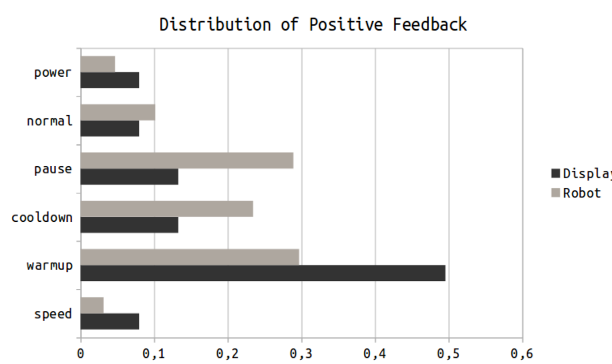

Figure 4. Distribution of positive feedback given to the user in the robot condition vs. hypothetical computed feedback for the display condition.

implications are that participants in the robot assisted condition received more positive feedback than the users in the display condition would have received.

In general the users in the robot condition received 51 corrective feedback while users in the display condition would have received 76. The feedback distribution in the robot condition is slightly better distributed over the different feedback classes. However, we can not actually compare the differences between those groups, because users in the display condition did not receive any feedback. Nevertheless, we can compare how often the same feedback has also been given on each following day. This indicates that missing assistance in terms of global feedback leads to repetitive bad performances. We counted how often the same feedback is given twice in a row for both conditions and compared them against how often a feedback was given at least one time. If the value is higher for the display condition it shows that the missing feedback has an effect on the user's training performance and that the feedback would have been an important advice to improve the training compliance. Figure 5 shows the computed feedback ratios for each participant (Robot: M=.227,SD=.006, Display:M=.8,SD=.36). This indicates that feedback was followed by the same feedback on the next day every 0.227 times in the robot condition and every 0.8 times during the display condition with significant difference ($p < 0.05$). A score of 0 means that there was no feedback given two times consecutively. This analysis only holds in the special case for a two-day time window. If we want to compute the persuasiveness for the system in the course of the complete study of 18 days we can use the proposed persuasive score (compare Eq. 4)

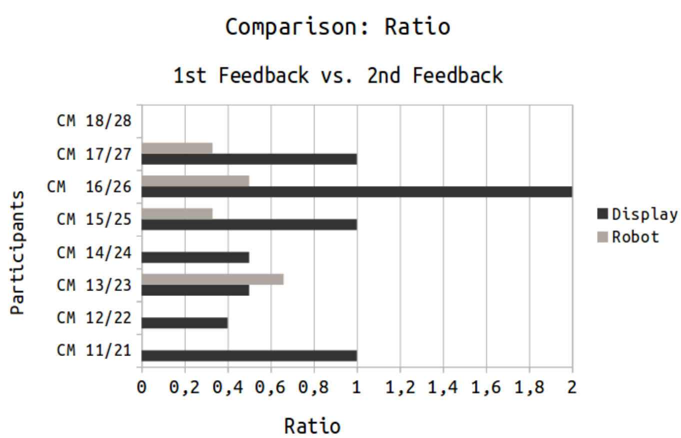

Figure 5. The average often a feedback had to be given a second time depending on it was also given the first time. The y-axis groups the matched participant pairs of our isolation studies indicated by ther identification number.

However, we can only compute the score for our robotic system, because this score takes all days into account and the display system did not have any feedback. The persuasive score is 1.56, which means that the robot had to give the same instruction on average 1.56 times in order to elicit a behavioral change of the user.

For reader's comprehension of the impact and discussion of this study, we will shortly sketch findings that were already reported but have not yet been backed with objective analytical results from recorded workout observations. Those results were obtained by daily questionnaires and health monitoring of the participants [16]: Resting heartrate significantly decreased over the course of the study for both conditions. (Robot: $p < .001$, Display: $p < .01$); Training enjoyment did not differ and remained the same between the conditions; Participants in the robot condition perceived the physical training as more challenging ($p < .01$) and more motivating ($p < .05$)

DISCUSSION AND FUTURE WORK

The results show that the robot's presence has an effect on the user's training performance and that they comply with the system's instructions and feedback. We have significant results that the user's compliance during robot-assisted training is higher and that the system is able to engage the user in long-term interaction over 18 days. Since the resting heartrate is decreasing and the compliance is rising in both conditions and the enjoyment is not different, the significant difference in the average compliance has its origins in feedback and assistance of the robot and not in the heartrate adaption due to daily exercising.

The significant differences for the warmup, cooldown and pause intervals might be due to the local feedback that is given to the user. Users in the display condition who received no global or local feedback tended to cycle in their own pace and did not follow the instructions that were displayed. This can explain the fact that the compliance is significantly lower during the cooldown and pause intervals, because they cycled faster than they were instructed. On the contrary during warmup phase the participants cycled faster than they should and therefore reached their heartrate threshold more often than participants in the robot condition, who received corrective feedback to warmup slowly when they cycled faster than the target value.

Surprisingly, we could not find any indications for a decline of the novelty effect. Usually when users are familiar with the robot during repeated HRI experiments they get frustrated or bored to interact with the system. This occurs from false expectations users have on the capabilities of the robot. The consequences are that they no longer comply with the instructions of the system and stop using it. However, our current observation of training data does not show any signs for a drop in motivation to comply with the instructions. There are two possible explanations for this: Either the system and the task are so well designed and motivating for the users that they are indeed not frustrated; or we have some ceiling effect. It is possible that the participants are already so well intrinsically motivated due to the selection criteria and this special kind of study that they do not need any further motivational boost for exercising. To further investigate this issue, we also need to run a qualtitative video analysis to find signs for a vanishing of the novelty effect.

Furthermore, the presented study was conducted under highly controlled conditions under very specific isolation circumstances. Hence, the participants were supremely motivated to participate and comply with the rules and strict daily routine. Therefore, it is questionable how the system performs under everyday study conditions and regular study participants (e.g. students from campus). Eventually, the compliance and training effects of the system can be quite different under normal conditions. Regular participants might need higher motivational expertise by the system in order to evoke behavioral change and compliance towards instructions.

The further challenging part for an upcoming system evaluation is to distinguish which part of the system has the primary

effects on the user's performance. Since the system in the robot condition gives also instructional feedback during the workout, we have to analyse the contribution of the global and the local feedback. In order to do to so, we will also annotate the video material to see how the athletes reacted on the robots feedback from an interactional point of view.
Moreover, the fixed training plan could demand too little from the users over extended periods of time. Hence, the challenges need to be adapted to the user's training progression to keep motivation for good performance high.
To investigate the effects of the embodiment of our robot, we need to implement a control condition using a text-to-speech interface instead of a display. Lastly, to evaluate the effectiveness of the local and global feedback we need to implement these features in a display and text-to-speech condition also. The results of this studies will give us new insights on the effectiveness of embodiment and adaption in long-term Human-Robot Interaction.

CONCLUSION

In this work we have presented a global feedback mechanism for robot assisted long-term sport training. We have shown that people comply with the robot's instruction for a very long period of consecutive training (18 days). Participants benefit from the socially assistive robot which proves to be a valuable feature for prospective far-reaching missions under isolated conditions (i.e. space missions, artic exploration). Furthermore, we have proposed measurements to evaluate and quantify socially assistive robots for long-term HRI in order to evaluate upcoming design iterations.

ACKNOWLEDGMENT

This work was supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG) and by the German Aerospace Center (DLR) - support code 50RA1023 - with funds from the Federal Ministry of Economics and Technology (BMWi) due to resolution of the German Bundestag. The authors want to thank Maikel Linke and Jan-Frederic Steinke for their support in developing and implementing part of the system and robot software.

REFERENCES

1. M. Csikszentmihalyi. 2000. *Beyond Boredom and Anxiety: Experiencing Flow in Work and Play* (25th anniversary ed.). Jossey-Bass. http://www.worldcat.org/isbn/0787951404
2. E. L. Deci. 1975. *Intrinsic motivation*. New York: Plenum.
3. E. L. Deci and R. M. Ryan. 2002. An overview of self-determination theory: an organismic-dialectical perspective. In *Handbook of self-determination research*, Edward L. Deci and Richard M. Ryan (Eds.). The University of Rochester Press, Rochester, NY, 3–33.
4. J. Fasola and M. J. Mataric. 2012. Using Socially Assistive Human-Robot Interaction to Motivate Physical Exercise for Older Adults. *Proc. IEEE* 100, 8 (2012), 2512–2526.
5. Jeonghye Han, Miheon Jo, Vicki Jones, and Jun H. Jo. 2008. Comparative Study on the Educational Use of Home Robots for Children. *JIPS* 4, 4 (2008), 159–168.
6. D. Harel. 1987. Statecharts: A Visual Formalism for Complex Systems. *Sci. Comput. Program.* 8, 3 (June 1987), 231–274. DOI: http://dx.doi.org/10.1016/0167-6423(87)90035-9
7. G. Schnabel H.-D. Harre and J. Krug. 2008. *Trainingslehre Trainingswissenschaft*. Meyer and Meyer Verlag, Chapter 4.1.2, 207–211.
8. D. M. Herold and Martin M. Greller. 1977. "Research Notes. FEEDBACK THE DEFINITION OF A CONSTRUCT.". *Academy of management Journal* 20.1 (1977), 142–147.
9. C. D. Kidd and C. Breazeal. 2008. Robots at home: Understanding long-term human-robot interaction. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 3230–3235. DOI: http://dx.doi.org/10.1109/IROS.2008.4651113
10. D. Lagerstrm and E. Trunz. 1997. *IPN-Ausdauertest*. Vol. 13(3):. 68–71.
11. C. Midden and J. Ham. 2009. Using Negative and Positive Social Feedback from a Robotic Agent to Save Energy. In *Proceedings of the 4th International Conference on Persuasive Technology (Persuasive '09)*. ACM, New York, NY, USA, 12:1–12:6. DOI: http://dx.doi.org/10.1145/1541948.1541966
12. A. Ramaprasad. 1983. On the definition of feedback. *Behavioural Science* 28, 1 (1983), 4–13.
13. R. M. Ryan and E. L. Deci. 2000. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology* (2000), 54–67.
14. S. Schneider, I. Berger, N. Riether, S. Wrede, and B. Wrede. 2012. Effects of Different Robot Interaction Strategies During Cognitive Tasks.. In *ICSR (Lecture Notes in Computer Science)*, Vol. 7621. Springer, 496–505.
15. A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich. 2006. Common Metrics for Human-robot Interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction (HRI '06)*. ACM, New York, NY, USA, 33–40. DOI: http://dx.doi.org/10.1145/1121241.1121249
16. L. Süssenbach, N. Riether, S. Schneider, I. Berger, F. Kummert, I. Lütkebohle, and K. Pitsch. 2014. A robot as fitness companion: towards an interactive action-based motivation model.
17. Robert S. Weinberg and J. Ragan. 1979. Effects of Competition, Success/Failure, and Sex on Intrinsic Motivation. *Research Quarterly. American Alliance for Health, Physical Education, Recreation and Dance* 50, 3 (1979), 503–510. DOI: http://dx.doi.org/10.1080/00345377.1979.10615637