"Can you answer questions, Flobi?": Interactionally defining a robot's competence as a fitness instructor

Luise Süssenbach, Karola Pitsch, Ingmar Berger, Nina Riether, Franz Kummert

***Abstract*— Users draw on four sources to judge a robot's competence: (1) the robot's voice, (2) physical appearance of and (3) the interaction experience with the robot but also (4) the relationship between the robot's physical appearance and its conduct. Furthermore, most approaches in social robotics have an outcome-oriented focus and thus use questionnaires to measure a global evaluation of the robot after interaction took place. The present research takes a process-oriented approach to explore the factors relevant in the formation of users' attitudes toward the robot. To do so, an ethnographic approach (Conversation Analysis) was employed to analyze the micro-coordination between user and robot. We report initial findings from a study in which a robot took the role of a fitness instructor. Our results emphasize that the participant judges step-by-step the robot's capabilities and differentiates its competence on two levels regarding to the robot's role: a robot as a (1) social/interactional co-participant and as a (2) fitness instructor.**

I. INTRODUCTION

In human-robot interaction (HRI) users assess the robot's competence for example from the physical appearance of the robot [1] as well as from the ongoing interaction itself. In recent years a range of studies investigate attitudes toward social robots via questionnaires and/or interviews [2], [3], [4]. The interaction itself and the micro-coordination between user and robot are often neglected, although the user infers a robot's competence online during interaction. Video-based studies that pursue a qualitative approach enable a finely-structured analysis that may shed light on this interactional process. A number of studies [5], [6], [7] indicate that in the initial beginning of an interaction the user quickly forms a vague idea concerning the robot's competence and generates hypotheses accordingly (e.g., Can the robot talk? Can it see and hear me?). Therefore, the beginning of an interaction is a particular important phase for interactionally defining a robot's competence. This already suggests the single answer of a participant in our study's post-interview: *How often did you communicate with the robot?* - "Rarely. In the beginning a few times, then I gave up.".

In this paper, we investigate for the scenario of a robotic fitness instructor how users perceive and categorize the robot in and through the interaction with the system. We will present a micro-analytic case study of one participant (drawn from a larger video-based long-term study using an initial WoZ approach) interacting with the robotic head Flobi [8] set up as indoor cycling instructor.

For this task the robot needs to fulfill a set of requirements. In the role of a fitness instructor, the system needs to comply to standards of sports theory and fitness instructions. It needs to know about next actions (i.e., action sequences) and should communicate them to the trainee at the right moment in time. Concerning its interactional competence, the system needs to observe the trainee's behavior moment-by-moment, react appropriately in a multimodal way, and be consistent in its behavior and reactions.

Hence, the current paper adresses the following questions: (i) How do users categorize the robot and infer its competences (with special focus to the beginning of the interaction)?, (ii) Which factors influence the ascription of competence during an interaction? and, (iii) which robot conduct might influence the user's perception of the robot's competences?

Because of the importance of the first moments of an encounter for shaping the user's perception and expectations about the robot, special emphasis will be given to the beginning of the interaction. The analysis will be completed using fragments from later stages of the workout, thereby illustrating how the user teases apart the robot's different roles in a step-by-step fashion. We argue that the participant judges the robot's competence on two levels: (1) interactional capabilities and (2) task-related competence.

II. BACKGROUND AND RELATED WORK

Previous research has pointed to different aspects that influence users' perceptions. The user infers a robot's competence from a range of sources: the robot voice [9], the physical appearance [1] of and the interaction with the robot [3], [4]. Methodologically, these studies have generally investigated the users' categorizations using questionnaires and/or interviews. A recent study [2] shows that merely the anticipated HRI already affects the users' attitude toward robots. Here, the participants were asked to watch a short video clip (without sound) showing the robot's head movements and facial features followed by a textual description of the robot's typical behavior. To assess the acceptance of the robot participants were asked to report how much time per week they would be willing to spend with the presented robot and how much they would like to meet and talk to it. Another study [9] reveals that - apart from visual cues - also the voice

*We gratefully acknowledge partial support by the German Aerospace Center (DLR) with funds from the Federal Ministry of Economics and Technology (BMBF) due to resolution 50RA1023 of the German Bundestag.

**Karola Pitsch acknowledges the financial support from the Volkswagen foundation (Dilthey fellowship).

L. Süssenbach, K. Pitsch, I. Berger, N. Riether, F. Kummert are with the Applied Informatics Group and the Research Institute for Cognition and Robotics (CoR-Lab), Faculty of Technology, Bielefeld University, 33501 Bielefeld, Germany (corresponding author: luise.suessenbach@ uni-bielefeld.de).

affects the acceptance of HRI: Participants displayed greater acceptance of HRI when the robot and the participant shared the same gender (gender via female/male voice). In further studies [3], [4], attitudes toward robots were investigated via real interaction: Following an interaction with an robot, participants' attitudes towards robots were assessed.

Whereas outcome-oriented approaches via questionnaires provide information on the final outcome, process-oriented techniques enable researchers to identify factors relevant in shaping robots' competence on-line [5], [6], [7]. Conversation Analysis, a qualitative approach, provides an interactional approach to the topic of categorization [10]. Conversation analysis [10] identifies the practices and resources which participants use in their interaction with persons and objects. In the present investigation of participants' perception of a robot system this is relevant in two ways: (i) through their own pro-active actions (first turn) and (ii) through their reactions upon the robot's actions (second turn) users attribute certain properties are ascribed to the robot.

III. THE ROBOT SYSTEM: A WIZARD-OF-OZ SET-UP

As a first approach in a series of iterative steps toward designing an autonomous system, we realized a basic wizard-of-oz set-up. This allows us to investigate user reactions toward a robot system at an early stage while - in parallel - continuing to develop the technical basis for creating an autonomous system. In this initial stage of the project, the robot's behavior is partly autonomous and partly controlled by a human, who is positioned in an adjacent room. Next to two HD cameras (Panasonic) and two Kinects (Microsoft) we recorded the participant's heartrate and power.

A. The robot

To carry out our investigation, we used the humanoid robot head Flobi (see Fig.1), which can communicate (using an external loudspeaker positioned next to the robot head) and to display a range of facial expressions [8]. We made the robot perform a set of preconfigured verbal utterances which were linked to three basic facial expressions (neutral, positive, negative). We chose a facial design that unpublished research from our lab indicated as being suitable for the task.

Fig. 1. Anthropomorphic robot head Flobi

B. Controlling the robot's action

1) Wizard-of-Oz: The robot system was controlled by a researcher acting as wizard: A research assistant who was positioned in the adjacent room (see Fig.2) monitored the human participant (speech, activities, heart rate, cycling frequency and performance) in real-time on a computer screen and controlled the robot's action. To do so, the wizard had at his disposal an interactional interface that included 123 preconfigured verbal utterances. All utterances were categorized in three types (neutral, positive, negative) and therefore linked with three kinds of facial expressions (neutral, positive, negative). For comparability reasons Flobi's

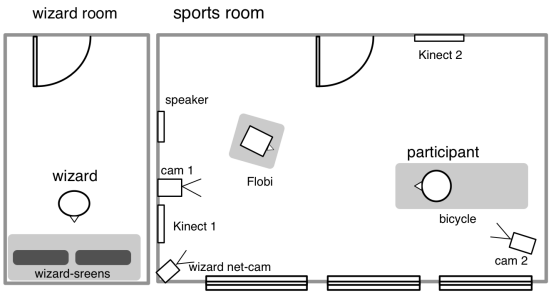


Fig. 2. Study Set-Up

interaction was restricted and limited. In general, there was a strict script which the wizard followed in every session for every participant. Thus, the wizard used a paper script, which included the interactional framing (the entry, greeting and farewell) and the workout itself. The wizard had to select specific verbal utterances depending upon the point in time during the workout. Furthermore, if the human initiated a dialogue or if the wizard had to react on human conduct he/she could do so via utterances and feedback (e.g. "yeah", "right this way", "pedal faster/slower").

2) Autonomous behavior: Besides verbal utterances that were controlled by the research assistant, the robot performed random background behavior autonomously, that included gaze and head movement. This background behavior was only active when the human wizard was not selecting any utterance.

IV. STUDY DESIGN

This wizard-of-oz study on long-term interaction effects includes 27 participants with 6 fitness sessions over 3 weeks. Before the study, all participants underwent a performance diagnosis. The group, in which participants interacted with the robot, consisted of 15 participants whereas the control group, in which participants only received general information concerning cycling, constisted of 12 participants. However, in a first step, this paper only focuses on one single case addressing the following questions: (i) How does the user infer the robot's competence in the beginning of interaction? , (ii) What factors influence the ascription of competence during on-going interaction? and, (iii) which specific skills (i.e., the robot's multimodal conduct) influence the user's perception of the robot's competences?

A. Procedure

When the participant arrived for the very first time he or she was asked to fill out the questionnaires regarding their

(i) attitude towards robots and (ii) their intrinsic motivation for doing sports in a separate room. In order to adjust the bicycle and to introduce the robot, the experimenter accompanied the participant to the sports room. To avoid a triadic situation (i.e., robot, experimenter and participant) the robot was turned off for that time. Nevertheless, the robot's eyes were wide open and it's gaze was oriented toward the bike. The experimenter introduced the robot by "That is Flobi and it will accompany you during the workout. Everything else will get clear in a moment. Enjoy yourself". After this explanation and the adjustment of the bike the experimenter left the room and entered the wizard room. With the closing of the door the wizard initiated the robot's first verbalization: "Hello. My name is Flobi."

B. Background: Flobi's behavior

In order to get an idea how human fitness instructors design an exercise unit and try to motivate their trainees, we first analysed how human fitness instructors motivate their trainees in everyday real-life situations [11]. Some of these methods observed in human-human-interaction (HHI) were implemented:

1) Greeting, farewell, questions: in order for Flobi to take the role of a fitness instructor the robot not only needed to organize the exercise unit but also needed to get into a situated interaction. To do so, the robot needed to (1) organize its gaze direction and (2) realize strongly ritualized interaction sequences such as the entry to an interaction [12] as well as the farewell. Throughout the interaction the robot posed questions to the participant which were very likely to evoke an answer (e.g., "what's your name?" or "how are you?"). These questions were time-locked (i.e., at a fixed point in time the question was posed).

2) Information: the robot verbalized simple information regarding the present state of the workout (e.g., "hill number one. 700 metres uphill.", "just 200 metres left", "you are almost there").

3) Feedback and repair: at fixed moments Flobi provided two types of instruction. The first type focused on the participant's heart rate: during non-challenging parts of the workout (plain land) Flobi verbalized for example "try to keep your pulse between 135-145." If the trainee was not able to do so Flobi tried to repair this by saying "change down a few gears" or "slower". The later instruction adressed the participant's movement itself. Likewise, the robot asked the trainee to get up and to pedal standing up. If the trainee did not follow the robot's instructions, it repeated its instruction and reacted to the situation with additional utterances like "come on, get up!".

V. METHOD

In order to investigate how the robot's conduct forms the way in which the participant perceived the robot, we used an analytic approach that enables insights into the sequential structure of the interaction [10], [12], [13]. This analytic approach is qualitative being based upon Conversation Analysis (CA). CA describes the structure, order and sequential patterns of interaction and the micro-coordination between the participants and detects analytical phenomena "from the data themselves" [13]. This allows us to investigate the interrelationsship between the robot's and trainee's actions and how they respond to each other on the level of structural features. The aim of this approach is to reconstruct the participant's view: We investigate the trainee's perception of the robot's actions and analyze how and especially why it changes in interaction.

We start with an explorative, in-depth qualitative analysis of one single case to detect the organizational features of various, naturally occuring, interactional phenomena in order to reveal the relevant analytical issues and categories using ethnomethodology and CA. In order to describe the detail of interaction and the timing and relationship of the events of all interaction partners this approach demands a repeated inspection of video-taped data and the transcribing of all events in interaction. The aim of this analysis is to find the structural organization of a number of actions (e.g., changing gaze, utterance, noticeable prosody) and how one action is contingently relevant for another action. Due to these structural properties, also the absence of an (otherwise expected) action is relevant.

In order to verify the findings made in one single case it is of course neccessary to apply the decribed approach in a set of cases in the future.

VI. OVERVIEW

In a first step we present some preliminary findings from the qualitative analysis taking a close look at eleven fragments of the first exercise unit of one participant. The fragments range from the very beginning (experimenter-phase # 00:00:06) till the last third of the workout (3rd hill # 20:30.04) (s. Fig.3).

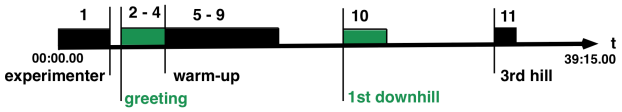


Fig. 3. Interactional fragments 1 - 11

VII. OPENING AN ENCOUNTER

In order to clarifiy the step-by-step change of categorization we have to analyse the initial pre-beginning phase, where the participant sees the robot for the first time.

A. The trainee's initial definition: the robot can talk

The participant (P) is enteres the sports room, accompanied by the experimenter (I). At this time the robot's eyes are wide open and they look at the bike. While the trainee gets on the bike and adjusts his feet in the pedals, he looks to the robot and asks the experimenter: "It will speak to me, right?" (l. 02). Only the question marker "right?" marks the declaration as a question. This suggests that the participant already has a competence profile of the co-present robot at this initial moment - *the robot can talk.*

Fragment 1: 00:00:06 - 07:47.26

```
01 E:   |SO::-
         well
   P-g: |@F...

02 P:  der spricht gleich mit mir; ne,
       it will speak to me, right?
   P-g:.............@F.............

03 E:  ja genau; also das |is floBI,|=
       yeah right, that's floBI,
   E-a:                  |pointing |
                         |#06:57.74|
                         |*Fig.3   |
```

Fig. 4. # 06:57.74

B. Introducing the robot: the robot as a technical artefact

The experimenter affirms this and introduces the robot's name: "yeah right, that's Flobi" (l.03). In this moment, the experimenter points to the robot (s. Fig.3). The trainee immediatly takes the robot's name and repeats it (l. 04). Up to this point, the trainee keeps looking at the robot (l. 01-05). Only when the experimenter starts to extend the explanation the participant changes his gaze and looks to his feet and then to the display (d)(l. 05).

```
04 P:  flobi,=
       flobi
   E:  |= das is:| ja genau flobi-
           that's: yeah right flobi

05 P-g: |@down.........................
   E:   |und das is quasi dein- der- äh
        and that's your- a- euh
        begleitet dich heute durchs training;
        it accompanies you during the workout today
        und wird dich dann über alles weitere aufklären.
        and later, it will explain everything else

06 P-g: |@F | @d.............................
   E:   |so;| jetzt müssen wir mal gucken hier-
        well; now let's take a look at this here-

(explanation of the bike's display)

 [...]

07 E:   viel spaß.
        enjoy yourself
   P-g: @F
```

Then, the experimenter explains the robot's role in depth. At this time neighter the participant nor the experimenter is looking at the robot. The following "well" (l.06) marks the next sequence and ends the introduction of the robot. The ritualized phrase "Enjoy yourself" closes the phase of robot's introduction.

This initial analysis suggests that introducing the robot generated first vague hypotheses concerning the robot's competence. In this situation - the first contact - the robot is not an interaction partner, but rather a technical artefact.

VIII. OPENING: DEFINING THE ROBOT'S SPEECH COMPREHENSION

Let us know consider the entry into the interaction: the trainee is already sitting on the bicycle and the experimenter has just left the room. The participant looks down on the display, where he can see the route, his own heart rate, cycling frequency and performance. Now Flobi changes its gaze to him and says "hello" (l.08).

Fragment 2: 07:48.50 - 07:57.34

```
08 F:         HALlo;
              hello
   F-g: @P
   P-g: @down       |@F
```

At this moment, the participant changes his gaze orientation from the display to Flobi. To the participant it seems that this eye-contact is the first moment of mutual awareness (l.08). The robot continues, introduces itself and asks for the participant's name: "Hallo. My name is Flobi and who are you?", which makes an answer from the human structurally relevant (l.09). Indeed, the participant reacts immediatly by giving his name: "Eugen." (l.09). The participant looks at the robot, waits for about 3 seconds, then repeats his answer in a more explicit form: "My name is Eugen".

```
09 F:   ich bin FLObi und wie heißt du?
        my name is Flobi and who are you?
   P:   Eug'n; (2.7)
        Eugen

10 P-f: smile
   P:   |ich HEIße EUgen; ||(2.5)    |
        my name is Eugen
   P-g: |@F...................||@display|

11 F:   schön, dass du hier bist.
        nice to see you
```

On the syntactical level, he updates the original one-word utterance to a full sentence; prosodically all syllables are stressed. With this reformulation, the user defines the 3 seconds of silence following his first answer "Eugen" as a slot, in which he would expect a reaction from the robot. Indeed, Flobi reacts on it and continues the opening sequence with "nice to see you". Thus, the trainee may suggest that the robot needs syntactically full sentences, that are articulated accurately.

IX. CONSTITUTING ROLES: INSTRUCTOR & TRAINEE

After that opening, the robot begins to suggest the specific role it is supposed to assume during the workout (l.12). This explanation consists of a sequence of declarative sentences. During this explanation the trainee changes his gaze and focuses on the display (l.14-16). This gaze organisation indicates the initial establishment of an interactional competence asymmetry. In the beginning phase eye-contact should be established [12]. However, in the end of the explanation the participant changes his gaze and looks to the robot. After the explanation he smiles and makes clear that he got it right: "All right, Flobi." (l.17).

Fragment 3: 07:57.36 08:12.78

```
12  F:  ich werde dich durch||die sechs
        I am going to support you for six
    P-g:|@F                 ||@d.......

13  F:  Trainingseinheiten begleiten un|d gebe
        exercise units and provide
    P-g: ............@d.................|@F....

14  F:  dir waehrend des Trainings Ratsch|läge.
        advice during the exercise.
    P-g:.........@F.....................|@d....

15  F:  solltest du mich einmal nicht verstehen,
        If you don't unterstand me
    P-g ....@d.................................

16  F:  gib mir bitte beschEId;
        please let me know.
        natürlich kannst du mir zwischendurch
        gerne sagen; wie du dich |fühlst. (1.5)
        You may tell me how you feel, of course.
    P-g: ..................@d....|@F.....

17  P:   .hhh (-) alles klar, Flobi.
                  All right, Flobi
    P-f: |smile|     |
    P-g: ...@F |@down|
```

A. Defining a fitness instructor: task-related requirements

The robot continues and describes what is scheduled for the first exercise unit (l.18). The fact that the trainee again is changing his gaze from the display to Flobi indicates that the robot's pause of three seconds is too long. This change of gaze organization and the fact that he until now does not begin to pedal suggests that the participant expected an instruction.

In HHI the instructor marks the start of the workout by starting to pedal [11]. The robot head does not fulfill this requirement, which thus has to be substituted by verbal means. In the following fragment this becomes apparent:

Fragment 4: 08:13.78 - 08:35.42

```
18  F:  |HEUte| stehen zehn kilometer
        today ten kilometres
    P-g:|@down......

19  F:  radfahren auf dem plan. (1.5)
        pedaling are scheduled
    P-g:.....@down...          |

20  P:         gut|
               fine
    P-g: |@F      |
```

Just two seconds later Flobi expands its explanation and clarifies the route (l.21-23). Until now the participant sits on the bike and does not begin to pedal. During the following explanation the trainee starts to move his legs, but not to pedal (l.23).

```
21  F:   erst wä|rmen wir uns AUf,
         first we warm up
    P-g:          |@down....

22  F:   dann berg- und talfahrt; (-)
         then up and down the mountain
    P-g: ...@down....

23  F:   am ende wärmen wir uns wieder ab (2.0)
         in the end we us cool down
    P-a: move legs

24  P-a: nod
```

He does not start to pedal before the instruction "we warm up now" (l.25). After that the robot verbalises the next notification (l.26). Immediately with "we start easy" the participant stops pedaling and looks at the robot. Shortly after Flobi has finished its turn the participant continues pedaling. This "wait and see"-behavior is also an indicator for Flobi's role: It suggests on the one hand that the participant waits for an instruction thereby illustrating Flobi's role as a fitness instructor. On the other hand the inadequacy of a robot head as an indoory instructor becomes apparent.

```
25  F:   wir |wärmen wir uns !JETZT! |AUf, (1.0)
         we warm up now
    P-g:     |@F.....................||@d..

26  P-b: starts pedaling (2.0)
    F:   |wir starten ganz leicht
          we start easy
    P-a: |stops p  | |starts p   |
    P-g: ....@d    | | @F        |
```

Up to this point, the interaction shows that the participant starts to design a competence profile of the robot head: the robot is interactionally restricted, but does have instructional influence. However, with ongoing interaction the participant's profile of Flobi's competence emerges:

The following fragment - immediatly after the opening - shows that the participant rephrases Flobi (l.28). The robot's anouncement "activate your circulation" (l.27) is a simple notification and a reaction on it is not necessary or rather cannot be expected. The participant catches Flobi's announcement and rephrases it: "activate your circulation, yo.". The added "yo" is a comment on an interactional meta-level. We suggest that this reaction may comment the discrepancy between semantic content and the prosodic design of Flobi's utterance, that may sound too stoccato.

Fragment 5: 08:42.26 - 08:47.62

```
27  F:   |bring den kreislauf in schwung (1.0)
          activate your circulation
    P-g: |@d.........................

28  P:   bring den kreislauf in schwung, jo;
         activate your circulation, yo
    P-g: ..........@d .....................|

29  P-f: smile
```

This user's rephrasing indicates that the robot's phonetically realized instruction "activate your circulation" does not fit it's task-related requirements and hence reduces Flobi's task-related competence. This shows that also the phonetical realization of instructions influences the evaluation of the robot's competence.

B. Monitoring & Testing

It is striking that the participant monitors Flobi for about nine seconds (Fig.5, l.30). In this moment Flobi is in its random behavior state and orients its gaze anywhere but not to the participant.

Fragment 6: 09:55.16 - 10:10.14

```
30  F-g: |@right.....|
    P-g: |@F.........| (9.0)
         |# 09:55.16|
         |*Fig.5     |
```

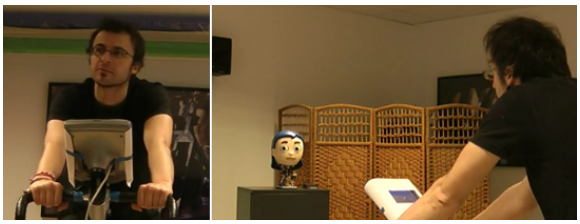


Fig. 5. # 09:57.78

```
31  P:   FLO::bi
         Flobi
    P-g: ...@F.....

32  F-g: @down left|
    P-f:           |smile      |
                   |# 10:06.46|
                   |*Fig. 6    |
```

Nine seconds later the trainee verbalizes "Flobi" (l.31). The intonation is remarkable and suggests that the participant in fact does not address the robot in this moment. Immediatly after the participant's announcement the robot changes its gaze orientation and moves its gaze elsewehere. The participant reacts on this action and starts to smile (l.32). His mouth is wide open and it seems as if the participant is surprised. This suggests that he interprets the change of gaze as a reaction to his announcement (Fig.6).

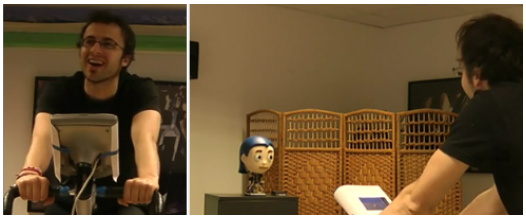


Fig. 6. # 10:06.46

C. Defining an interactional partner: interactionally inconsistent behavior

Flobi is still in its random behavior state and does not react to the participant. Hence, the trainee monitors Flobi again for about 16 seconds. Then he asks the robot "Can you answer questions, Flobi?" (l.33). Again the robot changes its gaze from the right to the left side (l.34). Due to this the human suggests that the robot is not able to answer questions (l.34), laughs and looks at the display.

Fragment 7: 10:20.12 - 10:38.36

```
33  P:   |Kannst du FRAgen beantworten, FLo:bi?| (3.0)
          Can you answer questions, Flobi?
    P-g: |@F.................................
    P-f:                                     |smile

34  F-g: @left |
    P:         |NEIn; kannst du NICHt. (-) (lacht)(-)
                No, you can't. (-) (laughs)

35  P-g: @d
P. starts to manipulate the gear
```

Just two seconds later Flobi orients its gaze to the participant and says "no" (l.36). Immediatly after the robot's answer, the participant stops manipulating the gear and changes his gaze from the display to the robot. The trainee responds to the robot's answer and replies laughing "no?" (l.37).

```
P. cancels manipulating the gear
36  F:   |nEIn; (-)
          no
    F-g: |@P...

37  P:   |NEIn?     |(lacht)
          no?
    P-g: |@F.......||@down (2.0)

38  P:   |is aber nen bisschen paradox (lacht)
          that's a bit paradoxal (laughs)
    P-g: |@F
```

This laughter comments the robot's behavior which is illogical. A few seconds later the participant makes clearer why he is laughing and comments on the robot's inconsistent interactional behavior (l.38).

D. Expending the robot's skills: Flobi remembers my name

During the warm-up phase the robot says "You do it well, Eugen." (l.39) and picks up the participant's name for the first time. The participants reacts and thanks. The participant thanks Flobi and smiles at him (l.40).

Fragment 8: 11:15.48 - 11:19.86

```
39  F:   |das ma|chst du sEHr gut; EUgen. (-)
          you do it well, Eugen.
    P-g: |@d...||@F.....................|

40  P:   oh:, gu|t; danke- (1.0)
         oh, well, thank you
    P-g:        |@d..............

41  P-f: smile
```

E. Defining an interactional partner: interactionally inconsistent behavior AGAIN

Afterwards the participant says "you remember my name, don't you?" and orients his gaze to the display (l.42). Due to the intonation in the end and the change of gaze toward the display we suggest that he does not expect any answer.

Fragment 9: 11:20.64 - 11:31.24

```
42  P:   den NAmen haste dir wohl gemerkt, ja,| (3.0)
         you remember my name, don't you?
    P-g: ..................@d...................|
```

Also the fact, that he already has asked if Flobi is able to answer questions, and the robot negated it, indicates that he does not expect any answer. However, the wizard and thus, the robot, respond to it (l.43). The participant's laughs may be due to the robot's inconsistent behavior.

```
43  F:   ja;

44  P-g: @F
    P:   (lacht)
          laughs
    P-m: smile
```

Up to this point, the analysis shows that the participant conceptualizes Flobi's capabilities as interactionally restricted and limited (e.g., inconsistent behavior, too long pauses). Furthermore, the participant's conduct suggests that he does not define the robot as a sociable partner (e.g. laughing about, gazing at the robot).

F. Defining a fitness instructor

That the participant may do not perceive the robot as a sociable co-participant also becomes apparent in the following fragment: The participant monitors the robot for about four seconds and adresses it with "do something with your eyes." (l.46). Of course, in an everyday-life instructor-trainee scenario such a comment can not be observed (e.g., because of role allocation and social conventions of our culture).

Fragment 10: 20:05.86 - 20:30.04

```
45  P-g:     @F (4.0)

46  P:      |mach mal was mit deinen augen (-) (3.0)
            do something with you eyes
```

The participant monitors the robot's conduct. His facial expression indicates that he waits for a reaction. After three seconds the participant starts to formulate an utterance again, but the robot interrupts him. With the robot's verbalization he stops his one (l.46).

```
46  P:(|xxx)-
    F: |kontrolliere mal deine TECHnik auf dem bike, EUgen.
        check your posture on bike, Eugen.
    P-a:straightes up  (3.0)
```

Ignoring the participant's instruction the robot instructs the participant by saying "check your posture on the bike, Eugen." (l.46). This utterance is scripted and not a responsive action. From the participant's point of view the robot formulates its advice in the right moment in time and syntactically in the same way as he adressed the robot just a few seconds before. Thus, the participant immediatly starts to move his upper body (l.46). Afterwards he adresses Flobi with the question "what do you mean by this?" (l.47).

```
47  P:    was meinst du damit? (-)
          what do you mean by this?
    P-g: |@down.........||@F...
```

The robot continues instantly and explains what the participant is supposed to do (l.48). With the instruction "your upper body is fixed the participant fixes his upper body (s. Fig.7, l.49). The fact that the participant follows this instructions shows that he shapes the robot's competences as suitable for the role of an fitness instructor.

```
48   F:  dein BAUch ist fest angespannt
         your belly is strained
         # 20:20.36
         dein oberkörper ist aufrecht (-)
         your upper body is fixed

49   P-b: |fixes upper body  |
          |*Fig.7 # 20:22.58|
```

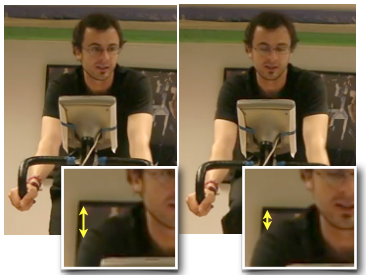


Fig. 7. # 20:20.36 & # 20:22.58

```
50 F:   die schultern sind locker
        your shoulders are relaxed
        und deine arme liegen parallel und leicht gebeugt
        and your arms are parallel and slightly bent

        (P moves his body, changes his posture)
51 P:   aha
        aha
```

In the following fragment this also becomes apparent: The workout is in the final stage (last hill) and the robot instructs the participant to get out of the saddle and to pedal a while standing up.

```
28:37.00
52  F:    |    |steh auf. (.) komm aus dem sattel.| (.)
           get off. get out of the saddle
    P-g:  |@d..|@d.............................|
          |# 28:37.46
          |*Fig.8|

53  P-a:  |gets up  (3.0)
          |# 28:40.00
          |*Fig.8
```

Immediatly, the participant gets up and pedals standing up (Fig.8, l.53). That the participant follows the robot's instruction indicates even clearer that he accepts the robot's role of an instructor.

Fig. 8. # 28:37.46 # 28:40.00

In this moment, Flobi does not mark the end of the exercise accurately and only says: "pedal a while standing up." (l.54). This means that the participant is able to complete the exercise by his own, whereas - of course - the top of the hill would be the final aim for doing so.

```
54  F:     fahr ein stück im stehen
           pedal a while standing up

55  F:    |gib alles| (-)
           do your best
    P-g:  |@f.......|
          |# 28:52.00 (1733 sec)

56  P-a:  |pedals faster
          |*Fig.8
[...]

57  F:     and sit down. (-)
58  P-a:   sits down
```

During the participant is still pedaling standing up the robot instructs him saying "do your best". Immedialty the participant starts to pedal faster. Figure 9 displays that the participant pedals faster and thus, his power increases starting with the instruction "do your best.". The participant does not sit down until the robot intsructs "and sit down".

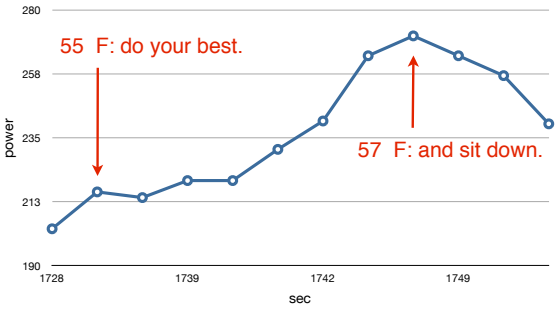


Fig. 9. Participant's power 1728 sec - 1764 sec

The analysis suggests that the participant differentiates the robot's role. On the one hand, he perceives the robot as a technical system that tries to adhere to communicative standards (e.g., question -> answer). Up to this point, the analysis indicates that the particiant does not perceive the robot as sociable or as a co-participant, because he does not follow social norms of politeness (exception: Fragment 8 "thank you"). On the other hand, the robot undertakes instructional task that are characterisic for a fintess instructor. Furthermore, the participant evaluates these both roles separated: although he perceives the robot as a limited technical system that acts partly illogical, he follows the robot's instruction.

X. CONCLUSION & IMPLICATIONS

The analysis suggests that the participant establishes a competence profile of the robot from the first contact - when he enters the room: "It will speak to me, right?". Furthermore, the pseudo-triadic (participant, experimenter and robot) situation in the beginning already affects this profile: at this time the robot is not an interactional partner, but rather a technical artefact that is *turned off*. With ongoing interaction it becomes apparent that the participant differentiates the robot between two roles: (1) sociable co-participant (2) technical system, that instructs. The participant judges the robot's conduct regarding these both roles.

However, these initial observations on an interactional level already sketch out a first set of requirements: (i) the robot may not be introduced by the experimenter or (ii) the robot has to be turned off - recognisable from the participant's point of view (e.g., closed eyes, lowered head) and (iii) the robot's conduct has to be consistent and adequate to the task (e.g., for an instruction to be motivating we need an appropriate phonetic design).

In a next step we need to analyze how this dual competence profile emerges in the following workout sessions. First preliminary analysis suggests that the participant starts to test the system not only on the level of interaction: The participant starts to count seconds between robot's verbal utterances, speaks syncronically to it, he waves his hand to the robot when it looks away and he rejects some of the robot's instructions.

APPENDIX: TRANSCRIPTION CONVENTIONS

Each tier/line represents the annotated conduct of robot (F), participant (P) and experimenter (E), and shows either verbal utterances, gaze (-g), actions (-a) or facial expressions (-f). Verbal utterances use the GAT standard, i.e. general spelling in lower case, upper case signifies stressed syllables and punctuation denotes prosodic features (',' = rising; ';' = falling). Important annotation symbols are F = robot, E = Experimenter, d=display, down=down; @ = at. Video stills are linked to the transcript via their time code.

ACKNOWLEDGMENT

The authors would like to thank Florian Berner, Marcel Riehle, Iris Sossalla, Hendrik Stellmacher, Lisa Viktoria Skutella, Jörn Weitz and Jana Mayer for their support in conducting the study.

REFERENCES

[1] F. Hegel, M. Lohse, & B. Wrede, "Effects of Visual Appearance on the Attribution of Applications in Social Robotics," in *ROMAN (IEEE International Symposium on Robot and Human Interactive Communication),* Toyama, Japan, 64-71, 2009.

[2] D. Kuchenbrandt, F. Eyssel & S. Bobinger, "Effects of anticipated human-robot interaction and predictability of robot behavior on perceptions of anthropomorphism" *6th ACM/IEEE Conference on Human-Robot Interaction,* Lausanne, 2011.

[3] M. Heerink, B. Kröse, V. Evers and B. Wielinga, "Studying the acceptance of a robotic agent by elderly users," in *International Journal of ARM, International Journal of ARM,* VOL. 7, NO. 3, 33-43, 2006.

[4] D. S. Syrdal, K. Dautenhahn , K. L. Koay , M. L. Walters, "The Negative Attitudes towards Robots Scale and Reactions to Robot Behaviour in a Live Human-Robot Interaction Study," 2009.

[5] K. Pitsch, B. Koch, "How infants perceive the toy robot Pleo: An exploratory case study on infant-robot-interaction," in *Proceedings of the Second International Symposium on New Frontiers in Human-Robot-Interaction (AISB)* , Leicester, UK, 80-87, 2010.

[6] K. Pitsch, H. Kuzuoka, Y. Suzuki, L. Süssenbach, P. Luff & Ch. Heath, "The first five seconds": Contingent stepwise entry into an interaction as a means to secure sustained engagement in Human-Robot-Interaction," in *ROMAN (IEEE International Symposium on Robot and Human Interactive Communication),* Toyama, Japan, 985-991, 2009.

[7] K. Pitsch, K. Lohan, K. Rohlfing, J.Saunders, C. Nehaniv & B. Wrede, "Better be reactive at the beginning. Implications of the first seconds of an encounter for the tutoring style in human-robot-interaction," in *ROMAN (IEEE International Symposium on Robot and Human Interactive Communication),* Paris, France.

[8] I. Lütkebohle, F. Hegel, S. Schulz, M. Hackel, B. Wrede, S. Wachsmuth & G. Sagerer, "The Bielefeld Anthropomorphic Robot Head "Flobi"", *2010 IEEE International Conference on Robotics and Automation, Anchorage,* Alaska, IEEE, 03/05/2010, 2010.

[9] F.A. Eyssel, D. Kuchenbrandt, S. Bobinger, L. de Ruiter, & F. Hegel, "If You Sound Like Me, You Must Be More Human: On the Interplay of Robot and User Features on Human- Robot Acceptance and Anthropomorphism," in *Proceedings of the 7th ACM/IEEE Conference on Human-Robot Interaction (HRI 2012),* Boston, USA, 2012.

[10] E. A. Schegloff, "A tutorial on membership categorization" *Journal of Pragmatics*, 39(3), 2007, 462-482.

[11] L. Süssenbach, "Motivation im Sport: Eine konversationsanalytische Rekonstruktion interaktiver Motivationsverfahren im "Indoor Cycling"", M.A. thesis, Fakultät für Linguistik und Literaturwissenschaft, Bielefeld Univ., Bielefeld, Germany, 2011, unpublished.

[12] Schegloff, E. A., "Opening Sequencing" In *Perpetual Contact: Mobile communication, private talk, public performance* J. E. Katz & M. Aakhus Eds. Cambridge: Cambridge University Press, 2002, pp. 326-385.

[13] H. Sacks, "Lectures on Conversation," Oxford: Blackwell, 1992.

Interactional Coordination and Alignment: Gestures in Indoor Cycling Courses

Luise Süssenbach, Karola Pitsch

Bielefeld University, Applied Informatics & Research Institute for Cognition and Robotics (CoR-Lab), Bielefeld, Germany, http://cor-lab.de

luise.suessenbach@uni-bielefeld.de, karola.pitsch@uni-bielefeld.de

Abstract

Interactions provide opportunities for coordination that can be enacted via gestures. Interactions in a sports setting (in our case indoor cycling classes) provide opportunities and demands for coordination that go beyond typical face-to-face interaction. In this paper, we investigate how a trainer motivates the trainees and with a focus on the methods that are applied. The approach presented here conceptualizes "gestures" as an interactional phenomenon rather than the mere utilization of specific body parts. Our analysis shows that pedaling in indoor cycling courses can be understood "gestures" getting interactional functions.

Index Terms: gesture, interaction, multimodal actions

1. Introduction

In general, research on the topic of "motivation" is undertaken by psychologists focusing on the individual level. However, motivation is also an interactive phenomenon. Such interactional motivation processes can be studied especially well in fitness courses, such as indoor cycling, which involve a trainer and several trainees. Indoor cycling is a form of exercise with classes focusing on strength, endurance and high intensity that involve using a stationary exercise bicycle with a weighted flywheel in a classroom setting and loud music. Due to the exercise, the participants' hand, arms and legs are engaged in the physical task and therefore not available for communicational purposes such as gesturing. Thus, this scenario poses a challenge for traditional gesture analysis that predominantly focuses on the communicative aspects of hand and arm movements [1].

In cycling classes the beat of the music and the participants' coordinating pedaling play a fundamental role for the analysis of gestures in this setting. In order to signal the pedaling rhythm and to support the trainees' coordination, the trainer introduces a set of resources such as music with a steady beat. Video recordings from such classes show that - in addition to this beat - some trainees need more support. In such cases, the trainer provides gestures: He starts pedaling himself to the rhythm of the beat (i) and he verbalizes his actions (ii). Thus, under specific conditions, part of the action is endowed with gesture-like qualities. Such a setting and practical task is interesting for gesture research: It suggests that the concept of "gesture" might not best be conceived of as an a priori defined construct that involves a number of well-describable semantic, semiotic or visual features. Instead, gestures appear as being constructed on-line, in the course of the action and fulfill a certain *interactive* goal. In this view, leg movements in indoor cycling classes may be conceived of gestures, if the participants' involved in the interaction understand them as such. The present paper aims at illustration (1) how the trainer establishes leg movements as gestures in a multimodal way and (2) how the meaning of leg movements is changed by the interpretation of the actors.

These questions and conceptual issues arise from an interdisciplinary project founded by the German Aerospace Center which aims at developing of a robot system that should motivate astronauts to increase the effectiveness of their daily fitness training. This is particularly important, because of the loss of muscular tissue due to long exposition to zero-gravity conditions, astronauts need to do sports for at least two hours a day. However, due to high workload this is not always possible. In the scope of the project, we investigate if a robot system may fill the role of a fitness trainer, when adequate modules for interaction and motivation are realized.

At a basis and inspiration for the design of the human-robot-interaction we investigate how human fitness trainers motivate their trainees during the training. The coordination of pedaling and the music's beat is one of the trainer methods.

2. Background

Previous research has often described gesture as a phenomenon that accompanies speech and which manifests itself in hand and arm movements. Whereas McNeill recognizes that gestures accompany speech as holistic and non-conventionalized, Kendon also covers conventionalized gestures with the concept of recurrent gesture. These are found in similar forms in specific contexts, but did not achieve lexical status. The concept of gesture in classical gesture research seems to be a non-dynamical construct describing specific movements of the hands and arms related to accordingly produced speech. Thus it is often conceived as isolated from interaction. The focus is thus on the producer of gesture. In this sense research has shown that and how gestures allow the speaker to economically describe spatial scenarios [7], location plans [8, 9] as well as abstract relations like functional hierarchies in the compound of sentences [10] in the gestural modality in gesture space.

Next to this classic conception of gesture research ethnomethodology and conversation analysis brought a holistic conception of non-verbal communication, which describes verbal and non-verbal communicative phenomena as diverse, equally valued resources of communication. In the last years concepts were suggested which tie to traditional concepts of gesture but view gestures in an interactive context as a holistic collaborative production [3]. In particular Goodwin's concept of an "ecology of sign systems" [12] has been influential. Instead of channels of behavior Goodwin speaks of „semiotic fields". These include syntactic structure, prosody, body posture, gestures which are embedded in a particular situation, state of participants, material structures in the environment. This combination of different signs plays a fundamental role in the constitution of interaction. The current combination of relevant resources is continuously changing during the interaction, so that specific „contextual configurations" emerge. By use of the concept of „ecology" Goodwin refers to the fact that different signaling systems may adopt different functions in the course of the interaction: „The term ecology is used to note the way in which these separate systems function as differentiated, interdependent components of a larger whole that can adapt to changing circumstances." [15]. His analysis on interaction expands the classic notion of gestures. He provides a framework for the collaborative constructions of meaning. Goodwin also shows that resources currently not available are replaced with other, available resources [12]. This dynamic

notion is relevant for the interactional setting investigated in this paper as it provides a framework to systematically describe the phenomenon that when certain communicational resources are not available (e.g., hands, arms, legs when cycling) their function might be taken up by other resources. Furthermore Streeck [13] suggests that certain gestures arise from everyday manual actions. This provides a basis to also think of pedaling on a bike as becoming - under certain conditions - a gesture.

The ethnomethodolgical tradition describes gesture as no definite concept of movements. Gesture is viewed as an interpretative category of interaction. This means that, in principle, every body movement could become a gesture, if the actors treat it as such. Starting with this consideration a differentiation between observable body gestures (facial expression, posture, body movement) and audible body gestures (articulation, prosody) [3]. Dausendschön-Gay and Krafft describe two functional areas of body gestures. The first functional area is assigned to face-to-face-situations following the assumption that as long as humans are in the same room, they behave. Through the intensity of gestural behavior they mark their status as being participants of this situation. Therefore the body gesture becomes a basic function of human interaction. The second function describes the processing of a form of an expression. The focus here also is on body gestures in relation to the joint production and assurance of comprehension. Gesture therefore describes no fixed concept or inventory of hand movements, but an interactive effort of all actors of an interaction to assure comprehension, which can be manifest in every body movement.

3. Study Design and Data

In order to investigate interactive strategies of motivation we are conducting a set of studies (09/2010 – ongoing) investigating the interaction between trainer and trainees in indoor cycling courses. These start from investigating authentic indoor cycling classes as they are carried out in the everyday life in fitness centres (corpus 1) and include – in a second step – a semi-experimental design to manipulate certain parameters and transfer the setting to the requirements of the envisioned human-robot-interaction (HRI) in the SocioRob-project (corpus 2).

Corpus 1 investigates every day group interaction in spinning courses as they occur in fitness centres. These group interactions comprise a trainer and several trainees (between x and y participants). Three different trainers are recorded in order to be able to abstract from potential personal differences in the communication strategies exhibited by diverse trainers.

Corpus 2 constitutes a semi-experimental design and reproduces the situation of a personal training (1:1). As – in the future – our robot system is supposed to interact with one individual trainee at a given time, we need to understand the particular differences between the group and the individual situation. This HRI adapted setting compares effects of the training situation (group vs. personal training) and investigates effects of different training methods and situations respectively. The setting contains one trainer and one trainee, with five consecutive training sessions per trainer and participant. Within these five sessions three training courses are recorded which only differ from a normal session in the number of participants. The goal of the sessions three and four is to manipulate two independent variables (availability of trainer bicycle, rhythm and music respectively), which are motivated by context of Human-Robot-Interaction: in session 4 only the trainee uses a bicycle which enforces the trainer to design the training on basis of different interactive resources.

4. Method

In a first Study data of three trainers in every day group situations was acquired. The data material covers video and audio material of approx. 180 minutes. In Study B (personal training) data of five training sessions with a total of 300 minutes was recorded with two video cameras so far.

Our analytic approach is qualitative being based upon Conversation Analysis [14]. Conversation Analysis describes the structure, orderliness and sequential patterns of interaction and the micro-coordination between the actors. This mixed approach enables to start with explorative, in-depth qualitative analysis of a small collection of cases drawing on ethnomethodology and conversation analysis to detect analytical issues and phenomena.

In a first step we produced a transcript of the trainer's and trainee's speech and body movement. Not only does the collected data show that annotation of activities (drinking, removal of sweat) is necessary, but also each pedaling is important to analyse. Pedaling and its frequency in relation to the rhythm of the music is relevant in all the collected data on the level of interaction. A systematic annotation of pedaling by the trainer as well as by the participant and the annotation of the rhythm of the music is required (Fig.1).

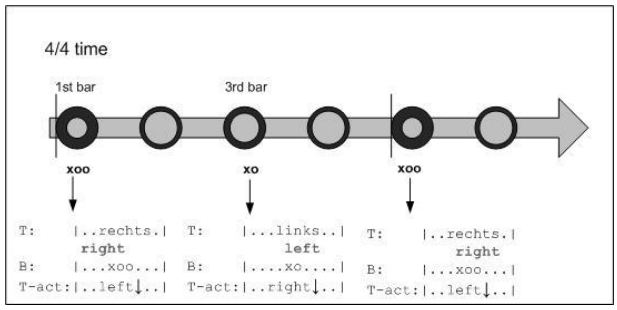


Fig.1: Annotation of multimodal actions to the beat

Figure 1 displays the pattern of intonation (i.e., the beat: 4/4 times) in relation to the multimodal acts of the trainer. On the first bar (noted here in an additional tier B as: xoo) the trainer says: "rechts" (right in German). Simultaneously to the exclamation "rechts" he pedals with the left foot. Consequently speech and act of the trainer are not congruent, but adapted to the perspective of the exercising participants.

5. Alignment and Coordination

5.1. Rhythm as resource for interactional coordination

Coordination plays a fundamental role in interaction in general as well as in fitness interaction like in indoor cycling classes. During indoor cycling the athletes not only coordinate their behavior to the trainer and other athletes but also to the beat of the music. Thus, this setting requires coordination of all persons involved. To do so, the trainer has a variety of resources at his/her disposal, such as the beat of the music, the simultaneous paddling and verbal utterances.

Let us consider the following fragment taken from one of the authentic group spinning courses (corpus 1, about 32 sec.): At the beginning of this exercise the trainer establishes the beat of the music as a shared attention in a multimodal way and therefore makes it relevant on the level of interaction. Not only the verbalization "just listen" (l.01) but also the timely related pointing gesture towards the ceiling which points to the

music in a metaphorical way, he establishes the relevance of the rhythm in this part of the exercise.

24:31.150

01 T: <<all> GEnau hinhörn? > (-)
listen carefully
02 T-act: |……iconic gesture……..|
03 GANZ GANZ einfach
very very simple

Through his pointing the trainer suggests the music as a focus of shared attention to the participants. Then, consecutively, he delivers a turn on every beat XO and XOO (once right once left; l.04), which refers to the rhythm of the pedaling. This method provides an orienting device for the participants' to synchronize their pedaling with those of the group and the rhythm easier. At the beginning of an exercise those verbal-rhythmic turns are repeated on every relevant beat XO and XOO, but are semantically and syntactically simplified during the progress of the exercise.

In the beginning the specification of the treading foot to each relevant beat XO and XOO by the trainer can be recognized, i.e. his left foot is treading on beat XO and his right foot on beat XOO. Verbal rhythmic advices are also verbalized simultaneously (l. 04, 05).

04 T: <<all> einmal rechts> ↑EINmal links (-)
one right **one left**
05 B: |……xo…….| |…….xoo…….|
06 T-act: |.....left↓…..| |…….right↓.....|
*Fig.2 *Fig.3

Making the display of the tread action relevant for the co-participants is also revealed in a particular the change of perspective: The trainer treads with the left foot, but verbalizes the right one, because of the training situation which is also defined by the alignment of the persons in the room. Spinning training is a classic sport, which is mediated in the training situation by a front alignment of trainer to participant(s) so that the trainees can directly align with and imitate what they observe.


Fig.2:" one right"


Fig.3: " one left"

These verbal-rhythmic devices change systematically during the progression of the exercise. While the trainer mediate *every* movement of pedaling to the participants at the begin of the session, he constitutes the rhythm only through *one side* of treading during the progression of the exercise (right right right right, l.07) until he finally uses any exchangeable word (hop, l.12).

[…] 25:00.800

07 T: <<all> RECHTS- RECHTS- RECHTS-
right right right
08 B: |.....xo.....| |.....xo.....| |.....xo.....|
09 T-act: | …..left↓…...| |….left↓. | |….left↓...|
RECHTS- RECHTS- >
right right
10 B: |.....xo.....| |.....xo....|
11 T-act: |….left↓ ….| |…left↓…. |

[...]
12 T: HOPP HOPP
hop hop
13 B: |..xo...| |..xo...|
14 T-act: |.left↓.| |.left↓|

During this systematic reduction of announcements the trainer permanently observes the group which mostly obeys the tact standard prior established.

5.2. Pedaling: An interactionally relevant gesture

The coordination of pedaling and beat poses a practical problem to the participants. In the case of non-coordination the difficulty involved in this task becomes apparent. In such a case, the trainer has to intervene, and subsequently repair strategies can be observed that go along with a change of the trainer´s gaze orientation.

In the process of the sequence it becomes clear that these verbal-rhythmic devices are not only understood as adressed to the entire group, but also as personally addressed towards one participant. While the trainer pays no attention to gaze organization in the beginning and watches all participants equally often with an equal duration in the average, he starts to watch participant S2 more often later in the exercise and restricts his field of gaze until eye contact is finally established. This focusing is carried out because participant S2 does not pedal synchronously to the rhythm of the music and to the homogenous pedaling of the group (s. Fig.4).
With continued practice (and an according difficulty increase) there is an increased demand for coordination. Participants now not only have to pedal to the beat but furthermore need to pedal standing up versus sitting down in a certain rhythm. S2 is the participant who has difficulties with this new task, which becomes particularly visible in the data comparing her body movement with that of her neighbors (as suggested in fig. 4).

[…] 25:45.100

Trainee S2 isn't pedaling in rhythm *Fig.4

15 T: up (.) UP (-)
16 B: |..xo..| |..xo..|
17 T-act: |left↓| |left↓|
18 S-act: |right ↓| |right↓|
S2


Fig.4: Trainee is not pedaling in rhythm

During the instruction "yeah yeah yeah (.) stay in rhythm" (l. 18) eye contact between trainer and participant takes place. The fact the participant adjusts her tread rhythm immediately after the trainer's turn shows her interpreting herself as the addresser of his advice.

18 T: <<cresc> ja: jaja ja:> (-) <<ff> bleibt im rhythmus?>
yeah yeah yeah **stay (plr.) in rhythm**
19 T: |@S2|
29 S: |.....@T.....|

(3.0) Trainee S2 is pedaling in rhythm *Fig.5

Fig. 5: Trainee S2 is pedaling in rhythm

In this phase the trainer uses a repair mechanism through a change in his gaze organization. Because of the trainer's eye-contact the trainee feels addressed by his announcement and synchronizes her movements to the group's homogenous pedaling rhythm of the group (Fig.5).

5.3. Body movement: An interactionally relevant gesture

If repair is not possible via gaze organization, the trainer employs other strategies.

After a couple of beats the participant S2's pedaling rhythm again is not synchronous to the music and hence not to the groups's rhythm. The trainer reacts upon it by focusing the trainee again and repeating his verbal-rhythmic advices (one more right (.) left, l. 20).

Trainee S2 is pedaling in rhythm, but losing it
20 T: noch EINe (-) <<p> RECHTS (.) LINKS.
one more **right** **left**
21 B: |.....xo.....| |....xoo...|
22 T-act: |……..right ↓……..| |..right↓.|
23 S-act: |…….right ↓……| |…..left↓……..|
24 T-gaz: |@S2|

However, the trainee is still not able to align her movements during the following bars. (Fig.6).

Fig.6: Trainee S2 is losing rhythm

As a reaction the trainer gets off his bike. Once standing he turns towards the group and constitutes the rhythm verbally once again. It is interesting to observe that he now uses his hands as a means to indicate the pedaling rhythm. This way, he manages to realize an optical indicator even though he is not sitting on the bike pedaling (l.25, l.27). Since this gesture achieves the same effect as the pedaling it provides an additional functional argument - including the participant's perspective - that the action of pedaling itself can constitute a gesture.

Trainee isn't pedaling in rhythm
Trainer is leaving his bike

25 T: RECHTs. (.) LINKs.
right **left**
26 B: |.....xo.....| |....xoo...|
27 T: |.fist left..| |fist right|
*Fig.7 *Fig.8

Fig. 7: "right"

Fig.8: "left"

The trainer then walks towards the trainee who is not synchronized. Arriving at the trainee he adjusts the resistance of her bike. At the same time he constitutes the rhythm via "right (.) left" (l. 30) once more. After a few beats the trainee is able to synchronize to the rhythm indicated by the trainer (l. 32).

Trainer walks towards Trainee S2

28 T: Rechts. (.) Links
right **left**
29 B: |...xo..| |.xoo.|

Trainer manipulates Trainees adjustabler resistor,
is leaving S2

30 T: RECHTs. (.) LINKs. (.) Genau.
right **left** **correct**
31 B: |.....xo....| |…xoo..|
32 S2: |..right ↓.| |.right↓.|

Depending on their state of participation the other trainees may utilize the trainer's additional advice for trainee S2 as an offer for self correction purposes. After adjusting the resistance he steps back again in order to gain an overview of the whole group (s. Fig.9). By actively intervening the trainer managed to establish a homogenous rhythmic synchronization. After observing the group for about six seconds he remarks "better better" (l.33). This remark displays an explicit praise and reveals the relevance of a homogenous pedaling rhythm.

Trainer is watching the group for about 6 seconds. All trainees are pedaling correctly respecting the rhythm.

33 T: << BESSer. BESSer.>
better better

34 B: |...xoo...| |...xo...| |...xoo...| |...xo...|
35 S2: |..left↓..| | right↓| |...left↓.| |right↓|

*Fig.9

Fig.9: Trainer is observing the group

The analysis shows that the ascribing of the gestural status to a multimodal action takes place on an interactional level. Therefore, the gesture can be considered as the result of an effort all actors of the face-to-face-interaction are involved in. In conclusion this means that the multimodal action of pedaling will achieve gestural status if an actor in the interaction ascribe this status to it and treat it as a gesture.

5.4. Production of Meaning: Replacement of pedaling gesture

As a first step, we presented results from the qualitative analysis taking a close look at the way a trainer is doing a indoor cycling course. The analysis shows that the coordination to music is an essential element of spinning. Those coordination activities do not merely take place on a macro level (e.g. the coincidence of low bpm count low cadence) but we can furthermore observe local coordination activities. The trainer constitutes the desired pedaling rhythm with the help of the music's beat. If a trainee is not able to pedal in the established rhythm, specific repair strategies initiated by the trainer can be observed. We noted that the trainer mobilited other resources (e.g., he used his hands) if he had no bike to demonstrate the rhythm. To investigate such strategies more systematically, we conducted a second study in which we varied whether personal trainers had versus did not have a bike at their disposal.

Let us consider the following fragment: In this situation we can see that the trainer is not only pedaling. He coordinates his pedaling moves to the music's rhythm. On a beat xo he pedals with his left foot. This observation can be made throughout the whole data. This mere coincidence of body movement and beat would not be referred to as a gesture yet. It is the fact that the trainer takes part in a social situation, a face-to-face-interaction, which puts this phenomenon into a different context. Additionally in this face-to-face-interaction the trainer makes a verbal announcement ("step") while pedaling to the relevant beat.

Fragment 1

01 T: TRITT- TRITT-
step step
02 B: |...xo.....| |...xoo..| |...xo.....|
03 T-act: |...left ↓..| |right ↓| |..left ↓..|
*Fig. 10 *Fig.11 *Fig12

Fig.10: 34:24.40

Fig.11: 34:25.102

Fig.12: 34:26.35

The importance of the pedaling display through the trainer's pedaling shows the following fragment: in this situation the same trainer does not have a bike. We can observe that he tries to replace the marked leg actions through his arms and hands. So he can recoup a missing interactive resource through another one.

Fragment 2

01 T: TRITT- TRITT-
step step
02 B: |...xo.....| |...xoo..| |...xo.....|
03 T-act: |...left ↓..| |right ↓| |..left ↓..|
*Fig. 13 *Fig.14 *Fig.15

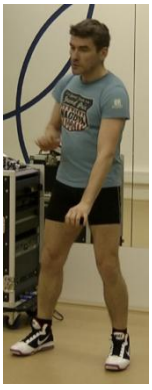


Fig.13: 30:04.40

Fig.14: 30:05.12

Fig.15: 30:05.92

This analysis leads to the conclusion that pedaling to the music's beat together with the trainer's verbal-rhythmic devices, constituting the rhythm, that make the pedaling movement a multimodal action and ascribe the status of an interactionally relevant gesture to it.

6. Results and Implications

In this paper, we have presented the first results of the analysis of a study aiming at identifying interactive strategies for motivation.

We observed that the pedaling movement combined with the simultaneous announcement of the foot to pedal with constitutes an offer for coordination that can have a motivational impact on the trainees. On an interactional level we could observe that the actors in this sports interaction ascribe a gestural status to the pedaling movements of trainer as well as the trainees. The relevance of this communicated gesture becomes describable on an interactional level especially when

an athlete is not able to carry the pedaling movement demanded to the beat into execution: The trainer gets off his spinning bike and adjusts the resistance of the trainee's bike. At the same time he gives verbal-rhythmic advice to the other trainees as well but this time it is not realized via his pedaling moves but through other communicational resources: his hands.
This observation leads to two conclusions:

1) The gestural status of a body movement does not depend on the utilization of specific body parts. Only if the body movement takes place in a social situation and results in a multimodal action it may be described as a gesture. The mere pedaling e.g. is not a gesture. It is the social situation - the interaction - in combination with the synchronicity of the pedaling to the music's beat together with the trainer's verbal-rhythmic advice, constituting the rhythm, that make the pedaling movement a multimodal action and ascribe the status of a gesture to it.
2) The ascribing of the gestural status to a multimodal action takes place on an interactional level. Therefore, the gesture can be considered as the result of an effort all actors of the face-to-face-interaction are involved in. In conclusion this means that the multimodal action of pedaling will achieve gestural status only if all actors in the interaction ascribe this status to it and treat it as a gesture.

So, what is gesture? Gesture is not limited to extremities such as arms and hands. Every kind of multimodal action can achieve the status of a gesture and accomplish the functions associated with this status [13, 3, 11]. With indoor cycling this especially holds true for the pedaling moves. Since the trainer establishes the pedaling moves multimodally they can be seen as multimodal actions. This form of multimodality (body movement, speech, music's beats) [12] shows that for the analysis of gesture not only speech has to be taken into account but also external factors such as music which have an influence on interaction and hence become part of it. The interaction and its associated interactive negotiation processes are essential for the definition of the term gesture. Gesture is what actors in an interaction treat as such. Interactive negotiation processes not only take place between trainer and trainee but also among the trainees. We can often observe that a trainee who is not able to synchronize to the pedaling rhythm not exclusively utilizes the trainer's movements for orientation but also their neighbors' movements as well as their interpretation of the audible beat. Further research examining this phenomenon has to be undertaken. The role of the beat in the accompanying music and its function as an external time base must be the subject of further analysis as well. We expect to gain eminently insightful information by the analysis of training courses where the trainer did not have music as an auxiliary means.

What does that mean for a robot system that is supposed to act as a fitness trainer both everyday situations and in the aerospace mission? A robot system would need to monitor the coordination efficacy of the participants (i.e., it needs to realize when pedaling is versus is not in congruency with the beat) in order to fulfill the role of an indoor cycling trainer. However, before such a system could be help of the athlete, it would need to successfully coordinate itself. To do so, it would need to recognize the beat and coordinate its movement accordingly.

7. Acknowledgements

This project was granted by the German Aerospace Center with financial support of the Federal Ministry of Economics and Technology due to resolution of the German Bundestag by the support code 50 RA 1023. Special thanks go also to the UniFit Bielefeld and Tennisland Dornberg.

8. References

[1] McNeill, D. "Hand and Mind", IEEE Chicago/US: University Press, 1992.
[2] Mondada L. "Pour une linguistique interactionelle", in Acta Romanica Basiliensia 8, 113-128, 1998.
[3] Dausendschön-Gay and U., Krafft , U.,"Text und Körpergesten", 2002.
[4] Müller, C., "Redebegleitende Gesten: Kulturgeschichte, Theorie, Sprachvergleich, Berlin, Berlin-Verlag A. Spitz, 1998.
[5] Schegloff, E.A. "On some gestures' relation to talk", in M. Atkinson, J. Heritage (Eds.), "Structures of Social Action. Cambridge: Cambridge University Press, 266-286, 1984.
[6] Kendon, A. "Gesticulation, quotable gestures and signs", in M. Moerman, M. Nomura (Eds.) Culture embodied, V.27, 53-77, 1990.
[7] Sweetser, E. "Looking at space to study mental spaces: Co-speech gesture as a crucial data source in cognitive linguistics", in M. Gonzales-Marquez, I. Mittelberg, S. Coulson, M. J. Spivey (Eds.), "Methods in Cognitive Linguistics, 201-224.
[8] Fricke, E. "Origo, Geste und Raum. Lokaldeixis im Raum." Berlin und New York, de Gruyter, 2007.
[9] Fricke, E. "Grundlagen einer multimodalen Grammatik des Deutschen: Syntaktische Strukturen und Funktionen." Manuskript. Habilitationsschrift Europa-Universität Viadrina Frankfurt Oder, 2008.
[10] Mittelberg, I. "Metaphor and Metonymy in Language and Gesture: Discourse Evidence for Multimodal Model of Grammar. Unpublished Doctoral Thesis, Cornell University, Ithaca, New York, 2006.
[11] Pitsch, K., Vollmer, A.-L., Fritsch, J., Wrede, B. Rohlfing, K. J., and Sagerer, G. "On the loop of action modification and the recipient's gaze in adult-child interaction", in GESPIN, Poznan, 2009.
[12] Goodwin, C. "Gesture, aphasia, and interaction", in D. McNeill (Ed.) "Language and Gesture: Window into Thought and Action", Cambridge: Cambridge University Press, 84-98, 2000.
[13] Streeck, J. "Gesturecraft. The Manu-facture of Meaning" Amsterdam: John Benjamins Publishing Company, 2009.
[14] Atkinson, J. M. and Heritage, J. (Eds) "Structures of Social Action: Studies in Conversation Analysis", Cambridge: Cambridge University Press. 1984.
[15] Goodwin, C. "The Semiotic Body in its Environment", in J. Coupland & R. Gwyn (Eds) "Discourses of the Body", New York, 19-42, 2003.

Interaction and Motivation in Fitness Communication

Luise Süssenbach, Karola Pitsch, Franz Kummert
CoR-Lab & Applied Informatics, Bielefeld University, Germany

Background: Project "SocioRob"

Research Goal:

+ To develop a robot system that motivates astronauts to increase the effectiveness of their daily fitness training
+ To investigate if a robot system may fill the role of a fitness trainer, when adequate modules for interaction and motivation are realized
+ To design the human-robot-interaction we investigate how human fitness trainers motivate their trainees during the training

Research Loop:

+ Interactional procedures of Human-Human interaction as an inspiration for the design of Human-Robot interaction [1]

Research Challenges:

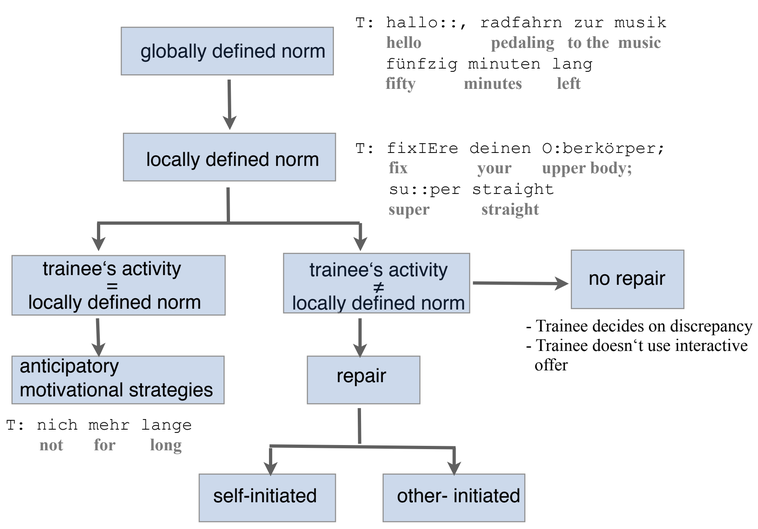
+ How can we deal with motivation on the level of sequence analysis?
+ What are the specific motivational strategies?
+ How does motivation work on the level of interaction?

Fitness Interaction: Indoor Cycling

+ Workout combining endurance, strength, intervals and high intensity
+ Exercised in a classroom setting
+ Special stationary exercise bicycle with flywheel
+ Trainer is positioned in the centre (s. spatial arrangement)
+ Music to allow for rhythmic coordination

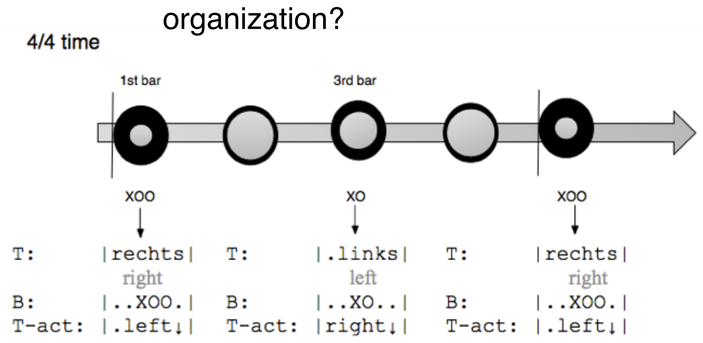
An interactional phenomenon: Motivation

+ Psychological topic vs. interactional phenomenon
+ **Challenge:** How can we deal with motivation on the level of sequence analysis? [2]



Indoor Cycling: Annotation

+ **Challenge**: How can we render and annotate the beat of the music as a part of the interactional organization?



Interactional Strategies for Motivation:

Gaze Organization - The Individual in the Group

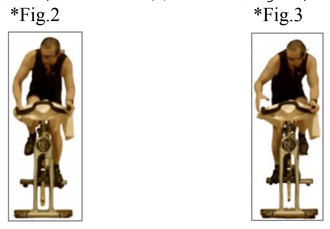

scene cameras

```
24:31.150

01 T:       <<all> GEnau hinhörn? > (-)
                   listen carefully
02 T-act:  |……iconic gesture……..|
03         GANZ GANZ einfach
            very  very  simple
04 T:       <<all> einmal rechts> ↑EINmal links (-)
                    one right          one left
05 B:       |.........xo.........||.......xoo......|
06 T-act:   |.......left↓........||......right↓....|
```

*Fig.2 *Fig.3

locally defined norm



```
[…] 25:45.100
Trainee S2 isn't pedaling in rhythm *Fig.4
```

S2

```
15 T:       up (.)      UP (-)
            up          up
16 B:      |..xo..|     |..xo..|
17 T-act:  |left↓|     |left↓|
18 S2-act: |right ↓|    |right↓
19 T:      <<cresc> ja: jaja ja:> (-)
                yeah yeah yeah
20          <<ff> bleibt im rhythmus?>
                stay (plr.) in rhythm
21 T:                  |@S2|
22 S2:      |.........@T.........|
```

trainee's activity ≠ locally defined norm

other- initiated repair

(3.0) Trainee S2 is pedaling in rhythm *Fig.5

trainee's activity = locally defined norm

S2

+ **Locally Defined Norm**
The trainer establishes the relevance of the rhythm through the verbalization „just listen" (l.01) and the timely related pointing gesture toward the ceiling which points to the music in a metaphorical way.
+ **Trainee's Activity ≠ Locally Defined Norm**
With continued practice there is an increased demand for coordination (not only pedaling to the beat, but furthermore pedal standing up vs. sitting down in a certain rhythm. S2 has difficulties with the new task (Fig.5):
Trainee S2 is not pedaling in rhythm
+ **Other-initiated Repair**
During the instruction "yeah yeah yeah (.) stay in rhythm" (l. 18) eye contact between trainer and participant takes place. The fact the participant adjusts her tread rhythm immediately after the trainer's turn shows her interpreting herself as the adresser of his advice.

Interactive Motivational Strategies:

Gaze Organization, Body Movement & Physical other-initiated Repair

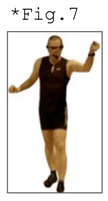
```
Trainee S2 is pedaling in rhythm, but losing it
20 T:          noch EINe (-) <<p> RECHTS (.) LINKS.
                one more              right      left
21 T:         |.............left↓.........||..right↓.|
22 B:         |.............xo.............||...xoo...|
23 T-gaz:                  |@S2|
Trainee isn't pedaling in rhythm
24 B:         |left↓||right↓|
25 S2:            |left↓||right↓|
```

trainee's activity = locally defined norm

trainee's activity ≠ locally defined norm

```
Trainer is leaving his bike

23 T:          RECHTs.          (.) LINKs.
                right                left
24 B:         |......xo......|    |....xoo...|
25 T:         |...fist left..|    |fist right|
                *Fig.7             *Fig.8
```



```
Trainer walks towards Trainee S2

26  T:     Rechts. (.) Links
              right        left
27  B:     |...xo..|     |.xoo.|
```

Trainer manipulates Trainees adjustabler resistor, is leaving S2

other- initiated repair

```
28 T:      RECHTs. (.) LINKs. (.) Genau.
            right       left        correct
29 B:     |...xo..|    |..xoo..|
30 S2:    |...xo..|    |..xoo..|
```

trainee's activity = locally defined norm

Trainer is watching the group for about 6 seconds.
All trainees are pedaling correctly according to the rhythm.

trainee's activity = locally defined norm

```
31 T:      << BESSer. BESSer.>
             better   better

32 B:    |...xoo...| |...xo...| |...xoo...| |...xo...|
33 S2:   |..left↓..| | .right↓| |...left↓.| |.right↓.|
```

+ **Trainee's Activity ≠ Locally Defined Norm**
After a couple of beats trainee S2 loses rhythm again
+ **Other-initiated Repair**
Arriving at the trainee S2: The trainer adjusts the resistance of her bike and he establishes the rhythm via "right (.) left" (l. 26) once more. After a few beats the trainee is able to synchronize to the rhythm indicated by the trainer
(l. 30) Depending on their state of participation the other trainees may utilize the trainer's advice for trainee S2 as an offer for self correction purposes.

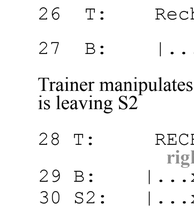
Discussion & Implications

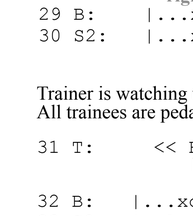
To fulfill the role of an indoor cycling trainer

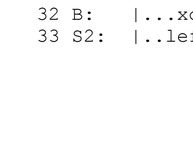
+ A robot system has to recognize the music's beat
+ It would need to monitor the coordination efficacy, so it needs to realize when pedaling is vs. is not congruent with the beat
+ It needs to coordinate its movement accordingly
+ It would need to give verbal-rhythmic devices to the beat („right")

Interactional procedures in a personal training scenario

+ How do these interactive motivational strategies work in a personal training set-up?







References

(1) K. Pitsch, H. Kuzuoka, Y. Suzuki, L. Süssenbach, P. Luff and Ch. Heath. 2009. "The first five seconds". Contingent stepwise entry into an interaction as a means to secure sustained engagement. *IEEE ROMAN 2009*, 985-991.
(2) L. Süssenbach 2011 "Motivation im Sport - Eine konversationsanalytische Rekonstruktion interaktiver Motivationsverfahren im *Indoor Cycling*", Masterarbeit.

Acknowledgments

+ This project was granted by the German Aerospace Center with financial support of the Federal Ministry of Economics and Technology due to a resolution of the German Bundestag by the support code 50 RA 1023.
+ CoR-Lab
+ UniFit Bielefeld, Sportland Dornberg

contact: luise.suessenbach@uni-bielefeld.de