FP6-IST-002020

COGNIRON

The Cognitive Robot Companion

Integrated Project

Information Society Technologies Priority

D 3.2.1
Report on user study on the role of posture and positioning in HRI

Due date of deliverable: 31/12/2004
Actual submission date: 31/01/2005

Start date of project: January 1st, 2004 **Duration** : 48 months

Organization name of lead contractor for this deliverable:
Royal Institute of Technology (KTH), Numerical Analysis and Computer Science

Revision: Final
Dissemination level: PU

Executive Summary

This report describes the investigation of the role of posture and positioning in Human-Robot Interaction (HRI) as part of the spatial management and interaction behavior between a human user and a robot. The term posture is defined as human body attitude and the non-verbal communication often signaled through it. Furthermore the term "positioning" and its significance in human-robot interaction are being discussed. Findings from studies in human to human interaction studies and social behavior are taken to establish parameters for posture and positioning in HRI. Especially the "intimate", "casual-personal", "social-consultative", and "public" *interpersonal distances* by E.T. Hall and the *spatial arrangements* of F-formations by Kendon are discussed and evaluated for their potential relevance.

Targeted at the Key experiment 1, "Robot Home Tour", a user study was performed with 22 subjects to study the spatial organization between a robot and a user during interaction. The study used the Wizard-of Oz methodology, and is described here in set-up, execution, and data collection. The analysis methods that incorporate annotation of the audiovisual data is explained and presented with selected examples of results. A discussion of preliminary findings for the first five trials is presented, pointing out that the current categorization and format of descriptions will need to be discussed further.

In the next phase of the COGNIRON research we will bring the analysis to a conclusion and start working with the spatial management and interaction as part of a joint topological map-building scenario in user studies. Previously teleoperated or simulated functions will be exchanged for real robotic components where possible.

Role of posture and positioning in HRI

If humans encounter a robot in the physical world they will need to determine how close they will get to this robot or how close they will allow the robot to come. Keeping a certain distance towards the system might be a safety issue, or at least an expression of how much the robot can or will be trusted.

An engagement with this robot in a cooperative task or interaction exchange might require for both the robot and the human to take certain positions suitable to the task. These positions will need to be negotiated, initially upon approach, during the interaction itself, and upon leaving one another, i.e. reciprocal positioning will require the active monitoring and dynamic reaction to each others' movement changes. Handing over objects, controlling the robot system, helping users or other kinds of manipulations might also require touching one another. All these activities by the two partners engaging in an interaction require that the user feels comfortable and in control in understanding the robot and that the robot is enabled to understand the significance of its movements for the purpose and in the context of interaction with its human user.

Posture and positioning in HRI are a prerequisite to read one another's signaling through joint spatial management. It is used in parallel to other communication modalities like spoken utterances.

Developing a "robot companion" as aimed at by the COGNIRON research program requires an understanding of the posture and positioning in HRI, e.g. identifying different interaction styles in individuals, and the crucial spatial factors in how a robot ought to position itself within the personal space of a human (and vice versa), recognizing user's intent in terms of observed actions and motions, as well as communicating in spoken dialogue the handling of space in interaction.

To find the relevant features of such physical interaction between a robot and a user it is necessary to let real robots interact with users and analyze the interaction for the spatial features. The role of pos-

ture and positioning will thus need to be driven by the real-world experiences and the data collected during user trials of human-robot interaction.

Relation to the Key Experiments

The explorative user study conducted and described in this report used the COGNIRON Key experiment 1 (KE 1) “Robot Home Tour” as defining scenario description. Based upon a dialogue pattern developed by the University of Bielefeld a scenario was designed that incorporated the user showing a robot around in a living room like environment and teaching it places and objects. Findings from the user studies on the role of posture and positioning in HRI are also expected to inform the COGNIRON key experiment 2 (“Curious Robot”) as the interaction between a curious robot and a human will be shaped and governed by the spatial positioning towards each other and in relation to objects dealt with for example.

Contents

1 Introduction to Posture and Positioning in HRI 5

1.1 Definition and relevance of posture for HRI 5

1.2 Related research fields 8

2 Social Interaction Studies and HRI 8

2.1 Introduction 8

2.2 Findings from human to human interaction study 8

2.3 Hall's Interpersonal Distances 10

2.3.1 Interpersonal distances 10

2.4 Kendon's F-Formation System 11

2.4.1 F-Formation System Arrangements 12

3 Role of Posture and Positioning – A User Study 14

3.1 Introduction 14

3.2 Experimental set-up 14

3.3 Data collection 16

3.4 Data analysis 18

3.4.1 Dense Representations 18
3.4.2 Statistical description of interpersonal distances and spatial formations ... 24

4 Discussion 25

5 Conclusion and Future Work 29

References 31

User study on the role of posture and positioning in HRI

Helge Hüttenrauch, Anders Green and Kerstin Severinson Eklundh
{hehu, green, kse}@nada.kth.se

1 Introduction to Posture and Positioning in HRI

1. 1 Definition and Relevance of Posture for HRI

To address the meaning, characteristics, and importance of posture and positioning in Human-Robot Interaction (HRI) some background on the terms themselves, their embeddedness in different disciplines, and a delimitation of the "role of posture and positioning in Human-Robot Interaction will be helpful.

According to the Merriam-Webster dictionary [28] a *posture* can mean

> "the position or bearing of the body whether characteristic or assumed for a special purpose"

i.e. it defines both the position and pose of a (human) body, as either a permanent attribute of the body structure or as an aspect in time to be used and changed for special purposes, e.g. communicational weaving of an arm.

One fundamental but often neglected purpose of posture, i.e. the position and carriage of the limbs and the body as a whole [18], is to enable movement under the condition of gravity. With continuous postural adjustments the changing center of gravity is managed and literally counterbalanced during movements, except when lying down. This requires that humans continuously and unconsciously perceive themselves, i.e. notice information about one's position and posture in space and actively, but unconsciously react to it as part of controlling, e.g. locomotion without falling. The struggle of infants in *learning* to walk upright demonstrates that these skills need to be trained before becoming automated, i.e. posture management functionality it is not an ability that humans are born with.

Posture has been studied in different fields, e.g. in human communication or dance choreography. In human communication studies a differentiation between *verbal* and *non-verbal communication* is made [25]. Posture as part of non-verbal communication can be understood as *body language*, i.e. bodily "signals" that communicate different messages [31]. Body language as a whole includes gaze, gaze-exchanges, facial expressions (mimics), different types of gestures, body attitude and orientation with the body and limbs, muscle contraction and relaxation, touch, or a combination of these.

Different postures such as standing, sitting, lying on face or back, kneeling, etc. can be further subdivided according to the specific expression *how this is performed*, i.e. a singular "posture" holds different describing granularities and it can be difficult to agree on a "correct" level of description [2]. An illustrating example would be a description of a *sitting posture* that can be further qualified by describing the position of the legs as *crossed*, *straight*, or *resting on a chair*. On the next level could a description of a *sitting posture* with *crossed legs* be informed about the direction feet are pointing at, e.g. *inwards, outwards,* or *straight.* The different granularity levels to describe a posture or body movement makes it difficult to find one generic description of postures. Consequently attempts differ to define vocabularies of body language in choreography, clinical medicine, industrial time and motion analysis, and animation systems [8]. Another term, *kinesics*, established by anthropologist Ray Birdwhistell is often used to describe the body movements as part of the nonverbal communication [5].

The study of posture in HRI needs to take into account that the physical appearance is rarely fully visible as clothing might cover parts of the body to a varying degree [2] depending upon situational context. This might affect, e.g. a robot's sensing, or the differentiation of body parts in their exact posture. An example might be a long-tailored rain coat or a skirt that covers most of the lower body and will make differentiation of individual leg positions by e.g. laser tracking difficult.

Body language is rarely used alone, i.e. without the simultaneous use of other modalities like spoken language [1, 25], i.e. posture can be said to contribute to the *multimodality* of human expression in communication. Consequently posture can be seen not only as a static description of a singular combination of body and body-part expressions, but also as one of multiple, in parallel used modalities in communication. If used in the latter sense, posture has the characteristics of a signaling system, probably best described by a spatial movement sequence in time. Posture thus has the duality of a static condition on one hand and on the other hand the ability to describe a body motion that incorporates movements in physical space in a time-continuous and dynamic manner.

Posture can also function as a *reflection of status* or a *reflection of emotional state* [2], i.e. postures can have a meaning of expressing a social status or they can be used to inform about the current mood of a person exhibiting this posture.

Instead of *non-verbal* communication, Dautenhahn et al. [12] advise to use the term *visual communication* to more precisely express that the communicative aspect of posture is based upon the visual channel [ibid., p. 413]. This can be argued about as *non-verbal* communication also includes bodily contact or the *tactile* modality [2, p. 92], which is achieved through a body motion. Jens Allwood suggests instead that the primary modes of perception in a face-to-face communication are *hearing* and *vision* as the spoken message "will normally predominate, while bodily gestures provide additional information." [1, p. 115]

The aspects of posture in the physical space affect the discussion about *the role of position* in HRI. As *position* we can understand a defining *orientation*, *distance*, and *height* towards another interaction partner or object. The physical entities of position are in this way defining the direction, space towards, and level of height according to an external reference point, e.g. in the eye of an external observer or interaction partner.

"Posture" thus already includes part of "position", as introduced above. The orientation of a body defines the direction (or orientation) towards another interactor or object as part of a position description. E.T. Hall [22] coined the term *proxemics* to model and to explain meanings of postures, communicative actions, and distances in social behavior (see below for "Hall's distances").

A posture can be described without referencing to something or somebody else. Therefore it is possible to describe a posture without giving physical measurements of orientation, distance, or height towards an external reference point. As a consequence it is justified to talk about a posture *and* a positioning in HRI, well aware that "position" can be used to describe parts of a "posture" while at the same time, "posture" can neglect defining elements of a "position".

Benford and Fahlén [4] present a *spatial model* for group interaction and collaboration in virtual environments. While targeted primarily to the domain of Computer Supported Cooperative Work (CSCW) and the co-operation in virtual 3-dimensional environments it is general enough to be applied to HRI. According to the authors an interaction occurs in a *medium* – for HRI this is the real, 3-dimensional physical world. The *aura* of an object (or interaction partner) can be understood as interaction enabling and more or less directed attention space around an object/interaction partner, enabling interaction if two (or more) auras collide or overlap. Only if this "collision" of potential spaces is accompanied by a mutual *awareness* can an interaction take place according to Benford and Fahlén. This is described as a relationship between *focus* and awareness and *nimbus* and awareness. The more some-

thing is within one's *focus*, the more one can be aware of it, and vice versa, the more someone or something else is within my *nimbus* the more this person or artifact can be aware of me.

The relevance of this spatial interaction model of Benford and Fahlén in HRI relies in the consequences for controlling spatial-temporal interaction: Movement and orientation of a robot and a human automatically influence aura, foci, and nimbus and through this also the level of awareness in interaction. Additionally this model can be used to check for expressions of movement and orientation that have the objective to influence, manage, or support the aura, focus, or nimbus of this interaction. Another usage of the spatial model by Benford and Fahlén can be its utilization to define design requirements: In this sense it should inform interaction design how to enable potential interaction partners to become aware of one another. The level of mutual attentiveness can furthermore be used to ground an interaction and communication situation according to the level of *awareness* the interaction partners have of one another.

Awareness, as central term for CSCW [16] affects perception of events and actions in cooperation between two or more interactors. Drury et al. [17] differentiate different awareness definitions in the CSCW literature and point out the relevance of awareness even for HRI.

In interaction situations between humans it is assumed and generalized that both partners have equal ability to *generate* and *perceive* one another's posture and positioning changes. Whether the same is true for an interaction situation between a robot and a human is to large extent determined by the robot's capabilities. Posture and positioning also assume an *embodied* agent as a prerequisite, i.e. a body that is able to express postures and can take different positions. Another requirement is the ability to sense and interpret signals that express meaning through posture and positioning, and possibly, act upon these appropriately.

In implicitly assuming embodiment the study of posture and positioning in HRI can be related to the field of Artificial Intelligence (AI). The AI research community has realized more than a decade ago [6, 7] the importance and pre-requisite of embodiment for artificial intelligent systems, founding what has become known as the *embodied artificial intelligence* discipline or approach [29]. Robots qualify as embodied systems both as research platforms and targeted applications to test with.

Humans who are embodied are also known to have a *personal space* that they feel comfortable in. Depending upon context and familiarity of others, this space might differ; it can thus be understood as part of the *territoriality* of humans as social beings [32].

As humans and robots are not equal in embodiment, perception, "thinking", and learning, an *unsymmetrical* relationship in body language expressions and cognition needs to be assumed. As robots neither come in one universal form nor with one universally defined functional capability set, it will be difficult to generalize findings in the posture and positioning behavior of robots. It is also important to realize that robots do not need to be comparable to a human-like body or to human sensing capabilities: No (manipulator-)arms might be available, the robot might or might not show a head, if a head exists the extent to which mimic signaling is enabled needs to be determined; wheel-based or multiple legged-robots exist as compared to the two legs humans have. In short, the human-like body and limbs in number and form are only *one possibility* for a robot's design-space, not necessarily the only one possible.

This fact is mirrored in the arguments about the necessity or characteristics of *anthropomorphic* or *humanoid* robots [11, 24], the discussion about the "uncanny valley" in design of robots, i.e. the non-linear relationship between the level of familiarity and similarity of robots (to humans or animals) [13] or the different possibilities in selecting design and functional features [19], all of which give an idea about the spectrum and complexity of questions to be addressed.

As a consequence of the unsymmetrical relationship between a human and a robot it is necessary to treat a robot as a unique specimen. Its abilities and limitations in posture and positioning need to be specified and tested for empirically. Studies that investigate the posture and positioning in interaction between a robot and humans are thus limited to a specific robot and the (cultural) setting of the study itself. Findings are therefore to be treated on a case basis and are unsuited for general conclusions.

1.2 Related Research Fields

This short introduction to the role of posture and positioning in HRI has excluded certain other research domains and perspectives that also have an interest in studying this area. Some of these are briefly mentioned here for reference.

Representations of postural body-movements in systematically structured form can be found in *choreography*, notably the *Labanotation* and *Eshkol-Wachmann* notation are well known, see Badler and Smoliar [3]. With the purpose of finding adequate digital representations for human posture and movements the authors discuss alternative representation forms.

Another research perspective that will be deferred to later studies is that of *group behavior* in posture and positioning. As the most simplified case, one robot and one human who meet can be investigated. An extension to more than one robot or more than one person would transfer this interaction and communication situation into one that will be governed by behaviors of *group behavior* and *communication*.

As this work deals with the *user's role* in the posture and positioning in HRI, the *technical aspects* are not covered either in this report. The technical components needed to enable for example the robot to change its position, to react upon or enact postures, or the exact sensing capabilities are thus not covered.

2 Social Interaction Studies and HRI

2.1 Introduction

This section introduces some findings of *social interaction studies* and their possible relevance for studying the role of posture and positioning in HRI. The fields of Human computer interaction (HCI) and usability engineering have a long tradition of utilizing both findings from cognitive science, psychology, and sociology in the design of improved interfaces for human-machine interaction or for computer supported cooperative work systems. A related and equally multidisciplinary approach might be helpful for HRI for the attempt to design robotic systems and their interfaces that are perceived as cognitive in their (spatial) behavior and signaling.

2.2 Findings from Human to Human Interaction Studies

Human to human communication and social behavior are studied as part of the social interaction studies. The studies have as objective the understanding of behavior and communication between two or more humans (or between groups of animals, e.g. primates, from an ethological perspective). In trying to apply findings from social interaction studies to HRI it is important to be aware that *human behavior*, *human communication*, or *human expectations* are transferred to a machine. Such approaches of attributing or imitating *human-like intelligence* or *behavior* as compared to machines has been profoundly criticized in the AI domain [15] or questioned for socially interactive robots [33].

Reeves and Nass [30] suggested taking findings of social interaction studies and applying selective findings as explorative test cases and for inspiration to the domain of interaction with machines. Their methodology of checking for appropriateness of social behaviors in the design of behavior for machines (and media) is based upon their theory of "social response to technology", claiming that humans naturally react even to technical systems as if they would be social beings.

The interaction encounters discussed in the social interaction studies deal – besides others – with a type of human-to-human encounter best characterized as a *face-to-face* meeting: The encounter happens in a shared physical space and enables synchronous communicative exchanges [14] between at least two partners, i.e. the communication and interaction is neither remote nor mediated technically.

In such a situation the (non verbal) communication can be regarded as partly unconscious, meaning that an exchange of signals in such an encounter is not necessarily a choice, but something that can not be avoided – it "simply happens". The extent to which this (bodily) signaling is more or less consciously conducted and perceived, possibly leading to a more elaborated exchange in communication or marking the beginning of a lengthened co-operation is dependent upon the exact circumstances. An example of this phenomenon can be the passing in a hallway in opposite directions without an exchange of spoken language utterances. If people see each other they can signal way in advance the planned for re-action of how the passing is to be handled without running into one another. This avoidance of bodily contact as goal is thus preceded by an exchange of signals that give the understanding to both how the situation is to be resolved spatially and in time. This can for example be done by one moving to the side of the hallway indicating that there will be enough room to pass each other.

As illustrated with this passing-in-a-hallway example, it can be assumed that even avoidance of spoken communication or observable forms of interaction in certain situations can be regarded as being based upon a previous, more or less unconscious, and nonverbal communication exchange.

Birdwhistell [5] believed that behavior of posture or bodily movements in relation to social and communicational processes can be understood and interpreted as an external visible and observable code which maintains and regulates relationships between humans. Goffman [20] proposed that the small behaviors of interactions to be studied to describe natural units of interaction to gain an *in situ* natural understanding of events that happen in encounters when people continuously exchange signals of behavior. This would aid the understanding of how "people routinely achieve order in their interactions with one another".

Three kinds of non-verbal communicational, body movement behaviors to be observed in face-to-face encounters were differentiated by Scheflen and Scheflen [31]: The *reciprocal exchanges* of kinesic behaviors are observable body movements or gestures that are displayed by both interaction partners as part of their joint exchange management. HRI still has to find these reciprocal supportive exchanges, i.e. what kind of movements by a robot can be expected to produce what kind of body movement responses by a user and signifying what intentions. Equally important are to find appropriate body movement responses by a robot, given that it observes certain movements or bodily expression by a human. The *territorial behaviors* allow or prevent the passage of people across a boundary and thus guide and frame the possibly for communication and interaction. For HRI this has a direct and important consequence: Studying the territorial behavior in humans interacting with robots could be a way to determine safety requirements on the robot or the robot motion behavior. It is also important to understand for a robot what kind of closeness to human interactors will be preferred in different situations. *Language-orientated* kinesic behaviors that can be regarded as part of the spoken communication are of interest to HRI as they support the verbalization of what is currently being talked about in a discourse between two interactors. A simple example would be a pointing gesture going along with the utterance, "Robot, this is a chair". Without the visual information of the pointing gesture the robot as listener would need to start searching for the object in question.

Humans are trained in social norms, taught to the young members of a social group [2]. For robot interaction behavior it remains an open question what social norms and rules robots should know and act upon in posture and positioning. Behavior in communicative and interactional encounters that are interpreted as orderly are said to be socially appropriate [25], i.e. they are characterized by being perceived as ordered affairs that go mostly unnoticed and are handled without consciously reflecting about them. The opposite are interaction behavior cases where interaction and communication breakdowns occur, e.g. by unsuited behavior that render a situation as socially impossible.

One of the difficulties of social interaction studies relies in finding the correct level of analysis, according to Kendon [ibid.]. Multiple levels of analysis (i.e. relevant units of analysis) might exist in parallel and thus make it difficult to find an *absolute unit of behavioral organization*. HRI studies on the spatial interaction behavior of robots have not yet agreed upon units of analysis and the attempt in this report on analyzing the observed human-robot interaction (see analysis section below) is in this respect to be treated as preliminary.

As a solution, Kendon [ibid.] advised to let the definition of a structural unit for analysis become apparent by the *next observable action* that can clearly be separated out from the stream of interaction events. By finding a clear beginning of another action the previous interaction event can be identified and classified.

In the next section two findings from social interaction studies, i.e. Hall's *social distances* and Kendon's *F-Formation system* are presented as potentially of interest to the role of posture and positioning in HRI.

2.3 Hall's Interpersonal Distances

Edward T. Hall [22] studied interpersonal distances and coined for his studies the term *proxemics*, i.e. "the interrelated observations and theories of man's use of space as a specialized elaboration of culture" [ibid, p.1]. In the human-robot-interaction context of posture and positioning, mainly three findings are of importance: The classification of interpersonal distances into 4 different classes, the realization of cultural differences in the spatial behavior of people from different countries, and last but not least man's perception of space.

2.3.1 Interpersonal distances

From his observations in the US, Hall concludes that social interaction is based upon and governed by four interpersonal distances: (1) *intimate*, (2) *personal*, (3) *social*, and (4) *public*. Each is further subdivided into "near" and "far", but that level of detail can be left to further discussions. The combination of measurable, spatial relationships, human ergonomic and kinetic capabilities, different social roles and interaction as well as typical characteristics and interaction situations make Hall's interpersonal distances interesting for HRI.

Intimate distance (range from 0 to 1.5 feet[1]): Humans can rely on smell and touch senses as vision is minimal (too close!). Examples of typical interpersonal activities/situations are given with making love, comforting others, i.e. intimate activities often including physical contact. Spoken communication is avoided and/or, almost involuntary, e.g. whispering. People need to be intimate with one another before this distance is acceptable to both partners.

Personal distance (range from 1.5 to 4 feet): While still possible, the importance of touch is reduced. Extremities can be used to get in physical contact with one another for a special purpose (handing over objects, shaking hands). The field of vision becomes broader (the angle widens), and one can com-

[1] 1 foot = 30,48 cm; in order to stay with the original distances given, the transformation into the metric system is not performed;

fortably have spoken conversations. The exact distance marks the level of familiarity of one person with another: Close friends are allowed closer than people one is less familiar with.

Social distance (range from 4 to 12 feet): this is the space used for more formal social interactions (i.e. one does not know each other or is acting in a more formal role due to the situational circumstances). The exact distance to one another and the surrounding noise-volume determine the voice volume in the conversation and to some extent the form of the conversation. As part of a more formal situation, the distance is often designed for, e.g. by conference-tables of a certain width (and people's sitting order). Hierarchies and social domination can be made explicit through varied distances.

Public distance (ranges larger than 12 feet): While communication (with raised voice) is still possible, interpersonal interaction ceases at distances larger than 25 feet. Speakers addressing an audience at distances > 25 feet only talk *to* the other part, but do not interact personally with the audience.

For a HRI exchange it might be postulated that the most interesting exchanges and reciprocal adaptations between a human and a robot will happen in the *social* and the *personal* distances. The *public* distance is of interest as this seems like an appropriate distance to perhaps try to signal that an exchange can or is about to happen. The *social* and the *personal* distance seem appropriate in theory to facilitate both the communication and the exchange of goods (for example the manipulation with a robotic arm). The *intimate* distance seems to be better suited for exchanges with, e.g. mental commit robots like the seal-robot Paro (Shibata et al., 2003 [34]), where touch is an intended interaction modality, resulting in the system giving off heat that can be felt.

2.4 Kendon's F-Formation System

Kendon's F-Formation system [25] is based upon the observations that people often group themselves in a spatial formation, e.g. in clusters, lines, circles, or other patterns. The term *formation* is used to express the dynamic aspect of this spatial arrangement, i.e. the need to actively sustain it during interaction. This takes often the observable spatial form of *small, well synchronized movements* of the participating interactors.

An *F-Formation* arises when two or more people form a shared space between them to which they have equal and direct access due to their sustained spatial and orientational configuration. The necessary behavioral organization and movement patterns which are used to sustain this F-Formation is called a *F-Formation system* [ibid, p.209].

The F-Formation system can directly be applied to the interaction encounter between a robot and human (see figure 1): Between the two a so called *transactional segment* or *o-space*, i.e. a space both to look and speak into, or in which they handle (shared) objects of interest is created and maintained.

Kendon claims that the *location* and *orientation of the lower body* is to be used to determine whether an o-space is maintained, i.e. the feet-placement to a large extent determines the actual o-space location and orientation.

This is somewhat counterintuitive to a notion of *focus of attention* which would instead specify an o-space location and orientation according to the direction of gaze, i.e. where the interactors are *looking*. Kendon defends his system by the possible *discomfort* which goes along with a sustained body orientation that differs from the lower body orientation. One can rotate one's head somewhat to look to the left or right of one's lower body orientation, but if one needs to look to a different direction than the one in which the lower body points over a longer period of time, the need to reorientate the lower body arises in order to stay comfortable.

The behavior of looking around and away from the o-space is called *facing out* in Kendon's system, i.e. if the interactor is "rotating his head so that a line projected from the midline of his face forms an angle of more than thirty degrees from the midline of his lower body" [ibid., p.212]

Figure 1: F-Formation system with illustrated transactional segment or *o-space* [semitransparent, white ellipse]

2.4.1 F-Formation System Arrangements

Some joint activities and spatial interactions are supported by certain F-Formation system *arrangements* according to [25], and are thus often encountered in prototypical situations. In the *vis-à-vis* arrangement (figure 2, top left) two participants normally face one another directly; a *L-arrangement* in which two contributors are positioned so that the frontal surfaces of their bodies fall on the two arms of an 'L' (see figure 2, top right) usually indicates a joint system in which something is shared in the o-space, e.g. an object of interest. The L-shape arrangement thus supports a *triadic* relationship between the two interactors and an object of shared interest.

As a last arrangement Kendon mentions the *side-by-side* configuration (figure 2, lower left) where the two participants are standing closely together and face the same way. This arrangement is said to occur often in situations were both interactors are face an outer *edge*, e.g. given externally by the environment in the form of a table, a wall, a kitchen sink, or similar. Another interaction situation that can be characterized by a side-by-side positioning of two interactors can be postulated as two people walking in parallel to one another and talking with one another. Kendon does not make this example explicitly, but there is no reason to exclude this situation from a side-by-side interaction categorization.

Note that the last situation illustrated in figure 2 (lower right), i.e. a so called *follow-me* situation is not a part of Kendon's system of arrangements. This situation has been added here to show a condition were people either stand in line, or to transfer the situation to the human-robot interaction domain, a

situation were a robot might follow a human user. However, it is doubtful that this depicted situation can qualify as an *arrangement* in the Kendon F-Formation system: The person being followed does not really have a continuous possibility to monitor in real-time what is happening behind her, thus the required *transactional* or *o-space* can not come into existence. Interaction is thus prevented between both parties as long as this formation is kept.

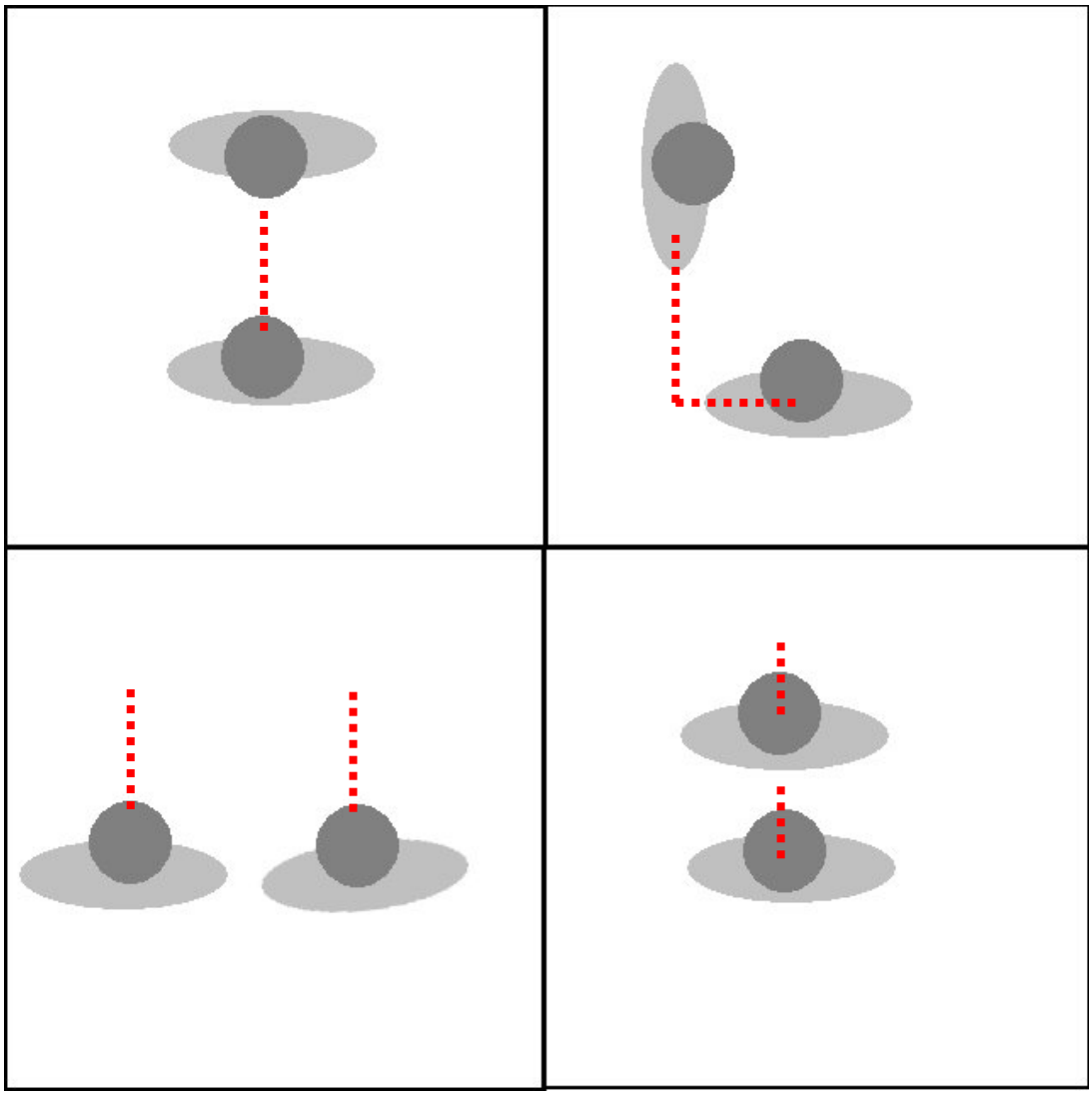

Figure 2: Two people depicted in F-formation arrangements, doted line indicating facing direction and/or direction of the interactional o-space;
Top-left: *vis-à-vis* arrangement Top-right: *L-shape* arrangement,
Lower-left: side-by-side arrangement; Lower-right: *Follow*-formation

In the following section we describe a study conducted to investigate the role of posture and positioning in a Home Tour Scenario.

3 Role of Posture and Positioning – A User Study

3.1 Introduction

While differing in the research questions aimed to study, the methodology, set-up, and data-collection that was used to study the role of posture and positioning in HRI is identical to the one described in the COGNIRON deliverable 1.3.1 on the "Evaluation Methodology of multi-modal dialogue" [10]. Thus the identical parts are here only briefly recapitulated in its main points.

In the Wizard of Oz research technique [21] a system that is not fully implemented can be simulated in parts, i.e. substituting missing system features by letting so called human *wizards* enact these functions as if they are already available. These simulated or *teleoperated* functions can then be presented to users of the system without disclosing the fact that not all system parts are actually implemented. The purpose of a Wizard of Oz study is to test interaction with a system. It is expected to make natural behaviors in trying a novel technical system observable. The role of posture and positioning as discussed above (section 1 above) is dependent upon the embedding in a natural interaction scenario with a high level of realism and believability.

For the human-robot interaction scenario of the COGNIRON key experiment 1, termed the "Robot Home Tour" [9], it is expected that a robot in a home-like environment needs to discover and compute an understanding of this environment. The initial learning can be achieved by the robot system through interaction with a human user who shows and names specific places or objects for the robot.

We combined the Wizard of Oz methodology and the interaction scenario of a Robot Home Tour: We invited trial users[2] and presented them with the task of teaching the robot initial places and objects in a living-room like environment (see figure 3 below). The emerging interaction between the robot and its user was expected to yield information about the posture and positioning used in the observed interaction.

3.2 Experimental set-up

The trial was based upon a scenario where a user has got a robot and is ready to use it for the first time. In order to make the robot familiar with the environment it needs to be shown around to learn places and objects that will be of interest to its operation. To ensure that the robot really has learned these important places and objects the user is also encouraged to test the robot about the newly learned artifacts and locations. This was done by encouraging users to send the robot on search or find mission to visit or find learned locations or objects. The task embedded in the scenario was thus for invited trial users to (a) get familiar with the robot and navigate it by letting it follow after him or her, (b) teach it places and objects, (c) validate already taught places and objects, and (d) handle interaction practically with a robot, including an initial opening and closing.

For the study of posture and positioning in HRI the modalities available for interaction and communication play a major role. The robot used in this study is an ActiveMedia Performance PeopleBot [www.activrobots.com]. It comes equipped with an on-board-camera with pan, tilt, and zoom capability. Trial users were told that this camera is employed by the robot for object and place recognition.

[2] The term "user(s)" can be discussed: People were invited to test a robot and interact with it based upon a scenario ("you got a new robot") in a setting characterized by a Wizard of Oz research environment. The purpose was that of *studying* interaction with a robot, thus beside the robot, a task to perform based upon a scenario, recording equipment was present. This qualifies the setting in itself as an experimental one – thus the term *subjects* might be appropriate. To stress that the setting at least aimed to be perceived as natural and as part of a Wizard of Oz technique the term "user" is adopted in this report, too.

They also were informed that the microphones placed upon the robot are used by the interactive speech system enabling the commanding of the robot by speech. A limited, initial vocabulary was provided, including a greeting, a follow-me initiation, a labeling of objects, and finally a closing phrase (for details on the speech dialogue system, cf. [10]).

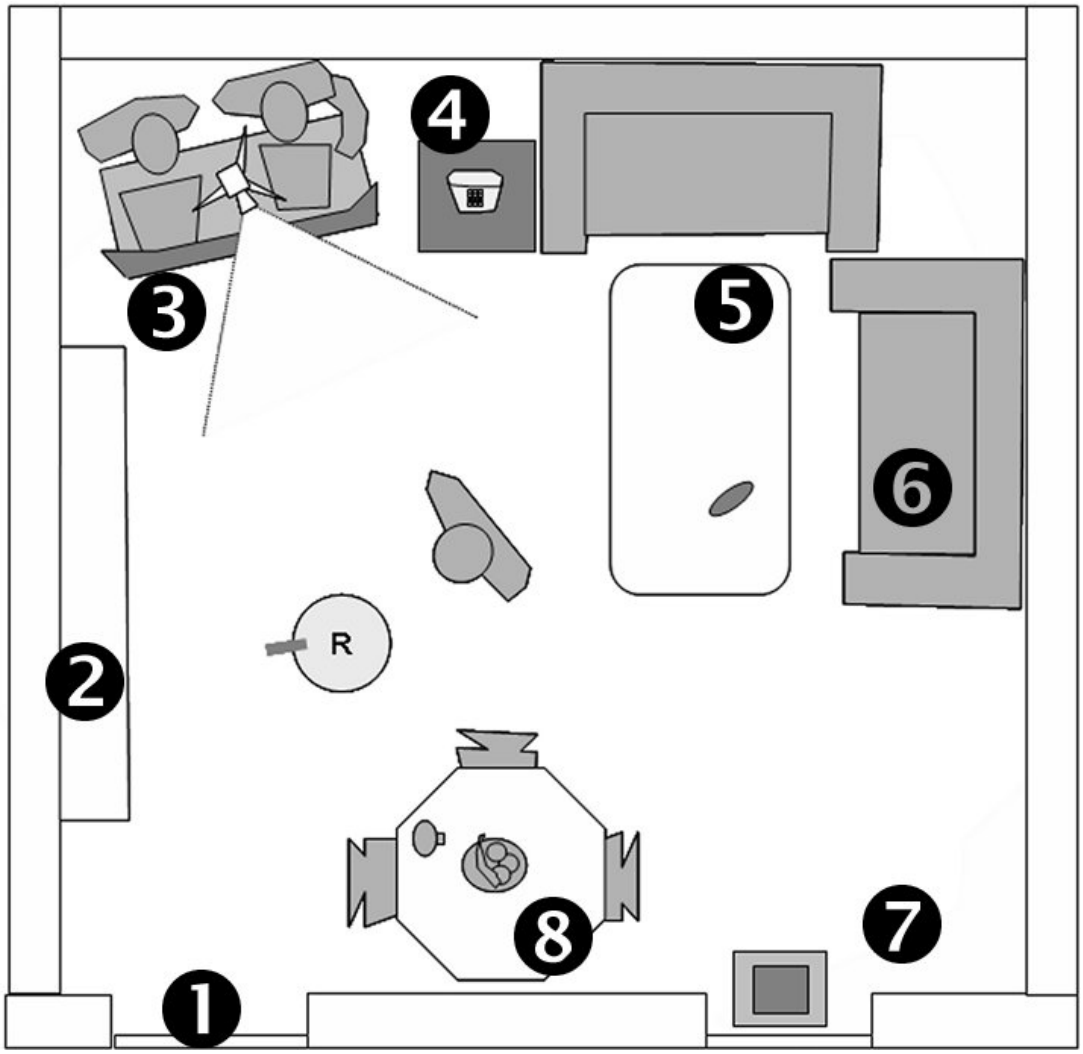

Figure 3: The CAS living room as experimental environment, see text for details

The trial was conducted in the so called CAS[3] living room at the Royal Institute of Technology in Stockholm, Sweden. A room five by five meter in size is furnished with IKEA living room furniture, including different tables, a bookshelf, and two sofas (see figure 3). Indicated with numbers are clockwise from the lower left hand corner ❶ the entrance, ❷ the bookshelf, ❸ the Wizard of Oz control station (with a DV-video camera), ❹ a small table with a telephone, ❺ a low coffee table upon which different objects, like a remote control and magazines were placed, ❻ a couch group, ❼ a TV and VCR combination on a small table, and finally, ❽ a small dining table with a fruit bowl, a coffee cup etc.

The trial subjects were recruited within the Royal Technical Institute of Technology, i.e. young technical students of both sexes. Requirements for selection were that that they *did not work or research in robotics* or *computer vision*, as this was judged to be the requirement of a robot encounter with novice

[3] CAS is the abbreviation for the Centre of Autonomous Systems, Royal Institute of Technology (KTH), see www.nada.kth.se/cas

and inexperienced users. We conducted 22 trials (after 4 initial pilot trials for trial-adjustments) with 9 women and 13 men. Participants of the study were rewarded a cinema ticket for their time and effort.

Upon arrival participants received an introduction to the robot and the task, both in written form and in the form of a short, but formalized demonstration of one of the experiment leaders. They were then free to use the robot to teach it new places and objects and validate these. Upon completion of the trial users were asked to fill in a questionnaire before they got debriefed about the simulated nature of the robot's behaviors in interaction and communication.

The actual robot behavior was teleoperated by two experiment leaders who used a wireless robot navigation and on-board-camera control for the robot and camera movement and a speech dialogue production tool to interact with speech [21]. For a thorough discussion of the Wizard of Oz method, see [10].

3.3 Data collection

Multiple data sources were captured and collected during the trial: An external video camera (DV) recorded the trial in audio and video from the experiment leaders' position and perspective. The room was furthermore surveyed by four webcams that were placed near the room's corners in order to ensure that user and robot movements, postures and gestures would be recorded independent on placement within the room (see figure 4). The images from the webcam were also intended to be used in the analysis of the spatial relationship and directional setting or formations between a user and the robot. This will be explained in the analysis section below.

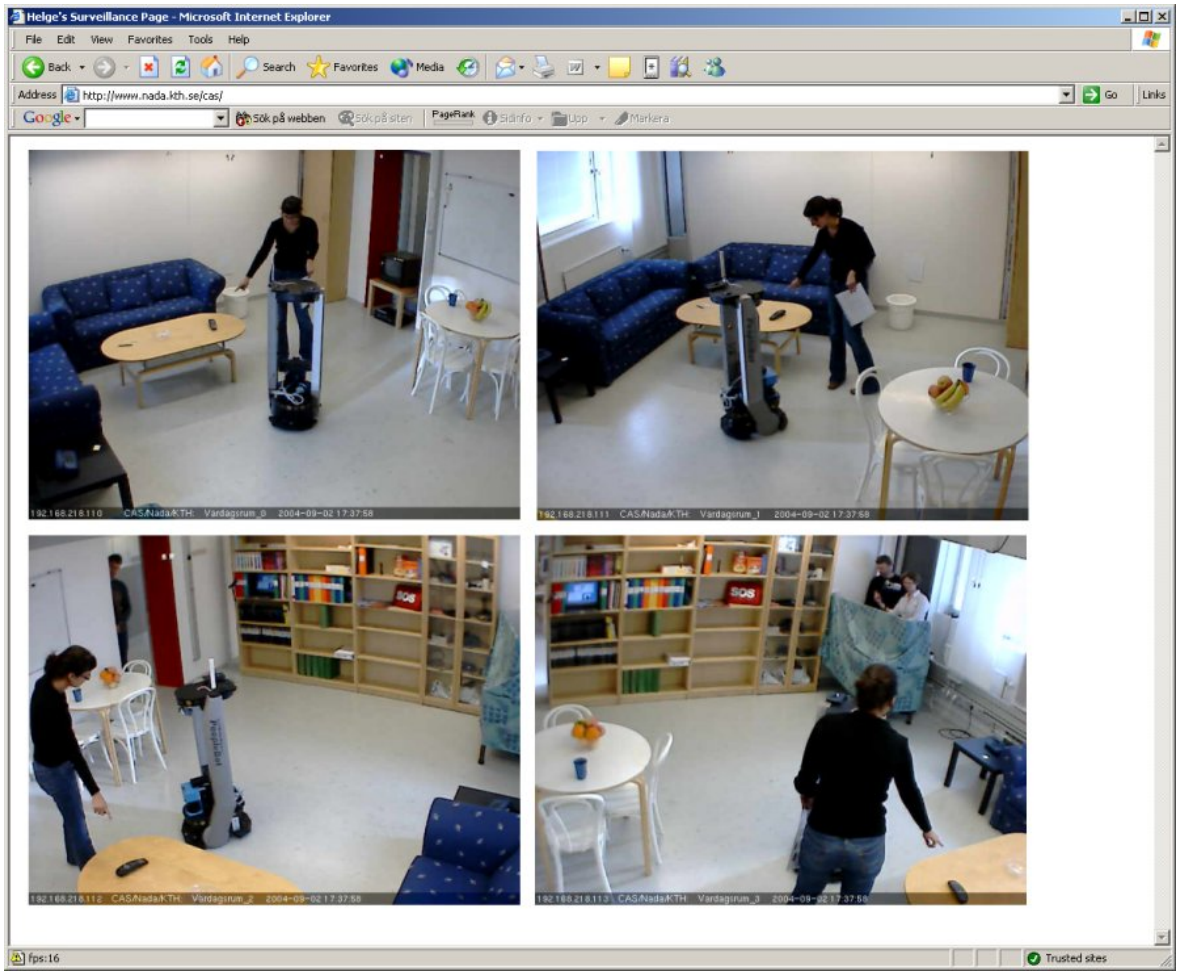

Figure 4: Online view of the different webcam perspectives surveying the CAS living room

The on-board video images from the robot were also saved to later check whether for example pointing gestures from the user could be seen. Figure 5 below is a picture from the on-board camera that shows "what the robot saw" during the interaction depicted in figure 4.

Figure 5: Example of the robot's on-board-camera view

A Sick laser range finder on the robot was tried for the collection of person tracking data[4]. This data is anticipated to allow for the gathering of numerical information about the spatial distance and positioning of the user under the condition that the user is in a 180° degree half-circle in front of the robot

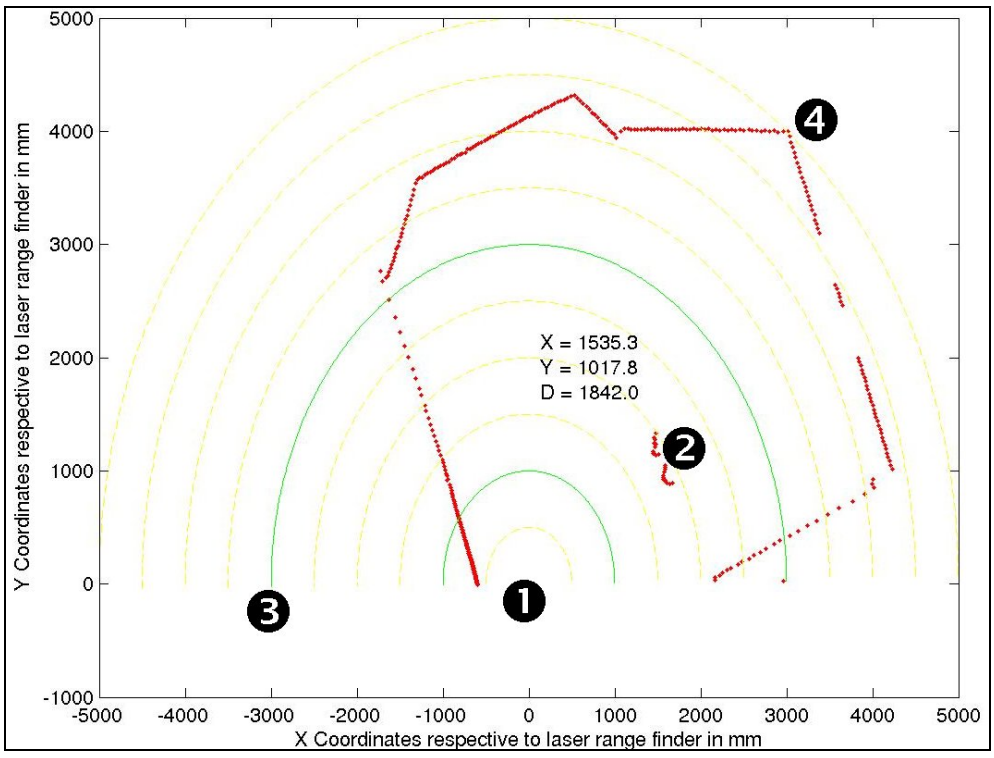

Figure 6: Example of a graphical representation calculated from the data of a laser range finder

[4] The laser range finder data collection and the person tracking system is work performed by Elin Anna Topp as part of the COGNIRON workpackage 5 on "Spatial Cognition and multimodal situation awareness"; it was used during the user study reported to aid positioning analysis.

(see figure 6). In figure 6 the robot ❶ "sees" a user ❷ standing to the right in front of the robot at a distance of 1.842 meters. The different half-circle lines surrounding the robot are added to support graphically the distance measurement (the line at ❸ for examples denotes the 3 meter distance). The room's walls as seen in the laser range finder are finally depicted as line ❹.

Data collected also includes a questionnaire that was given to participants once they had completed the task. The questionnaire was intended to assess how users experienced the robot.

Two other means of data recording during the trial are mentioned in brief: A system log was captured for all performed commands that were sent to the robot. Together with the timing information the robot trials can thus be run in a simulator at a later point of time. Timing (of the different systems involved) was synchronized against a local Network Time Protocol (NTP) server.

Finally a Marantz digital recorder was used to record the spoken commands on the robot for possible speech recognition training later. This recording is a step in the cooperation with the COGNIRON research group at the University of Bielefeld in the evaluation of the speech dialogue.

3.4 Data analysis

During the 22 trials conducted approximately 5 ½ hours of human-robot interaction were recorded. Unfiltered around 10 GB of raw data in about 230.000 files has been collected, excluding (digitized and compressed) video. The amount of data clearly points towards the need for a thoughtful strategy in the filtering, annotation, and interpretation of the data.

Another aspect in the analysis is the exploratory character of the study: To find the relevant interaction patterns and the ways to categorize them, it seems advisable to carefully experiment with the data of few trials first, scrutinize and discuss the approach before tackling the main data material in an analysis.

In the role of posture and positioning the first five trials were selected for an initial analysis. Only selected examples will be given here from these five cases to explain the methodology and a few very preliminary findings. A more detailed analysis will be deferred to another publication.

3.4.1 Dense Representations

The external video from the DV camcorder was captured as video-file and synchronized with the robot on-board recorded audio-files to provide for a master and timeline-based audiovisual recording of each trial. As part of the analysis of the spoken dialogue (for details, see [10]) the sessions were then annotated on the utterance level and each utterance's start and endpoint were determined by using a spectrogram view off the on-board-audio files [ibid., p. 26].

Once a file containing the utterances and their timing information was produced, this file could be imported for further annotation. In this next step annotations showing the overall progress in interaction, its different interactional acts, visible spatial relationships, social interaction behavior patterns, and different postures and gestures were added to give a detailed and fine-granularity description of the interaction between a human and a robot. For reference the absolute time was added to make a referential check with the webcam images easier. Resulting from this annotation is thus a so called *dense description* of the session: All spoken utterances are included, the unfolding interaction and progress of the task performed can be read out, and gestures and postures are documented and commented upon in a preliminary manner d.

The term *dense description* needs to be taken literally: The trial 3 from which an excerpt is given as table 1 below holds 270 individual observable interaction exchanges and/or commented observations in the 20 minutes and 20 seconds the trial lasted. The advantage of this notation is that one can move

back and forth between the observations made in the video and their descriptive explanation and interpretation in text format.

A short example from a dense description is given in table 1 below: Before the interaction event # 66 the robot and the user are standing close to the entrance (see figure 3 above). The robot has just successfully found a chair and the user is satisfied, giving a "thumbs-up" gesture (depicted in figure 8 below). The user now wants the robot to follow him (event # 66) to a new location (the phone table) to show a new object (a glue bottle), event # 72. The robot acknowledges this request after about 3 seconds (#67) and is following the user, driving for 11 seconds (# 69 / 70). At event # 68 the user's preparation can be seen and when the robot is explicitly commanded to "stop" (# 69) a rotating gesture and a demonstrative putting down of the glue bottle is observed. The robot acknowledges in a speech utterance the "stop" command (# 70) after about 4 seconds. In interaction event # 71 and # 72 the user shows the glue bottle to the robot and names it. The utterance is in parallel accompanied by a change

#	Trial t	Real t	Sp	Utterance	Task / comment / observation
66.	04:53:054	12:34:15	U:	follow me robot	R instructed while on the way towards R
67.	04:56:063	12:34:18	R:	Robot is following	
68.	05:07:000	12:34:28			Waits with putting down the glue-bottle until R has come quite close [to the phone-table] Glue bottle in <right hand>
69.	05:08:892	12:34:30	U:	robot stop following	U holds glue-bottle with <finger tips>; Glue-bottle is moved slightly, rotated, finally put down with a 'thumb'-sound repeatedly
70.	05:12:285	12:34:34	R:	Stopped following	
71.	05:14:069	12:34:36	U:	Robot	Posture: U stands bended over to reach low-table and points directly to object <right hand>
72.	05:16:770	12:34:38	U:	this is glue	gesture: takes up glue-bottle in a sweeping motion as if to grab R's camera attention, puts it down again <right hand>

Table 1: Excerpt from trial 3 – notation of dense description

of posture (bended over the low table) and a pointing gesture is directed at the object of interest (#71); furthermore, the glue bottle is swept in front of the robot's on-board camera as if to guide the robot's attention (#72), before it is put down on the phone table once more.

In the next step of the analysis a more abstract summary of the relevant features for the posture and positioning was extracted both from the dense description and the video material of the trial. The intention is to enable a later comparison between different trials, look for task performance differences and possible revealed strategies of interacting with the robot that may explain these. Especially interesting to the study of the role of posture and positioning are spatial interaction behaviors like movements or gestures that can be observed and qualified.

For trial 3 a summarizing description is given below as illustration[5]. The user in the trial three was obviously highly motivated and engaged, a clearly structured approach in speech was accompanied with equally strong movement patterns; gestures observed were repeated multiple times, but also accompanied by interesting novel and previously not seen ones, e.g. the "thumbs up" praising (see figure 8), or the "getting down on eye-level" with the robot (figure 6 below).

[5] The numbers in brackets indicate the interaction event for reference in the dense description; this enables finding the exact sequence in the video, the webcam pictures, and other trial materials; marked as "[clipped]" are sections that are of lesser interest– they are omitted for briefness.

Trial 3 Summary:
Trial number three was an intensive and long-lasting one: Taking 20 minutes and 20 seconds, 270 events of human-robot interaction, communication, and observations were observed and commented upon in the video analysis.

Seven different objects and places were taught or used for the robot learning and trying. The user started off by testing to teach a corner (9, 12), a wall (15, 16), and a paper (17, 20) all of which the robot had a hard time to recognize as *objects* or *places*. Thereafter a chair (44), a bottle of glue (72, 73) and a flashlight (114, 218) were used. Especially the flashlight played an important role in this trial as can be seen from the section about the *find and validation* behavior below.

As a final place, the robot was sent to the battery recharge station (268) after a sound indicating the *battery low* status, generated by the experiment leaders in order to signal a time limit for the trial.

The user made the robot follow on five occasions, i.e. towards the corner of the room (4), the door (24, 28, 38), the phone table (66, and again 256), and finally the bookshelf (211).
The subject let the robot go on with *find and validate* missions to an extent that is surprising: A chair (49) that had been shown and named for the robot was rather quickly found again by the robot after being put in a new location. Even with the glue (78) bottle which had been actively hidden at a different place after showing it to the robot, the *find mission* was resolved by the robot after "only" searching at four different places, every time actively encouraged by the trial subject to continue the search.

This needs to be compared to the *flash-light* (123, again: 226) that the robot was first shown and then sent off to find after placing it in a different location. The robot searched (unsuccessfully) at 10 (!) locations before the flashlight was finally found. This search and interaction with the user takes a total of 40% of this trial's duration, stretching from interaction event number 123 to 207, and again from interaction event 226 to 249. During this flash-light *search and validate* mission(s) the user hides the flash-light in the bookshelf (116, 117), puts it on the floor (175) so that the robot drives into it (181), into the robot's path (175, 185, 194, and 198), and finally onto a chair (221-226).

During this "search flashlight" interaction the user himself sits down on a chair to observe and wait for the robot (161). He also shows signs of frustration or desperation by lip-smacking (188), expresses his waiting status by putting both hands in his pockets (147, 186), but finally tries to be cooperative and almost *fatherly* toward the robot by cowering down and getting on eye-sight level with the robot's camera during a new showing of the flash-light to the robot (214).

[clipped]

During *search and validation* missions trial subject three can be seen to stay at the location he instructed the robot from, not accompanying the robot during its search at different locations. Only when the robot comes near a location where the object of interest is situated is the user actively coming closer to observe and then interact with the robot in close distance.

Other observations and problems of interest include the failure to teach the robot the objects / places (?) "corner", "wall", and "paper" (i.e. a white piece of paper on a white wall), trying to get the robot's visual attention by spoken command ("look here robot", 31), or waving of the arm and hand before the robot's camera (35, 36, 72). Another often observed interaction pattern is the active using of the robot's on-board-camera's posture to check whether the robot really has found or understood by closely observing the movements of the camera. The user is obviously actively seeking guidance for his decisions upon these camera movements by the robot, e.g. in the interactions situations 86, 99, 183-184, 202-203, 214, 218, 247-248, and 250.

A gesture of success and robot praise can be seen after a successful finding of an object searched by the robot: The user gives a two-handed "thumbs-up" gesture (62) and calls out the word "bra" (= good) in Swedish.
[clipped]
---End---

Selected scenes important for the role of posture and positioning in HRI have then been extracted as still images from the video as illustrated examples. For trial three above just a few of those are shown here to demonstrate the effect of this focusing on relevant observed postures, gestures, and social interaction behavior.

Figure 5: User monitoring and waiting for robot at distance; hands in pockets – legs crossed; note flash-light placed on floor

Figure 6: Cowering down to be on eye-height with robot during object labeling

Figure 7: User checking for robot's on-board-camera movement and orientation

Figure 8: Giving robot a "thumbs-up" sign of praise for finding chair

3.4.2 Statistical description of interpersonal distances and spatial formations

To check for the relevance of the interaction distances according to E.T. Hall (see above) and the F-formation system in human robot communication, a statistical analysis was performed.

In order to produce the data for this check a java analysis tool was programmed to present the distance and spatial configuration between the robot and the user. As seen in figure 9 the application can load two different views from the experimental room to allow for the analysis from different perspectives. The speech-annotations text-file (seen figure 9, lower right hand) can be loaded and informs about the ongoing interaction by providing a time-reference.

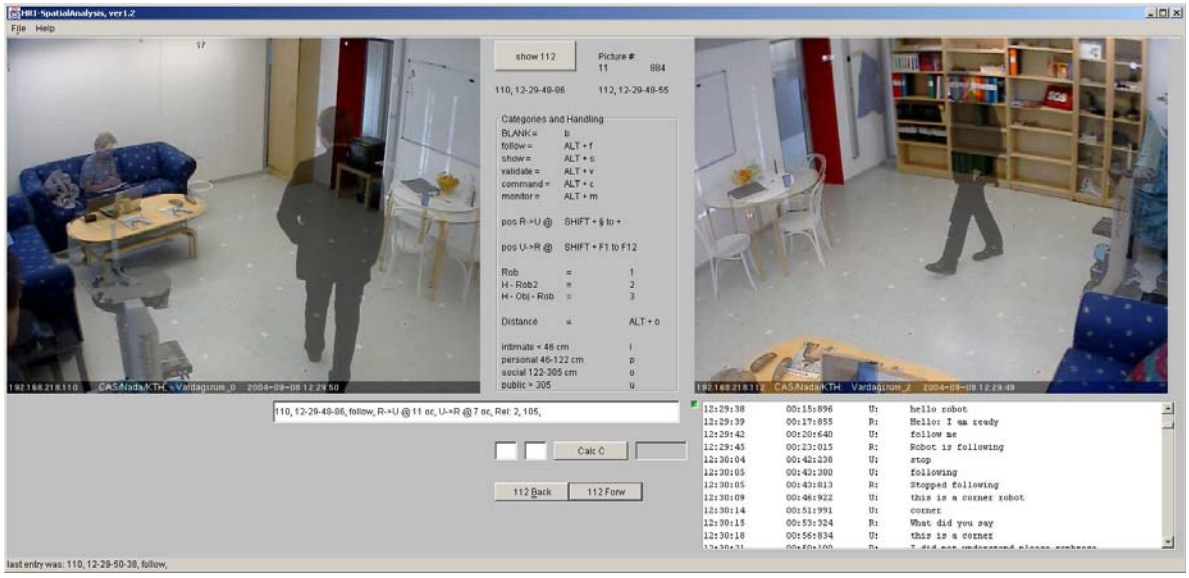

Figure 9: The spatial distance and formation annotation tool

The pictures are stepped through and annotated by hotkeys that write a data log file to be used for this analysis. Note that the floor in the picture seems to have dots and that the user depicted seems to be ghost-like. The reason for both the dots and the phantom-like appearance of the trial subject is the same: calibration pictures were taken in the empty CAS living room with dots on the floor marking .5 meters between them. These calibration pictures were after the trial fused (in a semitransparent mode) with the images collected during the trial. Noteworthy is that this allows for a low effort, low-fidelity visual measuring and annotation of the spatial relationship enacted during interaction – without having calibration dots on the floor *during the trial*. As the latter kinds of markers are rather seldom in normal living rooms and furthermore, might influence user's choice of positioning, our methodology of "measuring" seems like a practical choice.

The notation of the spatial annotation file (see table 2) on the spatial relationship needs to be explained: Separated by commas are from left to right, the camera-view that was analyzed, the time-stamp on the analyzed picture, the current status in the interaction, the relative position of the user according to a robot-centric view, and vice versa, the relative position of the robot according to a user-centric view, the current relationship, and finally, the distance between the robot and the user in centimeters (table 2, last column).

```
Cam; time;         state;   R has U @ pos ;   U has R @ pos;   relation; dist;
  …
[clipped]
  …
110, 09-28-11-54, command, R->U @ 12 oc,    U->R @ 12 oc,    Rel: 2,   100
110, 09-28-12-66, command, R->U @ 12 oc,    U->R @ 12 oc,    Rel: 2,   100
110, 09-28-13-79, follow,  R->U @ 12 oc,    U->R @ 2 oc,     Rel: 2,   130
110, 09-28-14-94, follow,  R->U @ 12 oc,    U->R @ 1 oc,     Rel: 2,   140
```

```
110, 09-28-15-86, follow,  R->U @ 1 oc,     U->R @ 2 oc,     Rel: 2,   150
110, 09-28-16-87, follow,  R->U @ 2 oc,     U->R @ 1 oc,     Rel: 2,   150
```

Table 2: Excerpt from trial number 2 – notation of the spatial annotation file produced by analysis tool

The camera can be switched and the positioning column(s) needs to be read as, e.g. “Robot has User at 12 o’clock” – using a pilot directional expression here it means the user is straight ahead in front of the robot. The relationship differentiates whether the user and the robot are acting together (denoted as a ‘2’ or *dyadic* relationship) or whether the robot, the user, and an object (denoted then as a ‘3’ or *triadic* relationship) is active. The last column holds the distance between the robot and the user; it was aimed to visually estimate this figure with a decimeter resolution (at best).

With such a log file of spatial distances and positioning different statistical descriptions can be produced. Here, only an example of the overall Hall distances and the different formations are given for trial number two.

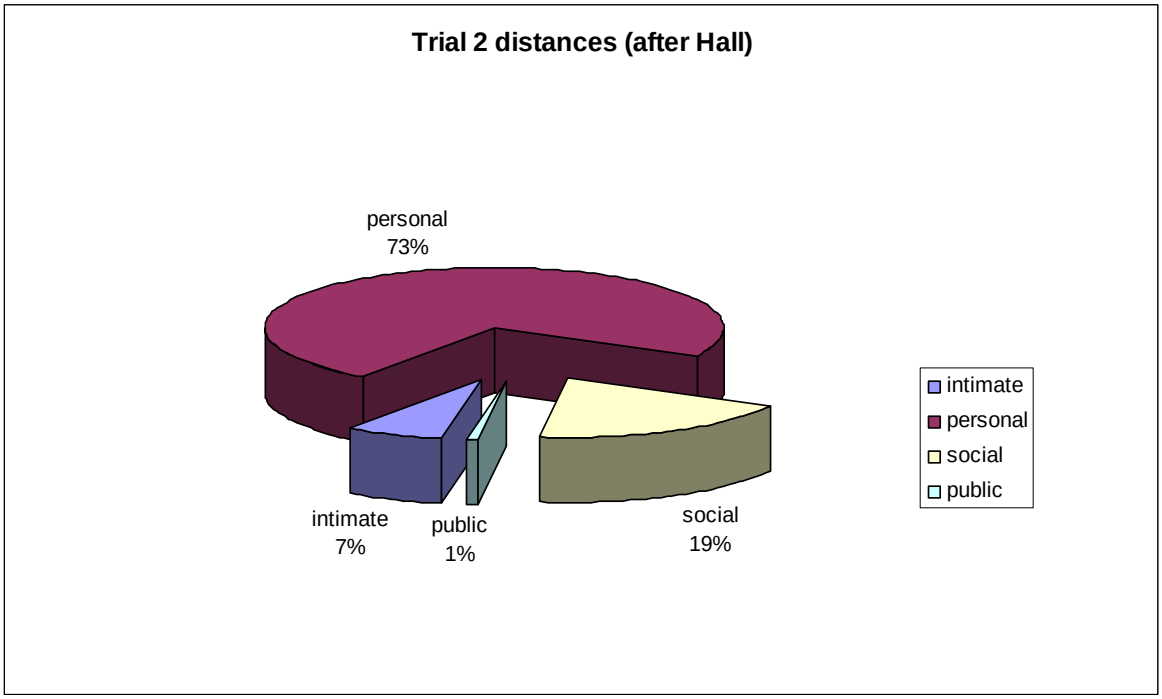

Figure 10: Social interaction distances according to Hall between robot and user during trial 2

In trial 2 the distance (see figure 10) between the robot and the user was mostly between 122 and 305 centimeters, i.e. the *personal* (73%) and the *social* distance (19%) were most often encountered. Higher than might have been expected is the percentage of moments where the robot and the user are less than 46 centimeters apart, i.e. almost touching one another: This so called *intimate* distance was encountered for 7% of all measured distances between the robot and its user. This might be interesting, e.g. for a safety concern.

The public distance does not play any role – this might be due to the setting of the room which in itself was “only” 5 * 5 meters, and furnished, too. This makes it quite unlikely to naturally keep a distance of more than 305 centimeters which Hall classified as the *public* distance in social interaction.

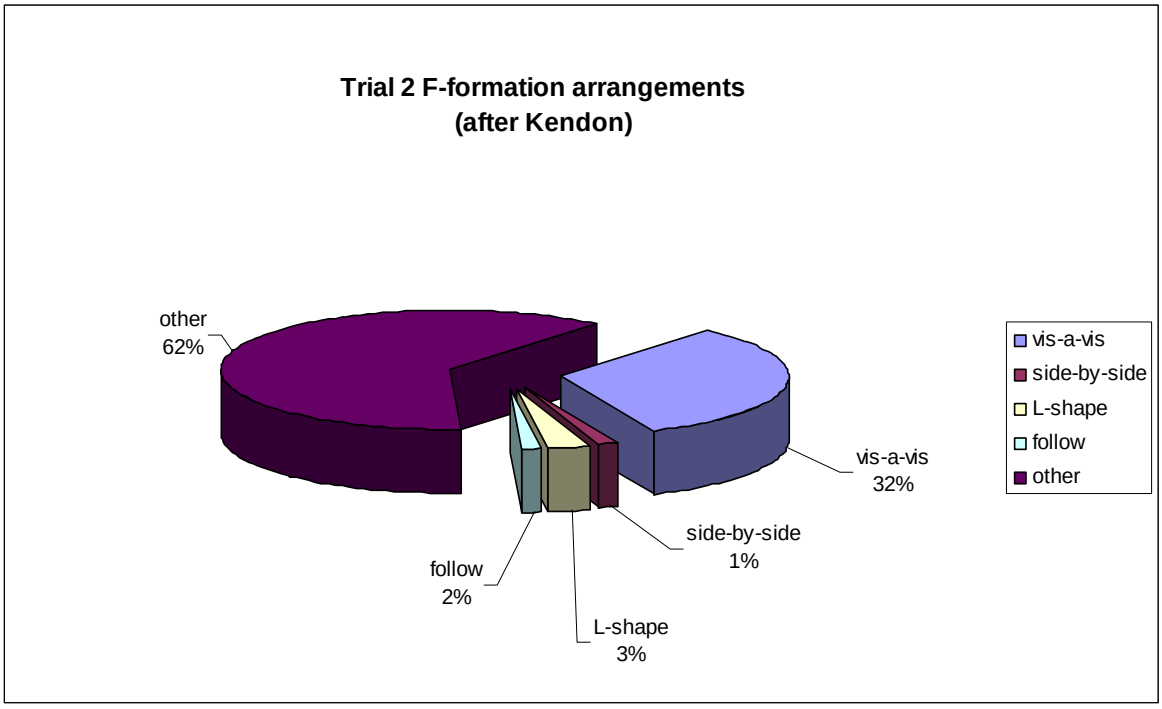

Figure 11: Social interaction F-formation arrangements between robot and user during trial 2

Checking the different F-formations arrangements (after Kendon [25]) the following scheme was tested: When both the robot and the user face one another, i.e. when they are judged to stand in a 12 o'clock – 12 o'clock position they were counted as having a *vis-à-vis* F-formation arrangement. A *side-by-side* F-formation arrangement was said to exist when the robot either has the user at his right or left side and vice versa. According to the positioning notation above, this would be the case if the robot / user are said to stand in a 3 o'clock – 9 o'clock position to each other.

An L-shape formation arrangement was more "relaxed" in its preconditions: Here all the analyzed interaction orientations where either the robot or the user were on one another's 10 or 11 o'clock and from the other side on the 1 or 2 o'clock position were included.

Finally the *follow* formation[6] is said to have occurred if the robot is having a user at its 11, 12, or 1 o'clock position and the user has the robot in his back, i.e. positions like the 5, 6, or 7 o'clock position. This is illustrated (figure 12) and discussed below in more detail.

If we apply this notation and analyze, e.g. the trial two according to this system we find almost 2/3rd of all interaction formations do not comply to the either of the formations postulated. However, 32% of all interaction positioning can be confirmed to happen in a *vis-à-vis*, facing one another fashion in trial 2. The other social interaction formations (side-by-side, L-shape, and following) play almost no role at all in this trial.

[6] As previously mentioned is *following* not a part of Kendon's scheme of F-formations.

4 Discussion

The discussion of the method and findings needs to take into account the preliminary nature of the reported work and the limited scope to which the data collected has been analyzed so far. Results were presented on example-basis to make both the research approach and methodology transparent as well as to show how observations can be treated to gain insights about the role of posture and positioning in HRI. In the following, assumptions, experiences with the explorative study and started analysis will be reflected upon. Critical points will be taken up for discussion to further the field and to prepare for the continued analysis of the collected data.

In [10] the Wizard of Oz research methodology conducted in this HRI study is discussed in detail. Among points raised are the cognitive demand on the wizards and the timing of utterance-responses.

The overall user trial set-up, trial execution, and data collection can be regarded as successful. The intended trials were conducted as scheduled; the scenario, user introduction, overall trial conduction, and data collection were on schedule and worked as planned without any major problems. As a result data from about 5 ½ hours of human-robot interaction have been collected. This is a valuable resource for analysis and an important result of the study itself.

The amount of data however also comes at a price: Analyzing this wealth of material is costly: The data needs to be saved, filtered, transcribed, annotated, categorized, converted into formats that are suitable both for analysis, a possible distribution among research colleagues as well as archiving as reference data of experienced interaction. Additionally, the explorative and qualitative nature of the trial itself and the data make a fully quantifiable statistical analysis seem unlikely or likely only in parts (e.g. the laser data). As a consequence we face a time-consuming manual analysis based upon annotating data which is necessary to gain insight on the research question of the role of posture and positioning in HRI.

A different approach is taken in [23] by putting infrared (IR)-markers both on the robot and the human user and then use an optical tracking system with IR cameras to collect data that can be numerically analyzed for the observed body movement data. Investigating “embodied communication” the detailed measurement of human and robot movement is achieved by the optical tracking system working with a frequency of 120 Hz and a 1 mm resolution of the movements measured. Comparing the data from this vision system to the subjective impressions by trial subjects the importance of well-coordinated behaviors between the robot and its users is suggested. Possibilities of such measurement systems for movements of the robot and a user need to be discussed for studies that will have a more quantitative and hypothesis driven aim and approach.

A major outcome and point for discussion from the initial analysis of the first five trials is the method of annotating the data. The time-line orientated transcription of spoken language utterances [10] was taken as the starting point to work with for the notations and comments upon the spatial interaction between the robot and a user. This emphasizes the spoken discourse as guiding and structuring unit of analysis and the influence shows in the annotation of the trial’s *dense descriptions* as presented in the “findings” section above where utterances are making up the main structure of interactional events.

Added to this transcription were then the observable tasks performed and the actions conducted by the human user and the robot. This included movements, their direction, gestures, and other spatial events. Many interaction events in addition to the pure spoken language transcription were thus added to the representation.

Also included in the dense descriptions are comments upon the observed interaction – these are of course subjective reflections, i.e. they are to be regarded as interpretations, not observable facts. The

motivation to include comments at all is that they are necessary to ensure that the *possible meaning* of the interaction observed is gained. Only in the context of the unfolding interaction and task performed do the observed bodily actions get a practical meaning. It is in this practical significance that the comments as interpretations of the observed events play an important role.

The dense description with its reference both to the video time-line as well as the real-world-time added can also be taken as a cross-referencing documentation of the trial. This means that it is possible to go from one medium and annotation to another data source within the trial, if in-depth check is necessary. The dense description can thus become an interesting starting point for the understanding and capturing of the detailed interaction history. Individual behavior can be tracked in different modalities and compared between trials. This helps finding interaction strategies of users and quantifiable units of analysis in the interaction itself. The summaries of the dense descriptions then provide the overall impression as well as references to particular events and strategies of interest to be further investigated.

There are however also concerns with this annotation format of the dense description as used so far: The prominence of speech utterances in the interaction sequence over the spatial interaction can be argued. Furthermore it could be helpful to have detailed and singled-out "tracks" for different observable entities, e.g. a column only for movements, distances, gestures with hands, head position etc. This would do more justice to the multi-modal nature of the interaction. A better form of presenting the different observable entities would perhaps be a sort of a composer's score with the different "instruments of interaction" playing together in parallel. Especially computer based tools, like the Anvil video annotation tool [26] can provide new perspectives (cf. [10] for our usage of Anvil).

Annotations of the used dense description type are also prone to inconsistencies. As there were no explicit rules of classifying interaction events and we are not aware of such being available in the HRI domain, the current scheme evolved over the first five trials analyzed. However, and this is a major problem, it is always difficult to maintain a consistent level of observational attention, transcription, and interpretation. One possibility to reach this is to annotate and analyze the same data by multiple persons, according to guidelines set up and iterated for consistency.

Another worry relies in the terms describing and categorizing the observed interaction. The human and the robot interact in a scenario with a certain tasks. For example the robot can by the user sent on a "validate mission", i.e. the robot is sent off to find places or objects it has previously learned by interacting with this user. The question resides now in the perspective that the annotation should take: The robot can be said to *drive and search an object*, while *at the same time* the user might a) *monitor the robot* or b) *do something else*. In this example, the robot and the user are doing seemingly different things that are not connected to one another – or at least, both actions need to be transcribed and evaluated. The question thus is how to describe the *interaction between robot and user* when two parallel activities can be observed for them. This can be understood as challenge of what view to take in: Should the human-robot interaction description take in an external observer's, a human- or a robot centric view upon the interaction and describe and interpret it accordingly?

Both the difficulty to come up with "correct" descriptive terms as well as the duality of the interaction nature can be further illustrated with the statistical analysis of the observed F-formation arrangements. The arrangements (vis-à-vis, L-shape, side-by-side) are based upon a certain spatial relationship. Annotating the two perspectives both as a user- as well as a robot-centric view (cf. table 2 above) relative positions according to a watch-hand notation, e.g. "12 o'clock = straight ahead", was introduced. There are at least a couple of problems with this categorization and notation: Numerically, the 360° degrees of a surrounding horizontal plane are divided with the 12 "hourly" sectors giving a theoretical resolution of 30° degrees for each position. The problem relies into truly determining whether, for example, a user is standing on the 1 o'clock or already on the 2 o'clock position by visually trying to evaluate the observed situation. Another approach could thus work with sectors instead, e.g. taking 45° or 90° degree wide sectors differentiating between "right-front", "right-back", "left-back", and "left-front". Again, such a system would make it on one hand easier, on the other hand it would be more

difficult to describe situations characterized by a true *side-by-side* positioning as it would fall exactly in between two sectors. Last but not least an attempt was undertaken to differentiate between for example, back and front, left and right. This on the other hand, seemed too coarse a description to analyze meaningful formations between the robot and its user according to the F-formation system.

In the near future there might also be technical aspects that determine the choice of categories as can be seen from the current implementation of the Sick laser range finder: It only monitors a half-circle 180° degrees in front of the robot, in effect making the robot "blind" to half of its surroundings if no other modalities are used to counterbalance this sensing limitation.

The F-formation system, as introduced above, makes a couple of assumptions which need to be taken up for discussion when transferred to the interaction behavior between a robot and a human (cf. figure 1). The system is based upon the precondition that the position of the feet determines the construction of the o-space between interactors as the lower and upper body can assume different orientations, i.e. a rotational axis in humans' hip exists. The robot used in this interaction study does not have this capability or degree of freedom (DOF) and the question thus arises what this means for the creation of the o-space between interactors, or more abstractly, which orientation formation from the robot may be counted as the one or other arrangement according to Kendon's scheme.

The categorization performed for the statistical analysis of F-formation arrangements (figure 11) has been made explicit, i.e. the 12 – 12 o'clock was counted as vis-à-vis, the side-by-side arrangement required a 6 – 9 o'clock position, the L-shape arrangement was accepted when either interaction partner had the other at a 10 or 11 o'clock position while the other opposite partner had the other on a 1 or 2 o'clock position.
This calculation of multiple possible positions of the interaction partner is illustrated in figure 12 for the *following* formation: The user can have the robot (dark circle lower half) directly behind him, i.e. the user could express to have the robot at a 6 o'clock position, while the robot might find the user at a 12 o'clock position; for the following formation however, the case depicted with lighter shaded robot "clones" to the right and left of the central robot, were also counted as being in the following formation. These robots have the user at a 1 or 11 o'clock position respectively, and vice versa, the user has the robot in his back in either a 7 or 5 o'clock position.

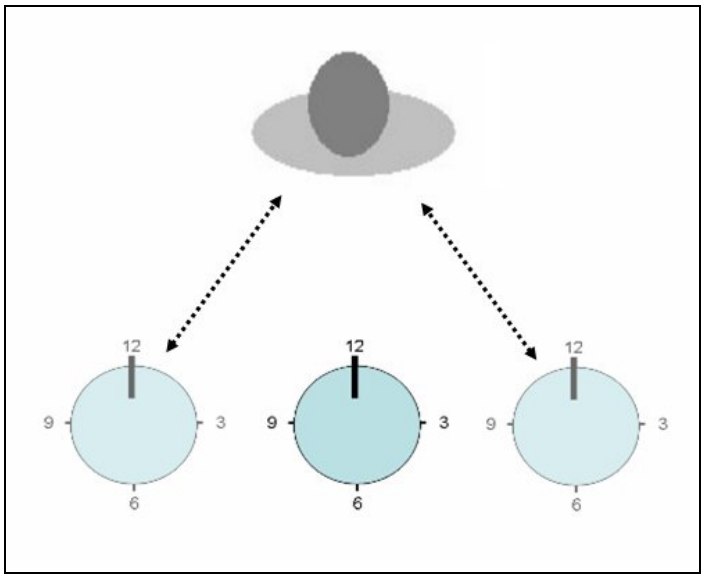

Figure 12: Tolerated formations calculated as "following"

This system seemed in theory stable enough to be applied to the observations made during the human-robot interaction study conducted and analyzed with the spatial annotation tool. Nevertheless the outcome is a surprise: During the trial 2 which was used as a case example, 62 % of the total trial duration

and interaction saw the robot and the user in another arrangement or formation than the ones expected (see above, figure 11). It is thus questionable whether this kind of statistical description can be performed on an overall trial. Instead one should focus on certain sequences of interaction, e.g. excluding times when the robot is sent on search missions without the user accompanying the robot. This needs to be studied in more detail.

More positive so far are the experiences made on the strategies for the spatial management and the role of posture and positioning observed during the first analyzed trials in other aspects: Already during the first 5 trials analyzed individual strategies to interact with the robot and its relevance to the spatial management were detected, i.e. certain patterns of interaction with a robot appeared. Monitoring for example was repeatedly seen to be performed without accompanying the robot, i.e. standing left at the previous interaction position the robot was allowed to drive off without the user accompanying it. Only once the robot came close to a new position of interest, users followed after the robot and positioned themselves in a distance and orientation suitable to either observe the robot or be in the position to engage in a speech dialogue with it. Another recurring pattern is that the initiation of the *follow-me* failed, i.e. people requesting the robot to follow, and then either not waiting for the robot's acknowledgement and simply walking away, or, if the robot had replied, walking off too quickly so that the robot could not follow. Users thus had a hard time to adapt to (and remember throughout the trial) the fact that the robot needed to "see" them in the following mode and to move accordingly.

This points towards an important issue in the spatial management: Users looked eagerly for the onboard-camera's movement to make an inference about the robot's point of attention and behavior. Users for example positioned themselves in order to monitor the robot's camera movement to judge for themselves what the robot was looking at. This possibility of the robot to give feedback and to guide interaction actively through "body movements" should be pursued as promising approach for the spatial management. This observation of the possibility to direct user's spatial attention with help of a directed robot camera (head) is in line with the observations made by Kuzuoka et al. [27].

In the introduction to the term posture (see above, p. 5) *height* was introduced as one of the defining characteristics to define a position. Furthermore the interaction space was defined as 3-dimensional, including height naturally. Despite this we were unprepared that different interaction height actually would lead to observable interaction patterns that saw users actively change their posture in height. Examples of this behavior as part of the body movement changes for interaction can be found in the figures 6 and 7 above. These observations and realization agrees with findings reported by Kanada et al. [23] who equally found that interactors adopt different heights in interaction and even lowered themselves as if to get on "eye-level" with the robot, it is to the height of the robot's (camera-) "eyes", pointing towards the significance of different heights or the overall perspective in interaction.

5 Conclusion and Future Work

The role of posture and positioning in Human-Robot Interaction (HRI) has been studied as part of the COGNIRON research activity 3 on "Social Behavior and Embodied Interaction". This report focused on the spatial management in posture and positioning between a user and a robot participating in a joint activity as part of the COGNIRON key experiment 1 "Robot Home Tour".

Above the terms *posture* and *positioning* have been defined and discussed. Social interaction studies of human communication and interaction behavior were used to introduce concepts and guide the formation of categories and interesting parameters in HRI. A user study with 22 participants has been presented in methodology, set-up, trial conduction, and data collection. The analysis so far has been limited to the first five trials to gain an understanding for the relevant units of analysis and develop the

tools and formats of the analysis. These preliminary findings will need to be discussed and iterated before being extended to the whole data set.

In the next phase of the COGNIRON project we will finish the analysis and publish the findings accordingly. Furthermore we intend to initially perform user studies according to the same general framework and scenario, but to use real, implemented robot navigation capabilities rather than simulated, teleoperated robot movements. The robot movements will be expanded to include other co-operative locomotion behaviors to initiate, sustain, and terminate interactions. Candidates for prototypical situations are, for example, approaching and addressing of a potential interactors, joint management of passing narrow spaces, or leaving from an interaction in controlled and socially acceptable fashion.

References

1. Allwood, Jens. (2001). Cooperation and Flexibility in Multimodal Communication. In Cooperative Multimodal Communication Harry Bunt and Robbert-Jan Beun (Eds.). Lecture Notes in Computer Science 2155/2001. Springer Verlag, Berlin/Heidelberg, 2001, pp. 113-124.
2. Argyle, Michael. (1973). Social Interaction. Papperback. London: Tavistock Publications Ltd.
3. Badler, N and S. Smoliar. (1979). Digital representations of human movement, ACM Comput. Surveys 11(1), 1979, 19–38.
4. Benford, S. and Fahlén, L. (1993). A spatial model of interaction in large virtual environments. In: Proceedings of the Third European Conference on Computer Supported Cooperative Work (ECSCW'93), Milano, Italy, September 1993.
5. Birdwhistell, R. L. (1952). Introduction to kinesics: an annotation system for analysis of motion and gesture. Louisville, Ky.: University of Louisville Press.
6. Brooks, R.A.. (1991). Intelligence without reason. In Mylopoulos, J., Reiter, R., eds.: Proceedings of the 12th International Conference on Artificial Intelligence (IJCAI-91), San Mateo, CA, Morgan Kaufmann (1991) 569–595
7. Brooks, R.A. (1991). Intelligence without representation. Artificial Intelligence 47 (1991) 139–159
8. Calvert, T.W., Chapman, J., and Patla., A. (1980). The integration of subjective and objective data in the animation of human movement. In: Proceedings of the 7th annual conference on Computer graphics and interactive techniques ACM Press, 1980, pp. 198-203
9. COGNIRON. (2003). Annex 1 “Description of Work”, EU Project document, Contract number FP6-IST-002020.
10. COGNIRON. (2005). Deliverable 1.3.1 “Report on evaluation methodology of multi-modal dialogue”. Forthcoming.
11. Dario, Paolo, Guglielmelli, Eugenio, and Cecilia Laschi. (2001). Humanoids and personal robots: Design and experiments. Journal of Robotic Systems, Volume 18, Issue 12, Date: December 2001, Pages: 673-690.
12. Dautenhahn, K., Ogden, B., Quick T. (2002). From Embodied to Socially Embedded Agents –Implications for Interaction-Aware Robots, Cognitive Systems Research 3(3), pp. 397-428.
13. DiSalvo, Carl F., Gemperle, Francine, Forlizzi, Jodi, Kiesler, Sara. (2002). All robots are not created equal: the design and perception of humanoid robot heads. Proceedings of the conference on Designing interactive systems, London, ACM Press, pp. 321-326.
14. Dix, A., Finlay, J, Abowd, G., Beale, R. (1998). Human-Computer Interaction, 2nd Edition. Prentice Hall Europe, Pearson Education Ltd.
15. Dreyfus, Hubert L. (1992). What Computers Still Can't Do - A Critique of Artificial Reason. 1992, 6th printing (1999). MIT Press.
16. Dourish, P. and V. Bellotti. (1992). Awareness and coordination in shared workspaces. In: Proc. of CSCW '92, Toronto, Canada.
17. Drury, J.L.; Scholtz, J.; Yanco, H.A. (2003). Awareness in human-robot interactions. Systems, Man and Cybernetics, 2003. IEEE International Conference on , Volume: 1 , Oct. 5-8, 2003, pp. 912 - 918
18. Encyclopædia Britannica on "nervous system, human.". (2004). Encyclopædia Britannica Online Service. 11 Oct. 2004 <http://www.britannica.com/eb/article?tocId=75598>
19. Fong, Terrence, Nourbakhsh, Illah and Dautenhahn, Kerstin. (2003). A survey of socially interactive robots. Robotics and Autonomous Systems, Volume 42, Issues 3-4, 31 March 2003, Pages 143-166, Elsevier.
20. Goffman, Erving. (1967). Interaction Ritual - Essays on Face-To-Face Behavior. New York: Anchor Books
21. Green, A.; Severinson-Eklundh, K.; Hüttenrauch H., (2004). Applying the Wizard-of-Oz Framework to Cooperative Service Discovery and Configuration. In: Robot and Human Interactive Communication, 2004. Proceedings. 13th IEEE International Workshop on.
22. Hall, E.T. (1966). The Hidden Dimension: Man's Use of Space in Public and Private. The Bodley Head Ltd, London, UK.

23. Kanda, Takayuki, Ishiguro, Hiroshi, Imai, Michita, and Tetsuo Ono. (2003). Body Movement Analysis of Human-Robot Interaction, International Joint Conference on Artificial Intelligence (IJCAI 2003), pp. 177-182, 2003.
24. Kawamura, K., Peters, R.A., Wilkes, D.M., Alford, W.A., Rogers, T.E.. (2000). ISAC: foundations in human-humanoid interaction. Intelligent Systems, IEEE, Volume: 15 Issue: 4 , July-Aug. 2000, Page(s): 38-45.
25. Kendon, Adam. (1990). Conducting interaction - Patterns of behavior in focused encounters. Studies in interactional sociolinguistics. Cambridge, NY, USA: Press syndicate of the University of Cambridge.
26. Kipp, Michael. (2001). Anvil – A Generic Annotation Tool for Multimodal Dialogue. In: Proceedings of the 7th European Conference on Speech Communication and Technology, Aalborg, pp. 1367-1370.
27. Kuzuoka, H., Kosaka, J., Yamazaki, K., Yamazaki, A., Suga, Y., (2004). Dual Ecologies of Robot as Communication Media: thoughts on Coordinating Orientations and Projectability, in Proc. of CHI2004, pp.183-190, Vienna, Austria.
28. Merriam-Webster online on "posture". 2004. Merriam-Webster Online Dictionary. 11 Oct. 2004 <www.m-w.com>
29. Pfeifer, R.; Iida, F. (2004). Embodied Artificial Intelligence: Trends and Challenges. In: Embodied Artificial Intelligence: International Seminar, Dagstuhl Castle, Germany, July 7-11, 2003. Revised Papers, Iida, F., Pfeifer, R. Steels, L, et al (eds.), Lecture Notes in Computer Science (LNCS), volume 3139 / 2004, Springer-Verlag, Heidelberg.
30. Reeves, Byron, and Clifford Nass. (1996). The Media Equation - How People Treat Computers, Television, and New Media Like Real People and Places. CSLI Publications and Press Syndicate of the University of Cambridge.
31. Scheflen, Albert E. and Scheflen, Alice. (1972). Body Language and the Social Order – communication as behavioral control. Engelwood Cliffs, N.J.: Prentice-Hall, Inc.
32. Schegloff, E.A. (1989). Harvey Sacks' Lectures on Conversation. The 1964-65 Lectures An Introduction/Memoir. Human Studies, 12:185-209.
33. Schmidt, Colin T. (2002). Socially Interactive Robots. Why our current Beliefs about them STILL work. In: Proceedings of the 11th IEEE Int. Workshop on Robot and Human Interactive Communication (ROMAN 2002), pp. 560-564, 25-27 Sept. 2002, Berlin, Germany.
34. Shibata, T.; Wada, K.; Tanie, K., (2003). Subjective evaluation of a seal robot at the National Museum of Science and Technology in Stockholm. In: Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. pp.: 397 – 402