Face-to-Face with a Robot: What do we actually talk about?

Article *in* International Journal of Humanoid Robotics · April 2013

DOI: 10.1142/S0219843613500114

CITATIONS
4

READS
210

8 authors, including:

Nicole Mirnig
University of Salzburg
36 PUBLICATIONS **155** CITATIONS
SEE PROFILE

Astrid Weiss
TU Wien
112 PUBLICATIONS **1,056** CITATIONS
SEE PROFILE

Gabriel Skantze
KTH Royal Institute of Technology
91 PUBLICATIONS **1,157** CITATIONS
SEE PROFILE

International Journal of Humanoid Robotics
Vol. 10, No. 1 (2013) 1350011 (23 pages)

DOI: 10.1142/S0219843613500114

World Scientific
www.worldscientific.com

FACE-TO-FACE WITH A ROBOT: WHAT DO WE ACTUALLY TALK ABOUT?

NICOLE MIRNIG[*,‡], ASTRID WEISS[*,§], GABRIEL SKANTZE[†,¶], SAMER AL MOUBAYED[†,‖], JOAKIM GUSTAFSON[†,**], JONAS BESKOW[†,††], BJÖRN GRANSTRÖM[†,‡‡] and MANFRED TSCHELIGI[*,§§]

[*]*HCI & Usability Unit of the ICT & S Center, University of Salzburg, Sigmund-Haffner-Gasse 18, 5020 Salzburg, Austria*

[†]*Department of Speech, Music and Hearing, KTH, Stockholm, Sweden*

[‡]*nicole.mirnig@sbg.ac.at*
[§]*astrid.weiss@sbg.ac.at*
[¶]*gabriel@speech.kth.se*
[‖]*sameram@speech.kth.se*
[**]*jocke@speech.kth.se*
[††]*beskow@speech.kth.se*
[‡‡]*bjorn@speech.kth.se*
[§§]*manfred.tscheligi@sbg.ac.at*

Received 29 August 2012
Accepted 23 January 2013
Published 2 April 2013

While much of the state-of-the-art research in human−robot interaction (HRI) investigates task-oriented interaction, this paper aims at exploring what people talk about to a robot if the content of the conversation is not predefined. We used the robot head Furhat to explore the conversational behavior of people who encounter a robot in the public setting of a robot exhibition in a scientific museum, but without a predefined purpose. Upon analyzing the conversations, it could be shown that a sophisticated robot provides an inviting atmosphere for people to engage in interaction and to be experimental and challenge the robot's capabilities. Many visitors to the exhibition were willing to go beyond the guiding questions that were provided as a starting point. Amongst other things, they asked Furhat questions concerning the robot itself, such as how it would define a robot, or if it plans to take over the world. People were also interested in the feelings and likes of the robot and they asked many personal questions — this is how Furhat ended up with its first marriage proposal. People who talked to Furhat were asked to complete a questionnaire on their assessment of the conversation, with which we could show that the interaction with Furhat was rated as a pleasant experience.

Keywords: Human−robot interaction (HRI); back-projected robot head; speech-based interaction; discourse analysis; unrestricted scenario.

1. Introduction

Imagine a retired person, slightly senile and needy, on the verge of moving into a nursing home — a social robot could be the key to preserve this person's independence for as long as possible. This is exactly the scenario that is taken on in the movie *Robot and Frank*,[a] which nicely shows that upon free exploration of technology, people come up with all sorts of approaches and ideas for application — some of which researchers have not thought before. As researchers, we often have a fixed course of events in our minds upon which we build scenarios and set up studies. Looking at instances of human−robot interaction (HRI) from a task-oriented perspective, helps to research and possibly enhance certain situations in which humans encounter and cooperate with robots. Framing studies from the beginning within a certain scenario, entails that pre-assumptions influence the way studies are set up. It seems valuable to take one step back every once in a while, in order to verify existing and future interaction paradigms. Such practice can help in two ways: First, new ideas that were not considered previously might be detected. Second, existing approaches, assumptions, and rules can be double-checked to see if they are indeed right.

So, what happens if the visitors to a museum are confronted with a robot but are not given a clear scenario on which they may base their interaction, e.g. a tour guide robot which would provide a certain frame for the interaction? It is interesting to explore how people, who are not briefed prior to an interaction, talk to a robot on an unbiased basis: Do people in fact know about what to talk to this robot? Would people smalltalk with such a robot — and if so, about what? Do people go beyond the obvious and freely explore the knowledge and capabilities of the robot by asking all sorts of questions? To make use of people's creativity, we installed a speaking robot head at a robot exhibition. The robot was set up to proactively invite passing visitors to engage in an interaction with it, whereas the dialog capabilities of the robot were modeled so that they allowed for and fostered nonstructured, free exploration by naïve users. With our study we wanted to find out about what people talk to a robot when encountered in public space without a predefined scenario and task.

The study was conducted at the London robot festival "Robotville", a special exhibition in the British Science Museum in December 2011. For the qualitative data analysis we annotated five hours of video material that was recorded on the opening day of the exhibition. The data was then analyzed to find out what the visitors talked about with the robot. To provide the visitors with some clues, a few guiding questions were provided. With our analysis we wanted to explore if the visitors rather stuck to these guiding questions or if they were willing to explore the conversational capabilities of the robot more freely. We furthermore wanted to find out if there are certain topics that the visitors frequently addressed with the robot. The results from the dialog recordings were completed with data from a short questionnaire that the

[a] http://robotandfrank-film.com/.

interaction partners of Furhat were asked to complete. Over the total duration of the exhibition, 86 visitors answered the questionnaire and rated the robot and the conversation with it.

In the next section, an overview on related research on how and about what humans talk to artificial communication partners, will be given. Next, the robot head will be introduced in terms of its technical specifications and the particular setup for the study at hand. Then, the study procedure will be explained. In the results section, first the content analysis of the dialog recordings will be presented, followed by the quantitative questionnaire data. The discussion is dedicated to the question of what we can learn from people's conversations with the robot head.

2. Motivation and Background

Up to now, empirical studies mainly focused on very "particular, corpus-, domain- or situation-specific results"[1] with a strong emphasis on how users talk to robots in a specific, predefined interaction context. For instance, the study by Lee *et al.*[2] explored the diverse ways in which people communicate with a robotic receptionist. They analyzed the dialog log data and demonstrated how the occupational background of the robot helped participants to ground the conversation. The researchers also found out that not all users always followed the social norms of human−human communication. Kim *et al.*[3] were interested in affective vocalizations provided to robotic learners by human teachers. They could show that people vary their vocal input depending on the learner's performance history and naïve users tend to spontaneously use intensely affective vocalizations.

Studies in HRI that focus on communication, often investigate how the design of the robot can foster natural communication, frequently following the trend of designing robots as similar as possible to a human. For many HRI researchers, natural language is considered the most effective and satisfying interaction strategy. Breazeal[4] stated for instance that natural communication is a core element to enable natural social exchange between a human and a robot. Following this assumption, Torrey *et al.*[5] conducted a study investigating the effects on communication, if the robot adapts to the user in a conversation. Their experiment was based on a predefined scenario, in which the robot was a chef explaining cooking tools. They could show that when the dialog of the robot was adapted for expert knowledge (naming tools instead of explaining them), expert participants found the conversation more effective. However, Clark[6] claimed that the goal of natural communication with an artificial communication partner is not a realistic one. He argues for an alternative goal, namely to design robots as dynamic depictions of other people to which they can talk as if they were actual people. Of course, new dialog principles for such conversations would be needed which go beyond imitating human−human talk.

Summarizing, much of the above-presented research is directed toward examining "how" people talk to a robot or an artificial communication partner in situations which happen in a predefined context or scenario. Kopp,[7] like us, was interested in a

freer exploration which is less bound to a specific context. The research focus, in this case, was centered around the question of what people are interested in when talking to a robot without any "pre-scripted actions". He analyzed utterances made by naïve users toward a conversational agent in an unrestricted scenario. Similar to the study presented in this paper, Kopp's study took place in the public setting of a science museum. The embodied conversational agent Max was applied at an information kiosk in the "Heinz Nixdorf Museums Forum", a public computer museum in Paderborn (Germany), where the agent should engage visitors in face-to-face small-talk conversations. Max was installed in 2004 and is to date applied in the museum where it provides visitors with information about the museum itself and about the current exhibition. The main differences between our Furhat study and the study with Max are that Max was not an embodied agent but a screen agent, that it had more of a traceable occupation, and that, in order to speak to Max, the visitors had to type their utterances on a keyboard (no speech recognition was applied). Similarly to Lee,[2] Kopp could show that many dialog partners applied communicative strategies from human−human communication to Max (such as greetings, farewell, small-talk elements), but that they rather used shorter, yet close to everyday natural language utterances. Kopp derived six content categories from the data corpus of the museum visitors' conversations with the robot Max: greetings/farewell, flaming, feedback to the agent, questions, answers, and requests. This study revealed that the visitors did not wonder much about the language capabilities of the system, but about its world knowledge and general intelligence. However, in how far this result was induced by the appearance of the agent Max or its role/occupation in the museum could not be answered with this single study.

Gustafson and Bell[8] applied their conversational agent August, without a given task and predefined context, in a public place in Stockholm (Sweden) to collect a large corpus of spontaneous speech from nonspecialists. Upon analyzing what people talk about to the agent, the researchers identified two main groups of utterance types with three sub-types each (*socializing* — social, insult, test; *information-seeking* — domain, meta, facts). Gender and age correlations revealed the following: Many adult users socialized with the system for a few turns before moving on to seek information, whereas children tended to stay in the socializing phase. Women seemed to be more inclined to begin their interaction with the system by seeking for information.

In the following, we will present our study with the robot head Furhat which was performed at the London robot festival "Robotville". Our results will extend the findings of Kopp and Gustafson/Bell, in further exploring "what" people want to talk about with a robot.

3. The Back-Projected Robot Head Furhat

To create a situation as natural as possible, a robot head seems to be a good way to research conversations between humans and robots. As the focus of a study on the communication between humans and robots clearly lies within the conversations as

such, there is no imperative need for a moving robotic platform. It seems more valuable to utilize a sophisticated robot head which can provide subtle cues. Over the last years, increasing effort has been put in creating projected robot heads[9,10] as they bring about the advantage that swift movements without actual motor activity and concomitant noise are possible.

The use of facial animation for interactive agents has been investigated over many years.[11,12] It has been found that in case of situated, multi-party interaction, the use of a flat screen with an animated head suffers from what is known as the Mona Lisa effect,[13] since the agent is not spatially co-present with the user. This means that it is impossible to establish exclusive mutual gaze with one of the observers and either all observers will perceive the agent as looking at them, or no one will. While mechanical robot heads are indeed spatially co-present with the user, they are expensive to build, inflexible and potentially noisy. The robot head Furhat[14] (see Fig. 1), developed by KTH Royal Institute of Technology in Stockholm (Sweden), can be regarded as a middle-ground between a mechanical robot head and an animated agent. Using a micro projector, a two-dimensional (2D) facial animation is projected on a three-dimensional (3D) mask that is a 3D printout of the head used in the animation software. The head is mounted on a neck (a pan-tilt unit), which allows the use of both head pose and gaze to direct attention. To cover up the technical details and give the human interaction partner more the impression of talking to a "complete" robot head, the top- and back-area of the head are covered with a fur cap.

The mask was painted with back-projection paint to improve the visibility of the projection, which makes it possible to use Furhat under normal light conditions. Using software-based facial animation in a robot head enables a flexible generation of

Fig. 1. The robot head Furhat.

advanced mimics that are crucial for dialog applications. It also provides the robot with real-time lip-synchronized speech, which has been shown to increase speech intelligibility in noisy environments.[15] The lip-synchronized and synthesized speech also lends a sense of authenticity to the head. We have previously shown in an experimental setting that such a 3D projection increases the ability of the system to regulate the turn-taking in multi-party dialog, as compared to a 2D screen.[16]

4. User Study

It was the overall aim of this study to explore what people speak about with a robot when the topic of the conversation is up to their choice. To explore this matter, we recorded conversations between a robot and the visitors to a robot exhibition, and in addition we asked the visitors who spoke to the robot to complete a questionnaire. Our main research questions were: (i) About which topics do the visitors of a robot exhibition speak with the robot (content analysis of the dialogs)? and (ii) How do the people rate the conversation with the robot (quantitative analysis of the questionnaire data)?

The robot head Furhat was placed in a science museum where it addressed passing visitors. The employed dialog manager was set up to first attract people to engage in an interaction with the robot and consequently to keep the interaction going. There was no scenario given, Furhat was asking people questions from a predefined set in random order, or was inviting the museum visitors to ask questions to it, respectively. The setup allowed for two people to engage in a shared conversation with the robot head by shifting the attention of the robot back and forth between these two people. Apart from shifting attention, Furhat was able to produce different kinds of feedback to keep the interaction partners actively involved in the conversation: Furhat could address a specific person, change its facial expression and change some of its features (e.g. change the color of its face). The data collected during the study was twofold: First, the conversations Furhat held with the museum visitors were recorded and the verbal utterances were automatically transcribed for a subsequent qualitative dialog analysis. Second, the interaction partners of Furhat were asked to fill in a short questionnaire on their impression of the conversation and their rating of the feedback and the performance of the robot.

4.1. *Study setup*

The setting of a public exhibition in a museum poses considerable challenges to a multimodal dialog system. In order to engage in a multi-party, situated interaction, the system not only needs to cope with the extremely noisy environment, but also be able to sense when visitors are present. In the lab, we have been using a Microsoft Kinect,[b] which includes an RGB camera and a depth sensor for visual tracking of

[b] http://kinectforwindows.org/.

people approaching Furhat and a multiarray microphone for capturing and localizing speech. However, in the crowded and noisy environment of the museum, with often dozens of simultaneous by-standers, a Kinect would not suffice. Instead, we used two handheld, close-range microphones put on podiums with short leads, forcing visitors to walk up to one of the microphones whenever they wanted to speak to Furhat. To sense whether someone was standing close to a microphone, we mounted ultrasound proximity sensors on the podiums. Furhat and the two podiums formed an equilateral triangle with sides of about 1.5 m.

The multimodal dialog system was implemented using a newly developed framework based on the notion of statecharts.[17] Statecharts are a powerful formalism for complex, reactive, event-driven systems, and lend themselves well to visual representations. Statecharts are based on finite-state machines (FSM), but with several extensions. The most notable difference is that the statechart paradigm allows states to be hierarchically structured, which means that several states may be active at the same time, allowing the designer to define generic and specific event handlers on different levels. The transition between states can be conditioned, depending on variables on different levels, as well as event parameters. This relieves statecharts from the problem of state and transition explosion that traditional FSMs typically lead to when modeling more complex dialog systems.

For the exhibition scenario, the dialog contained two major states reflecting different initiatives: one where Furhat had the initiative and asked questions to the visitors (e.g. "When do you think robots will beat humans in football?") and one where the visitors asked questions to Furhat (e.g. "Where do you come from?"). In the former case, Furhat continued the dialog (e.g. "Why do you think so?"), even though it often understood very little of the actual answers, occasionally extracting important keywords which enabled the robot to continue the interaction without understanding the whole utterance.

To provide the visitors with some assistance in talking to Furhat, a poster was mounted on the side wall of the booth explaining the microphone usage and indicating the following guiding questions for the visitors to ask (and the speech recognition was set to understand — occasionally, the speech recognition failed nevertheless due to problems with surrounding noise or pitch/pronunciation varieties):

- What is your name?
- Where are you from?
- Tell me a joke?
- Can you look different?
- Knock knock
- What is your favorite movie?
- Who made you?
- How old are you?
- Do I need an umbrella?

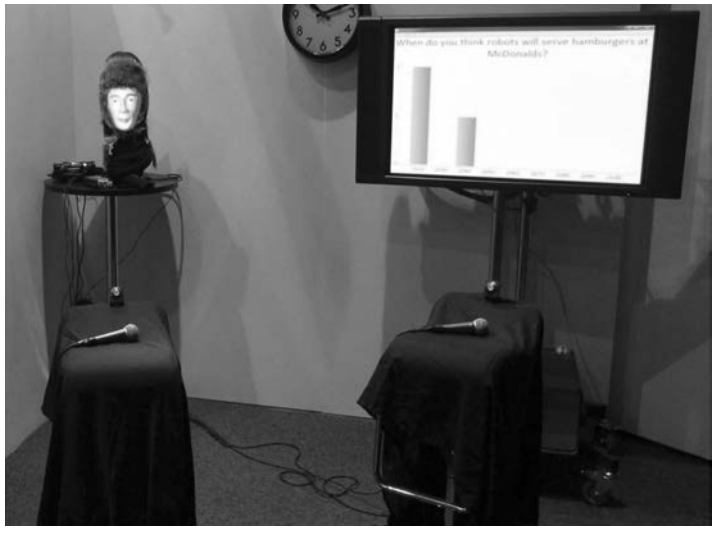


Fig. 2. Study setup with the robot head Furhat.

Certain sensory events were mapped to gesture actions in the statechart to exploit the possibilities of facial gestures that the back-projection technique allows for. For example, when the speech recognizer detected a start of speech, the eyebrows were raised to signal that Furhat was paying attention. A screen was mounted next to the robot head to provide the visitors with live extrapolations of the answers that Furhat received on its questions. The purpose of this was to make the whole exhibition more interesting for the visitors. The actual study setup with the robot head, podiums with microphone and proximity sensor, and the screen can be seen in Fig. 2.

For speech synthesis, we used the CereVoice system developed by CereProc[c] and lip-synchronized the verbal output with the facial animation. CereProc's text-to-speech system reports the timing of the phonemes in the synthesized utterance, which was used for lip-synchronization. The voice also contains nonverbal tokens like grunts and laughter that were used to give Furhat a more human-like behavior. For speech recognition, we used the Windows 7 automatic speech recognition (ASR), running in two separate modules, one for each microphone. This allowed the system to process simultaneous speech in both microphones. Each ASR engine also used two parallel language models, one context-free grammar with semantic tags (SRGS[d]), tailored for the domain, and one open dictation model. To interpret the dictation results, we have implemented a robust parser that uses the SRGS grammar to find islands of matching fragments. This allowed the system to recognize answers to very open questions and then pick out specific parts.

[c]http://www.cereproc.com/.
[d]http://www.w3.org/TR/speech-grammar/.

4.2. *Multi-party dialog*

A dialog involving multiple parties refers to an interaction with more than two partners talking to each other. In case of the Furhat user study, the setup allowed for two people talking to the robot at the same time (see Fig. 3). Multi-party dialogs enable a more complex interaction as the attention of the robot head has to be shifted back and forth between the two people. In case of Furhat, attention shifting was realized by means of turning the robot head and addressing a person via shifting the gaze.

An example dialog of two visitors talking to Furhat is given in Fig. 4, which illustrates a number of typical interaction patterns. As soon as Furhat was approached by a visitor, the robot immediately took the initiative and started to ask questions, as can be seen in turns 1 to 4. The example also illustrates how the system was able to extract partial results from the ASR. When the system actually understood an answer, it gave some relevant feedback (as in turn 6), but if it did not understand, it simply continued (as in turns 9 and 17). All answers were recorded and information about the corresponding questions was logged, which made it possible to annotate all answers later on. After each question, the system made an elaboration request (as in turns 6 and 15). All utterances from the system (including questions) were randomly selected from a set of possible utterances, resulting in a varied output.

With two users present, Furhat could either ask a directed question — with the head posed in direction toward the addressee, and eyes looking forward (establishing eye-contact) — or an open question to both of them — with the head directed between the users, while alternating gaze between them (as in turn 17). Furhat then turned to the person who answered the question. When speech was detected in both

Fig. 3. Children talking to Furhat.

1. U1	*(Approaching microphone 1)*
2. F	*(Looking up at U1)* Hi there. Could you perhaps help me?
3. U1	Yes [S: **yes**]
4. F	I have some questions for you. When do you think robots will beat humans in football?
5. U1	Maybe in 10 years [D: that see **in 10 years**]
6. F	That soon! Could you elaborate on that?
7. U1	Well, they are already pretty good at it [D: while they are already predicted owners]
8. U2	*(Approaching microphone 2)*
9. F	*(Still looking at U1)* Yeah... I have another question for you. Would you like robots to be like humans?
10. U2	Yes, absolutely [D: **yes** that see]
11. F	*(Turning to U2 with gaze only)* Could you just wait a second? *(Looking back at U1)* I'm sorry, where were we. Would you like robots to be like humans?
12. U1	Sure [S: **sure**]
13. F	Yeah? *(Turning the head towards User-2)* Do you agree?
14. U2	No [S: **no**]
15. F	Mhm. Could you say something more about that?
16. U2	Well, I want robots to be like robots [D: I want robots to be like Reynolds]
17. F	Yeah... *(Posing head between users, shifting gaze back and forth between them)* I have another question for you. When do you think robots will serve hamburgers at McDonalds?
18. U1	Do I need an umbrella? [S: **do I need an umbrella**]
19. F	*(Turning to U1)* What do you think, you are in London! Maybe you have another question for me?

Fig. 4. Example interaction. ASR results in brackets, using S: SRGS or D: dictation grammar (U1 = User1; U2 = User2; F = Furhat).

microphones at the same time, the audio levels were compared in order to choose who to attend to. If a question was directed to one of the users and the other user tried to take the turn, Furhat would acknowledge this by shifting the gaze toward this user and request "Could you please wait a second", while keeping the head directed toward the original user. Furhat would then shift the gaze back and continue with the interaction it was previously involved with (as in turn 11). Another option for Furhat in the multi-party setting was to follow up an elaboration request with an agreement request, by turning to the other interlocutor (as in turn 13). This could sometimes result in longer sequences on the same topic, such as the one shown in turns 9 to 16: question-elaboration-agreement-elaboration. Although Furhat always started the interactions by asking questions, it was also possible for the visitors to ask questions (as in turn 18) and thereby take the initiative and shift the topic of the dialog. This would trigger Furhat into awaiting further questions, resulting in mixed-initiative interaction patterns. In order to prepare the system for these questions, we used corpora from interactions with agents previously on display at museums,[18,19] as well

as from pre-tests in the lab. After answering some questions, or after too many nonunderstandings, Furhat again tried to take over the initiative and started to ask questions to the interaction partners.

Using the statechart framework, we defined generic states, such as idle and dialog, with sub-states to handle specific question types (e.g. questionYN, questionOpen, requestElaborate). The generic dialog state then defined event handlers to handle questions from the users regardless of the current sub-state, allowing mixed-initiative interaction. Low-level sub-states, such as speaking, attending and listening, were also defined with relevant event handlers, for example to handle situations where someone left while Furhat was speaking or listening.

5. Results

In the following section, the results of the study are presented. For the qualitative data analysis, we examined five hours of video recordings, which added up to 3.301 dialog lines. From this corpus we could extract 98 conversations between Furhat and visitors of the robot exhibition. Five people talked to Furhat twice and one visitor even three times. As the recurring conversations were much shorter than the first interactions, we decided to add these conversations up, what results in 91 conversations from individuals talking to Furhat. 54 of Furhat's interaction partners were male, 37 female. The mean dialog duration was 2'30" (SD 1'34", min=18", max=9'52"). 20 of the visitors talked to the robot on their own, with no other person talking to it. For the conversations in which Furhat talked to two people at a time, in 23 cases the second person was at least temporarily a researcher if no other visitor was willing to participate in the conversation.

For the quantitative data analysis, we collected 86 questionnaires over the whole exhibition from visitors who interacted with Furhat themselves. The mean age of the people who filled in the questionnaire was 35.49 years (SD 16.17), ranging from 12 to 80 years. 46 of the respondents were male, 39 female (one participant did not fill in the demographic data section of the questionnaire).

5.1. *Discourse analysis*

From the 3.301 dialog lines that were transcribed from the five-hour video recording of original conversations between Furhat and visitors of the Robotville exhibition, 2.106 lines (64%) were uttered by the robot, 1.195 (36%) by the visitors. During these five hours Furhat uttered a total of 12.168 words, whereas the visitors produced only 3.771 words. By means of analyzing the word quota, we could also find out that Furhat produced longer sentences: An average sentence from Furhat adds up to six words, whereas the mean length of the sentences spoken by the visitors was four words. Upon analyzing the data, we had the impression that the visitors occasionally tended toward monosyllabism (meaning that they frequently produced one-word sentences like e.g., "Maybe" or "Hello"), therefore we counted the number of the

one-word sentences. Although the number is almost the same (Furhat produced 240 one-word sentences, the visitors 244), keeping in mind the difference in the total number of utterances, Furhat produced only 12% of one-word sentences, whereas the visitors produced twice as many (25%).

As the conversation was designed around asking questions with either having the robot randomly ask by-stopping visitors from a set of predefined questions or the robot inviting the visitors to ask questions to it, a large part of the dialogs happened in the question-and-answer mode. From the data corpus we could extract a total of 311 questions that the visitors asked Furhat. 199 questions (64%) were based on the guiding questions, 112 (36%) were questions upon which the visitors were freely exploring the conversational skills of Furhat. Based on Kopp's[7] content analysis of conversations between humans and a 2D conversational agent receptionist, we grouped the questions in the following categories: anthropomorphic questions, questions concerning the robot, questions concerning the exhibition, commonplace phrases, questions to test the robot, requests, and feedback to the agent. Table 1 gives an overview on the categories and their frequency of occurrence.

Anthropomorphic questions dealt mainly with topics such as the current condition of the robot (e.g., "Are you cold?"), things the robot likes (e.g., "What is your

Table 1. Overview on the questions the visitors to the exhibition asked Furhat ($N = 91$).

Question type	Examples	N (N of free questions)
Anthropomorphic questions 17%	Are you happy? Do you have a girlfriend? What is your favorite color?	53 (35)
Questions concerning the robot 32%	What is your name? Who is your master? How do you understand me?	98 (9)
Questions concerning the exhibition 1%	What are you doing here today? What do you do?	2 (2)
Commonplace phrases 4%	How are you? Good morning. Thank you.	12 (12)
Questions to test the robot 14%	What is the meaning of life? How far is it to Australia? What day is it?	44 (32)
Requests 26%	Tell me a joke. Can you look different? Can you make an angry face?	80 (0)
Positive feedback to agent 3%	I like your hat. Your look is great. Wicked styling!	11 (11)
Negative feedback to agent 3%	That's not funny. Stop winding me up! Your face looks disgusting.	11 (11)

favorite type of electronic equipment?"), and the love life of the robot (e.g. "Are you married?"). Furhat also received its first marriage proposal, which the robot, however, left uncommented.

The questions concerning the robot were to a large extent centered around the guiding questions. However, the visitors were also inquiring about the design of the robot and the technology behind it. One visitor for example for whom the speech recognition was working quite well, was very astonished about the answers Furhat provided, to the extent that he was questioning the autonomy of the robot ("Is there someone talking behind you?"). The fact that there were only two questions about the exhibition as such underlines the visitors' interest in Furhat itself and also the nonrestrictedness of the dialog as it could be shown that people talked about a variety of topics other than the exhibition.

The visitors were quite inventive when it came to challenge the conversational skills of Furhat. The variety of questions ranged from substantial inquiries ("What is love?") to questions on robots (e.g., "What is your definition of a robot?"). Some visitors tested the spirit of the robot by asking somewhat conspiratorial questions (e.g., "When is the robot uprising?", "Do you plan to take over the exhibition?"). Since the robot, on request to tell a joke, asked some silly conundrums, some visitors also sought to tease Furhat (e.g., "How high is noon?"). In one case, a visitor walked straight up to Furhat and without starting the conversation with whatsoever, asked "What's the square root of Pi?" — a question which Furhat was not able to answer. When we asked the visitor after his conversation with Furhat about his reason for asking this question, he said: "Since this is a scientific exhibition I expect a robot being here on display to be much cleverer than I am!"

Many requests addressed toward Furhat were related to the guiding questions, most of which dealt with requests to change facial features, or to tell a joke. Furhat was also given direct feedback from the interaction partners. Some visitors told the robot that they liked how it looks and especially the hat was well-liked, some people even asked where Furhat got the hat from. Other visitors, however, thought that the jokes of the robot were not funny, some did not like the looks of the robot, some felt annoyed and one person was even scared.

5.2. *Survey data*

After the conversation with Furhat, we asked all communication partners to fill in a short survey (14 Likert-scaled items) on how they experienced talking to the robot head. We asked the participants furthermore to indicate their age and gender (as reported in the beginning of this section), their interest in technology and if they had any prior experience with a robot.

The participants rated their interest in technology in general on a mean of 4.42 (SD 0.798) and in robots on a mean of 4.28 (SD 0.954); both on a five-point Likert scale (5 "very much" to 1 "not at all"). 26 participants stated to have interacted with a robot before. It must be stated at this point that, as the study took place in the

framework of a robot exhibition in a science museum, the sample is of course more technophilic then if the data was collected in a different context.

5.2.1. *Descriptive results*

80 participants were convinced by the performance of Furhat and said that they think that the robot is able to respond to a human; 79 participants thought that Furhat actually responded to them. Table 2 gives an overview on what feedback the participants thought the robot had given them as opposed to what kind of feedback they would like to receive. In asking the participants how Furhat respond to them, we wanted to gather which modalities were recognized by how many people. In addition, we also wanted to know if the participants rate the behavior they observed as appropriate. Whereas most participants mentioned the verbal and gaze feedback, less than 20% stated to have received feedback via head gestures, mimics or other (e.g., feedback via a screen or affirmative sounds such as "ah", "ok"). Interestingly, not all participants indicated that they would like Furhat to talk to them, neither would all participants want Furhat to look at them. Approximately half of the participants would like Furhat to provide head gestures and/or mimics. For both questions, no distinct patterns could be detected that show that certain modalities were mentioned together or ruled each other out.

Table 3 provides an overview on the participants' mean ratings for each of the 14 questionnaire items (five-point Likert scale ranging from 5 "very much" to 1 "not at all").

5.2.2. *Factor analysis*

The dimensionality of the 14 items from the measure on how the participants liked Furhat and the conversation with it was analyzed using principal component factor analysis. Three criteria were used to determine the number of factors to rotate: the *a priori* hypothesis that the measure was unidimensional, the scree test, and the interpretability of the factor solution. The scree plot indicated that our initial hypothesis of unidimensionality was incorrect. Based on the plot, four factors were rotated using a Varimax rotation procedure. The rotated solution as shown in

Table 2. How Furhat responded and how the participants would like it to respond ($N = 86$).

Feedback	How did Furhat respond to you?	How would you expect Furhat to respond to you?
Verbally	80	74
Gaze	75	64
Head gestures	11	41
Mimics	17	30
Additional response (e.g., via a screen)	13	24

Table 3. Participants' rating of Furhat and the conversation with the robot ($N = 86$).

Question	Mean rating	SD
How much do you like Furhat?	4.08	0.775
How much do you like Furhat's response behavior?	3.80	0.708
Did you enjoy talking to Furhat?	4.13	0.838
Did you find Furhat uncanny?	2.99	1.206
Did Furhat respond quickly enough?	3.57	1.056
Did Furhat interrupt you?	3.00	1.414
Could you understand what Furhat said?	4.25	0.890
Did Furhat understand what you said?	2.99	1.048
Was the conversation with Furhat fluent?	2.94	1.101
Was the conversation with Furhat easy?	3.17	0.985
Was the conversation with Furhat frustrating?	2.67	1.083
Did you have to concentrate to talk to Furhat?	3.38	1.162
Was it easy to know what to say to Furhat?	2.99	1.047
Did Furhat provide enough feedback to you?	3.28	0.067

Table 4, yielded two interpretable factors: conversational fluency and perceived enjoyment. The conversational fluency factor accounted for 19,8% of the item variance, and the perceived enjoyment factor for 17,0% of the item variance. No item loaded on both factors. Table 4 provides an overview on the factor analysis.

Table 4. Correlations between the coping items and the coping factors ($N = 86$).

	Conversational fluency ($\alpha = 0.819$)	Perceived enjoyment ($\alpha = 0.729$)	Disturbing factors ($\alpha = 0.581$)	Ease of conversation ($\alpha = 0.336$)
How much do you like Furhat?	0.149	**0.843**	0.061	−0.137
How much do you like Furhat's response behavior?	0.324	**0.610**	−0.058	0.009
Did you enjoy talking to Furhat?	0.197	**0.779**	−0.106	0.074
Did you find Furhat uncanny?	0.009	0.313	**0.663**	0.101
Did Furhat respond quickly enough?	**0.720**	0.079	−0.040	−0.009
Did Furhat interrupt you?	−0.153	−0.161	**0.764**	0.063
Could you understand what Furhat said?	0.248	−0.037	−0.032	**0.419**
Did Furhat understand what you said?	**0.791**	0.237	0.039	0.264
Was the conversation with Furhat fluent?	**0.802**	0.280	−0.095	−0.016
Was the conversation with Furhat easy?	**0.752**	0.315	−0.280	0.071
Was the conversation with Furhat frustrating?	−0.135	−0.321	**0.663**	−0.277
Did you have to concentrate to talk to Furhat?	0.114	0.146	**0.409**	**−0.712**
Was it easy to know what to say to Furhat?	−0.032	0.069	0.224	**0.724**
Did Furhat provide enough feedback to you?	0.367	0.449	0.008	0.465

An internal consistency estimate of reliability was computed for all factors. The coefficient alpha was 0.819 for "Conversational fluency" and 0.729 for "Perceived enjoyment", both indicating satisfactory reliability. The item groupings "Disturbing factors" and "Ease of conversation" did not indicate satisfactory internal reliability and thus are not regarded as factors as such. Nevertheless, they will be interpreted as regards their overall common ground in terms of content. Based on the factor analysis, the following conclusions can be drawn: The factor "Conversational fluency" (mean = 3.17, SD 0.84) indicates that the participants rated the conversation with Furhat as medium fluent, which provides much room for further improvement. The factor "Perceived enjoyment" (mean = 4.00, SD 0.63) indicates that the conversation with Furhat was regarded as rather enjoying.

A further analysis as regards content can be drawn from grouping the following items that all report on disturbing factors: 4, 6, 11 and 12 are all rated around 3, indicating that the conversation with Furhat was neither really disturbing nor completely normal. Regarding the ease of the conversation items 7, 12 and 13 show a slight tendency toward rating this criterion as slightly better than average (mean item 7 = 4.25; 11 and 12 around 3).

5.2.3. *User-specific reactions toward Furhat*

As the data is not normally distributed, only nonparametric calculations can be performed (Kolmogorov−Smirnov test was significant). We correlated the answers from the questionnaire with the independent variable age to see if the participants' age influences their answers. Therefore, the age variable was split into three groups as can be seen in Table 5.

Only the question "Was the conversation with Furhat fluent?" resulted in significant differences between the age groups, $H(3) = 6.497$, $p = 0.039$. Mann−Whitney tests were used to follow up on this finding and detect between which age groups there is a difference. A Bonferroni correction was applied and so all effects are reported at a 0.0167 level of significance. A significant difference could be found for the groups "up to 25 years" (mean = 3.20, SD 1.031) and "26 to 38 years" (mean = 2.46, SD 1.103), $U = 226.00$, $r = -0.33$. We can conclude that people up to an age of 25 years rated the conversation with Furhat more fluent than people between 26 to 38 years. In comparison to the oldest group there was, however, no significant difference ("39 years and older": mean = 3.08, SD 1.129).

Table 5. Categorized age variable ($N = 82$).

Age group	No. of participants
up to 25 years	30
26 to 38 years	24
39 years and older	28

Note: Four people did not state their age and thus are not included in this analysis.

Based on the fact that age only influenced one single question with a medium effect, it can be said that the Furhat study seems to be quite inclusive as regards age. This means that the overall positive experience that could be proven by the overall analysis of the results is equally distributed in all three age groups.

In terms of gender, significant differences could be detected for two questions: The answers on the question "Are you interested in technology?" differed significantly between male (mean = 4.64) and female (mean = 4.18) participants, $U = 610.00$, $z = -2.586$, $p = 0.010$, $r = -0.28$. This result shows that male participants were more interested in technology than female ones — keeping in mind that the overall interest was very high (overall mean = 4.42, SD 0.798). The answers on the question "Was the conversation with Furhat easy?" differed significantly between male (mean = 3.36) and female (mean = 2.92) participants, $U = 621.00$, $z = -2.102$, $p = 0.036$, $r = 0.23$. This result indicates that the male participants considered the conversation with Furhat easier than the female ones, which can at least in part be explained through the fact that the speech recognition worked better for men (due to differences in pitch between male and female voices).

We could hardly detect any differences in our results for age and gender which is promising in the sense of an inclusive interface that enables interaction not just for a limited group of certain users. The results are limited due to the circumstance of the study being performed in the framework of an exhibition in a science museum, which means that it may be a given fact that people who attend such an exhibition bring more interest and have a higher tolerance toward technical systems in the first place. This accounts also for the age groups as, given the special setting in which the study was performed, it may be assumed that older visitors are also more interested in technology than the average person would be.

6. Summary and Discussion

This article reported on a study which was performed in the framework of a robot exhibition which took place in the British Science Museum in London in December 2011. The study was aimed at taking one step back and question the application areas of a social robot in terms of what people actually talk about with a robot that they encounter in a public place, but which is not part of a fixed scenario. The robot head Furhat was used to actively engage visitors to the exhibition in a conversation with the robot, the progress of which, however, was at least in parts subject to the visitors.

We analyzed the content of five hours of original conversations between Furhat and visitors to the exhibition, to find out what people want to know from a robot if it is up to them to direct the interaction. As the dialog pattern of the robot was set to either asking questions to people or inviting people to ask questions to the robot, large parts of the conversations were made up of question-and-answer dialogs. About one third of the questions concerned the robot itself. People showed much interest in the technology behind the robot and the robot as an individual. The second largest

part of questions was directed toward requests. Despite the fact that the people making these requests were inspired through the provided guiding questions, the visitors nevertheless showed great interest in playing with the robot (jokes) and changing it to their liking (change the appearance), which indicates that a customizable robot might be to users' liking. The third largest set of questions was centered around the anthropomorphic quality of the robot. The questions posed by the visitors impressively showed that people do care about a being (even if it is obviously artificial) which in turn shows interest in them. That they cared, could also be proven by means of examining the direct feedback that the visitors gave to Furhat. Positive feedback was especially expressed toward the looks of the robot and the furry hat. Some people were annoyed by not-funny jokes, recurring questions or also the looks of Furhat, and they were willing to tell it directly to the robot. Finally, the visitors showed interest in challenging the robot, especially regarding its comprehension, knowledge, wittiness, and most strikingly about its (ill) intentions.

The quantitative questionnaire data could show that the overall design was quite inclusive regarding age and gender. A factor analysis on 14 Likert-scaled items through which the visitors could rate the robot and the conversation with it, resulted in the findings that on the one hand there is room for improvement regarding the flow of conversation. On the other hand, the participants rated the conversation with Furhat as an overall enjoying experience. A further analysis as regards content of those items which did indeed not result in a factor but showed some weaker connection, could show that even if people eventually detected the narrow-mindedness of the robot (limited number of questions resulted in early repetitions) they did not rate the conversation as disturbing. The questions upon ease of conversation resulted in a medium rating saying that it is slightly easier than average.

These results are liable to some limitations: The fact that the robot had a higher share of the dialog can in parts be explained through the dialog manager being set up in a way to keep the conversation going. So whenever the conversation was about to drop, the robot would take the initiative and change the perspective (either from asking questions to being asked or vice versa). The content of the conversations was in part influenced by the guiding questions that were provided by the researchers. We found the usage of these questions quite useful, given the crowdedness of such an exhibition with often dozens of people at the booth at once, which is when it is sometimes better that we as researchers stand back and thus cannot immediately provide first-hand assistance. We are aware that these results were gathered in the framework of a robot exhibition which explains the technophilic audience and the general positive attitude.

7. Conclusion

With our study we could show that a robot head as an artificial communication partner is interesting for people also in an unrestricted scenario, which is an important finding for the future in which social robots could serve as a communication partner

independent from a precise application. From our data we can refer the following implications to guide the design of interactions with social robots. It is good to make people marvel about the robot to create curiosity but they should also be allowed to understand the robot, the functionality and the interaction with it. Customization is highly desirable and playful features spice up any interaction. People express care and cooperativeness but also curiosity, which may be satisfied with providing a more round robot character by adding "personal features" (such as creating a life around the robot by means of likings, dislikings, being in love, etc.). Finally, in our study we could show that it was important for people who interacted with Furhat, to on the one hand, test the knowledge of it, but on the other hand, to question the intentions of the robot. Future work should be directed toward further exploring how the communicative capabilities of the robot can be enhanced, by e.g., advancing the dialog manager to create the impression of a more "knowledgeable" robot. Since Furhat provided extensive feedback via mimics and head movements, this discrepancy between which feedback the participants reported to have observed and what they wished for also remains to be explored in further experiments.

Acknowledgments

This work is supported within the European Commission as part of the IURO project, see also www.iuro-project.eu.

References

1. K. Fischer, *What Computer Talk is and isn't: Human-Computer Conversation as Intercultural Communication*, Vol. 17 (AQ-Verlag, 2006).
2. M. K. Lee and M. Makatchev, How do people talk with a robot? An analysis of human-robot dialogs in the real world, in *ACM SIGCHI Conf. Human Factors in Computing* (*CHI*) (ACM, 2009), pp. 3769–3774.
3. E. S. Kim, D. Leyzberg, K. M. Tsui and B. Scassellati, How people talk when teaching a robot, in *ACM/IEEE Int. Conf Human-Robot Interaction* (*HRI*) (ACM, 2009), pp. 23–30.
4. C. Breazeal, Robot in society: Friend or appliance, in *Proc. Autonomous Agents Workshop on Emotion-Based Agent Architectures* (1999), Seattle, pp. 18–26.
5. C. Torrey, A. Powers, M. Marge, S. R. Fussell and S. Kiesler, Effects of adaptive robot dialog on information exchange and social relations, in *ACM/IEEE Int. Conf. Human-Robot Interaction* (*HRI*) (ACM, 2006), pp. 126–133.
6. H. H. Clark, Talking as if, in *ACM/IEEE Int. Conf. Human-Robot Interaction* (*HRI*) (ACM, 2008), pp. 393–394.
7. S. Kopp, How people talk to a virtual human-conversations from a real-world application, *How People Talk to Computers, Robots, and Other Artificial Communication Partners* (2006), Bremen, Germany, p. 101.
8. J. Gustafson and L. Bell, Speech technology on trial: Experiences from the August system, *Natural Lang. Eng.* **6**(3–4) (2000) 273–286.
9. B. Pierce, T. Kuratate, A. Maejima, S. Morishima, Y. Matsusaka, M. Durkovic, K. Diepold and G. Cheng, Development of an integrated multi-modal communication

robotic face, in *IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)* (IEEE, 2012), pp. 101−102.
10. F. Delaunay, J. de Greeff and T. Belpaeme, A study of a retro-projected robotic face and its effectiveness for gaze reading by humans, in *ACM/IEEE Int. Conf. Human-Robot Interaction* (*HRI*) (ACM, 2010), pp. 39−44.
11. D. Bohus and E. Horvitz, Facilitating multiparty dialog with gaze, gesture, and speech, in *Int. Conf. Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction* (*ICMI-MLMI*) (ACM, 2010), pp. 5:1−5:8.
12. E. Moser, B. Derntl, S. Robinson, B. Fink, R. C. Gur and K. Grammer, Amygdala activation at 3T in response to human and avatar facial expressions of emotions, in *J. Neurosci. Methods* **161**(1) (2007) 126−133.
13. S. Al Moubayed, J. Edlund and J. Beskow, Taming mona lisa: Communicating gaze faithfully in 2d and 3d facial projections, *ACM Trans. Inter. Intell. Syst.* (*TiiS*) **1**(2) (2012) 11.
14. S. Al Moubayed, J. Beskow, G. Skantze and B. Granström, Furhat: A back-projected human-like robot head for multiparty human-machine interaction, in *Cognitive Behavioural Systems* (Lecture Notes in Computer Science, Springer, 2012), pp. 114−130.
15. S. Al Moubayed, G. Skantze and J. Beskow, Lip-reading Furhat: Audio visual intelligibility of a back projected animated face, in *Int. Conf. Intelligent Virtual Agents* (*IVA2012*) (Springer, 2012), pp. 196−203.
16. S. Al Moubayed and G. Skantze, Turn-taking control using gaze in multiparty human-computer dialog: Effects of 2d and 3d displays, in *Int. Conf. Auditory-Visual Speech Processing* (*AVSP*) (2011), Volterra, Italy.
17. D. Harel, Statecharts: A visual formalism for complex systems, *Sci. Comput. Program.* **8**(3) (1987) 231−274.
18. J. Gustafson, Developing multimodal spoken dialog systems, Empirical studies of spoken human-computer interaction, Doctoral Dissertation (KTH, 2002).
19. W. Swartout, D. Traum, R. Artstein, D. Noren, P. Debevec, K. Bronnenkant, J. Williams, A. Leuski, S. Narayanan, D. Piepol *et al.*, Ada and grace: Toward realistic and engaging virtual museum guides, in *Intelligent Virtual Agents* (Springer, 2010), pp. 286−300.

Nicole Mirnig is a Ph.D. research fellow in HCI at the HCI & Usability Unit, of the ICT & S Center, University of Salzburg, Austria. She holds a master's degree in communication studies from the University of Salzburg (2011). During her studies she specialized on the communication between humans and robots in social dialog situations. She is currently involved in the EU-project IURO (www.iuro-project.eu) that aims at developing interaction paradigms for an interactive urban robot. Her present research focuses on feedback in human−robot interaction. It will be the aim of her Ph.D. research to develop a taxonomy for multimodal robot feedback based on existing research, which she will in the next step in parts complete with the findings from her own studies on the topic.

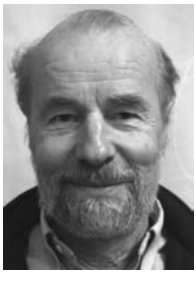
Astrid Weiss is a postdoctoral research fellow in HCI at the HCI & Usability Unit, of the ICT & S Center, University of Salzburg, Austria and she is also part of the Christian Doppler Laboratory on Contextual Interfaces at University of Salzburg. She holds a masters degree in sociology and a Ph.D. in social sciences from the University of Salzburg. During her studies she specialized on methodologies of empirical social research and applied statistics. Her current research focuses on user-centered design and evaluation studies for human−computer interaction and human−robot interaction. She is especially interested in the impact technology has on our everyday life and what makes people accept or reject technology.

Gabriel Skantze is a senior researcher (Docent) in speech technology at the Department of Speech Music and Hearing, KTH, Stockholm, Sweden. He holds a Ph.D. in speech communication from KTH. During his studies he specialized in error handling and miscommunication in spoken dialogue systems. His current research focus is on real-time models of spoken dialogue and empirical studies of human−robot interaction. He has participated in numerous EU projects related to dialogue systems and robotics, including CHIL, MonAMI and IURO. He is currently the principal investigator of a nationally funded project in the area of multimodal incremental dialogue processing.

Samer Al Moubayed is a postdoctoral researcher at the Department of Speech, Music and Hearing, at KTH, Stockholm, Sweden. He received his Ph.D. from KTH in 2012, for his studies and developments on human−robot face-to-face interaction. Samer has been part of several EU projects, including H@H, MonAMI, and IURO. His main work and interest are embodied dialogue systems, multimodal synthesis, and nonverbal social signal processing.

Joakim Gustafson has been a prolific researcher and active systems developer of spoken and multimodal dialogue systems since 1992. He has participated in several EU projects such as Onomastica, NICE, MonAmi, IURO and GetHomeSafe. He is currently the principal investigator in the nationally funded three-year research project introducing interactional phenomena in speech synthesis. He has an industrial background from TeliaSonera where, in addition to research, he was involved in launching of public speech applications. He is also member of the Editorial Board of the journal Speech Communication.

Jonas Beskow is an associate professor in speech technology and communication, with main research interests in the area of audio-visual speech synthesis, talking avatars and virtually and physically embodied conversational agents. He has participated in numerous EU projects related to multimodal speech technology in human−machine interaction and accessibility applications, including PF-STAR, SYNFACE, CHIL, MonAMI, HaH, IURO and LipRead. He is currently the principal investigator of two nationally funded projects in the area of sign language and gesture in face-to-face interaction. He is involved in two start-up companies in the domain of talking avatars, and is one of the developers of the open source speech processing tool WaveSurfer.



Björn Granström is the director of CTT, the Centre for Speech Technology and professor of Speech Communication at the Department of Speech, Music and Hearing, KTH, Stockholm, Sweden. Together with Rolf Carlson, he created the first multilingual text-to-speech system, with extensive use in the disability area. Granström is a founding member of ESCA (now ISCA) and ELSNET. He has organized several international meetings, such as ICPhS 1995, Eurospeech 2001, two ESCA workshops and a European summer school on Multi-modality in language and speech systems. Granström has published numerous papers in the speech research and technology area, including multi-modal speech technology. Present interests include multimodal verbal/nonverbal communication with applications in human−robot interaction, virtual language tutors and human-like spoken dialogue systems.

Manfred Tscheligi is a professor for the Human-Computer Interaction & Usability Unit at the University of Salzburg. He holds a master's degree in Business Informatics and a Ph.D. in Social and Economic Sciences (with a specialization in Applied Computer Science). Previously, he had been an Associate Professor for Applied Computer Science at the University of Vienna (Institute for Computer Science and Business Informatics) where he became a co-founder of CURE (Center for Usability Research & Engineering) in Vienna. Since December 2009, he has been directing the Christian Doppler Laboratory on "Contextual Interfaces" at the University of Salzburg and since 2011, he is also (co)-directing the ICT & S Center.

Manfred Tscheligi has been active in the area of Interactive Systems, Human-Computer Interaction, and Usability Engineering for more than 20 years. He has done pioneer work in this field and established it within Austria Universities as a research field and an industrially applied field. As a member of different expert groups, he is very active in the international research scene; among these, he has chaired and (co-)chaired several international conferences such as CHI2004, MobileHCI2005, ACE 2007, EuroiTV 2008, AmI2009, as well as AutomotiveUI2011. Manfred Tscheligi is the author of several publications and a distinguished speaker at conferences and seminars. He has been member of many programme committees and editorial boards (e.g. Book Series "Human-Computer Interaction" (Kluwer), ACM Interactions Magazine (ACM)). He has been responsible for more than 200 national and international projects (basic research, applied research, industrial co-operations) and several national and international initiatives. Among these, he has been actively involved in more than 20 EC projects.