Intentionality in Interacting with Companion Systems – An Empirical Approach

Dietmar Rösner[1], Rafael Friesen[1], Mirko Otto[1], Julia Lange[2], Matthias Haase[2], and Jörg Frommer[2]

[1] Department of Knowledge Processing and Language Engineering, Otto-von-Guericke University, Magdeburg, Germany
{roesner,friesen,miotto}@ovgu.de
[2] Department of Psychosomatic Medicine and Psychotherapy, Otto-von-Guericke University, Magdeburg, Germany
{joerg.frommer,matthias.haase,julia.lange}@med.ovgu.de

Abstract. We report about a WOZ experiment with a carefully designed scenario that allows to investigate how users interact with a companion system in a mundane situation with the need for planning, re-planning and strategy change. The data collection from the experiments comprises multimodal records (audio, video, biopsychological parameters) and transcripts of the verbal interaction, and all subjects fill out a battery of well established psychometric questionnaires about various aspects especially of their personality. This will allow to correlate observed behaviour and detected affects and emotions with measured aspects of the personality of subjects and is expected to serve as a basis for defining a typology of users. In addition, a subgroup of the subjects takes part in semiformal in-depth interviews that focus on retrospective reflexion of the users' subjective experience during the experiments and especially on the intentionality that users ascribed to the system during the course of interaction.

Keywords: Intentionality, Wizard of Oz, Companion Systems, Emotion, Multimodal.

1 Introduction

Companion systems are an important class of emerging future devices. These can be defined as: "[...] conversationalists or confidants – not robots – but rather computer software agents whose function will be to get to know their owners, who may well be elderly or lonely, and focusing not only on assistance via the internet (contacts, travel, doctors etc.) that many still find hard to use, but also on providing company and companionship, by offering aspects of personalization" [16].

In order to realize such systems, psychological and social issues of human-computer-interaction (HCI) have to be taken into account [16]. However, studies in this field concentrate on the users' emotion, reactance and cooperativeness.

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2011, LNCS 6763, pp. 593–602, 2011.

One aspect, which is furthermore relevant for the acceptance and usability of such systems, has not been investigated enough: the issue of intentional stance [3] towards companion systems. Hereby, explanations and predictions of the systems' behaviour are based upon the presumption that the system holds certain information, pursues certain goals and chooses a behavioural pattern on that background, which is rational and appropriate to the situation. This means that humans ascribe intentional states to the system such as hopes, worries, perceptions, emotions etc. We have therefore designed and implemented a Wizard of Oz (WOZ) experiment 'last minute' that allows to empirically investigate the following issues: intentional stance and user emotions during HCI, the risk of breaking off the dialogue, and the classification of different user-types. User types are discriminated by e.g. specific markers in spoken language, certain psychological constructs, socio-demographic characteristics and the users' subjective experience of HCI. This experiment is currently conducted with users from two age groups: young adults between 18 and 28 and seniors above 60, balanced in gender and educational level.

In the following we elaborate on the design rationale of the szenario of the WOZ experiment and indicate first findings from the ongoing experiments.

2 The Intentional Stance in Interacting with Companion Systems

The philosopher Daniel C. Dennett described three stances humans can adopt towards a biological (including people) or non-biological system to explain or predict its behaviour: physical, design and intentional stance [3].

By using the term "intentional" Dennett refers to the philosophical term "Intentionality" as described by Franz Clemens Brentano [4,3]. He writes "the intentional strategy consists of treating the object [...] as a rational agent with beliefs and desires and other mental stages exhibiting what Brentano and others call intentionality" ([4] p. 15).

Companion systems are designed for reacting individually and empathetically to users. Their construction plan and the physical processes inside the systems are complex in a way that an average user is unable to predict or explain the behaviour of the system in the design or physical stance. Hence, for an average user adopting the intentional stance is the most natural and pragmatic way of predicting and explaining the behaviour of such systems successfully. It is desirable that users assume positive intentions in companions system like helpfulness, trust-worthiness and empathy. These are able to increase the acceptance and usability of companion systems, whereas assumptions of negative intentions like malice, pursuit to dominance and poor willingness to cooperate may contribute to their decrease.

3 A Wizard of Oz Experiment

The WOZ experiment was designed to investigate how users interact with a companion system in a mundane situation with the need for planning,

replanning and strategy change. The dialogue processes experimentally induce critical events that provoke negative emotions and as a consequence possibly break off the dialogue. Our aims are (1) to find out features in users' speech content that allow a prediction of dialogue crisis, e.g., manifestations of negative emotions, (2) to investigate the usefulness of an empathic intervention in dialogue crisis and (3) to reconstruct how users experienced the interaction and its critical events, especially which emotions and intentional ascriptions occurred. Critical events are not induced by a "dysfunction" of the companion system but by an enforced strategy-change and additionally occurring constraints that complicate task completion. The software architecture is described in [13]. The WOZ experiment comprises two submodules [6]. In the first submodul "initial narrative" the subject undergoes an initialisation by a personalization dialogue. This simulates that the CS is individualized for one user to adapt to him or her, to stay with him or her for a long time and to support him or her in everyday life tasks - the most relevant characteristics of companion systems. The second submodule "last minute" is described in detail in the following paragraph.

3.1 Wizard of Oz Scenario

The systemic functional linguistics model of Halliday [8] analyzes three metafunctions of language: ideational, interpersonal and textual. The ideational metafunction bears the content of speech. Speech in current HCI is focussed primarily on this metafunction. To be reliable and empathic companion systems have to adapt to the user's situation, skills, preferences, aims, needs and emotions. Thus companion systems have to communicate not only on the ideational metafunction but employ also the textual and interpersonal metafunction.

This is simulated and investigated in the second submodule "last minute" consisting of six stages (Fig. 1).

1. The user is asked to imagine a situation in summer when he as a surprise gets informed to have won a two weeks vacation. The prize includes as well the opportunity to select all necessary items for the travel from an online catalogue organized in a number of categories (e.g. coats, underwear, etc.).

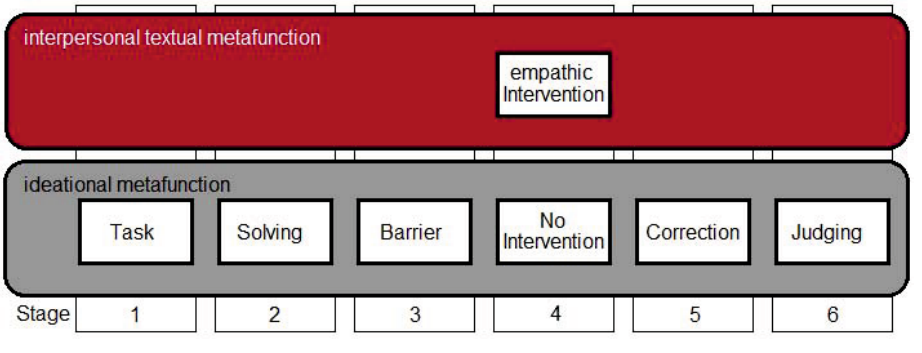

Fig. 1. Stages of the WOZ experiment

For this selection fifteen minutes are available, because then the taxi to the airport will arrive and the packed suitcase will be available at check in. This cover story is provided very detailed to stimulate ego involvement and imagination of the subject. The task to choose equipment and to pack a suitcase is introduced and some (time limit), but not all (weight limit of suitcase) limitations are given.

2. The subject can choose items in a uniform procedure from 12 categories in a predefined sequential order. The system informs the subject when a limitation (weight limit of the suitcase) is infringed and provides information about the task's status.
3. The system informs the subject that the destination is in the southern hemisphere and the forecast estimates temperatures around 0 °C instead of the summer temperatures assumed by the subject. This information lets the subject realize that he had aimed a wrong goal and has now to change his strategy. This might lead to frustration and anger.
4. The system addresses the subject using the interpersonal and textual metafunctions with increased intensity asking questions about the subject's feelings and motivation. Such selected dialogue strategies create the chance for reflection and expression of anger and provide empathic help based on the principles of Rogers' paradigm of client centered psychotherapy [15].
5. The subject can select from chosen categories a second time to readjust the equipment to the cold weather. The system's speech is focussed primarily on the ideational metafunction again.
6. After finishing the selection from the remaining categories, the subject is asked to rate his own performance: how content he is with the packed suitcase and if the subject would start the trip with the suitcase's content.

3.2 An Example Interaction

The following excerpts are anonymized and translated examples of a dialogue in German from our first experiments. Please note that we tried to preserve the subject's talking style. (U=user, S=system, [...] parts left out)

Stage 1: Personalisation (Task)

```
   [...]
S: At the beginning some information about you is necessary. Can you
   please introduce yourself?
U: Yes. My name is Max, coming Magdeburg, and doing this test.
S: The system needs more information to adapt itself to you. Can you
   please make a statement about the following points: Your age, your job,
   your jobs location, your family, your dress size, your size of shoe.
U: Yes, my size of shoe is 42, I am 22 years old, my birthday is
   1.1.1990, I am single, have no girlfriend, work in construction,
   everywhere in Magdeburg, and yes.
S: Please complete the information about your body size.
   [...]
```

S: Please describe a recent event that you enjoyed.
[...]
S: Please describe a recent event that annoyed you.
[...]

Stage 2: Selection (Solving)

[...]
U: Next section.
S: You can now choose from the section 'coats and cloaks'.
U: Yeah. I would take a macintosh, or two.
S: A macintosh was added. You can continue.
U: And a jacket.
S: A jacket was added. You can continue.
[...]

"Weight Limit"

U: Two swimming trunks.
S: Two swimming trunks can not be added, otherwise the bag's maximal weight limit imposed by the airline would be exceeded. You have to care for enough space in the bag before further items can be chosen. This can be done by unpacking already packed items. If requested you get a listing of the already packed items.
U: One swimming trunk, does it fit?
S: The item swimming trunk can not be added, otherwise the bags maximal weight limit would be exceeded.
U: What do I have in my bag?
S: The following items were already selected: One macintosh, one jacket, ...
[...]

Stage 3: Barrier "Climate Explanation"

S: The transmission of additional information about the holiday destination Waiuku is completed. Waiuku is located on the northern island of New Zealand in the Franklin district. [...] Do you wish this information to be repeated?
U: No.
S: The months November until February are the summer months in New Zealand. With an average temperature of 20 degree Celsius they are the warmest. In the winter months June until August average temperatures of only five to eight degree Celsius are reached. [...] For the time of your holidays the forecast gives temperatures about two to five degree and felt minus four to one degree Celsius. For the time of the forecast strong rain and hail are predicted.

Stage 4: Intervention (Note: Half of the participants get an intervention like this.)

S: Because of a broken data line the information about the destination could not be provided earlier. This may have changed your situation surprisingly. The selected items show that you expected different weather. If you knew the weather at the destination, you might have had selected different items. I am interested in your opinion about this.
U: Yes, I would have.
S: Did this situation evoke negative emotions? If yes, can you describe them?
U: [...]
S: I hope, your motivation to work on this task did not decrease too much. [...]

3.3 Questionnaires

To validate and extend the data collected during the experiment different psychological constructs are measured post hoc with the following questionnaires (german versions), too:

- Attributional-Style-Questionnaire (ASF-E [14]) measuring individual differences in attributional style,
- NEO-FFI [2] based on the Five Factor theory of personality,
- Inventory of Interpersonal Problems (IIP [10]) based on Sullivan's model of personality and focusing interpersonal relationships,
- Stress-coping questionnaire (SVF [11]) measuring different types of response to an unspecific selection of situations that impair, adversely affect, irritate or disturb the emotional stability or balance of the subject,
- TA-EG [12] measuring the technophily,
- Emotion Regulation Questionnaire (ERQ [7]) to ask for the two concepts of emotion regulation, the strategy of cognitive reappraisal and the suppression of emotional expression,
- BIS/BAS [1] based on Gray's theory of the Behavioral Inhibition System (BIS) and the Behavioral Approach System (BAS),
- AttrakDiff [9] measuring the subjective rating of usability and user experience.
- In addition to these psychometric instruments sociodemographic variables are collected like age, gender, educational level, experience with computers (e.g. years overall, hours per day/week) and in what context the subjects use the computer.

Based on the data from the experiments and the self-rating questionnaires different user types shall be differentiated.

3.4 Interviews

Furthermore a subgroup of participants undergoes an additional semi-structured interview of 60 to 160 Minutes after the experiment to determine their intentional stance on the computer system, their emotions during HCI and to identify further characteristics in their subjective experience of the HCI that support the differentiation of user-types.

At the beginning a so called narrative stimulus is given to evoke a narration about the subjective experiences during the WOZ-experiment: "*You have just done an experiment. Please try to recall this experiment. Please tell me, how you did during the experiment. Please tell me in detail, what you thought and experienced in the different situations!*". Afterwards, the interviewer tries to fill up gaps in the narration by so called immanent questions. Further exmanent questions focus on: occurred user emotions, intentional ascriptions towards the CS, the experience of the speech based interaction, the experience of the intervention (if given), the role of technique in autobiography and the general evaluation of the simulated CS. Methods of reconstructive qualitative social research that aim at the comprehension of the subjective experience of social reality and the related subjective meaning (here with respect to HCI) [5] will be used to analyse the interview material.

4 Evaluation

4.1 First Observations

As first inspections of the recordings show users very quickly adapt to restrictions that they experience. For example: wizards deliberately do not process conjunctions. If a user uses a conjunction in a selection (e.g. "*Three t-shirts and two jeans.*") then only the first conjunct is processed and the system gives resp. feedback ("*Three t-shirts have been added.*"). Most, but not all, users seem to quickly adapt to such implicit restrictions.

First unsystematic analyses of the interviews' material showed, that different feelings occur during the interaction with the system. Many subjects report annoyance, e.g., because the system did not warn about the weight limit in advance but always informed about its exceedance only after the subject spent time on trying to pack another item. Other subjects report uncertainty, e.g. when the system wants them to introduce themselves ("*I thought my god what shall I tell now*").

Furthermore different intentional ascriptions occur, e.g., one subject experienced the system as trustworthy in informing about the weather conditions ("*I trusted the computer when he said that it is that cold there, because of that I repacked my suitcase*"). Another ascription is malice, especially when the system gives additional information about the weather conditions at the holiday resort ("*I thought, well he wants to screw me, doesn't he*", "*now he wants to make fun of me*").

Metaphors, that many subjects use to explain how they experienced the system, seem to be very interesting to get insights into the user's subjective experience of

the system, too. For example, one subject experienced the system as "*a puppet, that wants to speak with someone*" (in regard to the initial phase of the experiment).

Most of the subjects seem to try to adapt to the system, e.g., one subject reported that she had to think about what to answer when the system asked for events in the use of technical devices she experienced as positive and as negative ("*I had to pore over what I can tell now, what the machine is able to understand*").

4.2 Interaction Analysis of Transcripts

The subjects in our WOZ experiments only get told that they will interact in spoken language with a new type of system that shall be personalised and that therefore will ask some questions and pose some tasks. They do not get explicit instructions about the linguistic constructions and interaction patterns that are possible or not possible to use in the interaction.

How do people with differing technical background interact in such a situation?

One might speculate that there are a number of different possible reactions in such a situation. People could on the one hand actively try to find out the possibilities and limitations of the system by starting with – from their perspective – simple constructions and then gradually further trying more complicated ones. On the other hand people could simply rely on their intuitive assumptions (their internal picture based on whatever prior experiences or projections) about such possibilities and limitations. In the latter case no active exploration will be tried and behaviour – even if unnatural or inefficient – will be pursued as long as it proves to be successful. An example: One subject e.g. told in the post-hoc interview that in a situation where he did not know the term "*Konfektionsgrösse*" (engl. size of clothing) that he was asked about by the system he would have preferrred to do a Google search. "*But you could not ask the system for this!*", he continued, without having ever tried.

For the evaluation of the records and transcripts of the interaction there are many interesting questions. In the following we will name just a few.

Do users differ in their curiosity and there openness for experimentation? If so, how does this relate to results from the psychometric questionnaires?

Do users – although system initiative is dominant – at least try to take initiative by e.g. trying to ask questions? If so how do they react when the system is ignoring such attempts?

Do users e.g. try to interrupt lengthy system outputs (barge-in)?

Do users mirror the language of the system, e.g. on the lexical or syntax level? The system e.g. uses the general and somewhat uncolloquial term "*hinzufügen*" (engl. to add) in its feedback for selection operations. Do users mirror this usage?

How quickly do users adapt to observed limitations of the system? An example: are there any re-trials once a user experiences that from a conjunction only the first conjunct gets processed by the system?

How quickly do users employ efficient communication strategies like the use of ellipses in consecutive dialogue turns? Are there any subjects that stick to full blown sentences in a sequence of dialogue turns with a high potential for ellipses?

Do users employ synonyms (or hypernyms or other semantic equivalents) when they choose objects from the selection menus or – even more to be expected – when they unpack items from the suitcase? How do they react when the system refuses to understand a semantically correct unpacking command that does not use the term from the selection menu but a semantically equivalent one (e.g. "*remove three pairs of socks*" instead of "*remove three pairs of stockings*")?

Do users employ a kind of simplified language, e.g. without proper inflection, as they would use towards foreigners without sufficient knowledge of German?

These are some of the questions which will be focussed on in the first round of in-depth analyses.

5 Conclusion

At the time of writing ca. 60 of the planned 120 experiments of the current series are performed (including questionnaires and post hoc interviews). According to our schedule all 120 experiments will be completed in April 2011. The verbal interactions are transcribed in close relation to the resp. WOZ sessions.

Besides the aspects explained above the collected data will be used to explore the users' subjective experience during the interaction. These findings will also contribute to the description of user types in the interaction with companion systems on the one hand, and on the other hand to the development of adaptive dialogue strategies. Both shall contribute that the users' willingness to cooperate even in critical situations of the interaction is preserved.

On the basis of the reconstructed subjective experience of the interaction (especially the occurring emotions and ascriptions) dialogue strategies of the system should be identified, that are experienced as empathic or cooperation-hindering. This will contribute to develop dialogue strategies to support the development of positive ascriptions and to avoid the development of negative ones.

The current series of experiments will provide us with data and insights that stipulate future experiments. Based on the technical infrastructure and on the framework set up we will be able to explore systematically variants of the szenario. The overall issue will be, how subjects do experience the interaction with the (simulated) system when relevant system parameters are changed (e.g. quality of the TTS employed, allowed language fragment, varying NLU capabilities, talking style, initiative type, exhibited intelligent behaviour of the system).

In sum: We expect that the findings, insights and conclusions from our experiments will provide input for the ongoing efforts to develop the next generation of companion systems.

Acknowledgment

The presented study is performed in the framework of the Transregional Collaborative Research Centre SFB/TRR 62 "Companion-Technology for Cogni-

tive Technical Systems" funded by the German Research Foundation (DFG). The responsibility for the content of this paper lies with the authors. Special thanks goes to the participants of the Workshop "Linguistische Analysen in der Mensch-Maschine-Interaktion" (Linguistic analyses in human-machine-interaction) in Magdeburg, Oct. 15–16, 2010.

References

1. Carver, C.S., White, T.L.: Behavioral Inhibition, Behavioral Activation, and Affective Responses to Impending Reward and Punishment: The BIS/BAS Scales. Journal of Personality and Social Psychology 67, 319–333 (1994)
2. Costa, P., McCrae, R.: NEO-PI-R Professional Manual. Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI). Psychological Assessment Resources, Odessa (1992)
3. Dennett, D.C.: Intentional Systems. Journal of Philosophy 68, 87–106 (1971)
4. Dennett, D.C.: The Intentional Stance. The MIT Press, Cambridge (1987)
5. Flick, U.: An Introduction to Qualitative Research, 4th edn. Sage Publications Ltd., London (2009)
6. Frommer, J., Haase, M., Lange, J., Rösner, D., Friesen, R., Otto, M.: Project A3 'Prevention of Negative Dialogue Courses' Wizard of Oz Experiment Operator Manual, SFB-TRR-62 working paper (2010) (unpublished)
7. Gross, J.J., John, O.P.: Individual Differences in Two Emotion Regulation Processes: Implications for Affect, Relationships, and Well-Being. J. Pers. Soc. Psychol. 85, 348–362 (2003)
8. Halliday, M.: A Brief Sketch of Systemic Grammar. System and Function in Language, 3–6 (1976)
9. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Ziegler, J., Szwillus, G. (eds.) Mensch & Computer 2003, pp. 187–196. B.G. Teubner, Stuttgart (2001)
10. Horowitz, L., Alden, L., Wiggins, J., Pincus, A.: Inventory of Interpersonal Problems Manual. The Psychological Corporation, Odessa (2000)
11. Jahnke, W., Erdmann, G., Kallus, K.: Stressverarbeitungsfragebogen mit SVF 120 und SVF 78, 3rd edn. Hogrefe, Göttingen (2002)
12. Karrer, K., Glaser, C., Clemens, C., Bruder, C.: Technikaffinität erfassen – der Fragebogen TA-EG. Der Mensch als Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme 29, 196–201 (2009), zMMS Spektrum, Reihe 22
13. Otto, M., Friesen, R., Rösner, D.: Message Oriented Middleware for Flexible Wizard of Oz Experiments in HCI. In: HCI International 2011. vol. this volume (2011)
14. Poppe, P., Stiensmeier-Pelster, J., Pelster, A.: Attributionsstilfragebogen für Erwachsene (ASF-E). Hogrefe, Göttingen (2005)
15. Rogers, C.: A Theory of Therapy, Personality and Interpersonal Relationships as Developed in the Client-centered Framework. In: Koch, S. (ed.) Psychology: A Study of a Science, Formulations of the Person and the Social Context, vol. 3, McGraw Hill, New York (1959)
16. Wilks, Y.: Close Engagements with Artificial Companions: Key Social, Psychological and Design issues. John Benjamins, Amsterdam (2010)