Perception of Artificial Agents and Utterance Friendliness in Dialogue

Sascha Griffiths[1] and **Friederike Eyssel**[2] and **Anja Philippsen** and **Christian Pietsch** and **Sven Wachsmuth**[3]

Abstract. The present contribution investigates the construction of dialogue structure for the use in human-machine interaction especially for robotic systems and embodied conversational agents. We are going to present a methodology and findings of a pilot study for the design of task-specific dialogues. Specifically, we investigated effects of dialogue complexity on two levels: First, we examined the perception of the embodied conversational agent, and second, we studied participants' performance following HRI. To do so, we manipulated the agent's friendliness during a brief conversation with the user in a receptionist scenario.

The paper presents an overview of the dialogue system, the process of dialogue construction, and initial evidence from an evaluation study with naïve users ($N = 40$). These users interacted with the system in a task-based dialogue in which they had to ask for the way in a building unknown to them. Afterwards participants filled in a questionnaire. Our findings show that the users prefer the friendly version of the dialogue which scored higher values both in terms of data collected via a questionnaire and in terms of observations in video data collected during the run of the study.

Implications of the present research for follow-up studies are discussed, specifically focusing on the effects that dialogue features have on agent perception and on the user's evaluation and performance.

1 Introduction

Research within the area of "language and emotion" has been identified as one key domain of innovation for the coming years [40, 20]. However, with regard to human-machine communication, we still need better speech interfaces to facilitate human-robot interaction (HRI) [30, 31]. Previous work on human-human communication has already demonstrated that even small nuances in speech have a strong impact on the perception of an interlocutor [1, 38].

In the present work, we have therefore focused on the role of dialogue features (i.e., agent verbosity) and investigated their effects on the evaluation of an embodied conversational agent (ECA) and the user performance. We designed a receptionist scenario involving a newly developed demonstrator platform (see Section 3.2) that offers great potential for natural and smooth human-agent dialogue. To explore how to model dialogues efficiently within actual human-robot interaction we relied on a Wizard-of-Oz paradigm [16, 17].

This HRI scenario involved an embodied conversational agent which served as a receptionist in the lobby of a research center. A similar set-up has been realized in previous studies [2, 24, 25]. Moreover, we draw from existing research on dialogue system design [33] and the acceptance of artificial agents [13, 22].

The question that we seek to answer arises frequently during the implementation of a robot scenario (such as this receptionist scenario) [26], and can also be phrased as how the system should verbalize the information that it is supposed to convey to the user. Obviously, a script has to be provided that covers the necessary dialogue content. The relevant issue is that each utterance can be phrased in a number of ways. This brings up several follow-up questions such as: *Can the perceived friendliness of an agent be successfully manipulated? Is the proposed script a natural way of expressing the intended meaning? Are longer or shorter utterances favourable? How will the user respond to a given wording? Will the script elicit the appropriate responses from the user?*

For the purpose of investigating these questions, we will first discuss related literature and relevant theoretical points. The following section will describe the system. We then turn to the dialogue design and first empirical evidence from a user study.

2 Dialogue Complexity and Perception of Artificial Agents

Obviously, the issue of how to realize efficient dialogue in HRI has been of interest to many researchers in the area of human-machine interaction and principles of natural language generation are generally well understood [39]. However, this is less so the case when taking into account communication patterns between humans and embodied conversational agents and robots.

2.1 Dialogue Complexity and Social Meaning

As Richard Hudson notes, "social meaning is spread right through the language system" [23]. Thus, there is a clear difference between interactions if one commences with the colloquial greeting "*Hi!*" versus one initiated with a more polite "*Good Morning*". However, this does not only concern peripheral elements of language such as greetings, but also syntax. Hudson uses the following example to illustrate this:

1. *Don't you come home late!*
2. *Don't come home late!*

Both sentences differ in terms of syntax and their social meaning. The syntax varies as the first sentence explicitly refers to the subject,

[1] Queen Mary University of London, UK, email: sascha.griffiths@qmul.ac.uk
[2] New York University, Abu Dhabi, email: fae5@nyu.edu
[3] Bielefeld University, Germany, email: anja.philippsen, christian.pietsch, sven.wachsmuth@uni-bielefeld.de

whereas the second sentence does not. The first sentence in the example also appears more threatening in tone than the latter. These subtle differences in the statements' wording lead to a fundamentally different interpretation. Analogously, we assume that in human-agent dialogue subtle manipulations of aspects of that dialogue can result in changes in agent perception. Concretely, we will investigate the role of this kind of linguistic complexity [11] within human-machine interaction.

The impact of changing a dialogue with respect to the social meaning communicated has already been tested in the REA (an acronym for "Real Estate Agent") system [9, 5]. In a study [4] of users' perception of different versions of REA's behaviour, a "normal REA" was tested against an "impolite REA" and a "chatty REA". Results indicated that in the condition in which REA was able to produce a small amount of small talk REA was judged more likeable by participants. In further studies with the system the authors concluded that the interpersonal dimension of interaction with artificial agents is important [8]. It has been shown that implementing a system which achieves task goals and interpersonal goals as well as displaying its domain knowledge can increase the trust a user will have in a system [3]. Cassell [7] also argues that equipping artificial agents with means of expressing social meaning not only improves the users' trust in the domain knowledge that such systems display but also improves interaction with such systems as the users can exploit more of their experience from human-human dialogue.

2.2 Interaction Patterns

The dialogue flow used in the present study was implemented with PaMini, a pattern-based dialogue system which was specifically designed for HRI purposes [32] and has been successfully applied in various human-robot interaction scenarios [35, 36, 37]. The dialogue model underlying the present system (see Section 3.1) is therefore based on generic interaction patterns [33]. Linguistically speaking these are adjacency pairs [29, 10]. In these terms, a dialogue will consist of several invariant elements which are sequentially presented as pairs with one interlocutor uttering one half of the pair in his turn and the other interaction partner responding with an appropriate response.

The full list of generic interaction patterns which are distinguished according to their function given by Peltason et al. [34] includes the following utterance categories: *Greeting, Introducing, Exchanging pleasantries, Task transition, Attracting attention, Object demonstration, Object query, Listing learned objects, Checking, Praising, Restart, Transitional phrases, Closing task, Parting*.

For all these dialogue tasks one can see the interaction as pairs of turns between interlocutors. Each partner has a certain response which fits to the other interlocutor's utterance. Examples of this kind of interaction can be found in Table 1.

Table 1. Examples of adjacency pairs in human-robot interaction (adapted from [34])

Purpose	Example interaction
Greeting	*User: Hello, Vince.* *Robot: Hi, hello.*
Introducing	*User: My name is Dave.* *Robot: Hello, Dave. Nice to meet you.*
Object query	*Robot: What is that?* *User: This is an apple.*
Praising	*User: Well done, Vince.* *Robot: Thank you.*

The problem one faces is that while such dialogues are based on generic speech acts, there is the remaining problem of how the individual items need to be worded. Winograd [46] distinguishes between the ideational function and interpersonal function of language. The ideational function can loosely be understood as the propositional content of an utterance whereas the interpersonal function has more to do with the context of an utterance and its purpose.

3 System Architecture

In the following, we present the system which was constructed both as a demonstrator and as a research platform. We will present the entire set-up which includes an ECA, Vince [42], and a mobile robot platform, Biron [21]. Both of these use the same dialogue manager but only the ECA has been used in this pilot study.

Figure 1 illustrates the architecture of the complete system in autonomous mode. Communication between the components is mainly implemented using the XML-based XCF framework and the Active Memory structure [47]. Three memories are provided for different kinds of information: The short term memory contains speech related information which is inserted and retrieved by the speech recognizer, the semantic processing unit and the dialogue manager. The visual memory is filled by the visual perception components, it contains information about where persons are currently detected in the scene.

The system is designed to provide the visitor verbally with information, but also to guide them to the requested room if necessary[4]. For this purpose, the agent Vince communicates information about the current visitor and his or her needs to the mobile robot Biron via a shared (common ground) memory.

Although Biron is omitted in the present study to reduce complexity, we present the complete system, as Vince and Biron use the same underlying dialogue system. Note that the study could have been conducted also with Biron instead of Vince. Such a study is subject to future work.

3.1 Dialogue Manager

The dialogue manager plays a central role in the overall system as it receives the pre-processed input from the user and decides for adequate responses of the system. A dialogue act may also be triggered by the appearance of persons in the scene as reported by the visual perception component.

Speech input from the user is recognized using the ISR speech recognizer based on ESMERALDA [14]. The semantic meaning is extracted via a parsing component which is possible due to the well defined scenario. Additionally, this component retrieves missing information from an LDAP server that the human might be interested in (e.g. office numbers). The dialogue manager PaMini [35, 36, 37] is based on finite state machines which realize interaction patterns for different dialogue situations as described in Section 2.2. Patterns are triggered by the user or by the robot itself (mixed-initiative). The dialogue component sends the selected response and possibly gesture instructions to the Vince system which synchronizes the speech output and the gesture control internally [28, 27]. Exploiting the information from the visual perception component, Vince attends to the current visitor via gaze following [24].

Biron incorporates a separate dialogue which is coupled with the Vince dialogue. The Biron dialogue at the moment receives input

[4] A short video demonstration of the scenario is provided in this CITEC video: `http://www.youtube.com/watch?v=GOz_MsLel1Y#t=4m32s`. Accessed: March 2, 2015

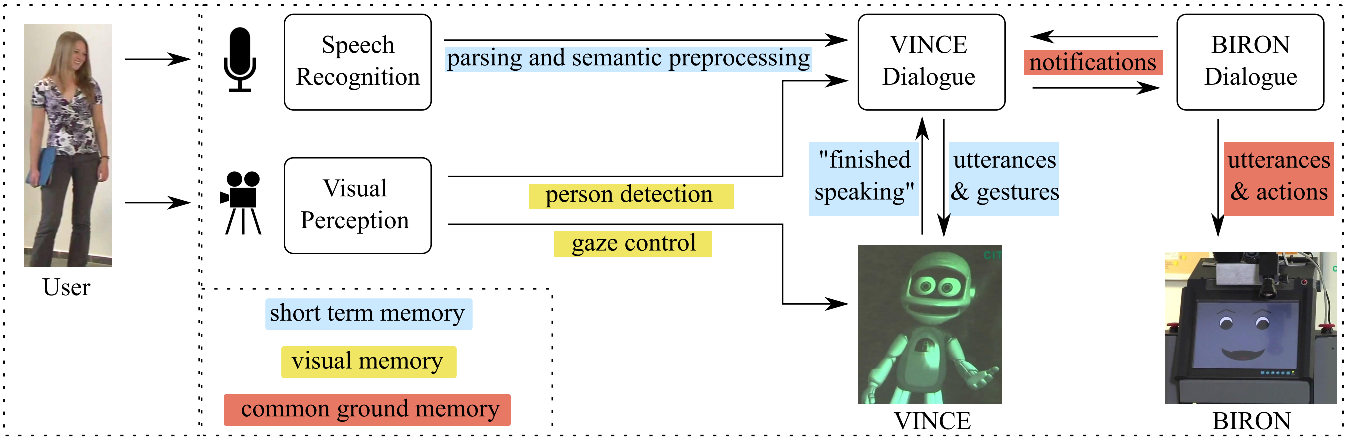

Figure 1. Overview of the architecture of the system in autonomous mode. The colors of the three memories indicate which information is stored in which memory. See Section 3.1 for a thorough description of the information flow.

solely from the Vince dialogue component (not from the user) and communicates the current state to the user. If the visitor wishes, Vince calls Biron and orders him to guide the visitor to the requested room. This feature is currently limited to offices on the ground floor, if visitors are looking for a room on the first or second floor, Biron guides them to the elevator and provides them with information about how to find the room on their own.

3.2 Demonstrator Platform

The embodied conversational agent Vince is installed on a workstation. An Apple Mac Mini is used for this purpose. The system runs a UNIX based operating system (Linux Ubuntu 10.04 32bit). The user interface is controlled by a wireless bluetooth mouse and keyboard or via remote access. The ECA is displayed on a holographic projection screen (i.e. a HoloPro Terminal[5]) in order to achieve a high degree of perceived embodiment. A microphone records speech input and video data are recorded using two cameras. Two loudspeakers are connected to the Mac Mini workstation to provide audio output.

4 Study Design and Realisation

We set up a simplified version of the CITEC Dialogue Demonstrator for the purpose of the study. One difference is that we do not make use of the mobile robot Biron here. Secondly, we rely on Wizard-of Oz teleoperation [12, 45] to trigger interaction patterns by means of a graphical user interface that was designed for our case study.

4.1 Preparation of Dialogues

The dialogues were prepared bottom-up. We tried to leave as little as possible to design by the researchers or a single researcher.

To investigate human-machine dialogue in the context of a receptionist scenario, we initially simulated such dialogues between two human target persons who were given cards which described a particular situation (e.g. that a person would be inquiring about another persons office location).

We recorded two versions of eight dialogues with the two participants, who were asked to take the perspective of a receptionist or a visitor, respectively. The dialogues were then transliterated by a third party who had not been involved in the staged dialogues.

To model the receptionist turns, we extracted all phrases which were classified as greetings, introductions, descriptions of the way to certain places and farewells. We then constructed a paper-and-pencil pre-test in order to identify a set of dialogues that differed in friendliness. 20 participants from a convenience sample were asked to rate the dialogues with regard to perceived friendliness using a 7-point Likert scale.

These ratings were used as a basis to construct eight sample dialogues which differed both in friendliness and verbosity. In a subsequent online pre-test, the sample dialogues were embedded in a cover-story that resembled the set-up of our WoZ scenario.

We used an online questionnaire to test how people perceived these dialogues. On the start screen participants were presented with a picture of the embodied conversational agent Vince and told that he would serve as a receptionist for the CITEC building. On the following screens textual versions of the eight human-agent dialogues were presented. Participants were asked to rate these dialogues with regard to friendliness in order to identify dialogues that would be perceived as either low or high in degree of perceived friendliness of the interaction.

The dialogue with the highest rating for friendliness and the dialogue with the lowest rating for friendliness were then de-composed into their respective parts and used in the main study. The two dialogue versions are presented in Table 2.

4.2 Study

In the main study, the participants directly interacted with the ECA which was displayed on a screen (see Figure 1).

We recruited students and staff at the campus of Bielefeld University to participate in our study on "human-computer interaction". 20 male and 20 female participants ranging in age from 19 to 29 years (M = 23.8 years, SD = 2.36) took part in the study. Before beginning their run of the study, each participant provided informed consent. Each participant was then randomly assigned to one of two conditions in which we manipulated dialogue friendliness.

The study involved two research assistants (unbeknownst to the participants). Research assistant 1 took over the role of the "wizard" and controlled the ECA's utterances, while research assistant 2 interacted directly with the participants.

[5] `http://www.holopro.com/de/produkte/holoterminal.html` Accessed: March 2, 2015

Table 2. Friendly and neutral dialogue version

Dialogue Act	Neutral version	Friendly version
Greeting	Hallo *Hello*	Guten Tag, kann ich Ihnen helfen? *Good afternoon, how can I help you?*
Directions	Der Fragebogen befindet sich in Q2-102. *The questionnaire is located in Q2-102.*	Der Fragebogen befindet sich in Raum Q2 102. Das ist im zweiten Stock. Wenn Sie jetzt zu Ihrer Rechten den Gang hier runter gehen. Am Ende des Gangs befinden sich die Treppen, diese gehen Sie einfach ganz hoch und gehen dann durch die Feuerschutztür und dann ist der Raum einfach geradeaus. *The questionnaire is located in room Q2-102. That is on the second floor. If you turn to your right and walk down the hallway. At the end of the floor you will find the stairs. Just walk up the stairs to the top floor and go through the fire door. The room is then straight ahead.*
Farewell	Wiedersehen. *Goodbye.*	Gerne. *You are welcome.*

Following the Wizard-of-Oz paradigm, research assistant 1 was hidden in the control room and controlled the ECA's verbalisations using a graphical user interface. A video and audio stream was transmitted from the dialogue system to the control room. The "wizard" had been trained prior to conducting the study to press buttons corresponding to the "Dialogue Acts" as shown in Table 2. Importantly, research assistant 1 only knew the overall script (containing a greeting, a description of the route to a room and a farewell), but was blind to the authors' research questions and assumptions.

To initiate the study, research assistant 1 executed "Greeting A" or "Greeting B", depending on whether the "friendly" or "neutral" condition was to be presented, then proceeded to pressing "Directions A" or "Directions B" and finally "Farewell A" and "Farewell B" once the user had reacted to each utterance.

The users then had to follow the instruction given by the agent. Research assistant 2 awaited them at the destination where they had to fill in a questionnaire asking for their impressions of the interaction.

The questionnaire investigated whether differential degrees of dialogue complexity would alter the perception of the artificial agent with respect to a) warmth and competence [15], b) mind attribution [19], and c) usability (system usability scale *SUS*) [6]. We consider these question blocks as standard measures in social psychology and usability studies.

The questionnaire was comprised of three blocks of questions. These do to some extent correspond to the four paradigms of artificial intelligence research listed in Russell & Norvig [41]: "thinking humanly", "acting humanly", "thinking rationally" and "acting rationally". As we were only looking at perception of the artificial agent, we did not look into "thinking rationally". However, warmth and competence are used in research on anthropomorphism, which one can regard as a form of "acting humanly". Mind perception can be related to "thinking humanly". Usability (SUS) is a form of operationalising whether an artificial agent is acting goal driven and useful which holds information on whether it is "acting rationally".

The first block of the questionnaire included four critical items on warmth, and three critical items on competence, as well as nine filler items. The critical questions asked for attributes related to either warmth, such as "good-natured", or competence, such as "skillful".

The second block consisted of 22 questions related to mind perception. These questions asked the participants to rate whether they believed that Vince can be attributed mental states. A typical item is the question whether Vince was capable of remembering events or whether he is able to feel pain.

Finally, the SUS questionnaire consisted of 10 items directly related to usability. Participants were asked question such as whether they found the system easy to use.

Upon completion of the questionnaire, participants were debriefed, reimbursed and dismissed.

5 Results

In the following, two types of results are reported. In Section 5.1, we present results from the questionnaire, in Section 5.2, we present initial results from video data recorded during the study.

5.1 Questionnaire Responses

As aforementioned, 7-point Likert scales (for the warmth, competence and mind question blocks) and a 5-point Likert scale for the SUS questions block) were used to measure participants responses to the dependent measures. For each dependent variable, mean scores were computed with higher values reflecting greater endorsement of the focal construct. Values for the four blocks of questions were averaged for further analysis. The results for the questionnaire are shown in Figure 2.

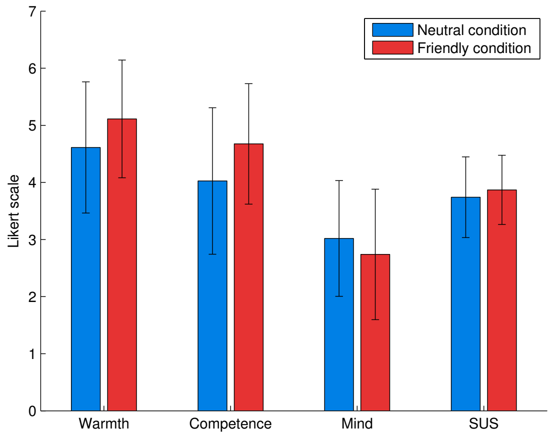

Figure 2. Mean response values for the questionnaire question sets. The mean for the dependent variables warmth, competence, mind and SUS are compared for the two categories neutral (blue) and friendly (red).

5.1.1 Warmth

The mean values for the warmth question set can be seen in Figure 2. It can be notices that the values for the friendly condition are mostly higher than for the neutral condition. The descriptive statistics confirm this. The friendly condition has a maximum value of 7 and a minimum value of 3.25 whereas the neutral condition has a maximum value of 6.75 and a minimum value of 2.25. The mean of the friendly condition is $M = 5.11$ ($SD = 1.14$) and the mean of the neutral condition is $M = 4.61$ ($SD = 1.14$). The mean values suggest that within the population on which our system was tested the friendly condition is perceived warmer than the neutral condition.

5.1.2 Competence

Similarly, the values for the friendly condition are mostly higher than for the neutral condition. The descriptive statistics confirm this. The friendly condition has a maximum value of 7 and a minimum value of 2.75 whereas the neutral condition has a maximum value of 6.25 and a minimum value of 1.5. The mean of the friendly condition is $M = 4.68$ ($SD = 1.05$) and the mean of the neutral condition is $M = 4.02$ ($SD = 1.28$). The standard deviation shows that there is more variation in the values for the neutral condition. The mean values overall suggest that within the population on which our system was tested the friendly condition is perceived more competent than the neutral condition.

5.1.3 Mind Perception

As Figure 2 shows, the ECA is perceived slightly higher on mind perception in the neutral condition than in the the friendly condition. The neutral condition has a maximum value of 4.9 and a minimum value of 1.32 whereas the friendly condition has a maximum value of 4.93 and a minimum value of 1.09. However, the mean of the neutral condition is $M = 3.02$ ($SD = 1.01$) whereas the mean of the friendly condition is $M = 2.74$ ($SD = 1.14$). The standard deviation suggests that there is more variation in the values for the neutral condition. The mean values overall suggest that within the population on which our system was tested in the friendly condition the participants attributed less mind to the ECA than the neutral condition.

5.1.4 System Usability Scale (SUS)

The values on the system usability scale are slightly higher in the friendly condition than in the neutral condition. The friendly condition has a maximum value of 4.7 and a minimum value of 2.7 whereas the neutral condition has a maximum value of 4.9 and a minimum value of 2.5. The mean of the friendly condition is $M = 3.87$ ($SD = 0.61$) and the mean of the neutral condition is $M = 3.74$ ($SD = 0.71$). The standard deviation suggests that there is more variation in the values for the neutral condition. The mean values overall suggest that within the population on which our system was tested the friendly condition was rated slightly more usable than the neutral condition.

5.2 Further Observations

Further observations that could be made on the dialogue level resulted from the analysis of the video data collected during the runs of the study. The dialogues were transcribed and inspected by one student assistant[6] trained in conversation analysis [18]. The purpose of this was to examine the dialogues to find out whether there were any particular delays in the dialogues and whether participants conformed to the script or not.

[6] Taking this line of research further, we would use two annotators and check for agreement between them. However, this was beyond the scope of the current contribution.

5.2.1 Alignment

We looked at the mean utterance length (MUL) of the participants in interaction with the ECA. We take this as an indicator of how participants align their verbalisations with the agent's verbalisations. The differences between the two conditions can be seen in in Figure 3, the values for the friendly condition are mostly higher than for the neutral condition.

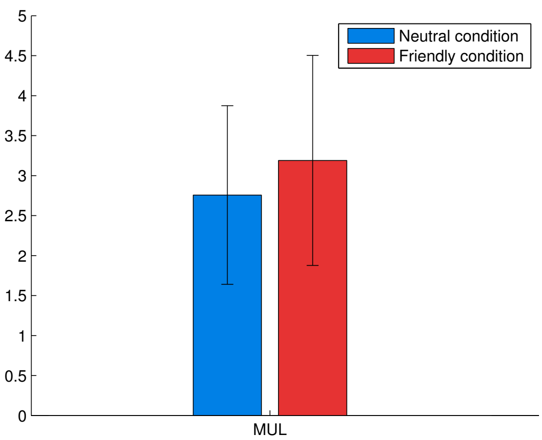

Figure 3. The mean utterance length averaged over the two conditions. The friendly condition has a slightly higher mean value than the neutral condition.

The descriptive statistics confirm this. The friendly condition has a maximum value of 5.5 and a minimum value of 1 whereas the neutral condition has a maximum value of 5.25 and a minimum value of 1. The mean of the friendly condition is $M = 3.12$ ($SD = 1.31$) and the mean of the neutral condition is $M = 2.76$ ($SD = 1.11$). The standard deviation suggests that there is more variation in the values for the friendly condition. The mean values overall suggest that within the population on which our system was tested the friendly condition showed more alignment with the ECA's MUL than the neutral condition.

5.2.2 Irregularities

The video data were reviewed and four types of noticeable effects on the dialogue were determined:

1. Participants returning because they did not understand or forget the ECA's instructions (22.5%, see Section 5.2.3),
2. deviations from the script, i.e. participants trying to do small talk with the ECA (5%, see Section 5.2.4),
3. timing difficulties causing delays in the interaction (25%), and
4. other ways in which the script was altered in small ways (22.5%, e.g. mismatches between the ECA's utterances and the participants utterances).

The overall number of irregularities accumulated across the two categories is summarized in Table 3. In interactions with the neutral condition irregularities can be observed in 75% of the cases, while in the friendly condition only 50% of the interactions show irregularities.

Table 3. Overview of occurred irregularities in the neutral and friendly condition.

	Neutral	Friendly
No irregularities	5	10
Irregularities occur	15	10

5.2.3 *Clarity of instructions*

Out of the 40 interactions in 9 cases (22.5%) the participants returned because they realized that they could not remember the room number correctly. Out of these the majority, namely 6, were in the neutral condition. Three participants came back for a second short interaction with Vince in the friendly condition.

5.2.4 *Small talk*

Only two participants (5%) deviated from the script of the dialogue by attempting to do small talk with Vince. Both of these were in the friendly condition. One participant asked the ECA for its name. Another participants tried three deviating questions on Vince during the interaction. The first question was "How are you?", the second "What can you tell me?", and finally the ECA was asked whether they were supposed to actually go to the room after the instructions were given.

6 Discussion

In reporting our results we concentrated on the descriptive statistics and no attempt will be made to generalize beyond this population. Within this first pilot study with the current demonstrator we tried to assess whether manipulating the degree of perceived friendliness has an effect on the interaction.

We now return to the questions asked in the introduction, the main question being how the manipulation affected the interaction between the user and the artificial agent.

6.1 Can the perceived friendliness of an agent be successfully manipulated?

We obtained slightly higher values regarding the perceived warmth in the friendly condition as opposed to the neutral condition. The differences are very small, though. The descriptive statistics point towards a "friendly" version of the dialogue actually being perceived as more friendly by the user. We propose that this will make users more willing to use the services the system can provide. Thus, further research into "friendly agents" seems a productive agenda.

The friendliness level also suggested higher ratings for competence, despite the fact that the friendly dialogue actually led to more misunderstandings. This failure was not reflected in the users judgements directly. Also, participants seem to prefer interacting with the friendly agent.

6.2 Is the proposed script a natural way of expressing the intended meaning?

The results which the video data analysis presented indicate that actually the majority of interactions conducted within this study were smooth and there were no noticeable deviations from the overall "script" in most dialogues. The operator was able to conduct most of the dialogues with the use of just a few buttons. This suggests that one can actually script dialogues of this simple nature quite easily.

However, the wording is crucial and the results suggest that the friendly version of the dialogue is more amicable to clarity. Only three participants did not fully understand or remember the instructions whereas twice as many had to ask for the room a second time in the neutral condition.

6.3 Are longer or shorter utterances favourable?

In a task-based dialogue the artificial agent will ideally demonstrate its knowledge and skill in a domain. However, the pilot-study did not find a very high difference between the two conditions regarding the competence question. The descriptive statistics, however, suggest that the longer utterances in the friendly dialogue received higher competence ratings.

Converse to the prediction, mind perception was slightly higher for the neutral dialogue, though. Thus, the friendly agent is not necessarily perceived as more intelligent by the user.

However, the longer utterances in the friendly version of the dialogue received higher ratings with respect to usability. Also, fewer participants had to come back and ask for the way again in a second interaction in the friendly condition. This suggests that the longer version of the dialogue better conveyed the dialogue content than the neutral version.

6.4 How does the user respond to a given wording?

In the friendly condition, users used longer utterances themselves when speaking to the friendly version of the ECA with more verbose verbalisations. This shows that the participants do align their speech with that of the artificial agent.

One can also tell from the video analysis that only in the friendly condition participants were motivated to further explore the possibilities the system offers. Two participants decided to ask questions which went beyond the script.

6.5 Will the script elicit the appropriate responses from the user?

Participants found it easy to conform to the proposed script. There was only a low percentage of participants who substantially deviated from the script and stimuli presented by the ECA (5% tried to do small talk with the agent). Most dialogues proceeded without the participants reacting in unanticipated ways and only a small percentage of participants failed to extract the relevant information from the verbalisations of the artificial agent.

7 Conclusion

We presented a pilot-study in which participants were confronted with dialogue exhibiting different degrees of friendliness.

While maintaining the same ideational function (see Section 2.2 above) we changed the interpersonal function of the dialogue by using sentences which were obtained through a role-playing pre-study and then rated by participants according to their friendliness.

The obtained dialogues (a friendly and a neutral version) were presented to participants in interaction with an ECA which was implemented via generic interaction patterns. Participants filled in a questionnaire after the interaction which was analysed along with further observational data collected during the study.

The results point towards higher perceived warmth, higher perceived competence and a greater usability judgement for the ECA's performance in the friendly condition. However, mind perception does not increase in the more friendly dialogue version.

Further research should replicate our findings using a larger sample size. Also, in a similar study the variation of friendliness in interaction had less impact on the participants' perception than the interaction context [43]. Thus, one would have to take a closer look at how politeness and context interact in future studies. In addition, related literature also suggests that anthropomorphic perceptions could be increased by increased politeness [44]. Thus, friendliness can generally be expected to have an effect on the perception of artificial agents.

The dialogue in the present study not only varied in terms of friendliness but also in terms of verbosity. It could be argued that this is not the same and a higher verbosity might have had an unwanted effect, especially on the user's task performance. Future studies could consider whether they can be designed to investigate the effect of friendliness without directly changing agent verbosity.

It would also be interesting to conduct a similar study to explore dialogue usage in the robot Biron. As he is supposed to guide the visitor to the requested room, he spends several minutes with the visitor without exchanging necessary information, thus, is can be expected that the usage of small talk affects the interaction in a positive way.

ACKNOWLEDGEMENTS

The authors would like to thank all colleagues who contributed to the Dialogue Demonstrator: Anja Durigo, Britta Wrede, Christina Unger, Christoph Broschinski, David Schlangen, Florian Lier, Frederic Siepmann, Hendrik Buschmeier, Jan De Ruiter, Johanna Müller, Julia Peltason, Lukas Twardon, Marcin Wlodarczak , Marian Pohling, Patrick Holthaus, Petra Wagner, Philipp Cimiano, Ramin Yaghoubzadeh, Sebastian Ptock, Sebastian Wrede, Thorsten Spexard, Zofia Malisz, Stefan Kopp, and Thilo Paul-Stueve. The research reported here was funded by the Cluster of Excellence "Cognitive Interaction Technology" (EXC 277). Griffiths is also partly supported by ConCreTe: the project ConCreTe acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 611733. The authors would also like to thank three anonymous reviewers of the contribution for their very useful and productive feedback.

REFERENCES

[1] Nalini Ambady, Debi LaPlante, Thai Nguyen, Robert Rosenthal, Nigel Chaumeton, and Wendy Levinson, 'Surgeons' tone of voice: A clue to malpractice history', *Surgery*, **132**(1), 5–9, (July 2002).

[2] Niklas Beuter, Thorsten P. Spexard, Ingo Lütkebohle, Julia Peltason, and Franz Kummert, 'Where is this? - Gesture based multimodal interaction with an anthropomorphic robot', in *International Conference on Humanoid Robots*. IEEE-RAS, (2008).

[3] Timothy Bickmore and Julie Cassell, 'Small talk and conversational storytelling in embodied conversational interface agents', in *Proceedings of the AAAI Fall Symposium on "Narrative Intelligence"*, pp. 87–92, (1999).

[4] Timothy Bickmore and Justine Cassell, '"how about this weather?" social dialog with embodied conversational agents', in *Proceedings of the American Association for Artificial Intelligence (AAAI) Fall Symposium on "Narrative Intelligence"*, pp. 4–8, Cape Cod, MA, (2000).

[5] Timothy Bickmore and Justine Cassell, 'Relational agents: a model and implementation of building user trust', in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 396–403. ACM, (2001).

[6] John Brooke, 'SUS – a quick and dirty usability scale', *Usability evaluation in industry*, **189**, 194, (1996).

[7] Justine Cassell, 'Embodied conversational agents: representation and intelligence in user interfaces', *AI Magazine*, **22**(3), 67–83, (2001).

[8] Justine Cassell and Timothy Bickmore, 'Negotiated collusion: Modeling social language and its relationship effects in intelligent agents', *User Modeling and User-Adapted Interaction*, **13**(1-2), 89–132, (2003).

[9] Justine Cassell, Timothy Bickmore, Mark Billinghurst, Lee Campbell, Kenny Chang, Hannes Vilhjálmsson, and Hao Yan, 'Embodiment in conversational interfaces: Rea', in *Proceedings of the CHI'99 Conference*, pp. 520–527. ACM, (1999).

[10] David Crystal, *A Dictionary of Linguistics and Phonetics*, Blackwell Publishers, sixth edn., 2008.

[11] Östen Dahl, *The Growth and Maintenance of Linguistic Complexity*, John Benjamins Publishing Company, Amsterdam/Philadelphia, 2004.

[12] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg, 'Wizard of oz studies: why and how', in *Proceedings of the 1st international conference on Intelligent user interfaces*, pp. 193–200. ACM, (1993).

[13] Friederike Eyssel and Dieta Kuchenbrandt, 'Manipulating anthropomorphic inferences about NAO: The role of situational and dispositional aspects of effectance motivation', in *2011 RO-MAN*, pp. 467–472. IEEE, (July 2011).

[14] Gernot A. Fink, 'Developing HMM-based recognizers with ESMERALDA', in *Text, Speech and Dialogue*, eds., Václav Matousek, Pavel Mautner, Jana Ocelíková, and Petr Sojka, volume 1692 of *Lecture Notes in Computer Science*, pp. 229–234. Springer Berlin Heidelberg, (1999).

[15] Susan T Fiske, Amy J C Cuddy, and Peter Glick, 'Universal dimensions of social cognition: warmth and competence.', *Trends in cognitive sciences*, **11**(2), 77–83, (February 2007).

[16] Norman M Fraser and G Nigel Gilbert, 'Simulating speech systems', *Computer Speech & Language*, **5**(1), 81–99, (1991).

[17] Dafydd Gibbon, Roger Moore, and Richard Winski, *Handbook of Standards and Resources for Spoken Language Systems*, Mouton de Gruyter, 1997.

[18] Charles Goodwin and John Heritage, 'Conversation analysis', *Annual Review of Anthropology*, **19**, 283–307, (1990).

[19] Heather M Gray, Kurt Gray, and Daniel M Wegner, 'Dimensions of mind perception.', *Science (New York, N.Y.)*, **315**(5812), 619, (February 2007).

[20] Sascha Griffiths, Ciro Natale, Ricardo Araújo, Germano Veiga, Pasquale Chiacchio, Florian Röhrbein, Stefano Chiaverini, and Reinhard Lafrenz, 'The ECHORD Project: A General Perspective', in *Gearing up and accelerating cross-fertilization between academic and industrial robotics research in Europe:*, eds., Florian Röhrbein, Germano Veiga, and Ciro Natale, volume 94 of *Springer Tracts in Advanced Robotics*, 1–24, Springer International Publishing, Cham, (2014).

[21] Axel Haasch, Sascha Hohenner, Sonja Hüwel, Marcus Kleinehagenbrock, Sebastian Lang, Ioannis Toptsis, Gernot A. Fink, Jannik Fritsch, Britta Wrede, and Gerhard Sagerer, 'Biron - the Bielefeld robot companion', in *Proc. Int. Workshop on Advances in Service Robotics*, eds., Erwin Prassler, Gisbert Lawitzky, P. Fiorini, and Martin Hägele, pp. 27–32. Fraunhofer IRB Verlag, (2004).

[22] Frank Hegel, Friederike Anne Eyssel, and Britta Wrede, 'The social robot Flobi: Key concepts of industrial design', in *Proceedings of the 19th IEEE International Symposium in Robot and Human Interactive Communication (RO-MAN 2010)*, pp. 120–125, (2010).

[23] Joseph Hilferty, 'Interview with Richard Hudson', *Bells: Barcelona English language and literature studies*, **16**, 4, (2007).

[24] Patrick Holthaus, Ingo Lütkebohle, Marc Hanheide, and Sven Wachsmuth, 'Can I help you? - A spatial attention system for a receptionist robot', in *Social Robotics*, eds., Shuzhi Sam Ge, Haizhou Li, John-John Cabibihan, and Yeow Kee Tan, pp. 325–334. IEEE, (2010).

[25] Patrick Holthaus and Sven Wachsmuth, 'Active peripersonal space for more intuitive HRI', in *International Conference on Humanoid Robots*, pp. 508–513. IEEE RAS, (2012).

[26] Patrick Holthaus and Sven Wachsmuth, 'The receptionist robot', in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pp. 329–329. ACM, (2014).

[27] Stefan Kopp, 'Social adaptation in conversational agents', *PerAda Magazine (EU Coordination Action on Pervasive Adaptation)*, (2009).

[28] Stefan Kopp and Ipke Wachsmuth, 'Synthesizing multimodal utterances for conversational agents', *Computer Animation and Virtual Worlds*, **15**(1), 39–52, (2004).

[29] Stephen C Levinson, *Pragmatics*, Cambridge University Press, Cambridge, 1983.

[30] Nikolaos Mavridis, 'A review of verbal and non-verbal human–robot in-

teractive communication', *Robotics and Autonomous Systems*, **63**, 22–35, (2015).

[31] Roger K Moore, 'From talking and listening robots to intelligent communicative machines', in *Robots that Talk and Listen – Technology and Social Impact*, 317 – 336, De Gruyter, Boston, MA, (2014).

[32] Julia Peltason, 'Position paper: Julia Peltason', in *6th Young Researchers' Roundtable on Spoken Dialogue Systems*, pp. 63 – 64, (2010).

[33] Julia Peltason, *Modeling Human-Robot-Interaction based on generic Interaction Patterns*, Ph.D. dissertation, Bielefeld University, 2014.

[34] Julia Peltason, Hannes Rieser, Sven Wachsmuth, and Britta Wrede, 'On Grounding Natural Kind Terms in Human-Robot Communication', *KI - Künstliche Intelligenz*, (March 2013).

[35] Julia Peltason and Britta Wrede, 'Modeling Human-Robot Interaction Based on Generic Interaction Patterns', in *AAAI Fall Symposium: Dialog with Robots*, pp. 80 —- 85, Arlington, VA, (2010). AAAI Press.

[36] Julia Peltason and Britta Wrede, 'Pamini: A framework for assembling mixed-initiative human-robot interaction from generic interaction patterns', in *SIGDIAL 2010: the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 229–232, The University of Tokyo, (2010). Association for Computational Linguistics.

[37] Julia Peltason and Britta Wrede, 'The curious robot as a case-study for comparing dialog systems', *AI Magazine*, **32**(4), 85–99, (2011).

[38] Rajesh Ranganath, Dan Jurafsky, and Daniel A McFarland, 'Detecting friendly, flirtatious, awkward, and assertive speech in speed-dates', *Computer Speech & Language*, **27**(1), 89–115, (2013).

[39] Ehud Reiter and Robert Dale, *Building Natural Language Generation Systems*, Studies in Natural Language Processing, Cambridge University Press, Cambridge, 2000.

[40] F Röhrbein, S Griffiths, and L Voss, 'On industry-academia collaborations in robotics', Technical report, Technical Report TUM-I1338, (2013).

[41] Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall International, Harlow, third int. edn., 2013.

[42] Amir Sadeghipour and Stefan Kopp, 'Embodied gesture processing: Motor-based integration of perception and action in social artificial agents', *Cognitive computation*, **3**(3), 419–435, (2011).

[43] Maha Salem, Micheline Ziadee, and Majd Sakr, 'Effects of politeness and interaction context on perception and experience of HRI', in *Social Robotics*, eds., Guido Herrmann, Martin J. Pearson, Alexander Lenz, Paul Bremner, Adam Spiers, and Ute Leonards, volume 8239 of *Lecture Notes in Computer Science*, 531–541, Springer International Publishing, (2013).

[44] Maha Salem, Micheline Ziadee, and Majd Sakr, 'Marhaba, how may I help you?: Effects of politeness and culture on robot acceptance and anthropomorphization', in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pp. 74–81. ACM, (2014).

[45] Aaron Steinfeld, Odest Chadwicke Jenkins, and Brian Scassellati, 'The oz of wizard: simulating the human for interaction research', in *Human-Robot Interaction (HRI), 2009 4th ACM/IEEE International Conference on*, pp. 101–107. IEEE, (2009).

[46] Terry Winograd, *Language as a cognitive process (Vol. 1)*, Addison-Wesley, Reading, MA, 1983.

[47] Sebastian Wrede, Jannik Fritsch, Christian Bauckhage, and Gerhard Sagerer, 'An XML based framework for cognitive vision architectures', in *Proc. Int. Conf. on Pattern Recognition*, number 1, pp. 757–760, (2004).