

# Conversational Assistants for Elderly Users – The Importance of Socially Cooperative Dialogue

Stefan Kopp  
Bielefeld University, CITEC  
skopp@uni-bielefeld.de

Katharina Cyra  
University of Duisburg-Essen  
katharina.cyra@uni-due.de

Franz Kummert  
Bielefeld University, CITEC  
franz@techfak.uni-bielefeld.de

Lars Schillingmann  
Bielefeld University, CITEC  
lschilli@techfak.uni-bielefeld.de

Mara Brandt  
Bielefeld University, CITEC  
mbrandt@techfak.uni-bielefeld.de

Farina Freigang  
Bielefeld University, CITEC  
farina.freigang@uni-bielefeld.de

Christiane Opfermann  
University of Duisburg-Essen  
christiane.opfermann@uni-due.de

Carolin Straßmann  
University of Duisburg-Essen  
carolin.strassmann@uni-due.de

Ramin Yaghoubzadeh  
Bielefeld University, CITEC  
ryaghoub@techfak.uni-bielefeld.de

Hendrik Buschmeier  
Bielefeld University, CITEC  
hbuschme@uni-bielefeld.de

Nicole Krämer  
University of Duisburg-Essen  
nicole.kraemer@uni-due.de

Karola Pitsch  
University of Duisburg-Essen  
karola.pitsch@uni-due.de

Eduard Wall  
Bielefeld University, CITEC  
ewall@techfak.uni-bielefeld.de

## ABSTRACT

Conversational agents can provide valuable cognitive and/or emotional assistance to elderly users or people with cognitive impairments who often have difficulties in organizing and following a structured day schedule. Previous research showed that a virtual assistant that can interact in spoken language would be a desirable help for those users. However, these user groups pose specific requirements for spoken dialogue interaction that existing systems hardly meet. This paper presents work on a virtual conversational assistant that was designed for, and together with, elderly as well as cognitively handicapped users. It has been specifically developed to enable ‘socially cooperative dialogue’ – adaptive and aware conversational interaction in which mutual understanding is co-constructed and ensured collaboratively. The technical approach is described and results of evaluation studies are reported.

## KEYWORDS

conversational assistants; elderly users; cooperative dialogue

## 1 INTRODUCTION

In recent years politics and society have placed emphasis on ways to enable a longer autonomous and self-determined life for elderly people. One approach is the development of assistive technology. However, this has often been focused on supporting physical tasks (e.g., fetching or lifting objects, moving around) and it has been struggling with questions of human-machine interaction and user acceptance. The goal of the KOMPASS project (which started in 2015) was to develop a virtual assistant (‘Billie’) to accompany and guide a user throughout the day. The system has been specifically designed for, and together with, two user groups: elderly users that live autonomously in their home environment but are on the verge of



Figure 1: Concept for an elderly user interacting with the virtual assistant ‘Billie’ in her home environment.

needing home assistance services, and cognitively handicapped users that are already supported by professional care-givers. What both user groups have in common are mild cognitive impairments that create a need for support with autonomously organizing and following a structured day schedule.

While technical means of supporting this are already available, many elderly users have little prior experience with using assistive systems. Applying such technology thus requires to overcome a ‘digital barrier’ both with the individual users as well as their care-providing environment. The KOMPASS project built on pre-studies [13] suggesting that natural spoken-language interaction with a virtual agent may be desirable and acceptable for these user groups.

Building and applying conversational agents for these user groups, however, raises its own challenges. Elderly users often have selectively impaired abilities, e.g., for auditory perception, articulation, adapting to a recommended interaction style, adhering to a clean turn-taking structure, or comprehending content of high information density [35, 37]. We thus set out to develop a conversational agent that provides a dialogue style that enables robust and reliable, yet acceptable spoken-language interactions with these user groups. We refer to this special quality as ‘socially cooperative dialogue’ [32].

In this paper we present our approach and report on results obtained in evaluation studies. After discussing related work in the next section, Sect. 3 points our requirements before Sect. 4 describes our approach to modeling socially cooperative dialogue in the virtual assistant ‘Billie’. Section 5 presents results and lessons learned from several evaluation studies carried out with users in the lab environment as well as in their real home environment (ongoing), showing how conversational agents can be built to achieve the interaction abilities needed to provide elderly users and mildly cognitively impaired persons with successful assistance.

## 2 RELATED WORK

Several conversational assistants have been developed for care-related settings such as companionship for people living alone [20], assistance in multilingual care-giving/receiving [31], or pain and affect management [28]. An increasing body of work suggests that spoken interaction with users with cognitive impairments seems, in general, to be feasible and accepted by the user group, though some requirements need to be met. Meis [15] noted that older subjects want a spoken-dialogue helper to have a name and to react contingently to social affordances such as expressions of gratitude. Miehle and colleagues [16] noted that conversational assistants are required to speak sufficiently loud and with an appropriate pace but were accepted as interlocutors in a study with elderly people. Bickmore and colleagues [2] analyzed long-term interactions of older adults with an agent-based coaching system using spoken language while user input was given through a touchscreen and found them effective on the short term. Sidner and colleagues [23] attempted to identify preferred domains of conversation or joint activity based on this system design. Yaghoubzadeh and colleagues [36] reported that older adults and people with cognitive impairments are able to successfully ground information. Explicit confirmation patterns and a low information density (one information unit per utterance) enabled the user to detect and repair more of the system’s language understanding problems and subsequent errors.

A well-discussed concept that is central for successful human-human dialogue is ‘grounding’ [6]. Researchers have attempted to model it computationally for conversational agents, discretely as a finite-state process [27] or probabilistically using Bayesian networks [19]. Recent work on real-time dialogue systems has focused on advanced issues so that discourse context is taken into account [24], partial and overlapping utterances can be grounded incrementally [29], groundedness can be estimated from multimodal feedback cues [4], or information from multiple modalities related to socio-emotional aspects such as attention and engagement are taken into account [14].

## 3 REQUIREMENTS

### 3.1 Interactional tasks

The goal of the KOMPASS project was to develop a conversational assistant that helps users organize and keep track of their schedule for the day. This goal was identified by our application partner (v. Bodelschwingh Foundation Bethel, Bielefeld, Germany) as an important need of their respective client groups. The conversational agent system ‘Billie’ thus offers several functions within the domain of schedule management. Users can enter their various kinds of appointments (single appointments, recurrent appointments with specification of recurrence, e.g., weekly, biweekly and for how long the recurrent appointment will be reiterated), they can choose being reminded of them including setting a time for the reminder, and they can edit already entered appointments. The editing of appointments comprises the following sub-tasks: Users can change any of the appointment values ‘start time’, ‘end time’, ‘topic’ and ‘duration’, they can delete or replace appointments within the calendar, and they can query their entered appointments of any point in time (same day, same week, forthcoming day or days and weeks, and previous days or weeks). Moreover, the agent system provides for user-tailored suggestions for leisure time activities [17, 18] in order to promote a more active life.

### 3.2 Socially cooperative dialogue

In line with other previous work, focus groups and pre-studies that we carried out within a user-centered design process made clear that assistive functions of the system preferably should be accessible and realized through easy-to-use spoken dialogue interaction [13]. Further analyses of several quasi experimental studies (run in a Wizard-of-Oz scenario in 2015, as a semi-autonomous study in 2017, and as a long-term study ongoing; see section 5) pointed to the fact that spoken-language interaction, while being generally preferred, raises a number of (well-known) challenges that tend to be amplified in our user groups. Therefore, the conversational assistant has to fulfill specific requirements for the users to be acceptable and successful with respect to the various sub-tasks of schedule management:

- The system has to be able to deal with long, extensive user utterances. Interruptions and barge-ins by the user must be possible at all times, in particular when they are instrumental in solving the current communicative task at hand. Overall, turn-taking has to be cooperative such that interruptions by the system should be foreshadowed through nonverbal behavior [33]. Simultaneously, the system must be robust to non-cooperative turn-taking behavior of the user such that turn fights are avoided (generally yielding the turn to the user).
- Generally, the system must work to ensure dynamic coordination of understanding and grounding in dialogue. Feedback by the system to user input must be provided timely to prevent long user turns, and clearly mark the system’s current level of understanding [8, 9]; user feedback must be continuously processed and interpreted for indicators of miscommunication.

- The handling of understanding problems on part of the system or the user is crucial. User-initiated displays of non-understanding (e.g., “Sorry” or “Can you repeat please”) must always be possible and handled by the system properly. For non-specific system displays of non-understanding, a variation of error handling strategies must be available, e.g., in form of reprompts and non-understanding notification, combined with more restrictive clarification sequences depending on the sub-task and local move. While reprompts may be beneficial for problem solving in its first issuing [17], a lack of progress in problem solving without a change in error recovery strategy may lead to further complications (cf. [17]). This holds especially in action contexts like appointment suggestions in which user responses can address social matters like willingness, availability, disposition and deontic authority. Thus, error handling on part of the system should employ strategies that clarify the user’s agreement or resistance [18].
- Topic shifts by the user must be possible and followed readily by the system. This requires the system to also keep track of non-settled discourse segments and to return to them at appropriate points in time (when the user does not pursue other discourse goals) and with a cooperative and gentle entrance strategy, possibly with repeating and rectifying parts of the discourse unit that have been discussed already. Generally, the system should avoid topic shifts. If they are unavoidable, e.g., because a previous discourse topic has not been settled yet, they should be of as close a distance as possible and must be marked explicitly.

The requirements listed above are (necessary, but most probably not sufficient) examples of a specific dialogue quality that we deem necessary for our user groups. We refer to this dialogue quality as ‘socially cooperative’ [32] and note that it goes well beyond classical notions like grounding in dialogue, as it implies a specific role that the agent has to fulfill consistently throughout the interaction. This role entails a range of collaborative-supportive action policies, e.g., for readily following topic shifts, yielding turns, adhering to rules of politeness, and adapting dialogue structures thoroughly to the needs of the user.

## 4 APPROACH AND IMPLEMENTATION

### 4.1 Architecture

To account for the requirements identified above, we have based the conversational assistant on an interaction framework that aims to support the features of incrementality (to quickly update and relay discussed information), provisions for representation and resolution of uncertainty (resulting from input and unclear grounding) with explicit representation of topics, structured hierarchically in units intuitive to laymen. The overall architecture is shown in fig. 2. It is built on top of the IPAACA middleware [21], a distributed, platform-independent implementation of a general model for incremental dialogue processing proposed by Schlangen and Skantze [22]. This provides the back-end for the connection of the core dialogue management components to input (including ASR, tagger and parser, eye tracker, keyboard/mouse/touch etc.), multimodal fusion, behavior planning (NLG, gaze, gesture, calendar) and output

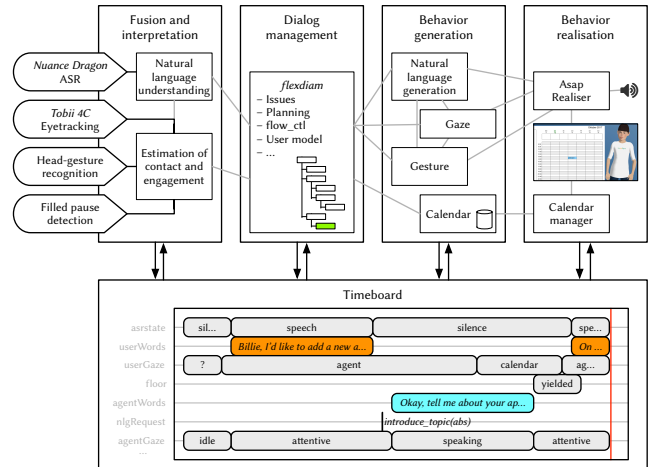


Figure 2: Overall architecture of the conversational assistant

realization (synthesis, graphical components / GUI changes, control of animated characters etc.).

The architecture is built around a ‘timeboard’, a central representation that captures temporal information of interactional events on different tiers. Importantly, these tiers hold rewindable representations of certain and uncertain variables (probability distributions) with generic metrics – like entropy – that serve as the basis for local decision heuristics. Event-driven observers are used to derive events from interval relations between existing ones, and trigger higher-level functions, most centrally the dialogue manager proper, but also a contribution manager, which schedules queued communicative intentions when the floor situation allows.

### 4.2 Dialogue management

To realize the required dialogue management abilities, the ‘flexdiam’ [34] system has been developed. Following an issue-based approach, it generally pursues a single joint task and discourse model for both interactants. The basic structure of the joint task and discourse model is a forest of independent but hierarchically interdependent agents termed ‘Issues’, as well as generic update rules to transform this forest after dialogue management invocations. When an Issue is instantiated, it is at the same time made a child of the Issue that created it. Any path from a leaf Issue to the root corresponds to a nested (sub-)topic of discussion. Any number of topics can be active at any one time and will be considered valid points of reference in parallel, if applicable according to their grounding state. To that end, any Issue can be in one of five states (new, entered, fulfilled, failed, obsolete).

Invocations that trigger local processing in Issues come in two flavors: input handling (e.g., prompt request, NLU parse) and plan structure updates (e.g., child issue progressed, completed or invalidated). Issues will decide along their local path in the hierarchy, and based on the current global context, whether they can provide a plan to handle an invocation. If an Issue cannot handle an input handling invocation locally, a preference is marked to let its parent handle it instead. Partial localized processing does not preclude propagation through the hierarchy, though. This allows for situated

partial interpretation and processing, which is most specific and situation-dependent in the leaves, and most generic and general in the roots of the forest.

If a user contribution does not fit well into any active Issue, a discourse transition based on user initiative can be assumed to have taken place. Depending on the situation, this could be construed as either a forward-looking contribution (if anticipated by the currently invoked entrance point or a direct ancestor) or a real topic jump. A new branch is created then, marked as entered, and moved to the top of the entry point priority queue.

Note that the present system is suited to quick, interactive approaches to spoken interaction and to modeling real-world applications within limited domains. Manual extension is quite straightforward. Incremental processing and the handling of uncertain input and information derived from it has received special focus, the ‘output’ side employs a similar notion of indeterminate state until evidence for communicative success provides a precondition for grounding being attested. Communicative plans are capable of employing several modalities and the implemented suite of basic Issues for grounding problems can be fine-tuned to cover a wide space of varying explicitness, verbosity, and conversational styles, which can be used to seed user models that best suit the estimated capabilities and preferences of our specific user groups. This extends to information density (configurable via different options for packaging and different approaches to confirmation requests), but also discourse structure: explicit ratification for topic jumps beyond a distance threshold (and implicit acceptance by means of contingent continuation by the user) are currently in development.

### 4.3 Socio-communicative signal processing

Human communication is highly multi-modal and thus the ability to process this variety of information is very important to facilitate communication with a virtual agent. Therefore, several modules in our architecture recognize visual communication signals and non-verbal socio-emotional speech cues. Confirmations play an important role in the dialog structure of the interaction with ‘Billie’. Therefore, we focused on the recognition of natural confirmation signals, like nodding and non-lexical confirmations like “mhm”, which are typically not recognized by automatic speech recognition. To detect non-lexical confirmations, the speech signal is segmented into speech intervals using voice activity detection (VAD). Our module detects non-lexical confirmations by extracting acoustic features and classifying the result using a Support Vector Machine (SVM). If confirmations are detected, the component sends messages via the IPAACA middleware to inform other components [3]. Further, the system is able to detect human head nods based on dynamic time warping and estimations of head pose angles from facial landmark features [30]. In addition, the face detection is used to verify contact with the user. A cue aggregation module combines signals detected in individual modalities to derive a higher-level interpretation using a Bayesian network (see Fig. 3). Currently, the system detects if the user signals confirmation by combining non-lexical confirmation and nod detection. User contact is detected by combining face detection and eye contact related cues.

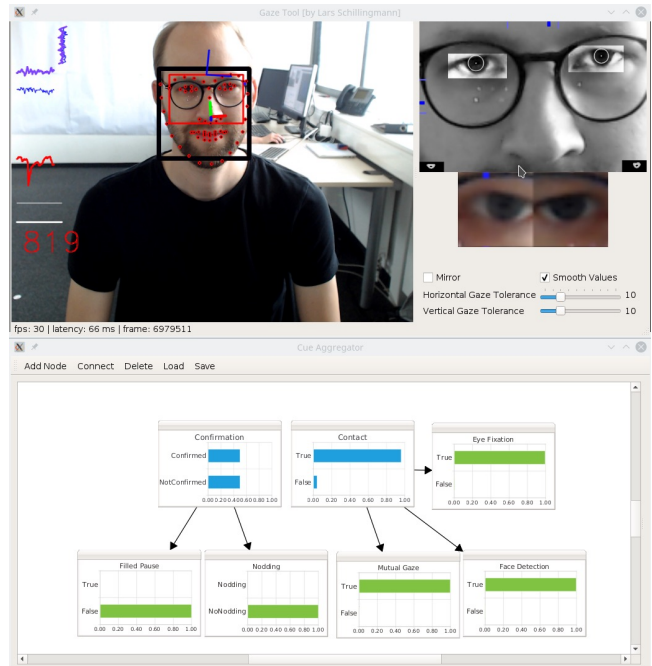


Figure 3: Screenshot of the system processing and fusing socio-communicative signals.

### 4.4 Multimodal expressiveness

Regarding output behavior, the conversational assistant ‘Billie’ is enhanced with multimodal cues such as gestures, facial expressions and head movements in order to develop a more natural behavior and making it easily accessible, understandable and helpful for the human user. The cues were selected based on an analysis of their form, function and frequency in natural interaction data [11]. Specifically, we simulate cues that serve the pragmatic functions to convey and mark emphasis, de-emphasis and (un-)certainty of the speaker. These functions have been successfully mapped onto the conversational agent [10, 12]. For the calendar domain, key phrases were selected and are accompanied by these multimodal functions if applicable (emphasis: “Let’s continue!”, de-emphasis: “Good, this is canceled”, uncertainty: “Did you say ‘swimming?’”; see fig. 4). The multimodal expressiveness has been evaluated with the participants in lab-based studies (e.g., showing better information uptake) and will be evaluated systematically during a long-term study.

Additionally to the assistant ‘Billie’ itself, we designed different states, visual and sound features of the weekly-based calendar that support the dialogue between user and agent multimodally. As previous analyses have shown [8], users orient toward the calendar area of the interface while entering appointments as this is the area of interest for the ongoing interactional task. So, to design a responsive and recipient system that supports the dialogue not only with verbal and non-verbal features of the agent, the calendar provides visual cues (mainly through highlighting) to comprehension hypotheses of the system even before words are uttered by the agent (see fig. 5). These visual updates represent the system’s

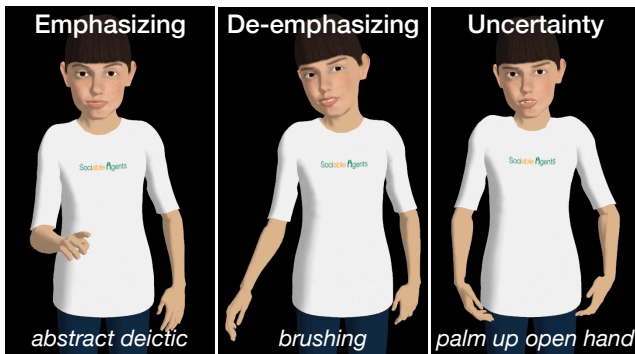


Figure 4: Multimodal cues used by the conversational assistant for different pragmatic functions.

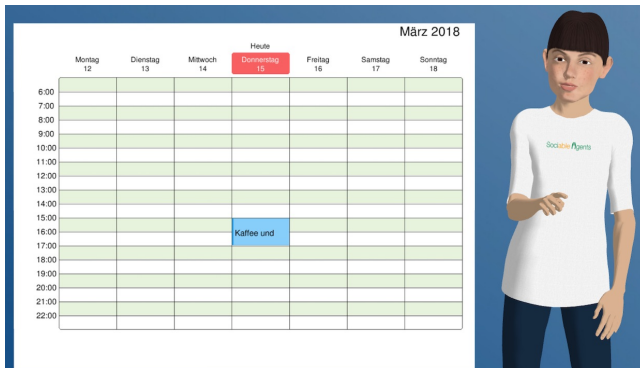


Figure 5: The calendar including a visual cue (highlighting of the current day), an appointment “Coffee and Cake” and the conversational assistant confirming the successful entry of the appointment.

status regarding the understanding of user input. Furthermore, a sound was added to mark the successful entry of appointments.

#### 4.5 Data Recording

Our system architecture is designed to be used in both lab and field environments. As manual recording control is not feasible in long term field studies, the architecture contains an automatic recording module which starts recording when the user starts interacting with the system by pushing a button. Recording is automatically stopped if the user says goodbye and, in order to ensure users’ privacy, when the user is not visible to the system and does not react to system prompts for an extended period of time, or when a system error is detected. The recordings comprise five video and five audio tracks which are compressed in real-time using hardware acceleration. One video track (depicted in Fig. 6) is a 4-in-1 overview of the other video tracks. System and interaction log files are archived after each session.

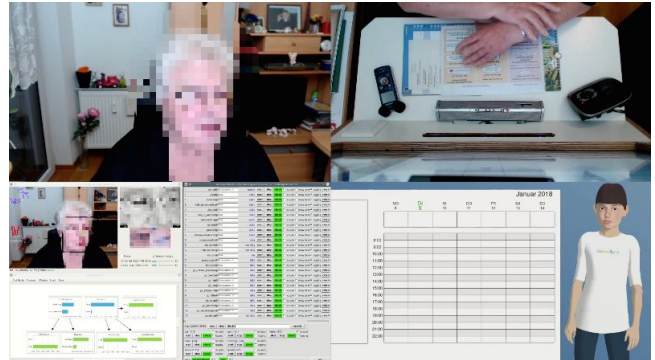


Figure 6: Example of the data recorded during the field evaluation study.

## 5 EVALUATION RESULTS

In our project we followed an iterative design-implement-evaluation approach that comprised a number of empirical evaluations of individual sub-systems of the agent [3, 5, 8–10, 17, 18, 25, 26, 30]. Further, to gather training data for the socio-emotional signal recognition components and to get insights into users’ reactions to potential interaction problems, an initial Wizard-of-Oz study with 53 participants from the different user groups (18 senior participants, 19 participants with cognitive impairments and, 16 controls from the local student population) was carried out. In order to elicit participants’ reactions, the agent’s behavior in this study followed a script that was designed to create a number of typical interaction problems. Participants could negotiate and enter their own appointments, but could also use previously prepared appointment cards. In the following, we report more recent studies that were carried out to evaluate the socially cooperative dialogue abilities of the full-blown conversational assistant in less restricted interactions.

### 5.1 Lab-based evaluation

Based on first insights gathered from the above described WOZ-study as well as on knowledge acquired in preparatory studies [13, 35], a first version of a semi-autonomous agent was evaluated in a laboratory setting. This study investigated whether participants from the different user groups – without specific instructions – were able to carry out calendar-related tasks through spoken interaction with the agent. Furthermore, the study’s objective was to investigate how participants manage transitions between bigger topics/issues, how the length of participants’ utterances vary given different confirmation strategies, how users react to communication of uncertainty, and whether the agent’s ways of guiding the users’ attention (via voice, manual gesture, gaze, calendar highlighting, and sounds) is effective.

We employed a system that autonomously handled dialogue management for entering appointments as well as for stating appointment suggestions. A human ‘wizard’ was included only for controlling transitions between global modes of user entering appointments, auto-generated partial suggestions of appointments by the agent (“Would you like to do something on Saturday?”), and closing the interaction. 44 participants took part in the study: 19

older adults (SEN) aged about 75+; 15 cognitively impaired adults (CIM) of working age; and 10 students serving as control group (CTL). The task was free-form entering of appointments. All subjects managed to enter the required number of appointments into the calendar. The number of final entries averaged 10.4, 8.5, and 8.9 for CTL, SEN and CIM, respectively (including up to two agent-recommended items). Older adults and the group with impairments on average spent about 20% longer on a topic than controls; some participants from the CIM group made long hesitations in isolated instances (up to tens of seconds; see [34] for a detailed discussion). Still, the socially cooperative dialogue abilities of the agent enabled the user to conduct successful repair with the agent and to settle in on acceptable solutions in every subtask.

The study was also conducted to gain qualitative insight into the repair, revision and meta-communicative patterns exhibited by the user groups. Further, we wanted to observe temporal aspects of the planning and verbalizing of an appointment to learn about participants' practices in different phases of appointment entry (to eventually design the timing and turn taking of the system). Analyses of the data are still ongoing and focus among others (1) on the analysis of gaze when interactional trouble occurs, (2) on the system's repair strategies and (3) on the different multimodal states and uptake strategies of the system (as a representation of its reciprocity) with effects on turn production of the users.

## 5.2 Field study

A long-term field study is currently ongoing to evaluate the system's performance and effects, as well as how participants adopt and handle it in their home environment over a 15-day period. For this study we implemented and apply a fully autonomous system with no additional aids by a human wizard (see fig. 7 for the setup). This study comprises an ethnographic component focusing on daily life management in the homes of seniors living alone and people with cognitive impairments in supported-living [1, 7]. These analyses aim to shed light also on broader questions such as: What are the challenges when a novel technology is brought into a household of the target groups? What are the participants' experiences and expectations of a technological assistant? What are the effects of the assistant on participants' daily routines and in particular their schedule management? And, an overall topic considered throughout the project, what are issues with regard to privacy and data protection with our special user groups?

Overall, the field study followed an iterative approach consisting of three phases. In the pre-pilot study, we aim to acquaint participants with the system and to identify their specific expectations and needs. Researchers lead semi-structured interviews in the apartments of the participants and discuss the possible placement of the system in the apartment using a full-size paper prototype. Next, a pilot study evaluates the feasibility of the main study with a special focus on the robustness of the system and its performance outside of the lab. Therefore the prototype system is set up in the apartments of the participants for a period of about 48 hours. In addition to the system evaluation, we want to learn about the acceptance of the system by the participants and their assessment of the dialogue design. This provides the basis for further optimizations of the dialogue design and preparation of the main study, for which

the prototype system is placed into the participants' apartment for a period of about 15 days (including setup and dismounting day). Participants are asked to manage their daily schedule together with "Billie" and to jot down their impressions about the system into a research diary (freely as well as to structured questions). After the period of applying the system, participants give a final rating in the form of a semi-structured interview.

*First results.* The above described study design has been carried out with one female senior person as the first long-term study. The system has been in use in the participant's home for 13 days (excluding a setup and a dismounting day), during which she interacted 61 times with the system, for a total duration of 284 minutes and 46 seconds. She used the system for a mean number of 4.8 times per day ( $SD = 1.7$ ,  $Min = 2$ ,  $Max = 8$ ). Although usage over the duration of the study varied between days, it did not differ much between the first seven days ( $M = 5.3$ ,  $SD = 1.8$ ,  $Min = 3$ ,  $Max = 8$ ) and the last six days ( $Mean = 4.2$ ,  $SD = 1.6$ ,  $Min = 2$ ,  $Max = 6$ ), suggesting that the participant did not lose interest.

Interaction durations varied greatly, with the shortest interaction lasting only 6 seconds and the longest interaction lasting 18 minutes and 34 seconds. Mean interaction duration is 4:26 minutes ( $SD = 5:11$ ). This is mainly due to the fact that interactions differ due to the type of activity. Interactions that are initiated by the user are typically longer, whereas agent-initiated reminder interaction can often be handled quickly and usually do not lead to longer dialogues. Pending a detailed analysis of the actual interaction logs, we differentiate between reminder-based interactions and other interactions by setting a threshold of 120 seconds. The 33 shorter – probably reminder – interactions have a mean duration of 45 seconds ( $SD = 0:33$ ). The 28 longer interactions have a mean duration of 8:57 minutes ( $SD = 4:37$ ). This differentiation gives further insight into daily usage of the system: A mean of 2.5 ( $SD = 1.7$ ) of the interactions per day were reminders, and a mean of 2.2 ( $SD = 1.0$ ) of the interactions lasted longer than 120 seconds.

The durations of the interactions indicate that the participant used the conversational assistant quite a lot. After the 13 days of usage she had entered 67 unique events in her calendar, five of which were serial events (yielding a total of 132 events displayed in the calendar). The large number of reminder-based interactions (33) also indicates that she successfully used this function of the system.

As the first main study has just ended, further analyses regarding the changes of daily life and routines are still due. However, the pre-pilot and pilot study already has been carried out with other elderly participants who could interact with the prototype system for three days at home. Different assessments or concerns were gathered from these participants as anecdotal feedback:

- Concerning the social presence and relationship: "We greet each other kindly every morning and he is asking what he can do for me. It's great."
- Concerning the dialogue: "Whenever we had problems in a conversation, we could resolve them. Was really nice."
- After coming out of coma and being isolated in hospital: "I would have been glad about having such a guy next to my bed." (in order to practice speaking and to have social interactions)



**Figure 7: The prototype system deployed in an elderly participant's apartment.**

- Concerning the size of the current setup and affordability: “Who wants to have this in the living room? Me not and I wouldn't buy it either, if it's a lot of money.”
- Concerning the duration of acquaintance: “He doesn't know anything about me. If that's possible, one has to use such a device for at least half a year to get used to it.”

## 6 CONCLUSIONS

The present work has explored how conversational agents can be used to provide cognitive or emotional assistance to elderly users. We have focused in particular on the use of spoken language dialogue as a preferred way of interacting with technical systems (as indicated by studies by others as well as ourselves). Yet, enabling successful and acceptable dialogue with these user groups raises several challenges and communication problems abound quickly with off-the-shelf dialogue system technology. However, our findings indicate that virtual assistants can still be an effective and acceptable help if they provide abilities for the kind of socially cooperative dialogue needed to resolve these issues. The key insight of the present project is how conversational agents can be built such that this is possible for the majority of issues, even for the special user groups of persons with a mild cognitive impairment (and often also additional motoric or perceptual handicaps). This requires numerous things, from the processing of subtle, multimodal and context-dependent communication-relevant signals, to generating them in combination with visual cues (calendar), to enabling a highly flexible dialogue with responsive turn-taking, communicative feedback, and (pro-)active strategies for avoiding communication problems as well as repairing them. One prerequisite to achieve this was a high degree of user involvement throughout

the design and implementation phases, which also helped a lot in increasing acceptance and willingness to participate in the project.

## ACKNOWLEDGMENTS

This research was supported by the German Research Foundation (DFG) in the Cluster of Excellence ‘Cognitive Interaction Technology’ (EXC 277) and by the German Federal Ministry of Education and Research (BMBF) in the project ‘KOMPASS’ (FKZ 16SV7271K).

## REFERENCES

- [1] Antje Amrhein, Katharina Cyra, and Karola Pitsch. 2016. Processes of reminding and requesting in supporting people with special needs. Human practices as basis for modeling a virtual assistant?. In *Proceedings 1st ECAI Workshop on Ethics in the Design of Intelligent Agents*. The Hague, The Netherlands, 18–23.
- [2] Timothy W. Bickmore, Rebecca A. Silliman, Kerrie Nelson, Debbie M. Cheng, Michael Winter, Lori Henault, and Michael K. Paasche-Orlow. 2013. A randomized controlled trial of an automated exercise coach for older adults. *Journal of the American Geriatrics Society* 61 (2013), 1676–1683. <https://doi.org/10.1111/jgs.12449>
- [3] Mara Brandt, Britta Wrede, Franz Kummert, and Lars Schillingmann. 2017. Confirmation detection in human-agent interaction using non-lexical speech cues. Presented at the *AAAI Fall Symposium on Natural Communication for Human-Robot Collaboration*. (2017).
- [4] Hendrik Buschmeier and Stefan Kopp. 2013. Co-constructing grounded symbols—Feedback and incremental adaptation in human-agent dialogue. *Künstliche Intelligenz* 27 (2013), 137–143. <https://doi.org/10.1007/s13218-013-0241-8>
- [5] Hendrik Buschmeier and Stefan Kopp. 2018. Communicative listener feedback in human-agent interaction: artificial speakers need to be attentive and adaptive. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems*. Stockholm, Sweden.
- [6] Herbert H. Clark. 1996. *Using Language*. Cambridge University Press, Cambridge, UK. <https://doi.org/10.1017/CBO9780511620539>
- [7] Katharina Cyra, Antje Amrhein, and Karola Pitsch. 2016. Fallstudien zur Alltagsrelevanz von Zeit- und Kalenderkonzepten. In *Mensch und Computer 2016 Kurzbeiträge*. Aachen, Germany, 1–5. <https://doi.org/10.18420/muc2016-mci-0253>
- [8] Katharina Cyra and Karola Pitsch. 2017. Dealing with long utterances: How to interrupt the user in a socially acceptable manner?. In *Proceedings of the 5th International Conference on Human Agent Interaction*. Bielefeld, Germany, 341–345. <https://doi.org/10.1145/3125739.3132586>
- [9] Katharina Cyra and Karola Pitsch. 2017. Dealing with ‘long turns’ produced by users of an assistive system: How missing uptake and reciprocity lead to turn increments. In *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication*. Lisbon, Portugal, 329–334. <https://doi.org/10.1109/ROMAN.2017.8172322>
- [10] Farina Freigang, Sören Klett, and Stefan Kopp. 2017. Pragmatic multimodality: Effects of nonverbal cues of focus and certainty in a virtual human. In *Proceedings of the 17th International Conference on Intelligent Virtual Agents*. Stockholm, Sweden, 142–155. [https://doi.org/10.1007/978-3-319-67401-8\\_16](https://doi.org/10.1007/978-3-319-67401-8_16)
- [11] Farina Freigang and Stefan Kopp. 2015. Analysing the modifying functions of gesture in multimodal utterances. In *Proceedings of the 4th Conference on Gesture and Speech in Interaction (GESPIN)*. Nantes, France, 107–112.
- [12] Farina Freigang and Stefan Kopp. 2016. This is what's important—using speech and gesture to create focus in multimodal utterance. In *Proceedings of the 16th International Conference on Intelligent Virtual Agents*. Los Angeles, CA, USA, 96–109. [https://doi.org/10.1007/978-3-319-47665-0\\_9](https://doi.org/10.1007/978-3-319-47665-0_9)
- [13] Marcel Kramer, Ramin Yaghoubzadeh, Stefan Kopp, and Karola Pitsch. 2013. A conversational virtual human as autonomous assistant for elderly and cognitively impaired users? Social acceptability and design considerations. In *Proceedings of INFORMATIK 2013*. Koblenz, Germany, 1105–1119.
- [14] Gregor Mehlmann, Kathrin Janowski, and Elisabeth André. 2016. Modeling grounding for interactive social companions. *Künstliche Intelligenz* 30 (2016), 45–52. <https://doi.org/10.1007/s13218-015-0397-5>
- [15] Markus Meis. 2013. Nutzerzentrierte Entwicklung eines Erinnerungssystemen. Presented at *Abschlusssymposium Niedersächsischer Forschungsverbund Gestaltung altersgerechter Lebenswelten*. (2013). <https://www.altersgerechte-lebenswelten.de/>
- [16] Juliana Miehle, Ilker Bagci, Wolfgang Minker, and Stefan Ultes. 2017. A social companion and conversation partner for elderly. In *Proceedings of the 8th International Workshop On Spoken Dialogue Systems*. Farmington, PA, USA.
- [17] Christiane Opfermann and Karola Pitsch. 2017. Reprompts as error handling strategy in human-agent-dialog? User responses to a system's display of non-understanding. In *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication*. Lisbon, Portugal, 310–316. <https://doi.org/10.1109/ROMAN.2017.8172319>

- [18] Christiane Opfermann, Karola Pitsch, Ramin Yaghoubzadeh, and Stefan Kopp. 2017. The communicative activity of 'making suggestions' as an interactional process: Towards a dialog model for HAI. In *Proceedings of the 5th International Conference on Human Agent Interaction*. Bielefeld, Germany, 161–170. <https://doi.org/10.1145/3125739.3125752>
- [19] Tim Paek and Eric Horvitz. 2000. Conversation as action under uncertainty. In *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*. Stanford, CA, USA, 455–464.
- [20] Lazlo Ring, Lin Shi, Kathleen Totzke, and Timothy Bickmore. 2014. Social support agents for older adults: longitudinal affective computing in the home. *Journal on Multimodal User Interfaces* 9 (2014), 79–88. <https://doi.org/10.1007/s12193-014-0157-0>
- [21] David Schlangen, Timo Baumann, Hendrik Buschmeier, Okko Buß, Stefan Kopp, Gabriel Skantze, and Ramin Yaghoubzadeh. 2010. Middleware for incremental processing in conversational agents. In *Proceedings of the 11th Annual Meeting of the Special Interest Group in Discourse and Dialogue*. Tokyo, Japan, 51–54.
- [22] David Schlangen and Gabriel Skantze. 2011. A general, abstract model of incremental dialogue processing. *Dialogue and Discourse* 2 (2011), 83–111. <https://doi.org/10.5087/dad.2011.105>
- [23] Candace Sidner, Timothy Bickmore, Charles Rich, Barbara Barry, Lazlo Ring, Morteza Behrooz, and Mohammad Shayganfar. 2013. Demonstration of an always-on companion for isolated older adults. In *Proceedings of the 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Metz, France, 148–150.
- [24] Gabriel Skantze. 2007. *Error Handling in Spoken Dialogue Systems. Managing Uncertainty, Grounding and Miscommunication*. Ph.D. Dissertation. Computer Science and Communication, Department of Speech, Music and Hearing, KTH Stockholm, Stockholm, Sweden.
- [25] Carolin Straßmann and Nicole C. Krämer. 2017. A categorization of virtual agent appearances and a qualitative study on age-related user preferences. In *Proceedings of the 17th International Conference on Intelligent Virtual Agents*. Stockholm, Sweden, 413–422. [https://doi.org/10.1007/978-3-319-67401-8\\_51](https://doi.org/10.1007/978-3-319-67401-8_51)
- [26] Carolin Straßmann, Astrid Rosenthal von der Pütten, Ramin Yaghoubzadeh, Raffael Kaminski, and Nicole Krämer. 2016. The effect of an intelligent virtual agent's nonverbal behavior with regard to dominance and cooperativity. In *Proceedings of the 16th International Conference on Intelligent Virtual Agents*. Los Angeles, CA, USA, 15–28. [https://doi.org/10.1007/978-3-319-47665-0\\_2](https://doi.org/10.1007/978-3-319-47665-0_2)
- [27] David R. Traum. 1994. *A Computational Theory of Grounding in Natural Language Conversation*. Ph.D. Dissertation. University of Rochester, Rochester, NY, USA.
- [28] Maria Velana, Sascha Gruss, Georg Layher, et al. 2017. The SenseEmotion database: A multimodal database for the development and systematic validation of an automatic pain- and emotion-recognition system. In *Proceedings of the 4th LAPR TC 9 Workshop on Pattern Recognition of Social Signals in Human-Computer-Interaction*. Cancun, Mexico, 127–139. <https://doi.org/10.1007/978-3-319-59259-6>
- [29] Thomas Visser, David R. Traum, David DeVault, and Rieks op den Akker. 2014. A model for incremental grounding in spoken dialogue systems. *Journal on Multimodal User Interfaces* 8 (2014), 61–73. <https://doi.org/10.1007/s12193-013-0147-7>
- [30] Eduard Wall, Lars Schillingmann, and Franz Kummert. 2017. Online nod detection in human–robot interaction. In *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication*. Lisbon, Portugal, 811–817. <https://doi.org/10.1109/ROMAN.2017.8172396>
- [31] Leo Wanner, Elisabeth André, Josep Blat, et al. 2017. KRISTINA: A knowledge-based virtual conversation agent. In *Proceedings of the 15th International Conference on Practical Applications of Agents and Multi-Agent Systems*. Porto, Portugal, 284–295. [https://doi.org/10.1007/978-3-319-59930-4\\_23](https://doi.org/10.1007/978-3-319-59930-4_23)
- [32] Ramin Yaghoubzadeh, Hendrik Buschmeier, and Stefan Kopp. 2015. Socially cooperative behavior for artificial companions for elderly and cognitively impaired people. In *Proceedings of the 1st International Symposium on Companion Technology*. Ulm, Germany, 15–19.
- [33] Ramin Yaghoubzadeh and Stefan Kopp. 2016. Towards graceful turn management in human-agent interaction for people with cognitive impairments. In *Proceedings of the 7th Workshop on Speech and Language Processing for Assistive Technologies*. San Francisco, CA, USA, 26–31.
- [34] Ramin Yaghoubzadeh and Stefan Kopp. 2017. Enabling robust and fluid spoken dialogue with cognitively impaired users. In *Proceedings of the 18th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Saarbrücken, Germany, 273–283.
- [35] Ramin Yaghoubzadeh, Marcel Kramer, Karola Pitsch, and Stefan Kopp. 2013. Virtual agents as daily assistants for elderly or cognitively impaired people. In *Proceedings of the 13th International Conference on Intelligent Virtual Agents*. Edinburgh, United Kingdom, 79–91. [https://doi.org/10.1007/978-3-642-40415-3\\_7](https://doi.org/10.1007/978-3-642-40415-3_7)
- [36] Ramin Yaghoubzadeh, Karola Pitsch, and Stefan Kopp. 2015. Adaptive grounding and dialogue management for autonomous conversational assistants for elderly users. In *Proceedings of the 15th International Conference on Intelligent Virtual Agents*. Delft, The Netherlands, 28–38. [https://doi.org/10.1007/978-3-319-21996-7\\_3](https://doi.org/10.1007/978-3-319-21996-7_3)
- [37] Victoria Young and Alex Mihailidis. 2010. Difficulties in automatic speech recognition of dysarthric speakers and implications for speech-based applications used by the elderly: A literature review. *Assistive Technology* 22 (2010), 99–112. <https://doi.org/10.1080/10400435.2010.483646>