Adaptive Grounding and Dialogue Management for Autonomous Conversational Assistants for Elderly Users

Ramin Yaghoubzadeh[1], Karola Pitsch[2], and Stefan Kopp[1]

[1] Social Cognitive Systems Group, CITEC, Bielefeld University,
P.O. Box 10 01 31, 33501 Bielefeld, Germany
[2] Institute for Communication Studies, University of Duisburg-Essen,
Universitätsstraße 12, 45141 Essen, Germany

Abstract. People with age-related or congenital cognitive impairments require assistance in daily tasks to enable them to maintain a self-determined lifestyle in their own home. We developed and evaluated a prototype of an autonomous spoken dialogue assistant to support these user groups in the domain of week planning. Based on insights from previous work with a WOz study, we designed a dialogue system which caters to the interactional needs of these user groups. Subjects were able to interact successfully with the system and rated it as equivalent in terms of robustness and usability compared to the WOz prototype.

Keywords: assistive technology, cognitive impairment, conversational agents

1 Introduction

Many older adults, when impacted by the aging processes, are able to continue living in their accustomed home environment if they receive support for their gradually declining capacities. The same holds true for people with congenital or acquired cognitive impairments [3] who, thanks to improved social integration and better-suited healthcare, nowadays can also enjoy the boon of independence, if given support. The factors that most often lead to an end of independent living, besides fall events, are cognitive problems in both user groups that lead to an erratic or completely lost day structure. This can range from forgetfulness to a total lack of the sense of time, hindering their management of activities like meals, medication, appointments or social events.

We present work towards an assistive conversational agent for self-determined daily schedule management and maintenance, a cooperation with one of Europe's largest health and social care providers for elderly people and people with various disabilities. We previously explored the principal suitability and acceptability of spoken-language virtual assistants for these users, identifying constraints for successful interaction [16]. In the present work we explored how interaction between an *autonomous* system and people from those groups could be made robust and effective, ensuring both mutual understanding and assessment as an acceptable interactant. After highlighting existing work in our field and the foundations from our previous work, we will describe our dialogue management framework `flexdiam` designed with these requirements in mind, and results of an initial evaluation of the autonomous system with older adults.

[3] In this paper, we adopt the terminology recommended by ACM SIGACCESS. [4]

In: Proceedings of the 15th International Conference on Intelligent Virtual Agents (IVA 2015), pp.28–38, Delft, The Netherlands.
POSTPRINT
The final publication is available at Springer via http://doi.org/10.1007/978-3-319-21996-7_3

2 Related work

Dialogue management has received much attention, and a number of established approaches exist; see McTear [10] for an overview. Of particular relevance here are flexible grounding mechanisms with error and repair handling. Larsson [9] proposed an issue-based dialogue management capable of ellipsis resolution and accomodating unaddressed questions; Skantze dealt with errors and repair in spoken dialogue and tracked the grounding status of concepts and confidence scores in the Galatea discourse modeller [14]. Roque & Traum [11] proposed a model to track the extent to which material has reached mutual belief in a dialogue by distinguishing between several discrete degrees of groundedness, demonstrating an improved appraisal of an agent's dialogue skills. Crook et al. [5] presented an approach for dealing with user barge-ins in a dialogue system, discerning between continue, abort and replan actions. Buschmeier & Kopp [3] proposed a Bayesian model incorporating evidence from back-channel feedback to infer continuous degrees of understanding or acceptance. These systems have generally not been confronted with the special interactional needs of elderly or cognitively impaired users. Interaction in that domain has mostly been text-based or multiple-choice, e.g. the Always-On Companion for isolated older adults [13].

There is a substantial body of work on the requirements and potentials of interactive assistive systems especially for elderly people. The GUIDE project [7] identified dimensions of user models for different impairments, and potential domains of support, remarking that spoken language is the preferred modality for older users without technical experience. The use of speech input by older adults has been explored systematically [17], taking into account the detrimental effect of dysarthria and articulation disorders on recognition quality, although even then limited speech interaction, e.g. for environmental control, can be realized [6,8]. Beskow et al. [2] evaluated a prototype multimodal reminder agent for people with cognitive problems that combined handwriting recognition and spoken dialogue. Studies for people with general cognitive impairments are still few; a first study showed that parallels to older adults exist in interactions with spoken dialogue systems, such as a diminished awareness of system errors [16].

3 Background and requirements

Lessons from previous work Previously [16], we engaged in a participatory design process with two user groups – older adults (n=6) living autonomously in apartments attached to a nursing home, and people with cognitive impairments (n=11) from an institution offering technological experience and training. The design process comprised interviews, focus groups, interaction experiments and concluding focus groups or interviews (depending on specific impairments, in accord with care personnel). In the interaction experiments, we used a WOz version of a spoken dialogue virtual assistant with a graphical calendar. Subjects were asked to enter fictional appointments using natural speech. They were able to use image cue cards depicting the day of the week, time, and topic of an activity. Subjects could at least read the days of the week and times, but images were required for topics. Crucially, we manipulated the "agent's" understanding of entered appointments by introducing systematic errors in the summaries at predefined

times. In a between-subject design with two conditions, we contrasted two methods for grounding and confirming information provided by the user: asking for confirmation for each "slot", or summarizing the whole appointment in one utterance. We found a noticeable difference in error detection rates, most pronounced for the group with at least mild mental retardation (APA-DSM-IV F70 [1]). We used a simplified process of usability rating involving nine interview questions and a graphical scale on which subjects with numerical or abstraction problems could visually point out ratings. There were no differences in rating between the conditions for any user group.

The central insight from the experiments is that the information grounding process in a multimodal spoken dialogue system for people with cognitive impairments must be highly adaptive, to avoid situations in which the users are overwhelmed by large chunks of information and fail to spot system errors. Grounding must thereby extend to very fine-grained and explicit strategies, in which each piece of information is best negotiated individually. This did not lead to negative usability assessment, in contrast to our inital assumptions. Anecdotally, we found in informal focus groups and meetings that participants looked forward to further studies, primarily stating their positive impression of the agent. This was especially pronounced in the group of people with cognitive impairments, but also with the older adults. One older participant hung a picture of the agent on her mirror and chatted about the agent, showing images to her grandchildren.

Requirements for an autonomous assistant The analysis of our initial interaction studies, related work, and interviews with subjects and care personnel, led to the following central requirements and design principles for robustness and interaction quality in a spoken-dialogue assistant for our user groups: **1)** The system must satisfy critical requirements for fluid dialogue: it has to present and process information in an *incremental* fashionr to facilitate the reception of the other party by maximizing both the duration of presentation and the *timeliness* of its feedback. Feedback and corrections must be flexibly processed *at any time*, in the form of barge-ins, but also as later revisions. The system must therefore be able to reason about changes in assumptions and evidence between *arbitrary points* in the past and the present and also to refer to past discourse units. **2)** The system has to be prepared for uncertainty in the interaction in two ways: firstly, input variables can be *uncertain*, e.g. due to inaccuracies in speech recognition, exacerbated by articulation disorders. Wherever possible, the system must maintain parallel *alternative hypotheses* and resolve them in a controlled manner. Secondly, even explicit feedback signals from the user cannot always be presumed to accurately reflect the result of a comprehensive assessment of the common ground by them. The system must hence be able to employ optimal strategies for information presentation that maximize the *incidence of meaningful feedback* but are not perceived as intrusive. **3)** Information should be structured in a way suited to the user groups in terms of its simplicity. When a hierarchical task structure is required (such as asking clarification questions or solving sub-tasks to contribute to a more complex task), the system must offer *transparency*. It should be able to summarize the current state of the hierarchical task, and to *inquire about the user's understanding* of it. Thus, when communication problems arise, such as a lack of contributions or evidence that the user is confused, it should attempt to successively *request explicit feedback* from the user by descending the task hierarchy, or back-track to a previously grounded point.

4 The `flexdiam` dialogue manager

In order to realize the aforementioned requirements, we developed and evaluated a dialogue management framework called `flexdiam`, written in Python and first presented here. The design of this framework was inspired by the analysis of the requirements and possible interaction problems of our variously impaired user groups, to allow for robust, flexible, and acceptable, mostly task-related, spoken-language interactions.

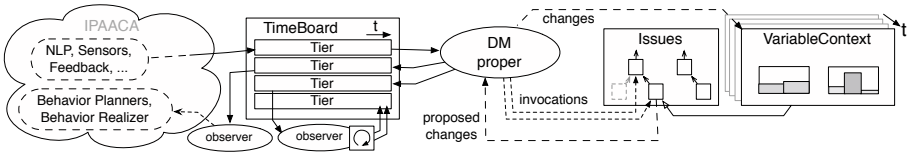

Fig. 1. Architecture overview of `flexdiam`

Information in the system is spread over three complementary data structures with specific capabilities contributing to this end (see Fig. 1 for an architecture overview). The **temporal structure** of dialogue is represented in a central data structure, termed `TimeBoard`, which stores all events, past, ongoing, or future (projected), in thematically grouped *tiers* (Fig. 3). It serves as the interface for communication between input processing, dialogue management proper, and behavior planning and realization. It is a globally modifiable blackboard, and it is by convention that strict governance of the flexible temporal aspects of dialogue is guaranteed. Entries on the board have a payload content (usually an *incremental unit* in our middleware IPAACA [12]) as well a as start and end time, possibly undefined to reflect open intervals. The board provides a set of interval relations enabling predicates over a set of tiers, describing their structure for any requested interval of time. The board receives any changes as write operations on the start, end, or content of events, storable with timestamps for full rollback capability. Any set of tiers can have attached *observers* to which all changes of event parameters on those tiers are relayed, and which can act as pipelines or concurrent processors. Data other than events with temporal extent, i.e. **knowledge and propositional information** are represented via a structure termed `VariableContext`, (Fig. 4, left), a blackboard satisfying two requirements: firstly, all information may reside there in the form of distributions with attached entropy values. Modules may analyze the distributions or entropies or instantiate maximum-likelihood versions. Secondly, the `VariableContext` has full temporal independence. All changes are stored as time-stamped deltas, providing the means both for rollbacks and for analysis between two points in time, e.g. changes in distributions. In order to represent **task and discourse states**, the system follows a hybrid approach that centers around a forest of structures called `Issues`, terminology adapted from Larsson [9], that represent (attributed) common current topics or current questions that have to be resolved cooperatively. In `flexdiam`, they are independent agents that encapsulate the structure of the task addressed so far, localized planning, as well as situated interpretative capability and situated capability for action,

encompassing both dialogue contributions and other side effects. Moreover, they contain a local success and failure test which the dialogue manager proper uses to decide on finalizing an Issue to defer processing back to the parent. The dialogue manager proper can invoke six operations on an Issue object: `interpret`, which is used to interpret an NLU result (Fig. 4, center), `can_interpret`, which performs a shallower check of whether a NLU result can be (partially) used by the issue object, `introduce`, which must be called and completed exactly once for any system-introduced issue opening a new sequence before the issue is considered a valid acceptor for user contributions (preemptive user contributions inside a sequence are however acceptable and expected), `reintroduce`, which is used to pick up the previously introduced issue when the system takes the floor from idle state, `child_closed`, which is invoked whenever a child issue has been finalized by the DM proper due to a flagged local success or failure condition, and `make_certain`, which is the meta-dialogue operation to summarize the current state of the issue object and explicitly ask for confirmation. Issues can themselves mark local progress using the `progressed` function. This invalidates all children (see below) and invokes `child_progressed` on the parent. Whenever an **Issue object is invoked**, it processes the respective input and proposes a comprehensive change set, termed `IssueReturn` (Fig. 4, right), composed of (1) a change set to merge into the VariableContext, (2) a set of *transmits*, which are objects that the DM proper should insert on the TimeBoard, publishing them to external components, (3) a flag indicating whether the issue object deems the invocation to have been an exhaustive operation on the input, and (4) optionally, a child issue that the DM proper should in the future pick as the preferential entry point for invocation in lieu of the presently invoked issue. In line with the general notion of temporal variability and uncertainty, all operations that do not have immediate effect are treated as **asynchronously performed operations**. New issue or transmit objects are not guaranteed to fully come to pass in the dialogue, nor assumed to be successfully completed a-priori. Both traverse state machines dependent upon external events, starting as incomplete and unreliable entities (cf. Fig. 2). The invocation of the **DM proper** is realized by observers on the input and structure tiers that call corresponding functions in the DM when new parses are received or a change in structure is detected (such as a prolonged silence after a user utterance).

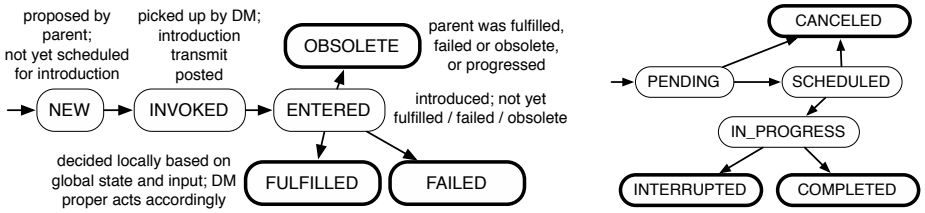

Fig. 2. Left: State machine for `Issue` objects. Terminal states (=issue inactive) in bold. **Right:** State machine for `Transmit` objects. State transitions closely correspond to planners / realizers.

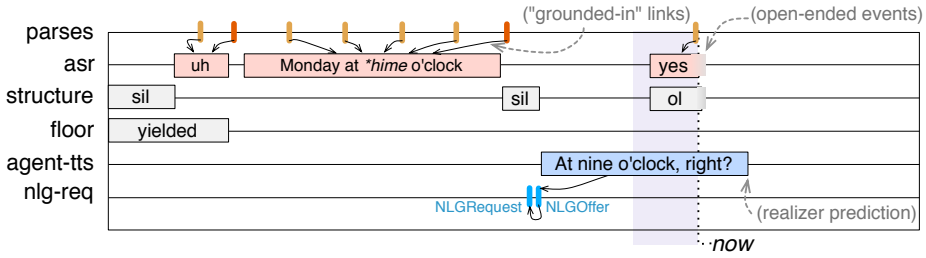

Fig. 3. The `TimeBoard` encapsulates the temporal structure and dependency of events. The typical setup encompasses input tiers (*asr* and dependent incremental *parses*), information derived by helper agents (*structure*: sil=silence, ol=overlap; *floor* is explicitly yielded by system after questions), and output tiers (*nlg-req* contains posted requests and offers received from NLG program, *agent-tts* contains agent utterances as reported by realizer feedback). **Example situation:** The user barges in with a reply after the agent asked a clarification question since the preceding user utterance was misarticulated. In the setup of the present work, the agent would interrupt himself.

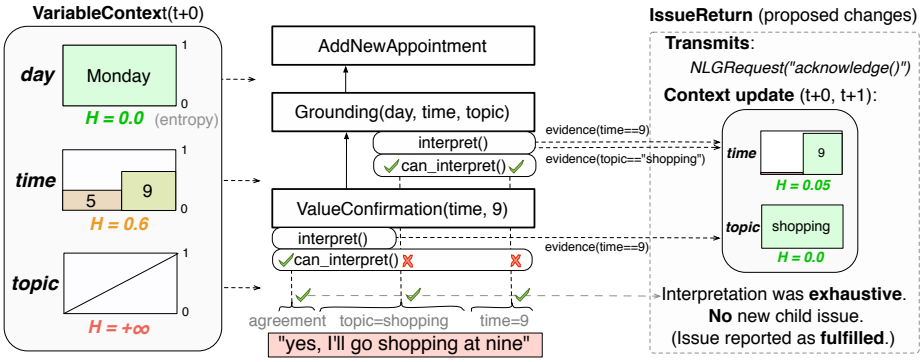

Fig. 4. Uncertain information in `VariableContext` and unfolded discourse in `Issue` forest. **Left:** Relevant subset of prior context: *day* has been established, *time* is uncertain, *topic* is unknown. The Grounding issue chose to propose a ValueConfirmation for *time*, primed for a yes/no answer, because the entropy was less than 1 bit (state from Fig. 3). **Center:** Hierarchical situated interpretation, moments after Fig. 3; the confirmation part of the utterance is understood as a reply to the most recent (bottom-most) issue, while the remainder is captured here by the immediate parent. **Right:** IssueReturn contains proposed updates from interpret(). Context is to be updated with new distributions. An *acknowledge()* will be requested from the NLG as contingent feedback. No new children are proposed and since the issue deems its success condition (processing a yes/no reply) fulfilled, it will henceforth report its fulfilled() predicate as True. **N.B.** Had the user omitted the "yes", the remaining evidence might, depending on configuration, have been sufficient for the parent to *progress*, silently deactivating the child to *OBSOLETE* state (cf. Fig. 2).

5 Evaluation

After the basic functionality tests accompanying the design of `flexdiam`, we carried out an early evaluation with our user groups, in line with the approach of an user-centered, inclusive design, with a system resulting directly from our previous work with them.

Study design and setup The task for participants was entering appointments into a calendar using spoken conversation with the virtual assistant "Billie". The setup comprised a dedicated ASR machine running Windows ASR and DragonNaturallySpeaking Client edition, with a desk microphone, and a touch-screen PC running `flexdiam`, parser, NLU, behavior planner, a 3D environment with a character driven by ASAPrealizer [15], and CereVoice TTS. During interactions, participants met the virtual agent as their conversation partner, next to a calendar with real-time visualization of intermediate results of the incremental interpretation of user input. The scene on the touch screen included a red button with a back arrow, introduced as an 'emergency button' should the agent constantly fail to understand the participant. The button was "wired" to reset the system to the top-most dialogue state (clarifying whether the participants wanted to enter anything else). We set up the system to cope with various expected verbalizations for the topic of calendar management, to be able to handle all types of repair patterns that we previously found, and set up the Issue structure accordingly. For all participants, we chose the slow-paced, explicit grounding strategy of asking for confirmation for every slot, which had been identified as the most reliable one for impaired participants. In contrast to a previous WOz experiment [16], we did not provide cue cards with completely predefined fictional appointments. Rather, participants were free to enter whatever they wished – but we provided a single cue sheet with textual and iconic representations of possible activities (Fig. 6, left). We prepared the NLU to preferentially spot those topics and respective synonyms. Yet, all other inputs were also valid; they were subjected to a simple heuristic extraction to reduce them to short entries.

Participants and procedure After an initial test run with three people from the group with cognitive impairments, in which additional failure modes were detected and addressed, we conducted a controlled small-scale experiment with the group of older adults (n=6, 4f, 2m, ages 77–86). The same group had already participated in the previous experiment, so they had cursory acquaintance with the task and the agent Billie from about 14 months earlier. After initial written consent, participants were asked to participate in a short speech recognizer training process by calmly reciting excerpts from a recipe book in three episodes of about 30s each. While the training program operated in the background, they were instructed about their task for the experiment proper. We split the interaction into two parts, an initial phase where participants should attempt to make up to three entries without further instruction as to the format, followed by an intermission in which we would be able to intervene should the system be totally overwhelmed by the user's interactional style. The primary hints provided where to avoid excessive verbosity and to use verbal instead of paraverbal feedback. A second interaction phase followed, for which participants were asked to make at least five entries. After this final phase, participants were asked to participate in a structured interview including the following usability questions (the same as in previous work; translated):

(1) How did you like to plan your week with Billie (B)? (2) Did B always do what you wanted? (3) Did B provide sufficient help? (4) Could you understand B's language well? (5) Did B express himself in an easy way? (6) Did B understand you correctly? (7) How did you like the image-based calendar? (8) Do you think B could be of help to you? (9) Do you think B could be a good appointment assistant?

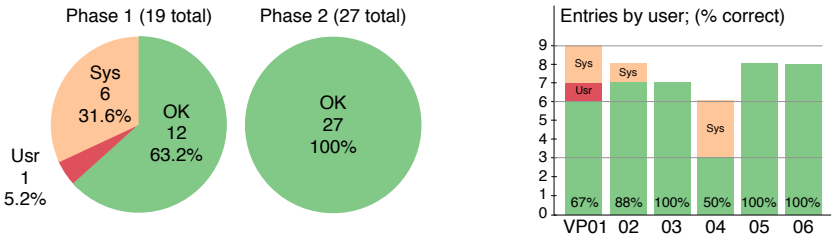

Fig. 5. Left: Completed topic negotiations in the two interaction phases. *OK*: entered verbatim as dictated or with acceptable paraphrase; *Sys*: ratified, but system failed to offer acceptable topic; *Usr*: false entry due to the user not detecting a system error (cf. previous work). **Right:** Total, correct and erroneous negotiations, by participant, with proportion of correct entries.

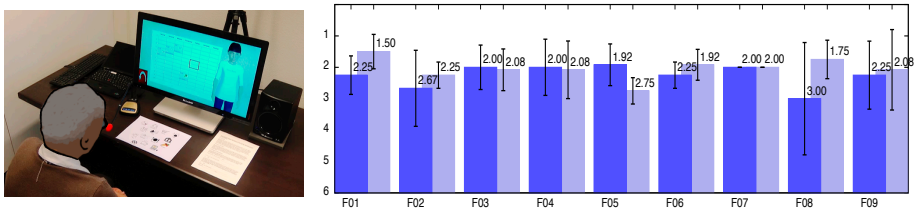

Fig. 6. Left: Interaction with autonomous system (older participant, anonymized). **Right:** Comparison of usability ratings (see text) between previous WOz experiment [16] (pale) and the present autonomous system (dark), for elderly users. German school grades (1–6, 1 is best).

Results All participants were generally able to enter appointments successfully. Concessions were made by participants in accepting entries that were altered by the system based on topic extraction heuristics (for topics not on the cue sheet). In total, 46 entries were negotiated, of which 7 (15.2%) did not correspond to the user's original wishes. All errors occurred in the first phase (Fig. 5). Topic negotiation was counted as successful if a suitable paraphrase was settled on (e.g. VP05 tried to enter "hairdresser", initially not recognized; she then elaborated "shampoo and cut", which she later ratified – we counted that as a success). Communication problems mainly resulted from the inability of the system to process long, convoluted utterances properly. In particular, one participant (VP04) used a very verbose style with back stories and indirect

replies. He produced noticeably more utterances compared to the average of the other participants (170 vs. 104), used more than twice the utterances per negotiation (28.33 vs. 13.16), while 50.6% were three words or longer (the maximum for the others was 27%). Several times, the ASR computer was causing lag due to processing of several conjoint utterances. The explanatory intervention in the intermission did not make the participant alter his interactional style in a lasting way. Further, as the study was conducted in the field (a care facility), there were fluctuations in noise due to simultaneous events in neighboring rooms, affecting portions of two other interactions. However, no participant made use of the "fallback" touch-screen reset button, which was the only supported non-speech interaction mode. Remarkably, the ratings of the usability questions differed only slightly from the WOz experiment conducted in our previous work [16] (overlaid in Fig. 6, right). That is, the *autonomous* assistant based on `flexdiam` achieved ratings similar to a WOz system in which the entire input processing and system response selection tasks were performed by a human wizard. The least favorable ratings came from VP04, for whom the system failed to provide a pleasant and effective interaction. Consequently, his intention-to-use (question F08) was minimal (grade 6).

6 Discussion and Conclusion

The general goal of the present work is to develop conversational assistants that can work autonomously and robustly with user groups like older adults or cognitively impaired people. Such groups often bring about special interactional challenges. Based on the results from previous WOz studies with these user groups, we have developed the dialogue manager `flexdiam` that specifically aims to enable flexible and adaptive grounding mechanisms, including barge-ins, repairs and interactively negotiated content. The evaluation study we have conducted with older adults showed that the general design of the system, along with the explicit grounding and ratification strategy selected for allowing users to detect system errors, is suitable to almost match the results obtained with the WOz version of the assistant – only one error was never detected. Participants with a relatively brief interaction style could effectively enter information error-free. This results lends support to the notion that autonomous conversational assistants in domains as confined as calendar management (although including relatively open sub-tasks like entering activities) are in principle possible.

In general, the challenge for such assistants is to maximize robustness of spoken dialogue, while at the same time ensuring acceptable interactions. This holds especially with elderly users who sometimes exhibit excessive verbosity, even when explicitly instructed to adhere to a brief interaction style. Consequently, the system must take initiative where it cannot cope, to actively "make life easier" for itself and preserve operability. Here, a virtual agent, by virtue of its ability to convey subtle interactional information, can be a socially acceptable "pace controller" by emitting successively more explicit turn management behavior instead of just barging in to grab the floor. Extending the assistant in this direction will be one of the areas for our future work. Moreover, we plan to extend the evaluation study to the other user group of people with congenital or acquired cognitive impairments.

Acknowledgements – This research was partially supported by the German Federal Ministry of Education and Research (BMBF) in the project 'VERSTANDEN', by the Deutsche Forschungsgemeinschaft (DFG) in the Center of Excellence 'Cognitive Interaction Technology' (CITEC), and by the Volkswagen Foundation.

References

1. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR, 4th ed. American Psychiatric Publ., Arlington, VA. (2000)
2. Beskow, J., Edlund, J., Granström, B., Gustafson, J., Skantze, G., Tobiasson, H.: The MonAMI Reminder : a spoken dialogue system for face-to-face interaction. In: 10th Annual Conf of the International Speech Communication Assoc, INTERSPEECH 2009. pp. 300–303. (2009)
3. Buschmeier, H., Kopp, S.: Using a Bayesian model of the listener to unveil the dialogue information state. In SemDial 2012: Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue, pp. 12—20, Paris, France. (2012)
4. Cavender, A., Trewin, S., Hanson, V.: ACM SIGACCESS Accessible Writing Guide. `www.sigaccess.org/resources/accessible-writing-guide`. Accessed 2015-03-10.
5. Crook, N., Smith, C., Cavazza, M., Pulman, S.,Moore, R., Boye, J.: Handling User Interruptions in an Embodied Conversational Agent. In: Proceedings of the AAMAS International Workshop on Interacting with ECAs as Virtual Characters. pp. 27–33. (2010)
6. Fager, S.K., Beukelman, D.R., Jakobs, T., Hosom, J.-P.: Evaluation of a Speech Recognition Prototype for Speakers with Moderate and Severe Dysarthria: A Preliminary Report. Augmentative and Alternative Comm., 26(4):267–277. (2010)
7. GUIDE Consortium: User Interaction & Application Requirements - Deliverable D2.1. (2011)
8. Hawley, M.S., Enderby, P., Green, P., Cunningham, S., Brownsell, S., Carmichael, J., Parker, M., Hatzis, A., ONeill, P., Palmer, R.: A speech-controlled environmental control system for people with severe dysarthria. Medical Engineering & Physics, 29(5):586-–593. (2007)
9. Larsson, S.: Issue-Based Dialogue Management. University of Gothenburg. (2002)
10. McTear, M.: Spoken Dialogue Technology: Towards the Conversational User Interface. Springer, London. (2004)
11. Roque, A., Traum, D.: Improving a Virtual Human Using a Model of Degrees of Grounding. In: Proc Int Joint Conference on Artificial Intelligence IJCAI-09, Pasadena, CA. (2009)
12. Schlangen, D., Baumann, T., Buschmeier, H., Buß, O., Kopp, S., Skantze, G., Yaghoubzadeh, R.: Middleware for incremental processing in conversational agents. In: Special Interest Group on Discourse and Dialogue. pp. 51—54. (2010)
13. Sidner, C., Bickmore, T., Rich, C., Barry, B., Ring, L., Behrooz, M., Shayganfar, M.: An Always-On Companion for Isolated Older Adults. SIGdial 2013, Metz, France. (2013)
14. Skantze, G.: Galatea: A discourse modeller supporting concept–level error handling in spoken dialogue systems. In Dybkjær, L., Minker, W. (Eds.), Recent Trends in Discourse and Dialogue. Springer. (2008)
15. van Welbergen, H., Reidsma, D., Kopp, S.: An Incremental Multimodal Realizer for Behavior Co-Articulation and Coordination. In: Proceedings of the 12th Int. Conference on Intelligent Virtual Agents. LNCS 7502. pp. 175–188. (2012)
16. Yaghoubzadeh, R., Kramer, M., Pitsch, K., Kopp, S.: Virtual agents as daily assistants for elderly or cognitively impaired people. In: Proc IVA 2013. LNCS 8108. pp. 79–91. (2013)
17. Young, V., Mihailidis, A.: Difficulties in Automatic Speech Recognition of Dysarthric Speakers and Implications for Speech-Based Applications Used by the Elderly: A Literature Review. In: Assistive Technology, 22(2):99–112. (2010)