flexdiam – Flexible dialogue management for incremental interaction with virtual agents (demo paper)

Ramin Yaghoubzadeh and Stefan Kopp

ryaghoubzadeh@uni-bielefeld.de, skopp@techfak.uni-bielefeld.de
Social Cognitive Systems Group, CITEC, Bielefeld University
P.O. Box 10 01 31, 33501 Bielefeld, Germany

Abstract. We present a demonstration system for incremental spoken human–machine dialogue for task-centric domains that includes a controller for verbal and nonverbal behavior for virtual agents. The dialogue management components can handle uncertainty in input and resolve it interactively with high responsivity, and state tracking is aware of momentary events such as interruptions by the user. Aside from adaptable dialogue strategies, such as for grounding, the system includes a multimodal floor management controller that attempts to limit the influence of idiosyncratic dialogue behavior on the part of our primary user groups – older adults and people with cognitive impairments – both of which have previously participated in pilot studies using the platform.

Keywords: incremental spoken interaction, uncertainty, dialogue management, floor management, virtual agents, nonverbal behavior, special user groups

1 Background

Spoken human-machine interaction affords access to modern technology to user groups that experience difficulties using other interactive modalities. Older adults unfamiliar with modern technology generally prefer spoken interaction [1]; and many people with cognitive impairments face challenges when interacting with the prevalent text-based interfaces. Regarding spoken-dialogue systems that offer an actual assistive function, many participants in these user groups report a preference for some degree of personification, embodiment, and social contingency [2]. Conversely, these systems can benefit from the effects of embodiment on interaction: offering additional output modalities that average interactants are already familiar with (such as gestures), and eliciting additional behaviors that can provide evidence about the dialogue situation (such as visually fixating the interlocutor as opposed to something else). Embodied virtual agents are an economic means to further these aims.

Previously, we explored the paradigm of the 'virtual assistant' for older adults and people with cognitive impairments, initially in a Wizard-of-Oz setup [5] in

Fig. 1. Left: Basic dialogue system setup with microphone and touchscreen (anonymized frame from autonomous study, older participant); **right:** example non-verbal behavior emitted by listening agent prior to a turn grab

the institutions of a large health-care provider, v. Bodelschwinghsche Stiftungen Bethel. In concert with Bethel, we chose assistance for appointment management and the maintenance of daily structure as the initial domain. We concluded that the approach found wide general acceptance among participants, particularly in the latter user group. We also concluded that dialogue structure for identical tasks ought to be adaptable to account for individual requirements: people with cognitive impairments in particular benefitted from fine-grained, explicit models of grounding information, leading to increased awareness of (simulated) system errors, while their self-reported usability ratings were not detrimentally affected by this. In subsequent experiments [6], we employed an autonomous prototype system using our `flexdiam` dialogue framework (see Fig. 1, left), and were able to replicate the results from the WOz studies. We found that one primary challenge of the existing system were overly long, verbose user turns. This was exacerbated when coincident with impaired articulation, and led to increased ASR delays and NLU confusion.

Building on research from conversation analysis and considering work on the perception of interruptions caused by agents, we fashioned a prototype multimodal interruption controller to emit nonverbal signals of gradually increasing urgency (see Fig. 1, right). Analyses of a small-scale pilot study with cognitively impaired users [7], where the controller was employed in parallel to a WOz-driven main task, indicated that these signals might be an effective – and acceptable – mechanism for managing the structure of user contributions.

The present demo showcases the current state of the dialogue framework, highlighting incremental processing of uncertain information, and incorporates the multimodal floor controller, modulating the listening behavior of the virtual agent in real time.

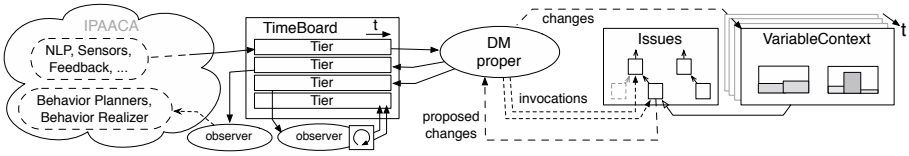

Fig. 2. High-level overview of `flexdiam` dialogue framework components. Input and output components are connected via IPAACA.

2 Framework overview

We provide a brief overview of the architecture here (see Fig. 2); for a more detailed look at the internal mechanisms, please refer to our previous work [6]. `flexdiam` is implemented in Python on top of the IPAACA middleware for incremental dialogue processing [3], which functions as the bridge for all input and output modules. Input modules include three different ASR modules, which can be run simultaneously, eye trackers, keyboard, mouse and touchscreen input. Output modules govern an embodied agent with synthesized speech output, dynamic GUI elements embedded in the agent scene, and various supplemental outputs, e.g. to control measuring equipment.

All events with temporal extents (i.e. with time of occurrence, or start and end times) are stored in a structure called TimeBoard that offers a categorized view of event tiers. Interval relations can be specified that trigger higher-level events. Microplanning and realization requests (and their status updates) are likewise placed on the board by the dialogue manager, and can be handled by external IPAACA modules. Factual information, and the state of the situation model, are stored in a structure called VariableContext, which can crucially treat any variable as a distribution and calculate derived statistics, such as entropy. The VariableContext is fully rewindable, enabling rollback in dialogue and also comparison between two points in time. The situation model is represented as a forest of Issue objects, which are encapsulated agents that have local interpretation and planning capabilities, i.e. they can autonomously introduce and retract sub-tasks and report on their capability of contributing to the interpretation of utterances. They also propose plans for output, in an abstract form that is rendered to surface form by external modules (like NLG). Information is processed hierarchically in an Issue tree until exhausted, from most specific to more general [6]. The DM proper encapsulates these propagation policies and governs modifications to both the Issue forest and the contents of the VariableContext.

Virtual agents are driven by the ASAPRealizer framework [4], which accepts the BML requests containing nonverbal behaviors and utterances generated by the NLG. The scene is rendered using the Ogre 3D framework, and speech synthesis is handled by the CereVoice framework.

3 Demo system

For the demo system, a subset of the projected initial domain for the assistive system has been implemented: going through a user's weekly calendar and allowing the entering and modification of events, combined with access to video telephony that is informed by the dialogue situation (such as calling participants of a tentative event). The system is personified by the virtual agent "Billie". Different models for information grounding can be selected, and the system strives to autonomously moderate the floor to its advantage. A live view into the attributed dialogue structure and information processing mechanisms is possible. Interactions with the demo system are to be conducted in English (note, though, that the target language of the project proper is German), using a table microphone and an eye tracker.

4 Acknowledgements

This research was partially supported by the German Federal Ministry of Education and Research (BMBF) in the project 'KOMPASS' (FKZ 16SV7271K) and by the Deutsche Forschungsgemeinschaft (DFG) in the Cluster of Excellence 'Cognitive Interaction Technology' (CITEC).

References

1. GUIDE Consortium: User Interaction & Application Requirements, Deliverable D2.1 (2011)
2. Meis, M.: Nutzerzentrierte Entwicklung eines Erinnerungsassistenten, Abschlussymposium Niedersächsischer Forschungsverbund Gestaltung altersgerechter Lebenswelten (2013)
3. David Schlangen, D., Baumann, T., Buschmeier, H., Buß, O., Kopp, S., Skantze, G. & Yaghoubzadeh, R.: Middleware for Incremental Processing in Conversational Agents, In: Proc. 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pp. 51–54, ACL (2010)
4. van Welbergen, H., Reidsma, D. & Kopp, S.: An Incremental Multimodal Realizer for Behavior Co-Articulation and Coordination, In: Proc. 12th International Conference on Intelligent Virtual Agents, LNCS (LNAI), 7502, pp. 175–188 (2012)
5. Yaghoubzadeh, R., Kramer, M., Pitsch, K. & Kopp, S.: Virtual Agents as Daily Assistants for Elderly or Cognitively Impaired People, In: Proc. 13th International Conference on Intelligent Virtual Agents, LNCS (LNAI) 8108, pp. 79–91 (2013)
6. Yaghoubzadeh, R., Pitsch, K. & Kopp, S.: Adaptive Grounding and Dialogue Management for Autonomous Conversational Assistants for Elderly Users, In: Proc. 15th International Conference on Intelligent Virtual Agents, LNCS (LNAI) 9238, pp. 28–38 (2015)
7. Yaghoubzadeh, R. & Kopp, S.: Towards graceful turn management in human-agent interaction for people with cognitive impairments, In: Proc. 7th Workshop on Speech and Language Processing for Assistive Technologies (in press)