Enabling robust and fluid spoken dialogue with cognitively impaired users

Ramin Yaghoubzadeh
CITEC, Bielefeld University
P. O. Box 10 01 31
33501 Bielefeld, Germany
ryaghoubzadeh@uni-bielefeld.de

Stefan Kopp
CITEC, Bielefeld University
P. O. Box 10 01 31
33501 Bielefeld, Germany
skopp@techfak.uni-bielefeld.de

Abstract

We present the `flexdiam` dialogue management architecture, which was developed in a series of projects dedicated to tailoring spoken interaction to the needs of users with cognitive impairments in an everyday assistive domain, using a multimodal front-end. This hybrid DM architecture affords incremental processing of uncertain input, a flexible, mixed-initiative information grounding process that can be adapted to users' cognitive capacities and interactive idiosyncrasies, and generic mechanisms that foster transitions in the joint discourse state that are understandable and controllable by those users, in order to effect a robust interaction for users with varying capacities.

1 Introduction

In recent years, politics and society have placed emphasis on ways to enable an autonomous and self-determined life for those who were previously automatic recipients of stationary care. This is most overtly the case for older adults whose capacities start to degrade but are still sufficient to organize their life given some help; but also for people with general cognitive impairments, who until twenty or thirty years ago were often regarded as unable of being afforded a lifestyle with a workplace and an independent living space, tailored to their individual strenghths and capacities.

In order to support these individuals in those areas where deficits might manifest, use of mobile personal help for organization and management is regularly employed. There has been heightened interest in offsetting some of the burden of common routine tasks to technological implementations. Most unexperienced users report spoken interaction to be their preferred modality, which they are also used to in those domains, due to interactions with personnel. Human-computer interactions have to be designed in a way that suits their experience and preferences, their prior and their attainable special knowledge, and their other capacities. There is regular comorbidity with impaired articulation which can complicate interactions (Young and Mihailidis, 2010) – although for mild cases automated speech recognition software has caught up in recent years to ensure suitable operation. Additionally, their capacity of adhering to a recommended interaction style, or their general capacity for learning, might be reduced. Information density in interaction is another issue: tightly-packed information might be overwhelming and lead to incomplete appreciation and inadequate reflection of the contents (Yaghoubzadeh et al., 2013). At the same time, and especially if comorbidity with impulse control disorders is present (Swaffer and Hollin, 2000), the frustration tolerance in adverse situations might be lowered, although their stakes – of obtaining assistance – can provide extrinsic motivation.

Altogether, we have to address several areas which assistive systems for these user groups have to be aware of and cope with: less reliable input, idiosyncratic interaction style such as verbosity, limitations to cognitive processing and adaptation on the user side, and less reliable adherence to implicit system expectations and overt instruction.

In this paper, we will first look at systems that aim to provide assistance or company for these people in their everyday life, and address existing approaches to dialogue management with respect to the above properties. Then, we will describe our approach to dialogue management that is tailored to meet these requirements. Finally, we present initial results from an evaluation with older adults and people with cognitive impairments.

Proceedings of the SIGDIAL 2017 Conference, pages 273–283,
Saarbrücken, Germany, 15-17 August 2017. ©2017 Association for Computational Linguistics

2 Related work

2.1 Assistive and accompanying systems

Technical assistance can be provided to the aforementioned user groups in several domains, striving to improve their quality of life: in enabling their control of their environment, in enabling them to communicate more readily, in aiding self-organization, in supporting and tracking therapeutic efforts, in ameliorating the effects of ennui and social isolation, among others. In the following overview, we omit those technologies that rely on physical support or that use non-interactive spoken control (keyword commands for smart homes etc.). However, there has been relevant work in domains that transcend these limited scenarios, and evaluations relating to all mentioned aspects.

If speech is chosen as a modality for an assistive system, the role of personification, involuntary attribution, and the social effect of help rendered must not be underestimated, Meis (2013) commented that older subjects, having interacted with a spoken-dialogue scheduling helper for an extended time, first and foremost wished for it to be given a name and to react contingently to social affordances such as expressions of gratitude.

Bickmore et al. (2013) analyzed month-long phases of interactions of older adults with a personified exercise coaching system – it used spoken language, but user input was selected from sets of touchscreen buttons. Sidner et al. (2013) addressed the social support aspect, attempting to identify preferred domains of conversation or joint activity based on the same system design.

An autonomous spoken dialogue prototype with a humanoid assistive agent for older adults and people with cognitive impairments has been analyzed by Yaghoubzadeh et al. (2015); they found that users with terse interaction styles from both groups were able to successfully ground information with their system, their earlier studies showing that explicit confirmation patterns and a preference for packing all pieces of information in separate utterances helped the latter user group in particular in detecting and repairing system errors.

More recently, Wargnier et al. (2016) have evaluated a low-level attention monitoring and management module with a small sample of older adults with mild cognitive impairment; their system performed as well as with the control group.

The two latter teams also mentioned that interactions were unsuccessful for only their respective participant with the most overtly noticeable impairments. However, spoken interaction with users with cognitive impairments seems, in general, to be feasible and accepted by the user group.

2.2 Relation to other DM approaches

As a preliminary, we want to establish what we consider the bounds of the safe action space for a robust, noise-resistant communication system – particularly, the case of potential categorial confusion of positive and negative evidence in key issues of ensuring mutual understanding. Clark and Schaefer (1989) stated that positive evidence for understanding generally arrives in five categories of increasing strength: 'continued attention' (i.e. without any repair initiation), 'initiation of the relevant next contribution', explicit 'acknowledgment' (possibly via back channels or multimodal signals), as well as 'demonstration' and 'display', referring to (partial) paraphrase or cooperative completion and verbatim repetition, respectively. However, for the assessment of the strength of evidence, a system has to take into account the risk of confusion with conflicting categories. In particular, we posit that in the case of verbatim display – nominally providing the strongest evidence – there is significant structural overlap with possible 'bare revisions' (i.e. unmarked other-repairs containing only the corrected information – which are abundant and should be handled by an SDS, cf. Larsson (2015)) or even incredulous return questions. These ambiguities only disappear if the confidence values (or suitable correlates) of the ASR process are on the level of near-certainty – and can be trusted – and, in the case of unmarked questions, prosody is also considered. In terms of negative evidence of successful grounding, spontaneous repairs and repeated requests are examples of explicit evidence, while multimodal modulations that indicate confusion or surprise (furrowed brows, 'double-checking' gaze patterns) are more subtle signals.

For a comparison of the present work to existing approaches and implementations of dialogue systems, we will consider the following taxonomical properties: globally accessible versus locally encapsulated state; rule-based versus statistically grounded decision making; human-authored vs. learned policies; approaches with or without strictly disjoint modeling of task and discourse models, with or without incremental processing,

and with or without modeling of probabilistic aspects or uncertainty in either their input, inner state, and/or output. Centrally, implementations differ in their presentation and modeling of revisions and repairs from the system or the user side.

With the information-state-update (ISU) approach, Traum and Larsson (2003) proposed a generic mechanism for the concurrent matching of a set of update rules to the current state of a globally accessible information blackboard – in contrast to plan-based or finite state machine-based approaches. `flexdiam` employs a hybrid approach, independent entry points can operate solely on the global state or in relation to their ancestors and children in the hierarchy. The designer and the domain define an emphasis on reliance on the global context for one globally active set of rules (flexible, but harder to scrutinize) or classical graph-based traversals (predictable, but rather rigid) – or a hybrid of both. The global context does not contain an additional logic-based representation of internal – or attributed – plans.

Larsson (2002) modeled the grounding process on earlier work by Ginzburg, implementing the 'questions under discussion' in the form of 'issues', with an explicit propositional model of the common ground between the parties and the system's short-term agenda and longer-term plan, and explicit signals on three levels (contact, semantic, and pragmatic understanding).

Skantze (2007) considered the effects of uncertainty on the grounding process, particularly in 'real-world' ASR scenarios. The approach included disjoint modules of (abstract) NLU and (contextualized) discourse model that performed contextual integration, and generic clarification request and display actions based on word and concept level estimations of confidence, driven by a rule-based decision policy. `flexdiam` features a similar dichotomy of NLU and discourse models for incremental processing, opting for hierarchical situative interpretation – enabling partial interpretation in the most specific context and additional interpretation (and forward-looking expansion) in the more general ones. Since we found our ASR to yield word confidence scores with domain dependent baselines, we decided to start with a pessimistic strategy to minimize the false-positive rate for assuming "certain" interpretation – thus, ambiguous slots from the lattice of hypotheses were weighted equally, producing the primary source of inherent low-level uncertainty. The basic grounding criterion for our first evaluations was likewise a rule-based one, operating on concept entropy values.

Bohus and Rudnicky (2009), with RavenClaw, proposed an logic-based approach that separated the task domain model, provided in a domain specific language to yield a hierarchical description of tasks and dependent subtasks, and a generic dialogue engine, configured with the task model and capable of employing two strategies for resolving detected ambiguity ('misunderstandings') and several more for non-understanding, including declaration of non-understanding, requests, reprompts, and help messages. `flexdiam` does also provide hierarchical task modeling, repairs and grounding strategy selection are however encapsulated in a library of reusable, specialized patterns that are configured[1] for specific situations.

Baumann and Schlangen (2012), with InproTK, provide a fully incremental dialogue management toolkit that builds a fine-grained graphical representation of sequences of incremental information in the system, including revoked and revised paths – that can thus also encode a full implicit discourse history. Notably, input and output sides can operate in an incremental fashion. In `flexdiam`, input and processing modules operate incrementally, but there is currently no provision for incremental adaptation in the NLG (although other output modalities do operate in an incremental fashion).

Skantze and Moubayed (2012), with IrisTK, presented another hybrid approach that combined a generic 'attention manager' with a hierarchical task and dialogue model (IrisFlow) based on a generalized, extended version of Harel statecharts, which can be conveniently authored. Their extension does take into account, and attempts to integrate, the asynchronous character of the relation of intention and actual spoken interaction. It has been employed in the autonomous robotic head FurHat. In `flexdiam`, authoring cannot be undertaken using an abstract modeling description that automatically transfers to code, as in IrisTK or RavenClaw. However, since it is written in Python, there is arguably little difference between the two anyway; graphical authoring might be attractive, though, especially since the existing live and off-line visualizations could serve as a basis.

[1] The system is tailored to incremental, multimodal referential behavior, hence the dynamics of promoting and retracting references is quite dependent on the domain.

Lison and Kennington (2015), with OpenDial, proposed another hybrid approach, combining logical and statistical methods. Probabilistic logical dialogue rules are parametrized with respect to probabilities of their outcomes and their estimated utility, and selected under consideration of uncertainty in their respective preconditions. The strength of the approach is the particular suitability for combining (or gradually replacing/adapting) hand-crafted parameters with learned ones. `flexdiam` presently foregoes any general representation of post-condition success estimations (although local planners are free to factor this in their plans opaquely). There is however a clearly defined way for monitoring the state of asynchronous output – and the user's closing of contingency pairs (or failure to) can be handled in the hierarchical situation model. Uncertainty in input and derived data is also represented.

As did most of the previous work, we also assert that our present system is a relatively loose framework that enables more than one philosophy to thrive within, though maybe not simultaneously.

3 Architecture and processing

`flexdiam` is an interaction framework that aims to unify 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 system is built on top of the `IPAACA` middleware, a distributed, platform-independent implementation Schlangen et al. (2010) of the 'general, abstract model for incremental dialogue processing' proposed by Schlangen and Skantze (2011). This provides the back-end for the connection of the core DM components to input (including ASR, tagger and parser, eye tracker, keyboard/mouse/touch etc.) and output modules (NLG, synthesis, graphical components / GUI changes, control of animated characters etc.).

An overview of the DM architecture is provided in Fig. 1. Temporal information, and the representation of `Events` is maintained in a functionally tiered structure called `TimeBoard`. 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 the contribution manager, which schedules queued communicative intentions when the floor situation allows.

Propositional information is, in the general case, resident in the global `VariableContext` (subsequently 'Context'), containing a rewindable representation of certain and uncertain (distribution) variables with generic metrics – like entropy – that serve as the basis for local decision heuristics. Other types of variables include watchdogs that update their state based on other values; one such use case is the recalculation of possible referents in a certain domain whenever information restricts or extends its determining variables.

In `flexdiam`, there is generally a single joint task and discourse model for both interactants (i.e. no explicit full Theory of Mind-like simulation of the other party); its presence in the actual common ground is on the other hand promoted by the update heuristics, below. The basic structure of the joint task and discourse model is a forest of independent but hierarchically interdependent agents termed `Issues`[2], as well as generic update rules to transform this forest after DM invocations. An Issue $I := I(Pattern)$ with $Pattern := (Cls, name, config)$ is defined by a functional class Cls that implements its input handling and planning dynamics, an abstract $name$ (used e.g. for mapping to specific verbalizations in the NLG module), and a $configuration$ that defines its initial internal state. If $Pattern$ is identical for any Issues I_1, I_2, they are defined to *match functionally*. When an Issue is *instantiated*, it is at the same time made a *child* of the Issue that effected its creation. Issues can have zero or one parent (root / non-root) and any number of children.

Any path from a leaf Issue to the root of its tree corresponds to a specific (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. Any Issue can be in one of five canonical states that correspond to its status with respect to the common ground and its continued relevance: NEW (it is on the system's agenda, but has never been raised by successful communication by the system or relevant contribution of the user), ENTERED (an initial communication attempt has been completed to introduce it to the common ground; it is presently considered a

[2] Terminology adapted from Ginzburg, via Larsson (albeit in a slightly less rigorous sense) – since the basal Issues do in fact correspond to grounding and acceptance questions.

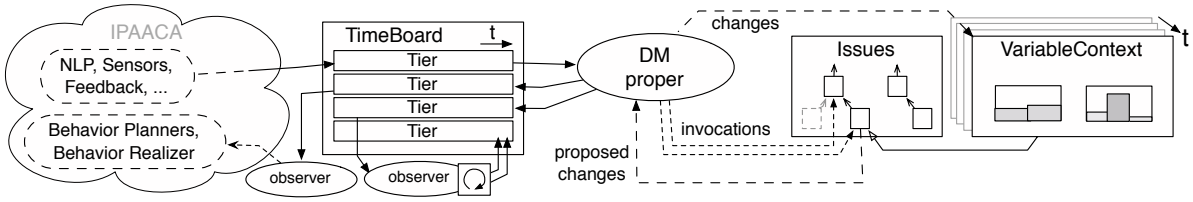

Figure 1: Overview of the architecture

valid target for DM invocations), FULFILLED or FAILED (terminal states decided locally by the Issue) or OBSOLETE (a terminal state which means that a replanning process in an ancestor has invalidated this instance explicitly, or implicitly through an intermediate ancestor).

3.1 Processing proper and plans

An invocation of the dialogue manager proper, triggered by the event structure on the TimeBoard, relays input records to all valid *entry points*. These refer to active topics (non-terminated leaves, see above), stored together with access time information to produce an implicit priority queue, similar to the 'partially ordered set' in Ginzburg (2012); however, rank is defined solely at invocation time since locally estimated utility is factored in.

Invocations that trigger processing in Issues come in two flavors: input handling and structure update handling. Under the umbrella of input handling, any abstract category of information can trigger a DM invocation (and 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 it). Two basic input categories for a general `flexdiam`-based SDS are `prompt_request` and `nluparse`, referring to calls for action at suitable points for contributions by the system, and partial incremental parses of user ASR, respectively. Under the umbrella of structure update handling, parent Issues are informed, and given the opportunity to contribute (or re-plan), when a plan is generated that involves either a child transitioning to a terminal state, or a child marking that it has made progress that might merit re-evaluation of the parent. Child Issues are informed, and given the opportunity for a final contribution, when they are invalidated (marked OBSOLETE) by an ancestor; the final contributions are usually limited to cleanup – especially retractions of situated referential behaviors.

For any invocation on an entry point E_x, starting at Issue I_x at time t, an individual clone overlay ('clover') C_{I_y} is generated for any contributing Issue I_y (Copy-on-Write access) (cf. Fig. 2); the global Context $\mathscr{C}(t)$ is also accessed via a CoW overlay $\mathscr{C} + \Delta\mathscr{C}_{I_x}$. This enables the generation of competing plans involving a common subset of Issues. Any modifications to the internal state of Issues is made to the clovers instead and later merged in after the DM commits to a plan. Prior to any overlay production and processing ($handle\_$), Issues may make a shallow assessment of the capability of handling the input in the given situation ($can\_handle\_$), for reasons of economy. For any invocation with input i that an Issue I_y can handle, it produces a partial plan $\mathscr{P}_{I_y} = \{C_{I_y}, O_{I_y}\}$ – with C the new 'clover' of the Issue, and O its *output record*. The latter may contain the following: a local *utility estimate*; a flag that signals *significant progress* to the ancestors; a preference for *propagation* of the input; a list of proposed *new child issues*; a list of *obsolete children* that are to be invalidated if the plan is selected; and, centrally, the current *communicative intentions*. The partial plans $\mathscr{P}_{I_{...}}$ contribute to the full plan for this input and entry point, $\mathscr{P}(E_x, i) = \{(C_{I_z}, O_{I_z})$ for all contributing $I_z\}$. Additionally, Issues may annotate (or even transform) the input record (primarily marking input keys as used and 'accounted for', thus also marking interpretation coverage). The modified record i' is reused for all contributions by other issues to the same plan. The Context overlay $\mathscr{C} + \Delta\mathscr{C}_{I_x}$ is also reused, progressively accumulating changes from all contributions to the same plan.

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, if flagged in the output record. A DM can enforce certain requirements beyond the marked propagation preferences in order to guarantee post-conditions (e.g. maximize opportunities that any prompt is generated).

Progressive propagation from the leaves

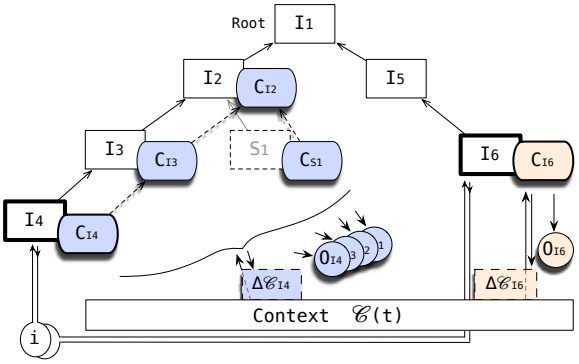

Figure 2: Invocation of the DM proper with input i leading to alternative plans starting at entry points I_4 and I_6, each yielding $i', \Delta\mathscr{C}, \{(C_{I_x}, O_{I_x})\}$. This example corresponds to two open topics (possible jump), the plan for I_4 includes deduced forward-looking agenda C_{S1}, contributed by shadow S_1.

through the ancestors thus allows for situated partial interpretation and processing; this is most specific and situation-dependent in the leaves, and most generic and general in the roots of the forest.

3.2 User-initiated agenda changes

Any Issue I_x can elect to define a set of anticipated Issue patterns that are not immediately on its local agenda (i.e. not actual children), but well-defined with respect to their arising at any time during the active life of I_x. This might include possible future child Issues, but also, crucially, anticipations about user behavior that stands outside the typical traversal though the local planning of I_x. In the former case, this is equivalent to defining precisely the opportunities of mixed-initiative approaches to subplan initiation. In the latter case, it simply affords offloading resources (and from the developers' perspective, code duplication and implementation time) to reusable patterns that are jointly servicing any number of issues with overlapping expectations. The anticipated patterns, implemented internally as specially-flagged Issues, are called *shadows*. Subtrees spanned by shadows must be cycle-free, and functionally matching shadows present in children and parents alike will always match only at the most specific location (the child). All shadows, leaf and non-leaf, are also defined to be valid entry points.

If a user contribution does not fit well into any active Issue, save for an existing shadow, 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 (when the shadow matches at another side branch of the current tree, opens a whole deep side branch, or belongs to an entirely different tree in the forest). From the point of time of plan selection using the DM policy, all employed shadows are copied into real instances and transplanted into their parents as proper children. The new branch is marked ENTERED and moved to the top of the entry point priority queue.

3.3 Decision making

The set of (non-empty) plans $\{\mathscr{P}(E_x, i)\}$ for all entry points E_x, with $\mathscr{P}(E_x, i) = \{(C_{I_z}, O_{I_z})$ for all contributing $I_z\}$, are ranked by a central policy using weighted criteria:

- local utility estimations placed in O_{I_z} by I_z,
- the coverage of the annotated input i', proportionally to the original,
- the recency of the topic, i.e. the latest invocation timestamp on the path from E_x to its root (freshly instantiated Issues are not considered),
- special rules (e.g. acting on estimated topic jumps can be deferred during an incremental interpretation phase).

The plan with the highest rank is selected for execution, which entails:

- merging the context overlay $\Delta\mathscr{C}_{I_x}$ into $\mathscr{C}$, producing the new global context (recalling that prior states remain accessible by obtaining a rewound view),
- merging the whole internal state of all clovers C_{I_z} into their respective Issue I_z – this also updates its canonical / grounding state,
- scheduling all *communicative intentions* from all O_{I_z} for the contribution manager to pick up, instantiate and post for asynchrononous micro-planning and execution,
- updating the winning entry point with the most recent invocation time, and
- instantiating any newly proposed children, and adding new entry points for them.

4 Summary of approach

In terms of the basic approach, and in relation to existing work, discourse modeling in `flexdiam` most closely resembles Ginzburg's approach and its incarnations, in a formally less rigorous fashion. Some features of the info-state approach are present in the system (and it can in principle be employed as such), but the structural confinement afforded by the forest of hierarchical Issue agents helps to alleviate problems of inscrutability when the domain size increases, while still remaining very flexible. The present system is most suited to quick, interactive approaches to spoken interaction (and notably not designed for rigorous logical representation or explicit simulation of the interlocutor's mind), and to modeling real-world applications with limited domains. Manual extension is quite straightforward and seems to scale if 'best practices' are honored[3]. 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 (small) 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 will be used in upcoming long-term experiments to seed user models that best suit the estimated capabilities and preferences of participants. 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) is currently in development. The system is modular; the central decision policy is exchangeable and could in the future be parametrized using machine learning.

5 Initial evaluation

We have recently performed an initial evaluation of the described architecture in a setup for diverse user groups. For this experiment, we recruited 44 participants: 19 older adults (SEN), aged about 75+, with age-typical perception and cognition; 15 cognitively impaired adults (CIM) of working age; and 10 university controls (CTL).

Figure 3: Scene from the first evaluation study with the present system; subject anonymized, and scene enhanced for clarity.

Participants were asked to enter at least five items into a fictional weekly schedule at their leisure, in spoken interaction with a virtual assistant agent who also offered external activity suggestions. The agent was presented alongside a graphical calendar; the DM was able to generate dynamic references in the calendar and referential behavior for the agent (Fig. 3).

We selected the activity / scheduling domain because it was on the one hand the support domain most requested by our corporate partner, *von Bodelschwinghsche Stiftungen Bethel*, a large health care provider, but also by merit of its interesting properties: it can be reasonably well constrained in certain dimensions (days, times, intervals), while being potentially boundless in another (the activity being discussed) - though possibly constrained implicitly by priming and suggestions. This provides a relatively safe starting point for shallow, heuristic understanding of the only unconstrained dimension, because attribution to the other domains is fairly exclusive. (On the down side, out-of-domain discrimination would then amount to deep pragmatic understanding, so prior instruction about the restrictedness of the system capacities were necessary). A full dictation language model was used for ASR (provided via Dragon Client SDK 12.5) [4] to realize the free-form entering of the appointment. NLU performed heuristic extraction of best guesses for this slot from ASR hypotheses. Specifically, the parser identified sentences that might contain an appointment declaration, both in elliptic form (such as "<day> <time> <comment>") and various explicit

[3] Proper provisions for authoring are on the wish list for a future open release of the framework.

[4] Our health care partner required that a client-only, offline solution be employed in the project to guarantee privacy.

forms (such as "I was planning to <comment> on Monday"). The rule-based heuristics attempted to reduce the comment to a coherent sequence of V-N or N-N, optionally with declared participants ("with <proper-name>").

Aside from the scaffolding of social interaction and calendar entry commitment, we designed the grounding problem for the schedule items in three Issues: `VariableSetGrounding`, for accepting in free form, and integrating in a frame-like manner, the variables of day of the week (*dow*), the *start* and *end* times, and the activity (*what*) alongside many types of revisions, marked and unmarked; `VariableSetSequentialRephrase`, representing a situation where the system rephrased the previously uttered understood partial information; and `VariableValueConfirmation`, for explicit need for ratification and disambiguation when information was too uncertain to proceed silently. For the agent-initiated suggestions, the same approach was used, but pre-seeded with one variable (the agent's suggestion), and with the additional possibility of handling outright rejection of the suggestion. A final ratification with full multimodal presentation was also required before any activity was actually committed to the schedule.

The autonomous dialogue system was overseen by an experimenter, who had three options to aid the system in strategy selection: initiate the raising of an auto-generated partial suggestion ("Would you like to do something on *Saturday*?"); proceed to two fully-formed possible activities if the user had stated, or was assumed, to be done with their entries; or initiate the final valediction sequence.

All subjects managed to enter at least the required number of appointments into the calendar. The number of negotiated entries ranged between 5 and 18; 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). The older adults spent 15% longer on average on a topic compared to controls, while the group with impairments spent 23% longer; some participants from the CIM group made long hesitations in isolated instances (up to tens of seconds). The number of required utterances was initially high especially for the older adults, but started to converge; most subjects from the CIM group relied slightly more on reacting to dynamically generated prompts, their performance compared to CTL indicates that the afforded structure was suitable for them (Fig. 4). As expected by us, most time per entry was spent on correcting the topic (*what*) of an activity, due to the heuristic extraction of possible topics from a multitude of alternative ASR hypotheses, which caused the majority of challenging situations. For the future, we aim to add deeper NLU capabilities to the system to better constrain the set of relevant candidates – currently, we are exploring the use of word embeddings to this effect.

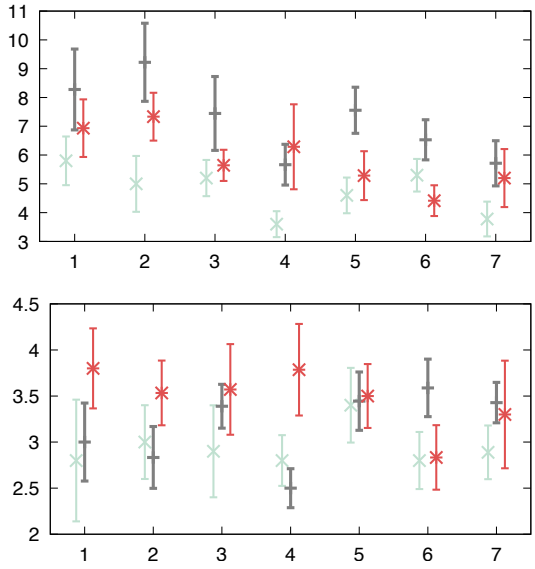

Figure 4: Top: user utterances per topic (for the first seven entered items, due to sample size); bottom: number of system variable prompts. User groups, ordered: CTL, pale; SEN, dark; CIM, red.

The experiment was conducted to gain qualitative insight into the repair, revision and metacommunicative patterns exhibited by the user groups; as such, there was no clearly delineated 'right' and 'wrong' with respect to final entries (hence there was no baseline reference to match). Detailed conversational analysis has only recently started (see appendix for two example situations with a view of DM internals)[5]; a statistical description of the language used - and of word error rates - can only be sensibly made based on a comprehensive transcription of the corpus, which is still pending at the time of writing. For upcoming experiments, we are currently scaling up the possible activities to include revisions and removal of older entries,

[5]Additional material will be made available here: https://purl.org/net/ramin/sigdial2017/

queries about specific topics or time ranges, and installing and managing reminders.

6 Discussion and conclusion

We have presented the principal approach and current state of our dialogue management framework `flexdiam`, which is being used to evaluate spoken interaction with people with cognitive impairments, informed by prior work in this domain. It is designed to handle uncertainty, interruptions, and many kinds of revisions in a robust manner in order to provide a stable interaction in task-oriented domains. The approach makes for flexible interaction dynamics that are also straightforward to analyze and scrutinize in detail by humans. With respect to the requirements for the specific user groups, confusion due to e.g. problems in articulation is resolved in place using generic recipes, information density can be configured for specific users, and the system can cope both with increased and reduced pace. Regarding idiosyncrasies in floor behavior, we observed long hesitations in specific users, which from the point of view of the system primarily entails non-standard assumptions in assessing engagement and disengagement; in previous work (Yaghoubzadeh and Kopp (2016)), we conversely explored multimodal preemptive floor management to reduce user verbosity in a socially acceptable manner; this module has been integrated into the architecture but not employed in the present study.

We regard our architectural requirements to be fulfilled and will integrate the results from the emerging qualitative analysis to refine the recipes in the system.

We strove to highlight the mechanics of `flexdiam`, and its novel combination of features for the target user groups, in comparison to existing approaches, and we have performed an initial evaluation with the target user groups in which subjects were generally able to solve the set task, and the system was able to reach successful grounding of the desired contents in most cases. Implementation of the domain and communicative behavior was straightforward, and has already been scaled up to include competing alternative actions. We would also like to employ learning approaches to seed and adapt utility estimations and policy weights in the system.

Acknowledgments

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

Timo Baumann and David Schlangen. 2012. The InproTK 2012 release. In *NAACL-HLT Workshop on Future Directions and Needs in the Spoken Dialog Community: Tools and Data*. Association for Computational Linguistics, Stroudsburg, PA, USA, SDCTD '12, pages 29–32. http://dl.acm.org/citation.cfm?id=2390444.2390464.

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(10):1676–1683. http://dx.doi.org/10.1111/jgs.12449.

Dan Bohus and Alexander I. Rudnicky. 2009. The RavenClaw dialog management framework: Architecture and systems. *Computer Speech and Language* 23.

Herbert H. Clark and Edward F. Schaefer. 1989. Contributing to discourse. *Cognitive Science* 13(2):259–294.

Jonathan Ginzburg. 2012. *The Interactive Stance*. Oxford University Press, Oxford, UK.

Staffan Larsson. 2002. *Issue-Based Dialogue Management*. Ph.D. thesis, University of Gothenburg, Gothenburg, SE.

Staffan Larsson. 2015. The state of the art in dealing with user answers. In *Proceedings of SemDial 2015*. pages 190–191.

Pierre Lison and Casey Kennington. 2015. Developing spoken dialogue systems with the OpenDial toolkit. In *Proceedings of SemDial 2015*. pages 194–195.

Markus Meis. 2013. Nutzerzentrierte Entwicklung eines Erinnerungsassistenten. Abschlusssymposium Niedersächsischer Forschungsverbund Gestaltung altersgerechter Lebenswelten.

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 SIGDIAL 2010 Conference, The 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 24-15 September 2010, Tokyo, Japan*. pages 51–54. http://www.aclweb.org/anthology/W10-4308.

David Schlangen and Gabriel Skantze. 2011. A general, abstract model of incremental dialogue processing. *Dialogue and Discourse* 2(1):83–111.

Candace Sidner, Timothy Bickmore, Charles Rich, Barbara Barry, Lazlo Ring, Morteza Behrooz, and Mohammad Shayganfar. 2013. An always-on companion for isolated older adults. In *14th Annual SIGdial meeting on discourse and dialogue*.

Gabriel Skantze. 2007. *Error Handling in Spoken Dialogue Systems*. Ph.D. thesis, KTH.

Gabriel Skantze and Samer Al Moubayed. 2012. IrisTK: a statechart-based toolkit for multi-party face-to-face interaction. In *Proceedings of ICMI '12*.

Tracey Swaffer and Clive R. Hollin. 2000. Anger and impulse control. In Rob Newell and Kevin Gournay, editors, *Mental health nursing*, Churchill Livingstone, chapter 15, pages 265–289.

David R. Traum and Staffan Larsson. 2003. The information state approach to dialogue management. *Current and new directions in discourse and dialogue* pages 325–353.

P. Wargnier, G. Carletti, Y. Laurent-Corniquet, S. Benveniste, P. Jouvelot, and A. S. Rigaud. 2016. Field evaluation with cognitively-impaired older adults of attention management in the embodied conversational agent Louise. In *2016 IEEE International Conference on Serious Games and Applications for Health (SeGAH)*. pages 1–8.

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 (SLPAT 2016)*.

Ramin Yaghoubzadeh, Marcel Kramer, Karola Pitsch, and Stefan Kopp. 2013. Virtual agents as daily assistants for elderly or cognitively impaired people - studies on acceptance and interaction feasibility. In *Intelligent Virtual Agents - 13th International Conference, IVA 2013, Edinburgh, UK, August 29-31, 2013. Proceedings*. Edinburg, UK, pages 79–91.

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.

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(2):99–112.

A Example interactions

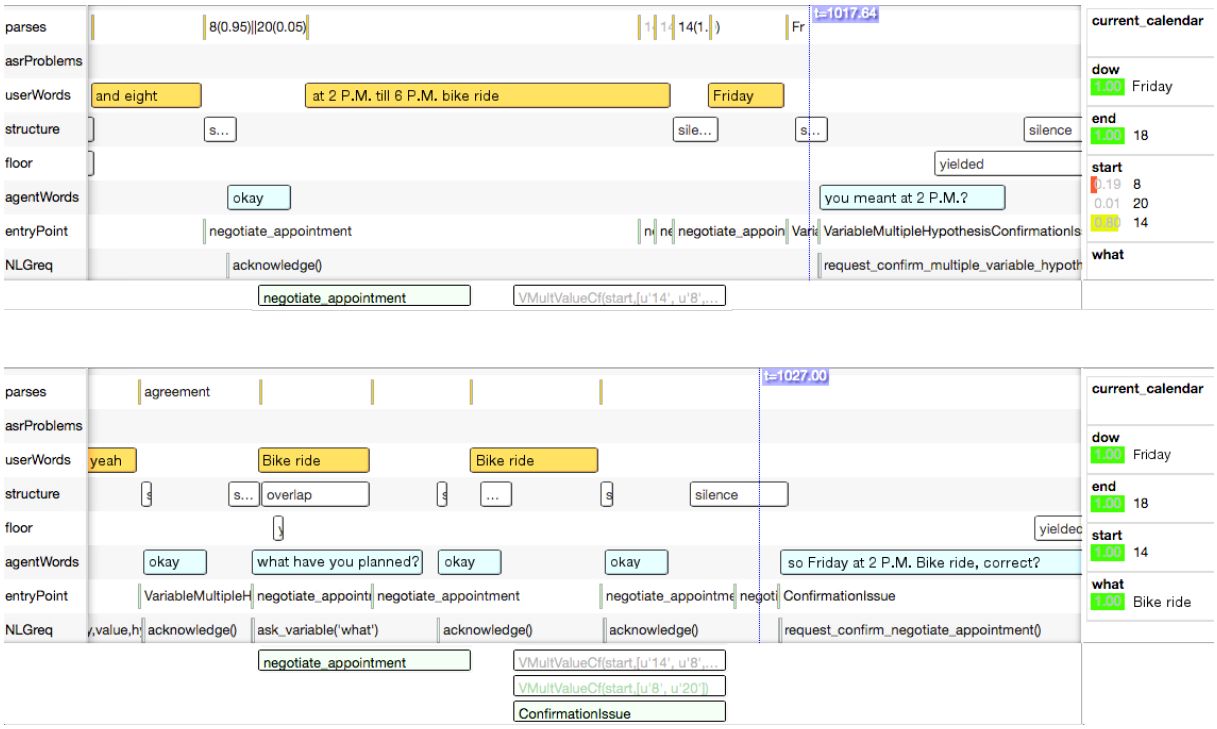

Figure 5: Example situation (via HTML transcript generated by `flexdiam`, and translated to English): top: user initiated new appointment, note that two possible start times were generated from the first fragment, and overridden by the second; bottom: final ratification phase after last information provided.

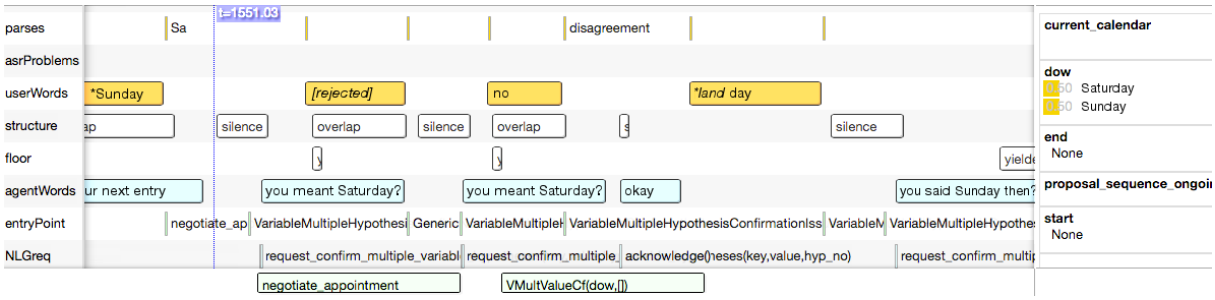

Figure 6: User with impaired articulation: cooperative repair. Prior to the blue cursor position (left), two equally valid hypotheses were generated for *dow* from the user's preceding utterance. The user provides negative evidence by rejection for the first grounding attempt, but their subsequent correction is not recognized – the system continues with the next hypothesis.