Towards an Information State Update Model Approach for Nonverbal Communication

Dirk Schnelle-Walka, Stefan Radomski, Stephan Radeck-Arneth, and Max Mühlhäuser

Technische Universität Darmstadt, Darmstadt, Germany

Abstract. The Information State Update (ISU) Model describes an approach to dialog management that was predominantly applied to single user scenarios using voice as the only modality. Extensions to multimodal interaction with multiple users are rarely considered and, if presented, hard to operationalize. In this paper we describe our approach of dialog modeling based on ISU in multiuser dialog settings employing multiple modalities, including nonverbal communication.

1 Introduction

Although the Information State Update Model (ISU) as introduced by Larsson and Traum [9] describes a general concept to implement dialog managers, most implementations like TrindiKIT [1] or Dipper [2] primarily focus on spoken dialog systems. The approach originated in Dynamic Information Theory (DIT) as introduced by Bunt [3]. While Bunt based his *dialog acts* as the "functional units used by the speaker to change the context" [3] on linguistic aspects, the concept is not restricted to spoken language. Extensions to multimodality have already been successfully employed, e.g. [6]. These systems, however, primarily target single user settings while ISU based systems dealing with more than a single user are still focused on spoken interfaces. Furthermore, they are more interested in understanding *who is speaking to whom* [5] [12] which is relevant to include the computer as a homologous conversational counterpart. This research question may also include nonverbal aspects like gaze and gestures in order to detect turn-taking cues in multiparty conversations [1].

In contrast to that we aim at interfaces where both input and output are multimodal and the computer takes a solely supportive role. The information is presented to the users by a combination of modalities that best suits the kind of information to deliver while taking into account the special needs of the users, in our case blind users in a brainstorming session together with sighted persons. They require a special sort of interface that allows them to follow the discussion and the visual artifacts that are employed. The interface should enable them to actively contribute to the discussion but the selection of available modalities is limited. In this paper we introduce a novel way to model information states

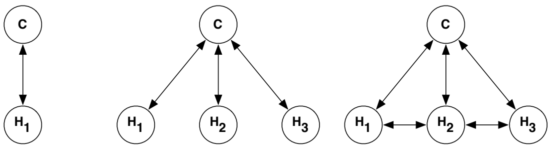
[1] `http://www.ling.gu.se/projekt/trindi/trindikit/`

K. Miesenberger et al. (Eds.): ICCHP 2014, Part I, LNCS 8547, pp. 226–230, 2014.

supporting multiple users with different communication needs with the help of grounding and reasoning capabilities. We describe how to derive the groups intention from the individual multimodal input and introduce concepts to present this information to the users. Moreover it can be used to communicate nonverbal communication elements.

2 Related Work

Most research focused on scenarios where a single user interacts with a computer (see Figure 1(a)) like in the original work from Traum and Larsson where they introduced the ISU concepts [9]. There are several conceptions for multiparty dialogs. The first obvious one was introduced in 2002 where Traum and Rickel [13] proposed a multiuser (see Figure 1(b)) and multimodal model for a game scenario, which also uses aspects of the information state theory from Traum and Larsson [9]. However Ginzburg and Fernandes [5] have relativized the work of Traum and Rickel since it does not model true multiparty dialogs (see Figure 1(c)) but rather multiuser dialogs. The multiuser model targets only parallel dialogs without any relations between the dialog flows (only one sender and one addressee). Therefore the communication between the participants and cross-related discussions about the same topic are not covered. Only with a multiparty dialog model the complex group discussion relations can be recognized.



(a) Single user (b) Multiple users (c) Multiparty

Fig. 1. Schematic view of communication modes

Ginzburg et al. [5] also introduced in 2005 three interaction principles for realizing a system as a dialog partner. In 2008 Kronlid [8] criticizes these principles as not flexible enough and tried to extend the ISU concept by including a full-fledged multiparty turn manager. With its base in the principles of Sacks et al. [11]. In general, a turn manager tries to coordinate dialogs with the goal of a fair time distribution among the participants. Kronlid had simplified Sack's turn manager by omitting some solve strategies, which cope with complex overlapping scenarios.

In 2010 Strauß et. al. [12] extended the ISU approach for multiparty dialogs with proactive functionality by using ontologies. This is relevant to answer current research questions from a linguistic perspective. However, in our scenario there is no need for a system with communication capabilities comparable to a human. This allows us to focus on nonverbal, i.e. multimodal, aspects of the interaction with multiple users rather than forcing us to fully understand parallel ongoing discussions.

3 Scenario

We aim at a scenario in a current project to improve the participation of blind people in a brain-storming process with sighted people. Here, a mind map is used on a touch table. Here we encounter, high dynamics in the communication, like speaker changes, deictic gestures as well as frequent changes of the artifacts.

4 Dialog Management

The W3C suggests to use SCXML [2] as a dialog manager within a multimodal system. However, SCXML seems to be suited *only* to implement finite state or frame-based/form-filling dialog management approaches since they lack grounding and reasoning. Fodor and Huerta [4] demand that dialog managers should feature: (i) a formal logic foundation, (ii) an interference engine, (iii) general purpose planners and (iv) knowledge representation and expressiveness. One solution to overcome these limitations is to introduce Prolog as a replacement of an embedded scripting language [10] similar to the approach by Kronlid and Lager [7]. Embedding it as a scripting language allows to use SCXML as an ISU dialog manger compliant to the W3C MMI Architecture [3] which has the potential for more natural and flexible dialog management.

5 Multiuser Multimodal Information States

We rely on a multimodal fusion component that emits the individual events representing the user's application specific interaction intent. The actual intents of the group are derived by the individual intents and made accessible in the group information state IS_{grp}. Ultimately, concepts like the social structure of the participants ought to play a role when inferring the group intent. Figure 2 introduces the global information state.

The information state IS_{grp} comprises inferred group-related attributes as well as the actual state of the mind map, along with a history of its changes. It unifies the individual information states IS_{ind} into a coherent view. Each IS_{ind} includes, besides personal beliefs, personal attributes which includes the user's presumed mental state of the mind map. The personal attributes contain information about output related settings like *blind user*, *available* and *preferred* modalities.

The mind map's actual nodes and their relations are loaded as Prolog facts in the IS_{grp}. The IS_{ind} contains a mental model of the global map as it is perceived by the individual users. This individual mental model might differ from the actual map especially in the case of blind users. If one participant changes a node in the mind map, his mental state (in IS_{ind}) and the global mind map (in IS_{grp}) is updated. The IS_{grp} will trigger updating all other participants via

[2] `http://www.w3.org/TR/2013/WD-scxml-20130801/`

[3] `http://www.w3.org/TR/mmi-arch/`

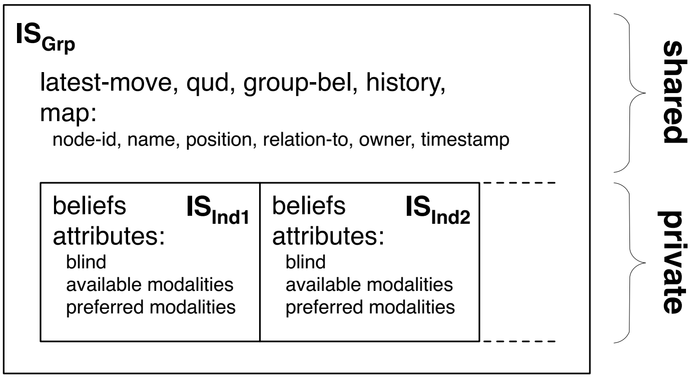


Fig. 2. Structure of adapted information state for multiuser scenarios

messages of an suitable type. Moreover, it can be used to reason when and how to deliver nonverbal communication elements.

For tracking the changes of the mind map and the discussion the IS_{grp} contains the attribute *history*. A timeout deletes expired events, because the scope or rather focus of discussion has changed.

This way we are able to provide an implementation for Strauß' and Minker's observation that not all participants share the common ground and the more participants the less common ground can be assumed [12].

6 Conclusion

This paper describes an extension of the ISU dialog management technique that integrates into the recently finalized W3C MMI recommendation. We exemplify the approach by employing it to a brainstorming session with blind and sighted participants.

The extension of SCXML via Prolog helps us to derive the groups intent and enables us to express ISU-based dialogs. In a next step, we will evaluate how this can be used to filter non-relevant events, especially for nonverbal communication, to avoid an information overload of the blind users.

Acknowledgments. This work has been partially supported by the DfG (German Research Foundation) and it has been produced out of the D-A-CH project. The D-A-CH project is a joined project between TUD, the ETH and JKU with the respective founding organizations DFG (German Research Foundation), SNF (Swiss National Science Foundation) and FWF (Austrian Science Found).

References

1. Bohus, D., Horvitz, E.: Facilitating multiparty dialog with gaze, gesture, and speech. In: International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction, p. 5. ACM (2010)
2. Bos, J., Klein, E., Lemon, O., Oka, T.: Dipper: Description and formalisation of an information-state update dialogue system architecture. In: 4th SIGdial Workshop on Discourse and Dialogue, pp. 115–124 (2003)
3. Bunt, H.: Dynamic interpretation and dialogue theory. The Structure of Multimodal Dialogue 2, 1–8 (1999)
4. Fodor, P., Huerta, J.M.: Planning and logic programming for dialog management. In: Spoken Language Technology Workshop, pp. 214–217. IEEE (2006)
5. Ginzburg, J., Fernández, R.: Action at a distance: the difference between dialogue and multilogue. Proceedings of DIALOR 5 (2005)
6. Honold, F., Schussel, F., Weber, M., Nothdurft, F., Bertrand, G., Minker, W.: Context models for adaptive dialogs and multimodal interaction. In: 2013 9th International Conference on Intelligent Environments (IE), pp. 57–64. IEEE (2013)
7. Kronlid, F., Lager, T.: Implementing the information-state update approach to dialogue management in a slightly extended scxml. In: Proceedings of the 11th International Workshop on the Semantics and Pragmatics of Dialogue (DECALOG), pp. 99–106 (2007)
8. Kronlid, F.: Steps towards Multi-Party Dialogue Management. Ph.D. thesis, Göteborgs Universitet (April 2008)
9. Larsson, S., Traum, D.: Information state and dialogue management in the TRINDI dialogue move engine toolkit. Natural Language Engineering 6(3&4), 323–340 (2000)
10. Radomski, S., Schnelle-Walka, D., Radeck-Arneth, S.: A Prolog Datamodel for State Chart XML. In: SIGdial Workshop on Discourse and Dialogue (2013) (preprint)
11. Sacks, H., Schegloff, E.A., Jefferson, G.: A simplest systematics for the organization of turn-taking for conversation. Language, 696–735 (1974)
12. Strauß, P.M., Minker, W.: Fundamentals. In: Proactive Spoken Dialogue Interaction in Multi-Party Environments, pp. 17–49. Springer (2010)
13. Traum, D., Rickel, J.: Embodied agents for multi-party dialogue in immersive virtual worlds. In: Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 2, pp. 766–773. ACM (2002)