Discourse-based Interaction Design for Multi-modal User Interfaces

Cristian Bogdan
School of Computer Science
and Communication
Royal Institute of Technology
10044 Stockholm, Sweden
cristi@csc.kth.se
and
Institute of Computer Technology
Vienna University of Technology
A-1040 Vienna, Austria
bogdan@ict.tuwien.ac.at

Hermann Kaindl
Institute of Computer Technology
Vienna University of Technology
A-1040 Vienna, Austria
kaindl@ict.tuwien.ac.at

Jürgen Falb
Institute of Computer Technology
Vienna University of Technology
A-1040 Vienna, Austria
falb@ict.tuwien.ac.at

Abstract
Current user interfaces do not sufficiently utilize multiple modalities. We developed a new approach to modeling discourse-based interaction design inspired by theories of human communication. From such an interaction design, we envisage to generate a multi-modal user interface. This paper presents our approach in the context of mixed-initiative interactions with a (semi-)autonomous robot.

CHI 2008 Workshop: User Interface Description Languages for Next Generation User Interfaces, April 5 – 10, 2008, Florence, Italy
ACM 1-xxxxxxxxxxxxxxxxxx.

Introduction
In previous work [2] we studied several theories of human communication from various fields to develop an approach for specifying discourse-based interaction design models. These design models are more understandable and possibly easier to build for humans with less technical background than user-interface models. Based on such an approach, we showed in [1] how graphical user interfaces can be rendered from high-level models.

Since the concepts of human communication are applicable to different modalities, we strive for rendering multi-modal interfaces that support mixed-initiative. As a benefit, modelers do not need to care about modality while specifying the interaction design. During rendering the system will suggest one or more modalities that a particular part of an interaction should be performed in. The modeler is still able to influence this decision making. This process should ease the development of multi-modal mixed-initiative interfaces for modelers, since they only have to specify one discourse-based interaction for all modalities.

Approach description
Our approach to multimodal communication consists of two distinct stages: the creation of the interaction

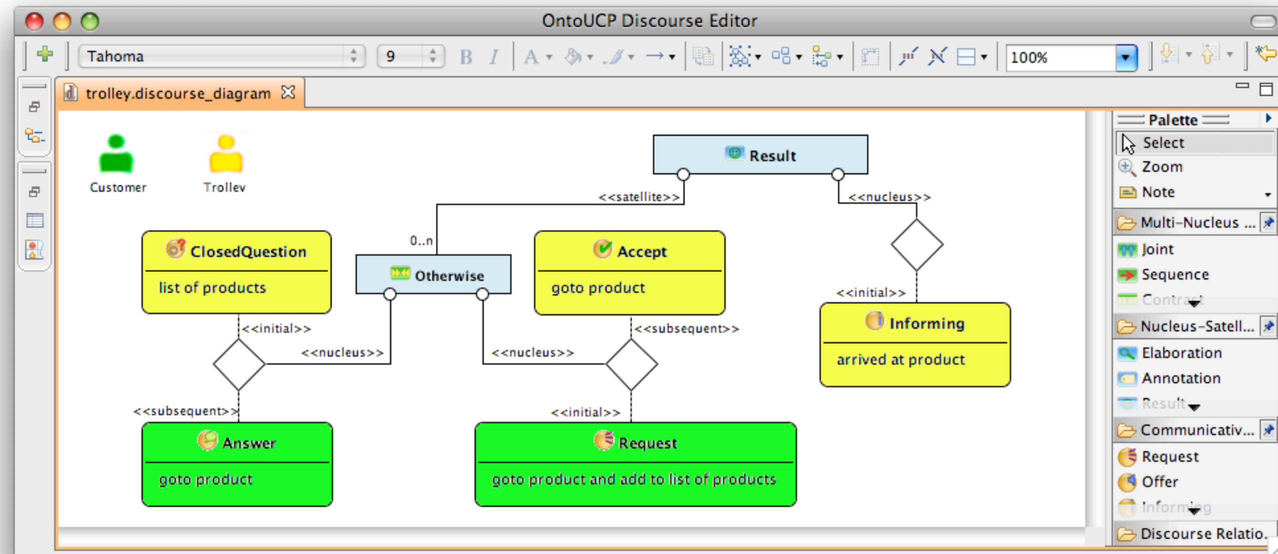

figure 1. The discourse model

model, which is modality-neutral, and the rendering where the modeller and possibly other people can assist the system in improving the interface by placement (spatial or temporal) of components within the constraints of the interaction model, choice of modality, etc. First we focus on the modality-neutral interaction design stage.

We describe our approach to model multimodal communication of humans with (semi-)autonomous robots through an example of a shopping trolley robot that helps the customer to process a predefined shopping list and to find items in a supermarket environment. Through the explanation we emphasize the concepts of human communication that our approach is inspired from. We have modelled (part of) an example interaction in figure 1 according to our discourse modelling approach.

A typical scenario covered by the discourse illustrated model goes as follows: First, either the trolley asks the customer to select a product from the shopping list to go to next, or the customer directly requests the trolley to go to yet another product in the supermarket. After specifying the next product, the robot shopping trolley starts moving to the indicated destination together with its assigned customer. When they get to the requested product, the trolley informs the customer about the arrival and removes the product from the shopping list. Our models describe classes of dialogues or scenarios, respectively, in a primarily declarative way. So, this model also includes e.g., that the customer can redirect the shopping trolley at any time, through requesting a new product as the current destination.

In the first step of the above scenario, the specification of a product can be accomplished in two different ways, either the trolley *asks* where to go, or the user *requests* to go somewhere. The two alternatives are an example of how our modelling framework can accommodate mixed-initiative interaction. We model these alternatives as two *adjacency pairs* (inspired from Conversation Analysis, details on the human communication concepts and the modelling language can be found in [1, 2]). These adjacency pairs are grouped together with a *rhetorical relation* (inspired from Rhetorical Structure Theory (RST)). All our models are, in fact, trees with adjacency pairs as leaves and rhetorical relations as the other nodes. In this case, since the two alternatives are of equal "weight", the adjacency pairs are grouped with an "Otherwise" RST relation, which is meant for such cases.

Indicating a destination from the part of the user is modelled at the bottom-centre, in the form of a "Request" *communicative act* inspired from Speech Act Theory. Communicative acts offer us an abstraction that is graphical-toolkit-neutral and also modality-neutral. Adjacent to the request, there is an "Accept" communicative act with which the machine confirms the new destination. The left side of the model offers the collected destinations for the user to choose from. This is modelled as a "Closed Question" communicative act, to which the user can respond, by way of the adjacent "Answer" communicative act, to choose from a defined list of possibilities. This list of possibilities is called *propositional content* in Speech Act Theory, and in our approach it is provided and refreshed by the application logic of the robot trolley. The "Closed Question" also helps the user to keep updated on what the shopping items are that were already added to the shopping list but not yet processed.

If there is no further user interaction and the robot reaches the destination currently specified, it informs the user about the status via the "Informing" communicative act at the right of our model, and the destination is removed from the shopping list by the robot's application logic. Since this is the main result of the interaction, the "Informing" is linked to the remainder of the dialogue model through a "Result" RST relation.

Multimodal Communication with a Robot according to this Model

Now let us focus on the rendering stage where the communication platform software will have to deal with modalities for expressing and receiving communicative acts. It is designed to do so based on heuristics, but the modeller and possibly other people may assist in choosing one or multiple modalities for improving the interface. Our robot trolley is designed to support three communication modalities and their combination: graphical interaction through a touch screen, speech input/output and movement.

Since the "Request goto product" communicative act is modelled to give the application logic data of a certain type (let's call it *destination*) and a speech input of type *destination* is available from the speech recognition, the render engine will recognize that the "Request" can be rendered in speech input. While assisting the rendering process, the modeller can decide that the request can also be done via the touch screen, in which case e.g., a widget providing alphabetical search for destinations can be rendered. Furthermore, our communication platform software can decide at runtime to fall back to the graphical input in a very noisy environment. In the case of accepting a "Goto Request", the trolley will utter the Accept communicative act in e.g., speech, since using the same modality for the adjacent communicative act improves clarity and answers the user's expectation.

The render engine will, in principle, render the "Closed Question" with the shopping list items only on the touch screen, as the speech medium is an expensive resource for a list. However, if desired at the rendering stage, speech could also be used in this case. This could be based e.g., on the level of ambient sound i.e., if the user appears to be alone in the shop, there is more "rendering space". And, maybe after periods of no communication with the user although sensed to be in the robot's proximity, the "Closed Question" can be uttered in speech as a suggestion. The user can interrupt the utterance via a predetermined speech utterance, to indicate that she chose the last destination uttered by the robot speech synthesis. In previous work, we have used the rendering *space* as a constraint for model rendering in GUI interfaces, but as exemplified here, a similar *temporal* constraint can be used for the speech modality.

Discussion and Conclusion

Our approach can be regarded as a very-high-level user interface definition language, or more precisely an interaction design language. We envisage that from the communicative acts, rhetorical relations and conversation analysis patterns it employs, a decent multi-modal interface can be generated. If a pre-rendering stage is added, our render engine will get even more guidelines for its runtime heuristics, resulting in higher-quality interfaces.

We have also shown that although our interaction models are modality-neutral, the modality can be inferred from the model in multiple ways: from data types involved, from the availability of widgets for the respective modality, from quasi-spatial constraints in the modality, and not the least from the "importance" of a certain communicative act as conveyed by the rhetorical structure of the discourse. If such inferences do not suffice, based on the interaction model, our system will be able to guide the modelers and designers to specify the modality of certain communicative acts.

Acknowledgements

We would like to thank Helge Hüttenrauch and Alexander Szep for their reviews and valuable comments.

References

[1] Bogdan, C., Falb, J., Kaindl H., Kavaldjian S., Popp R., Horacek H., Arnautovic E. and Szep A. Generating an Abstract User Interface from a Discourse Model Inspired by Human Communication. In *Proc. 41th Annual Hawaii Int'l Conference on System Sciences (HICSS-41)*, IEEE Computer Society Press (2008).

[2] Falb J., Kaindl H., Horacek H., Bogdan C., Popp R., and Arnautovic E. A discourse model for interaction design based on theories of human communication. In *Ext. Abstracts CHI 2006*, ACM Press (2006), 754-759.