

Decoupling of Modality Integration and Interaction Design for Multimodal Human-Robot Interfaces

Mathieu Vallée, Dominik Ertl

Vienna University of Technology
Institute of Computer Technology
Vienna, Austria
{vallee, ertl}@ict.tuwien.ac.at

1 Introduction

The development of a multimodal Human-Robot Interface (HRI) involving mixed-initiative and context-awareness is complex and laborious. The integration of individual modalities (e.g., gesture recognition or speech output) and the design of natural human-robot interaction are two different tasks that both require their own expertise.

In this paper, we consider three different roles that participate in the development of a multimodal User Interface (UI): the *interaction designer*, the *modality integrator*, and the *multimodal UI designer*. We present how the decoupling between these roles is facilitated by tools based on interaction modeling. We finally discuss how the decoupling is beneficial for introducing mixed-initiative and context-awareness in multimodal HRI.

2 Roles in Multimodal UI Design

Previous work shows that it is common to separate the task of UI generation from that of the application development [1]. We expect that the development of a multimodal UI is laborious enough to be split up in different tasks as well. So, we identify three different roles for a stronger decoupling of the needed tasks in UI generation. Figure 1 depicts the roles and their interactions:

- The *interaction designer* defines the modality-independent interaction between a human and a robot. It is responsible for defining the UI behaviour.
- The *modality integrator* provides a particular modality (or group of modalities), like speech input. This requires either adapting and configuring existing toolkits, or developing a new modality. Additionally, this role is responsible for defining the physical properties of the integrated modality.
- The *multimodal UI designer* considers the interplay between the modalities. This role considers the physical properties of the several integrated modalities in order to realize the desired interaction.

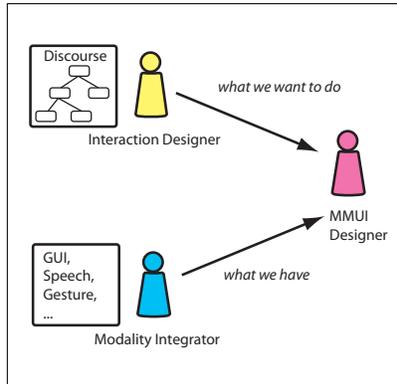


Fig. 1. Roles for Multimodal User Interface Design.

In current practice and in existing systems mostly a single person (or group of persons) realizes the task of the three roles with no clear separation of responsibilities. Often, the design of the UI is directed either towards demonstrating a particular modality or towards a single interaction description. In the first case, the risk is to limit interaction to what is supported by this modality, without considering usability. In the second case, the risk is to highly tailor the multimodal UI to the intended interaction, so limiting robustness and reusability. As a result, it is still difficult to understand the respective advantages of modalities (when to use a given modality) as well as the generic patterns governing the interplay between modalities (how to combine modalities for better usability). Furthermore, the role of the multimodal interface designer is particularly difficult, since it requires a good understanding of both the interaction design and the physical properties of particular modalities.

We propose to use supportive tools for the design of multimodal UIs. In particular, there is a strong need for: (i) languages to express the desired interactions, (ii) languages to express the physical properties of individual modalities, (iii) and tools for facilitating the mapping between desired interaction descriptions and physical properties of individual modalities.

A potential approach is the platform presented in [2] that uses a discourse model [3] and a communication platform for semi-automatic multimodal UI generation.

First, an *interaction designer* models the desired interaction scenarios as formal discourses between a human and a computer. The interaction is defined on a modality-independent high-level and supports modeling of mixed-initiative as well. At the same time, e.g., a *modality provider* couples speech input to the platform. For example, a freely available speech toolkit like Julius ¹ is manually integrated into the platform so that the platform can use the speech recognition functionality of the toolkit. Finally, the *multimodal UI designer* couples the dis-

¹ <http://julius.sourceforge.jp-visitedatthetimeofthiswriting>

course and the available modalities. The discourse model provides basic units of communication that are derived from the speech act theory. These are the intention, like an *Informing*, and the so-called propositional content. The propositional content is comparable to the meaning of the verb and the object in a simple English sentence. An example for a propositional content is *getNameOfUser*. Both, the intention and the propositional content form a basic unit of communication. The *multimodal UI designer* defines the pairs of intentions and propositional contents that a given modality can express. For example, the designer decides if speech input supports an *Informing-getNameOfUser* or not. This is a mapping of modality specific representations (e.g., a speech grammar or a set of gesture recognition symbols) into a more generic representation based on communicative acts. So, the *multimodal UI designer* “programs” the platform, extended with individual modalities (plugins), in order to realize the desired interaction.

3 Introducing Mixed-Initiative and Context-Awareness

The decoupled roles facilitate the definition of more natural multimodal HRI. In particular, the decoupling enables each role to concentrate on its main task. While the modality integrator concentrates on issues related to an individual modality (e.g., the performance of a speech recognition engine), the interaction designer focuses on more natural interaction for users.

For example, the interaction designer has to consider mixed-initiative and context awareness when a semi-autonomous service robot performs a task jointly with humans in a real-world environment. Mixed-initiative allows both, the human and the robot, to initiate the interaction. However, this often requires a way for the robot to attract the attention of the user. Speech and movement of the robot serve this purpose well, while GUI suffers from limitations (it needs to be visible to the user). A robot that attracts the attention of its user by moving towards products in a supermarket is described in [4]. With context-awareness, the robot takes its own context into account and the context of the user during the interaction. This is particularly useful when the robot has the capability to choose the right interaction modality depending on the context. Decoupling of the roles enables to delay the selection of a particular modality at runtime, so the robot can consider a specific context.

4 Discussion and Open Questions

The proposed approach has been studied successfully for the development of a multimodal UI of a semi-autonomous shopping robot with GUI, speech and gesture [5]. Previous work demonstrates interaction scenarios involving mixed-initiative [4] as well.

Despite these initial results, some open questions remain and are subject to future work. Regarding the interaction designer role, the interaction language

affects potential applications. The proposed discourse model focuses on process-oriented applications, like the one's for shopping, and may not be suitable for other types of applications. Further evaluation whether the interaction designer can really design "good" discourses without in-depth knowledge about modalities is under way. Regarding the modality integrator role, the effort for integrating a modality depends on the modality's type. For example, a transformation process for a GUI requires model-to-model transformations and is far more laborious than the barcode reader's integration where only a few method calls have to be programmed.

5 Conclusion

In this paper, we introduce a distinction between three roles involved in the design of multimodal UIs. This approach allows a better decoupling between tasks that require different expertise. Tools based on discourse modeling appropriately support this decoupling and facilitate communication between roles. We discuss how this decoupling and the accompanying tools facilitate the design of a more natural human-robot interface and point out the relationship to mixed-initiative and context-awareness. Although future works is necessary for evaluating the simplicity of interaction design and modality integration, this approach already enabled the development of a multimodal UI of a shopping robot.

6 Acknowledgement

This research has been carried out in the CommRob project (<http://www.commrob.eu>) and is partially funded by the EU (contract number IST-045441 under the 6th framework programme).

References

1. de Baar, D.J.M.J., Foley, J.D., Mullet, K.E.: Coupling application design and user interface design. In: CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY, USA, ACM (1992) 259–266
2. Ertl, D.: Semi-automatic generation of multimodal user interfaces. In: EICS'09: Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems, New York, NY, USA, ACM (2009)
3. Bogdan, C., Falb, J., Kaindl, H., Kavaldjian, S., Popp, R., Horacek, H., Arnautovic, E., Szep, A.: Generating an abstract user interface from a discourse model inspired by human communication. In: Proceedings of the 41th Annual Hawaii International Conference on System Sciences (HICSS-41), Piscataway, NJ, USA, IEEE Computer Society Press (January 2008)
4. Kaindl, H., Falb, J., Bogdan, C.: Multimodal communication involving movements of a robot. In: CHI '08 extended abstracts on human factors in computing systems, New York, NY, USA, ACM (2008) 3213–3218
5. Vallee, M., Burger, B., Ertl, D., Lerasle, F., Falb, J.: Improving user interfaces of interactive robots with multimodality. In: Proceedings of the International Conference on Advanced Robotics (ICAR 2009). (June 2009) 1–6