

Multimodal Interaction with a Virtual Guide

Dennis Hofs

Mariët Theune

Rieks op den Akker

University of Twente, P.O.Box 217, 7500 AE Enschede

Abstract

We demonstrate the Virtual Guide, an embodied conversational agent that gives directions in a 3D environment. We briefly describe multimodal dialogue management, language and gesture generation, and a special feature of the Virtual Guide: the ability to align her linguistic style to the user's level of politeness.

1 Introduction

At the University of Twente we have developed the Virtual Guide, an embodied conversational agent that can give route directions in a 3D virtual building called the Virtual Music Centre (VMC).¹ When navigating through the VMC, the user can approach the Virtual Guide to ask for directions. Currently the Virtual Guide is located at the reception desk of the VMC (see Figure 1), but she could be situated anywhere in the building. In fact, with only minor changes she could also be used for direction giving in actual environments.

2 The Virtual Guide

The first part of the interaction between the Virtual Guide and the user consists of a natural language dialogue in which the multimodal dialogue management module tries to find out the user's intended destination. This may involve subdialogues, in which either the Guide or the user asks the other for clarification, and the resolution of anaphoric expressions (e.g., *How do I get there?*).² Available input modalities are typed text or speech in combination with mouse pointing. To process the user's input, the Virtual Guide incorporates a speech recognizer (Philips SpeechPearl), a parser making use of a Dutch unification grammar, and a fusion module that merges deictic expressions with any co-occurring pointing gestures (e.g., the user asking *What is this?* while pointing at the VMC map). The results of input analysis are sent to a dialogue act classifier, which maps the user's utterance to one or more dialogue acts. Based on this, the dialogue manager chooses an appropriate action to be performed by the Virtual Guide, such as uttering a certain dialogue act (realised in natural language using one of a collection of sentence templates) or showing something on the map.

Recently, the Virtual Guide has been extended with an alignment module that enables dynamic adaptation of the Virtual Guide's linguistic style to that of the user. The grammar rules used for user input analysis have been associated with tags indicating the level of politeness of the user utterance, depending on the grammatical construction used. For example, an imperative such as *Show me the hall* is considered quite impolite, while indirect requests such as *I would like to know where the hall is* are considered very polite.

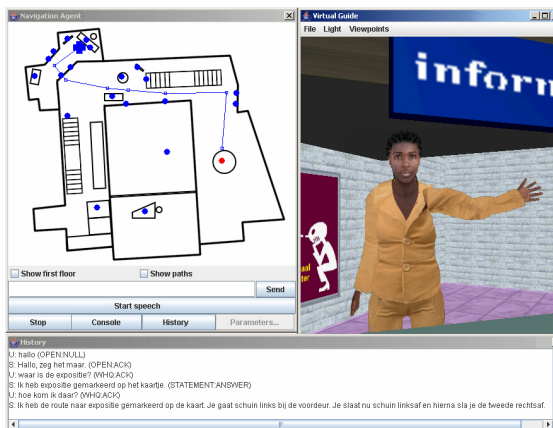


Figure 1: The Virtual Guide.

¹The Virtual Guide is accessible online via <http://wwwhome.cs.utwente.nl/~hofs/dialogue>.

²The actual language of the Virtual Guide is Dutch, but for ease of reading examples in this paper are given in English.

The templates used to generate system utterances have been similarly tagged, allowing the Virtual Guide to adapt the politeness of its replies to that of the user. Using different parameter settings for the system's initial levels of politeness, as well as the degree of alignment, allows us to model different professional attitudes or personalities for the Guide.

Currently, the alignment module is only used for the dialogue part of the interaction, not for the actual generation of the route description, which is presented in the form of a monologue when the user's destination has been established. The route description consists of a sequence of segments consisting of a turn direction combined with a description of a landmark where this turn is to be made. For example, *You go left at the information sign*. For the generation of the route description, a template-based realisation component has been built based on Exemplars [4].

Finally, the Virtual Guide's gesture generation component extends the generated text with tags associating the words in the route description with appropriate gestures. The marked-up text is sent to the animation planner (based on [3]), which realises the required animations in synchronization with the Guide's speech output. For text-to-speech synthesis, either Loquendo³ or Fluency⁴ can be used. The 3D model used for the body of the Virtual Guide was purchased from aXYZ design.⁵ In addition to being presented in speech and gesture by the Virtual Guide, the recommended route is also displayed on a 2D map of the VMC.

For more details on dialogue management, language generation and gesture generation in the Virtual Guide, see [2]. The linguistic alignment module used in the Virtual Guide is described in [1].

3 The demonstration

In the demonstration, which will last 10 to 20 minutes, we will carry out some scripted example interactions with the Virtual Guide to illustrate dialogue features such as multimodal fusion, resolution of anaphors and elliptic utterances, clarification subdialogues, error recovery and politeness alignment. In addition, visitors will be given the opportunity to interact freely with the Virtual Guide.

The system runs on a Windows computer with 2 GB of memory and a broadband Internet connection. It uses Java, Java 3D and Java Advanced Imaging.

Acknowledgements

We thank Martin Bouman and Richard Korthuis for their work on the language generation component, Marco van Kessel for his work on the gestures and embodiment, and Markus de Jong for developing the linguistic alignment component of the Virtual Guide. This work was carried out within the NWO project ANGELICA (grant no. 632.001.301).

References

- [1] Markus De Jong, Mariët Theune, and Dennis Hofs. Politeness and alignment in dialogues with a virtual guide. In *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, volume 1, pages 207–214, 2008.
- [2] Mariët Theune, Dennis Hofs, and Marco van Kessel. The Virtual Guide: A direction giving embodied conversational agent. In *Proceedings of Interspeech 2007*, pages 2197–2200, 2007.
- [3] Herwin van Welbergen, Anton Nijholt, Dennis Reidsma, and Job Zwiers. Presenting in virtual worlds: Towards an architecture for a 3D presenter explaining 2D-presented information. *IEEE Intelligent Systems*, 21(5):47–53, 2006.
- [4] Michael White and Ted Caldwell. EXEMPLARS: A practical, extensible framework for dynamic text generation. In *Proceedings of the Ninth International Workshop on Natural Language Generation (INLG-98)*, pages 266–275, 1998.

³<http://www.loquendo.com/>

⁴<http://www.fluency.nl/>

⁵<http://www.axyz-design.com/>