Engagement-based Multi-party Dialog with a Humanoid Robot

David Klotz and **Johannes Wienke** and **Julia Peltason** and **Britta Wrede** and **Sebastian Wrede**
Applied Informatics Group
Bielefeld University
{dklotz, jwienke, jpeltaso, bwrede, swrede}@techfak.uni-bielefeld.de

Vasil Khalidov and **Jean-Marc Odobez**
IDIAP Research Institute
{vasil.khalidov, odobez}@idiap.ch

Abstract

When a robot is situated in an environment containing multiple possible interaction partners, it has to make decisions about when to engage specific users and how to detect and react appropriately to actions of the users that might signal the intention to interact.

In this demonstration we present the integration of an engagement model in an existing dialog system based on interaction patterns. As a sample scenario, this enables the humanoid robot Nao to play a quiz game with multiple participants.

1 Introduction

Giving robotic systems the ability to join in conversation with one or multiple users poses many new challenges for the development of appropriate dialog systems and models. When a dialog system is situated in the real, physical world and used in more open settings, more effort needs to be spent on establishing and maintaining clear communication channels between the system and its users. E.g. the system first needs to detect that there are potential users with whom interacting would be possible, it needs to decide if a detected person wants to interact with the system at all and it needs to make decisions when and how it should try to start an interaction with that person.

Bohus and Horvitz (2009) have developed a model for representing the current relation of a user with such a system (their *engagement state*) and determining if they want to be involved in an interaction with the system (using explicit *engagement ac-tions* and the more abstract *engagement intention*). Each user can be *engaged* in specific *interactions* (denoting different "basic unit[s] of sustained, interactive problem-solving") and there can be multiple such interactions, each with potentially different users.

This demonstration shows how an engagement model inspired by these ideas was integrated into an existing dialog system and how it helps in realizing interactive scenarios with a robot that incorporate cues for the dialog from the system's environment. Section 3 gives more details about this model and how it is used by the dialog.

2 Scenario

As a scenario for this demonstration we chose a simple quiz game involving the robot Nao as a host playing with one or multiple human users. At first, the robot waits until one of the human interaction partners approaches. When the person opens the interaction (i.e. by greeting the robot), the system responds with an appropriate greeting. While the person continues to show the intention to interact with the robot (determined by the process described in section 3.1), the robot will ask questions randomly chosen from a predefined set and will try to judge if the person answered them correctly.

When another person enters the robot's field of view, the system also tries to determine if they have the intention to interact with it. If that is the case, the system suspends the current interaction with the first person and actively tries to engage the second person, encouraging him or her to join the ongoing quiz game. The prospective new player can then choose
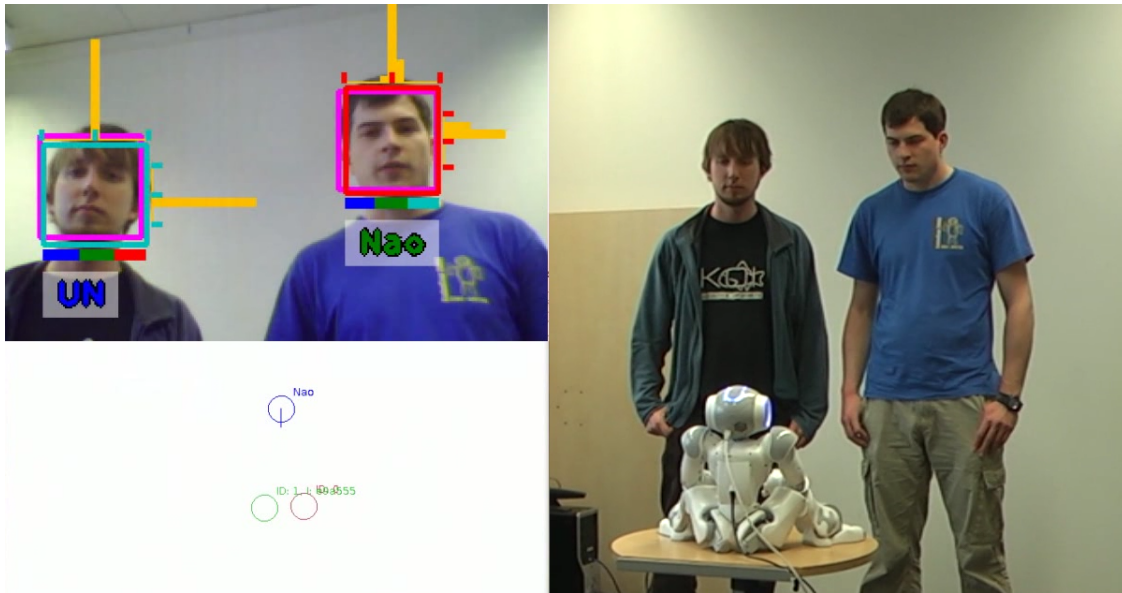
341

Figure 1: Two persons interacting with the developed system.

to join or decline the request.

As long as one of the engaged participants shows the intention to interact, the robot continues to ask questions which all participants can try to answer. The quiz game is stopped either by an explicit request of one the users or after all participants have left the scene.

This scenario serves as a good testbed for the integration of different cues for the engagement model and how that model affects the actions taken by the dialog system. The right-hand side of figure 1 shows two people interacting with the robot during the quiz game.

3 System Overview

Figure 2 shows an overview of the different components involved in the demonstrated system. This includes components for the perception (e.g. accessing images from the robot's camera and audio from its microphones), for generating actions (e.g. using the robot's text-to-speech system), the dialog system itself and a memory system for connecting these diverse components.

The dialog system used for this demonstration is called PaMini, which is short for "Pattern-based Mixed-Initiative human-robot Interaction" and is described in more detail by Peltason and Wrede (2010). This dialog system was modified in Klotz

(2010) with a model of engagement based on the ideas presented by Bohus and Horvitz (2009). In our adaptation of this model, there are extension points for integrating different sources of information about the user's engagement intentions and actions, described in the following section.

3.1 Determining the User's Actions & Intention

For determining the user's actions (e.g. if the user explicitly wants to start an interaction with the system), this demonstration uses a set of possible utterances which are simply matched against the results of a speech recognition module.

To get an estimation of the user's intention to interact, the image from the robot's camera is first used to detect the faces of users and to estimate their current visual focus of attention. A module based on a framework by Ba and Odobez (2009) is used to determine probabilities that the user is looking at each of a pre-defined list of possible focus targets, including the robot itself and other users visible in the scene. The upper left of figure 1 shows a visualization of this module's output. *Nao* denotes the robot as the focus target with the highest probability, while the designation *UN* is short for the "unfocused" target.

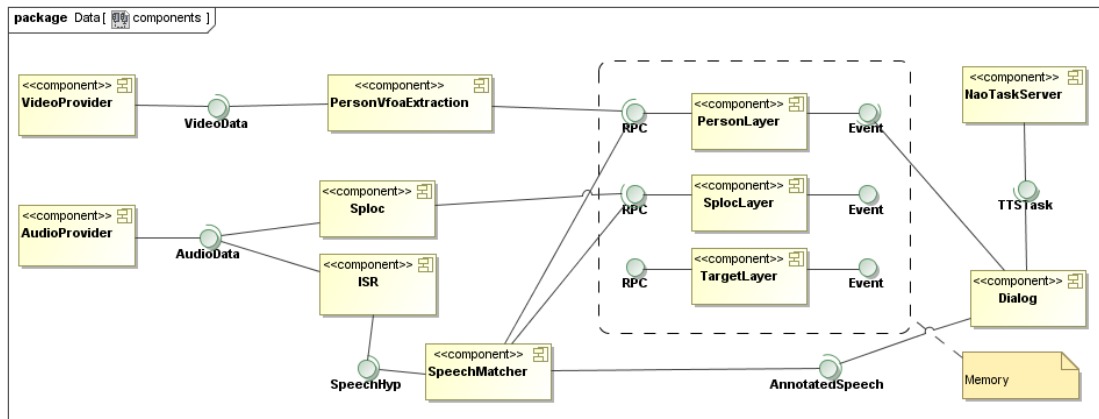This list of probabilities is then stored in a mem-

Figure 2: Components of the developed system.

ory system developed by Wienke and Wrede (2011). The memory system provides temporal query capabilities which are finally used to guess a user's current intention of interacting with the robot based on the history of the probabilities that the robot was the user's current visual focus of attention target. This result is also stored in the memory system together will all other information known about a user.

3.2 Engagement Cues for the Dialog

The dialog system receives the information about the user's state and intention from the memory system and uses it in several rules for controlling its own engagement actions. The intention is e.g. used to determine if there is a new user that should be persuaded to join the quiz game described in section 2 and if any of the users still shows interest so that a new question should be asked. The general state of the detected users is also used e.g. to observe when the users leave the robot's field of view for a longer period of time which causes the dialog system to close its current interaction.

4 Conclusion

We have shown how an existing dialog system that was enhanced using an explicit model of engagement can be used to realize interactive scenarios with a robot that is situated in the physical world. An estimation of the user's current visual focus of attention is used to gauge their intention to engage the robot in conversation.

A video recording of two people interacting with the developed system is available online at `http://youtu.be/pWZLVF2Xa8g`

References

S. Ba and J.-M. Odobez. 2009. Recognizing Visual Focus of Attention from Head Pose in Natural Meetings. *IEEE Trans. on System, Man and Cybernetics: part B, Cybernetics*, 39:16–34.

Dan Bohus and Eric Horvitz. 2009. Models for multi-party engagement in open-world dialog. In *Proceedings of the SIGDIAL 2009 Conference*, pages 225–234, London, UK. Association for Computational Linguistics.

David Klotz. 2010. Modeling engagement in a multi-party human-robot dialog. Master's thesis, Bielefeld University.

Julia Peltason and Britta Wrede. 2010. Modeling human-robot interaction based on generic interaction patterns. In *AAAI Fall Symposium: Dialog with Robots*, Arlington, VA, USA.

Johannes Wienke and Sebastian Wrede. 2011. A spatio-temporal working memory for multi-level data fusion. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*. submitted.