Universität Bielefeld

CORE Lab

idiap RESEARCH INSTITUTE

Robot-to-Group Interaction in a Vernissage: Architecture & Dataset for Multi-Party Dialog

D. Klotz, J. Wienke, B. Wrede, S. Wrede, S. Sheikhi, D. Jayagopi, V. Khalidov, J.-M. Odobez

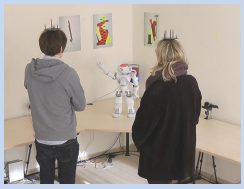
Overview

- **Overall Goal:** Improve robot-to-group interaction including natural dialog abilities
- Requires interactive system with appropriate dialog abilities and supporting architecture
- Collection of a multi-modal dataset in an HRI scenario for evaluation and analysis

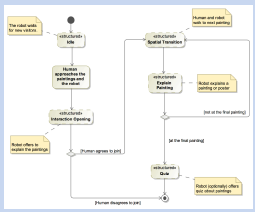
The humanoid robot Nao interacting with a group of museum visitors (from data collected for [7])

Research Topics & Scenario

- Robots should be able to work and interact with groups of users in populated spaces
- Abilities that are beneficial for smooth interaction include having a model of engagement ("Who wants to interact?", "Who am I already interacting with?")
- This enables the robot to decide when to engage specific users & to react to actions of the users that signal the intention to interact
- The "Vernissage": A scenario where the robot acts as a guide in an art exhibition



The robot explaining paintings to a group of visitors during the recording of the "Vernissage" dataset



Integration of Engagement in a Dialog System

PaMini dialog system:

- Models dialog based on configurable building blocks for HRI [3]
- Interaction patterns model recurring conversational structures
- Task-based communication with other components part of patterns

Engagement:

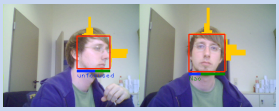
- Users have engagement state, intention (engaged, not engaged) & perform engagement actions
- Regulates establishment of communication channels, see [2]

Extensions for multi-party dialog:

- Explicitly modeling different interaction contexts (as a stack of interactions)
- Modeling the state of different users in those interactions (engaged / not engaged)
- Extension points for determining engagement actions (e.g. "Hello" from ASR), intentions (e.g. through visual focus), choosing appropriate responses

Focus of Attention as a Dialog Cue

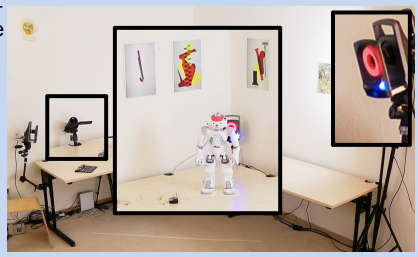
- Face tracking & pose recognition treated as coupled problems in a probabilistic framework, see [4]
- Head pose determines probabilities for different possible focus targets, see [5] for a similar approach
- Short term history of the robot being the most probable focus target used to estimate engagement intention, see [1]



Correctly detecting the "unfocused" state (left) and the visual focus on the robot (right), yellow bars indicate probability distribution over discrete set of pan and tilt angles, target probabilities & most probable target in blue/green

The Vernissage Dataset

- Setting: Small vernissage in one room of a rented flat. The robot stands on a table, explains paintings and performs a quiz with the visitors of the vernissage.
- Wizard of Oz controlling fixed utterances (incl. motion), free head gaze
- ~2h of recorded data, 20 subjects
- Motion capture of participants' and Nao's head, 3 external HD cams with sound, Nao video and audio, internal data of Nao, Speech of participants, Wizard commands
- Recording with RSB tools [6]
- Data will be available for research



Software Architecture

- Distributed system with Nao and external computers for processing
- Specific need: dataset recording for quantitative analysis
- New middleware RSB [6] facilitates using Nao in a research context
- Hierarchical bus architecture
- Generic record-replay tools available for RSB based on introspection support (no component changes for replay)
- Rich meta-data set facilitates research, e.g. synchronization of dataset

References

[1] D. Klotz, J. Wienke, J. Peltason, B. Wrede, S. Wrede, V. Khalidov, J-M. Odobez: *Engagement-based Multi-party Dialog with a Humanoid Robot*, In Proc. SIGDIAL 2010 Conference, pp. 341–343

[2] D. Bohus, E. Horvitz: *Models for Multiparty Engagement in Open-World Dialog*. In Proc. SIGDIAL 2009 Conference, pp. 225–234

[3] J. Peltason, B. Wrede: *Pamini: A framework for assembling mixed-initiative human-robot interaction from generic interaction patterns*. In Proc. SIGDIAL 2010 Conference, pp. 229–232

[4] E. Ricci, J.M. Odobez: *Learning Large Margin Likelihood for Realtime Head Pose Tracking*. In IEEE Int. Conf. on Image Proc, 2009

[5] S. Ba, J.M. Odobez: *Recognizing Visual Focus of Attention from Head Pose in Natural Meetings*. IEEE Trans. on System, Man and Cybernetics: part B, Cybernetics, Vol. 39. No. 1. pp. 16-34, Feb 2009.

[6] J. Wienke, S. Wrede, *A Middleware for Collaborative Research in Experimental Robotics*, IEEE/SICE International Symposium on System Integration (SII2011), Kyoto, Japan, IEEE, In Press

[7] K. Pitsch, S. Wrede, J.-Ch. Seele, L. Süssenbach: *Attitude of German Museum Visitors towards an Interactive Art Guide Robot*, In: Proceedings HRI 2011.

Supported by

Further partners

Inria INVENTEURS DU MONDE NUMÉRIQUE

CZECH TECHNICAL UNIVERSITY IN PRAGUE

ALDEBARAN Robotics

HUMAVIPS
Humanoids with Auditory and Visual Abilities in Populated Spaces

http://www.humavips.eu

Seventh Framework Programme
Theme ICT-2009.2.1:
Cognitive Systems and Robotics
Grant no. 247525