Heterogeneous Multi-Modal Mixing

Realizing fluent, multi-party, human-robot interaction with a mix of deliberate conversational behavior and bottom-up (semi)autonomous behavior

Dennis Reidsma, Daniel Davison, Edwin Dertien
University of Twente
Human Media Interaction / Robotics and Mechatronics
{d.reidsma|d.p.davison|e.c.dertien}@utwente.nl

Abstract

This project aims to work on a novel, state-of-the art setup for realizing fluent, multi-party, human-robot interaction with a mix of deliberate conversational behavior and bottom-up (semi)autonomous behavior. We approach this from two sides. On the one hand, there is the dialog manager requesting deliberative behavior and setting parameters on ongoing (semi)autonomous behavior; on the other hand, there is the robot control software that needs to translate and mix these deliberative and bottom-up behaviors into consistent and coherent motion. The two need to work well together in order to get behavior that is fluent, naturally varied, and well-integrated while at the same time conforming to the high level requirements as to content and timing that are set by the dialog manager. We will first look at the visual attention displayed by the robot in a multi person interaction scenario; once this works, we will extend the project towards other domains of expressive behavior as well. In order to prepare for an evaluation study that is to be carried out in follow-up to eNTERFACE'16, we will also design an experiment aimed at evaluation the core qualities of the system in a relevant scenario and carry out a pilot run of the study during the eNTERFACE'16 summer period.

1 Objective and Background

The main objective of this project is to bring forward the state of the art in fluent human-robot dialog by improving the integration between deliberative and (semi)autonomous / reactive behavior control. The interaction setting in which this will be done is one of multi-party interaction between one robot and several humans. The project will build upon interaction scenarios with collaborative educational tasks, as used in the context of the EU EASEL project [1], and will use and extend the state-of-the-art BML realizer ASAPRealizer [9].

Fluent interaction plays an important role in effective human-robot teamwork [3, 4]. A robot should be able to react to a human's current actions, to anticipate the user's next action and proactively adjust its behaviour accordingly. Factors such as inter-predictability and common ground are required for establishing such an alignment [6, 5]. Regulation of (shared) attention, which to a large extent builds upon using the right gaze and head behaviors [2], plays an important role in maintaining the common ground. In a multi-party setting, the matter becomes more complex. A mixture of conversational behaviors directed at the main interaction partner, behaviors directed at other people nearby to keep them included in the conversation, and behaviors that show general awareness of the surrounding people and environment need to be seamlessly mixed and fluently coordinated to each other and to actions and utterances of others.

For a robot that is designed to be used in such a social conversational context, the exact control of its motion capabilities will be determined on multiple levels: autonomous behaviors such as idle motions and breathing, semi-autonomous behaviors such as the motions required to keep the gaze focused on a certain target, reactive behaviors such as reflex responses to visual input, and deliberative behaviors such as speech or head gestures that make up the utterances of the conversation. Part of the expressions (and especially the deliberative ones) will be triggered by requests from a dialog manager. Other parts may be more effectively carried out by modules running in the robot hardware itself – especially the ones that require high frequency feedback loops such as tracking objects with gaze or making a gesture towards a moving object.

A dialog manager for social dialog orchestrates the progress of the social conversation between human and robot, and –based on this progress– requests certain deliberative behaviors to be executed and certain changes to be made to parameters of the autonomous behavior of the robot. Such requests are typically specified using a high level behavior script language such as the Behavior Markup Language (BML), which is agnostic of the details of the robot platform and its controls and capabilities for autonomous behaviors. The BML scripts are then communicated to the robot platform by a Behavior Realizer (in this project: ASAPRealizer [7]), which interprets the BML in terms of the available controls of the robotic embodiment. (Semi)autonomous behaviors may then be mixed into the deliberative behaviors, either by ASAPRealizer or by the robot platform itself. Since the behavior should respond fluently to changes in the environment, the dialog models as well as the robot control mechanisms must be able to adapt on-the-fly, always being ready to change on a moment's notice. Any running behaviour could be altered, interrupted or cancelled by any of the control mechanisms to ensure the responsive nature of the interaction. This multi-level control can include social commands like *maintain eye contact during conversation*, as well as reactive commands like *look at sudden visually salient movements.*

This project will work on such seamless integration of deliberative, reactive, and (semi)autonomous behaviors for a social robot. This introduces a challenge for an architecture for human robot interaction. On the one hand, the robot embodiment continuously carries out its autonomous and reactive behavior patterns. The parameters of these may be modified

on the fly based on requests by the dialog manager. On the other hand, the dialog manager may request deliberative behaviors that actually conflict with these autonomous behaviors, since the dialog manager does not know the exact current state of the autonomous behaviors. The envisioned architecture therefore should contain intelligence to prioritize, balance and mix these multilevel requests before translating them to direct robot controls. In addition, the robotic embodiment should send updates and predictions about the (expected) timing with which behavior requests from the dialog manager will be carried out, so the dialog manager can manage adaptive dialog [10].

The resulting system will be deployed in a setting in which fluent and responsive behavior can be shown off to good advantage. To this end we will setup a scenario centered around multiparty interaction with dynamic and responsive gaze behavior (see Work Plan), and evaluate this setup both in terms of user perception, and in terms of the capabilities offered by the system to realize natural and fluent human-robot dialog on the appropriate levels of abstraction.

2 Resources

To ensure rapid progress, we will build upon a number of resources that are readily available to the team. Besides the excellent lab facilities offered by the eNTERFACE'16 hosting institute, which include also a rapid prototyping workshop with the necessary tools for working on the hardware side of the project, we will build among other things on the following components.

- **Multimodal Behavior Specification: BML** The Behavior Markup Language standard (BML)[1] is a markup language that allows dialog level specification of the behaviors that a social agent or robot should display, together with their synchronization.
- **ASAPRealizer** is a BML compliant behavior realizer for generating multimodal verbal and nonverbal behavior for social agents from BML specifications. It is designed specifically for continuous interaction with capabilities for on-the-fly adaptation of planned behavior, and has already been extended to work with several different robotic embodiments.
- **A minimalistic proof-of-concept "social head"** developed by Edwin Dertien is available as the robot platform on which we can show the basic principles of the targeted architecture.
- **The Robokind R25 Zeno robot** is a small humanoid robot capable of gestures and face expressions, and is available for the project to show the capabilities of our solution for fluent multiparty interaction in a more advanced setting.

[1] http://wiki.mindmakers.org/projects:bml:draft1.0

- **A few basic perception modules** are available for basic audiovisual detection of multiple participants.

3 Project Management

The project management tasks will be shared by the three core members of the team. We plan to have a thorough preparation before the workshop takes place, communicating with the participants through email, so the participants do not enter the actual workshop completely blanco. One person of the team of project leaders will take responsibility for integrative activities at an early stage during the workshop, to ensure that at start of evaluation there is a full system to evaluate.

4 Work plan and Work Packages

WP1 Basic integration At the start of the project we will set up a basic integrated system that involves perception, dialog modelling, behavior planning and realisation, and control of the robot embodiment to be used. We will make use of the existing basic setup used in the EU EASEL project. This work will be carried out in the first few days in WP 1.

WP2 Scenario development and study design The project aims to achieve fluent multi-party interaction with a robot. Once this is achieved, we want to carry out a study to evaluate the effect of the achieved fluency on the interaction with users, and perception of the robot by users. WP2 is responsible for designing a relevant scenario and an evaluation study. To this end, the first two weeks WP2 will be spent working towards a suitable and specific scenario that displays this type of interaction to maximum advantage, and the last two weeks will be focused towards developing an HRI study and piloting with people from eNTERFACE. The larger scale user study will be carried out in follow-up to eNTERFACE and will be developed into a publication coauthored by the participants to this project.

WP3 Dialog modelling WP3 concerns the development of the necessary dialog models capable of orchestrating interactions in the scenario developed in WP2. The scenario developed in WP2 needs to be translated to actual dialog models that can carry out the necessary steps of the interaction. We will use the information state based dialog manager Flipper [8] for this.

WP4 Integrating deliberative and (semi)autonomous behavior control This work package focuses on the integration of the deliberative behavior controlled by the dialog manager and the (semi)autonomous behavior that runs in parallel. This work package is divided in two parts, that need to be carried out by different people yet must be tightly coordinated: the robot platform and control side, covering the aspects

implemented within the hardware architecture, and the behavior planning and realisation side that used ASAPRealizer to translate between dialog requests and robot control requests and feedback. To start with, we will focus on the integration of communicative and functional head and gaze behavior with reactive and idle gaze and head motions. Depending on the available time and the number of team members, we may extend this with other behavior types as well.

Other work packages Depending on the expertise and interests of people who join the project, there may also be additional work packages for topics such as audiovisual scene analysis or speech recognition.

5 Team

We are looking for team members with interest in taking a central role in topics such as (but not necessarily limited to):

- Development of dialog models for multi-party human-robot conversations
- Fluent low-level control of robot embodiments
- Design and execution of HRI studies
- Visual scene analysis in a multi-party setting for real-time perception for social robots
- Robust speech recognition in open settings

The lead team of this project consists of dr. ir. Dennis Reidsma, dr. ir. Edwin Dertien, and Daniel Davison (MSc). They will take responsibility for coordination as well as provide a major contribution to various work packages.

Dennis Reidsma is Assistant Professor at the Human Media Interaction group and Lecturer at the Creative Technology curriculum at the University of Twente. His PhD thesis, titled "Annotations and Subjective Machines – of annotators, embodied agents, users, and other humans", dealt with problems of annotation and reliability in large multimodal annotated corpora, and especially the relation between reliability and annotator agreement on the one hand, and the subjective nature of many annotation tasks in the field of human computing on the other hand. His current research activities focus on two main topics. He supervises a number of BSc, MSc, and PhD students on topics of computational entertainment and interactive playgrounds, carries out various research activities in this area, and is regularly involved in the organization of conferences such as INTETAIN and the conference Advances in Computer Entertainment. In addition, he has published many papers on behavior generation for social agents, and consolidated the results of this joint work with Herwin van Welbergen in the release of AsapRealizer (formerly Elckerlyc), a state-of-the-art Open Source software platform for generating continuous interaction with agents and robots. At the moment he is involved

in the EASEL project, in which robots are introduced in a learning task with children to facilitate the learning process.

Daniel Davison is a PhD student at the Social Robotics group of the University of Twente. He has a BSc in Computer Science and a MSc in Human Media Interaction. During his Master's thesis he developed a semantic knowledge base, which helps laboratory scientists store research data and manage their workflow. His current PhD research focuses on investigating child-robot interactions in the context of education. In this context, he aims to develop social robot behaviours for optimally supporting a child's learning process. As a member of the EASEL project he has been closely involved with other partners in the technical software integration, resulting in a robotics architecture capable of perception, reasoning, behaviour generation and actuation. This integrated architecture has been used to demonstrate a proof-of-concept, multimodal interaction between a user, a social robot and smart learning materials.

Edwin Dertien is assistant professor at the University of Twente. He carried out his PhD research at the University of Twente in the Mechatronics group, on the design of a pipe inspection robot and obtained his Ph.D. in 2014. He has a fascination for robotics, and has worked on several projects involving (autonomous) robots, both from an engineering or artistic interest. Since 2006 his company 'Kunst- en Techniekwerk' (making art work) has been providing the technical backbone in art-projects, ranging from robot-arms for sci-fi movies to autonomous GPS track drawing robots also realising various autonomous art works, big moving sculptures and weird installations for festivals and expositions. In 2008 he started at University of Twente as researcher / teacher for Creative Technology programme, developing and teaching courses in programming, physical computing, sensor technology, tinkering and explorative design. In 2013 he co-founded the ASSortiMENS foundation, running a FabLab inspired workshop tailored to people with autism, including a children's FabLab project (het kinderfablab). Skills include: mechatronic design, electronics, robotics, control engineering, creative technology, bottom-up creativity, personal fabrication, publishing, writing, instruction, teaching, digital fabrication (3D print, CNC, lasercutting) embedded design, microcontroller (AVR (arduino) ARM), programming (C, Java, Processing), web, metalworking (welding, lathe, mill, cnc).

References

[1] V. Charisi, D. P. Davison, F. Wijnen, J. van der Meij, D. Reidsma, T. Prescott, W. Joolingen, and V. Evers. Towards a child-robot symbiotic co-development: a theoretical approach. In M. Salem, A. Weiss, P. Baxter, and K. Dautenhahn, editors, *Proceedings of the Fourth International Symposium on "New Frontiers in Human-Robot Interaction", Canterbury, UK*, pages 331–336. Society for the Study of Artificial Intelligence & Simulation of Behaviour, April 2015.

[2] D. Heylen. Head gestures, gaze, and the principles of conversational structure. *International Journal of Humanoid Robotics*, 03(03):241–267, 2006.

[3] G. Hoffman. *Ensemble : fluency and embodiment for robots acting with humans*. PhD thesis, Massachusetts Institute of Technology, 2007.

[4] G. Hoffman and C. Breazeal. Effects of anticipatory perceptual simulation on practiced human-robot tasks. *Autonomous Robots*, 28(4):403–423, dec 2009.

[5] G. Klein and P. Feltovich. Common Ground and Coordination in Joint Activity. *Organizational simulation*, pages 1–42, 2005.

[6] S. Kopp. Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. *Speech Communication*, 52(6):587–597, jun 2010.

[7] D. Reidsma and H. van Welbergen. AsapRealizer in practice – A modular and extensible architecture for a BML Realizer. *Entertainment Computing*, 4(3):157–169, aug 2013.

[8] M. ter Maat and D. Heylen. Flipper: An information state component for spoken dialogue systems. In H. H. Vilhjálmsson, S. Kopp, S. Marsella, and K. R. Thórisson, editors, *Intelligent Virtual Agents - 11th International Conference, IVA 2011, Reykjavik, Iceland, September 15-17, 2011. Proceedings*, volume 6895 of *Lecture Notes in Computer Science*, pages 470–472. Springer, 2011.

[9] H. van Welbergen, D. Reidsma, and S. Kopp. An incremental multimodal realizer for behavior co-articulation and coordination. In Y. Nakano, M. Neff, A. Paiva, and M. Walker, editors, *12th International Conference on Intelligent Virtual Agents, IVA 2012*, volume 7502 of *Lecture Notes in Computer Science*, pages 175–188, Berlin, 2012. Springer Verlag. ISBN=978-3-642-33196-1, ISSN=0302-9743.

[10] H. van Welbergen, D. Reidsma, and J. Zwiers. Multimodal plan representation for adaptable bml scheduling. *Autonomous Agents and Multi-Agent Systems*, 27(2):305–327, September 2013.