How to Catch A Werewolf

Exploring Multi-Party Game-Situated Human-Robot Interaction

Lead-Organizers:

Catharine Oertel, KTH (PI)

Samuel Mascarenhas, INESC-ID

Zofia Malisz, KTH

José Lopes, KTH

Joakim Gustafson, KTH

Contributors and Team-Members:

Dimosthenis Kontogiorgos

Patrik Jonell

Joseph Mendelson

Todd Shore

Supporting Advisors:

Jonas Beskow

Gabriel Skantze

Team-Members:

You ? !☺

Introduction and Background Information:

With demographic changes in the last decade robots have become more and more important to cover demands as social companions in general or in the health and educational sector in particular. While there has been a lot of improvements in human-robot-interaction in general, robots are still lacking certain social capabilities. In this project we would like to address a small subset of these capabilities and in particular work on improving social awareness in a multi-party conversational setting.

In fact we would like to focus on the study of deception. Deception is a phenomenon which occurs every day in human-human interpersonal communication. It does not need to be associated with a negative intention at all but can simply encompass hiding anger in front of superior or fear in front of a child (De Rosis et al., 2003, Prendinger & Ishizuka, 2001). In general deceptive behaviour is not only restricted to induce false beliefs in others, but also for example the exchange of non-existent emotions or the communication of false preferences or opinions (DePaulo et al., 1996). Only if robots are also equipped with the capability to detect and also to generate deceptive behaviours will they be able to engage in social interaction on a long-term basis.

In previous studies the role-playing game "Werewolves" has been employed as a test-bed for studying deception (Hung & Chittaranjan, 2010, Dias & Aylett, 2013) and also for studying conversational dynamics in general (Oertel & Salvi, 2013). This game is particularly well suited for our purposes as it is engaging and does not use any objects which might distract people and alter their gaze patterns. Also, from a decision-making perspective playing "Werewolves" involves a lot of modelling what the other participants are thinking. Having that capacity in a robot can be valuable, as it makes robots closer to understand others. We therefore opted for adapting the same experimental set-up. However, in contrast to previous studies which investigated deception in an offline manner, we are aiming in this project to implement perception detection and perception generation in the robot-head Furhat (Moubayed et al., 2012) employing the IrisTk framework (Skantze & Al Moubayed, 2012).

For the ease of the reader the following paragraph will provide a short overview over the rules of the "Werewolves" game. The roleplaying game "Werewolves" is a game in which participants are given different roles. They can be for example be given the role of a "Werewolf" or the role of a "Villager". The participants do not disclose which role they have been given to each other. It is the aim of the game for the villagers to identify the werewolves, whereas it is the aim of the werewolves not to be identified by the villagers and to kill as many villagers as possible. If the villagers identify the werewolves before the game is finished then the villagers win, if the werewolves kills all villagers then the werewolves win.

In this project we will focus on the implementation of the roles of the "villager" and the "werewolves" using the IrisTK dialogue framework and the robot head Furhat. To be more precise, the aim of this project is to use multi-modal cues in order to inform the theory of mind model to drive the robot's decision making process. Theory of mind is a concept that is related to empathy and it refers to the cognitive ability of modeling and understanding that others have different beliefs and intentions than our own. In lay terms, it can be described as "to put oneself into another's shoes" and it is a crucial skill to properly play a deception game like "Werewolf".

Project Objectives:

In summary the objectives of this project proposal are:

- Explore the use of an array of multi-modal features for the detection and generation of deceptive behaviours
- Implement deceptive and non-deceptive behaviours in the robot-heat Furhat using the dialogue framework IrisTk
- Carry out human-robot game-situated data recordings
- Carry out user-evaluation

Research questions which will be investigated in particular are:

1. Is it possible to automatically detect the werewolf/villager amongst the participants ?
2. Is it possible to generate convincing werewolf / villager behaviour in Furhat ?
3. Is it possible to generate convincing deception behaviour in Furhat ?

In order to answer these research question several approaches are possible. One approach could be a rule-based implementation of multi-modal turn-taking behaviour for perception and generation. Another approach could be a machine learning/feature-engineering one. We will explore both in this project.

Technical Set-Up:

To implement the *Werewolf* scenario we aim to accommodate groups of 4(humans)+1(robot) playing the game. That is, 4 human participants and the Furhat robot. The scenario is implemented in a situated interaction setting where all participants are on a round table interacting. The robot's role is not to moderate or initiate the discussion, but to actively participate as a player. As, such we want to focus on a human like interaction where the robot is part of the group. Figure 1 illustrates a past set-up of a corpus recording carried out at KTH. For the current data-recording we intend to use a very similar sensor set-up. However, we plan to amend it by having 2 initial participants take part and not using the touch screen. Figure 2 illustrates the set-up we are planning for the summerschool.
In both set-ups several sensors are gathering data from the participants online and processing and syncing all signals to implement appropriate robot behaviour. The main setup of the equipment is gathering speech, motion and gaze. In order to gather such features we need sensors that will capture data and then combine the features to understand human behaviour. Table 1 provides an overview of the array of sensors we intend to use for this project.

Figure 1. Example of setup at KTH

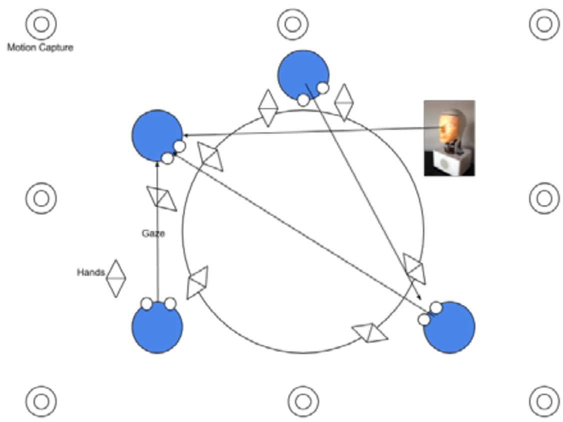


Figure 2. Enterface Set-Up

The software architecture, as detailed in Figure 3, is divided in 5 layers, namely; Hardware, low-level feature extraction, higher-level processing modules, decision making modules and a lastly robot output. A brief description of each component can be found in Table2.
In addition software which will be used in this project is ***IrisTK, OpenSmile, WebGL, Blender, Matlab***.

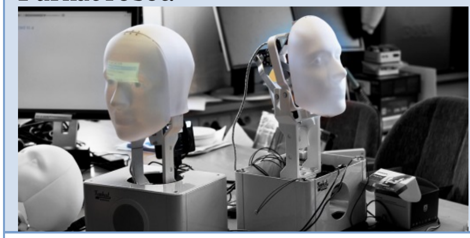
Initial Data:

Prior to the enterface workshop the team members will be given access to data of human subjects playing the "Werewolf" game in the target set-up (subjects are wearing motion-sensors, eye-tracking glasses etc.)

Resources Needed:

1) Access to Motion-Tracking Lab

Table 1: Equipment used

Equipment:	Explanation:
Motion tracking	We hope to use the motion capture facility provided by the Digital Creativity Centre. The mocap system consists of a Vicon Motion Capture System and sets of reflective markers. We capture motion from movement of the participants' head and hands by placing markers on the Tobii glasses they will wear and gloves.
Eye gaze tracking	We will also use eye trackers (Tobii Glasses 2) provided by KTH to capture online eye gaze information from the participants. Combining the coordinate systems of motion capture and eye trackers we get a trajectory of gaze from all participants in 3D space in real time.
Furhat robot. 	A Furhat robot provided by KTH will be used in the context of the Werewolf game. The robot uses information captured by the sensors to get verbal and non-verbal cues in order to improve its decision making module. A belief system is estimating probable roles per participant in the context of the game and actions are taken according to this accumulated knowledge of participants' behaviour.
Audio & Video.	Microphones are used to capture language and are sequentially processed by IrisTK and Google ASR. Cameras are recording the interaction as a whole, each participant's point of view (captured by the glasses) and facial expressions from each participant.

Software architecture

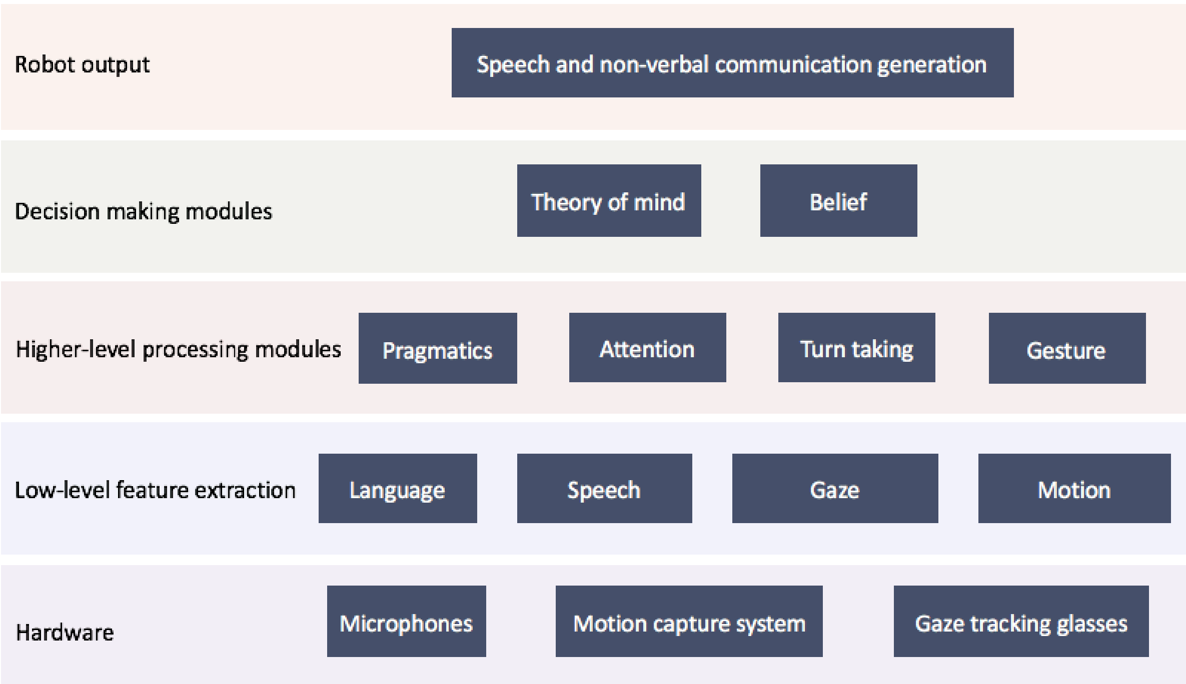


Figure 3. Software architecture

Table 2: Descripting of Modules illustrated in System Architecture

Module Name	Explanation
Speech and non-verbal communication generation	This module is responsible for the output of speech and non-verbal communication. It does so by utilizing the IrisTK API. The speech and non-verbal communication generation will be based on a data set recorded recorded prior to the eNTERFACE Summer School.
Theory of mind	In order to build an understanding about the belief about itself we will be using FAtiMA-Toolkit, which is a collection of agent-based tools with a focus on social and emotional intelligence. The toolkit will be used to add perspective-taking to the robot's reasoning as well as an emotional layer to the robot's behaviour.
Belief	The belief module is responsible to build an understanding about the robots opponents in the game. If the robot is playing as a villager it will try to assign probabilities of being the werewolf to each player. The belief is built on the data coming from the pragmatics, attention, turn taking and gesture modules.
Pragmatics	The pragmatics module is responsible for understanding the intention of what the other players are saying based on the input from the language and speech modules.
Attention	The attention module will utilize the data from the speech, gaze and motion feature extraction modules to determine each player's attentiveness.
Turn taking	The turn taking module will determine the turn taking during conversations based on the data from the speech, gaze and motion modules. A baseline model will be implemented before the summer school to be tested in the pilot experiments.
Gesture	Gestures will be derived from the motion and gaze features.
Language	The language model will be using google cloud speech api for ASR.
Speech	The speech module will extract features from the microphone input and from data available from IrisTK, such as for example pitch.
Gaze	The gaze tracking module will extract features based on the gaze tracking input.
Motion	The motion module will extract features from the motion capture and gaze tracking data.

Work-Packages:

The work packages will include

1. **WPI: Low-Feature Extraction**
   In WPI I participants will design the modules for the low-level feature extraction. This involves designing the API's for both input and output. Participants will familiarize themselves with the provided input devices.

2. **WPII: High-Level Processing Modules**
   This workpackage deals with developing rule based/ or machine learning based methods for modeling dialogue acts, attention, turn-taking and gesture both for understanding and generation.

3. **WPIII: Decision-Making Modules:**
   This workpackage will take data from the high-level processing modules in order to build a belief model of each participant's role including itself. This workpackage will make use of the Fatima toolkit for modeling the Theory of Mind capabilities of the robot.

4. **WPIV: Robot Output**
   WPiV will develop generation modules for speech and non-verbal behaviours.

5) **WPV: Data Recording and User Evaluations**
In WPV participant will gather data of humans engaging in the werewolf game with the robot. An initial user evaluation will be carried out.

Deliverables:

Week I: The first week will be developing to the final design of all workpackage, team building and the delivery of workpackage I.

Week II: At the end of week II workpackage II and IV should be finalised.

Week III: Week 3 should be devoted to data recordings and tuning of the workpackage 1-4.

Week IV: The last week of the eNTERFACE workshop will be devoted to data recordings, user evaluations and the writing of the final report and initial data-analysis as well as evaluation.

Benefits of the research

The here proposed implementation will be beneficial in many regards. For example, it will enable us to evaluate turn-taking models, belief tracking as well as gaze-patterns generation in multi-party dialogue. It will give all participants access to a powerful test-bed for an array of future studies. Moreover, the resulting human-robot interaction corpus will be novel in itself .

Why you wanna be joining this team!

This project requires a multidisciplinary team since it involves a number of different areas of knowledge. If you are a researcher that matches one or more of the following criteria, then you would most definitely love to join this project!:

- Students with some experience working with sensors
- if you have experience developing a dialogue system or you want to learn how to do it. It might be that you have interest in a dialogue system pipeline as a whole or in a specific module (e.g. language understanding, language generation,...)
- if you have a psychology background or you are interested in how to implement and evaluate psychology-based models in a robot situated in a physical environment.
- if you have experience dealing with multi-modal data streams, dealing with problems like synchronisation.
- if you have programming skills and you would like to see the result of your coding effort in a robot with social skills.

References:

De Rosis, F., Pelachaud, C., Poggi, I., Carofiglio, V., & De Carolis, B. (2003). From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. *International journal of human-computer studies, 59*(1), 81-118.

Prendinger, H., & Ishizuka, M. (2001). Social role awareness in animated agents. Paper presented at the fifth international conference on Autonomous agents.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*(5), 979.

Hung, H., & Chittaranjan, G. (2010). *The idiap wolf corpus: exploring group behaviour in a competitive role-playing game*. Paper presented at the 18th ACM international conference on Multimedia, Firenze, Italy.

Dias, J., & Aylett, R. (2013). *The Great Deceivers: Virtual Agents and Believable Lies.* Paper presented at the CogSci.

Oertel, C., & Salvi, G. (2013). *A gaze-based method for relating group involvement to individual engagement in multimodal multiparty dialogue*. Paper presented at the ICMI 2013, Sydney, Australia.

Moubayed, S. A., Beskow, J., Skantze, G., & Granström, B. (2012). Furhat: A Back-projected Human-like Robot Head for Multiparty Human-Machine Interaction *Cognitive behavioural systems* (pp. 114-130).

Skantze, G., & Al Moubayed, S. (2012). *IrisTK: a statechart-based toolkit for multi-party face-to-face interaction.* Paper presented at the ICMI, Santa Monica, USA.