DELIVERABLE D3.1 PROTOTYPE

Social State Processing Components (Initial)

Mary Ellen Foster, Alessandro Vinciarelli, Amol Deshmukh, Bart Craenen (Glasgow)

Beneficiaries:	**GLA**, HWU, IDIAP, LAAS, ALD
Work Package:	WP3: Social Signal Processing

Version:	Final
Nature:	Prototype (P)
Dissemination level:	Public (PU)
Pages:	2
Date:	2017-02-28

1   Description

This deliverable describes the initial social state processing software components for year 1. The goal for this year was to establish the initial components for the robot to understand the social signals of the users. In particular, WP3 adopted the technologies resulting from the completion of tasks T2.1 in WP2.

2   Software

Two software packages have been developed. The first uses the interface exposed by the software developed under WP2, and exposes an interface used by the software developed under WP4. The second exposes an interface used by the software developed under WP4, and uses an interface exposed by the software developed under WP5. Combined the two software packages provide a layer around the software developed under WP4, thus separating it from the software developed under WP2 and WP5. See figure 1 for a diagram of the software infrastructure.

The software packages combined provide a framework for seamlessly developing and deploying social signal processing software into the MuMMER software system in a way opaque to the surrounding software packages developed under WP2, WP4, and WP5.
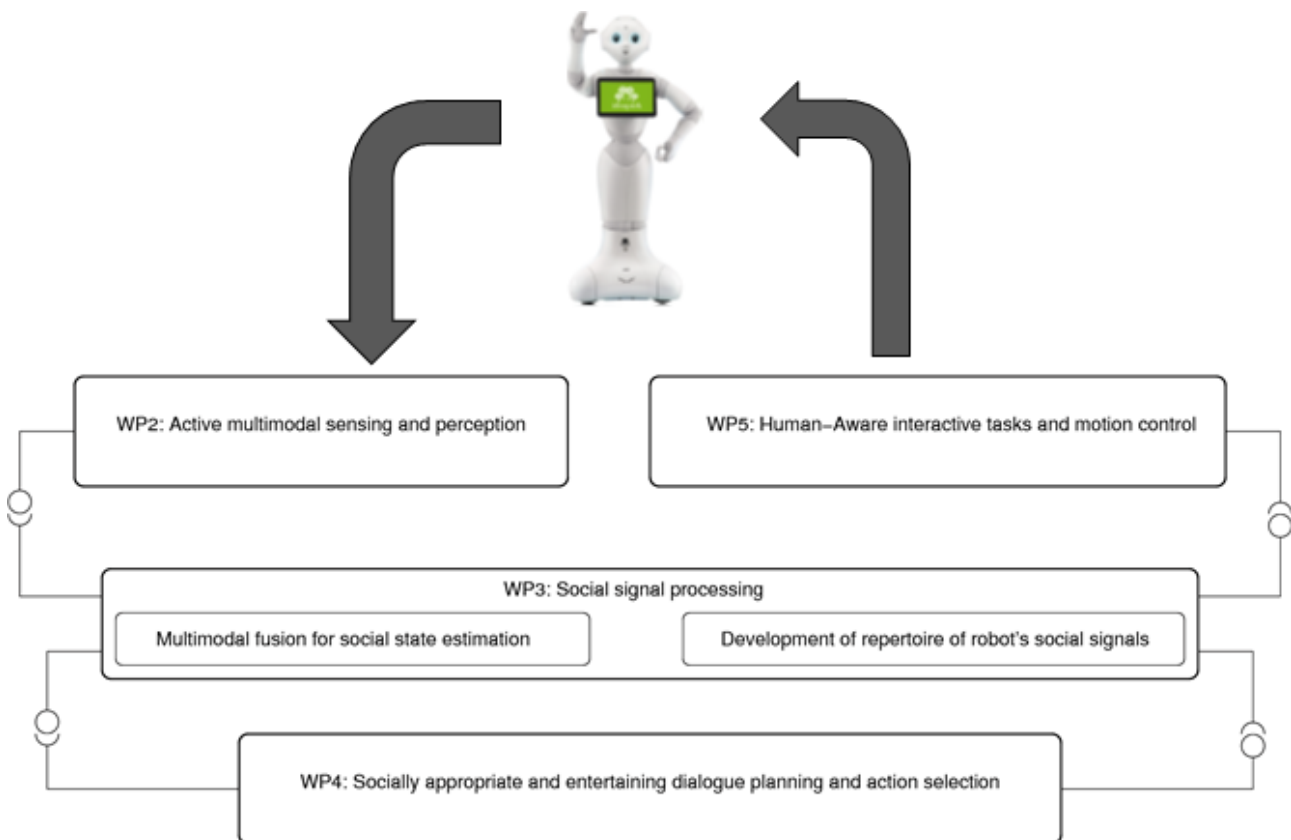

Figure 1: Software infrastructure of WP3 and surrounding software packages

2.1   Multimodal fusion for social state estimation

The software modules have been developed using the Robot Operating System (ROS) Indigo as agreed by all technical partners in the project. The module developed for year 1 in WP3, receives input from information published from WP2 modules. This module publishes messages for social signal processing. The engagement node publishes messages in the following format:

```
std_msgs/Header header
   uint32 seq time stamp
   string frame_id int32 id: User Id
   bool bAttention: Whether a user is looking at the robot, or not
   float64 dWillingnessToInteract : Probability of user looking at robot
   float64 dEngaged: Probability of user looking at robot (placeholder)
```

The messages published are bundled as an array for number of persons perceived by WP2 during that time frame. The messages published are further used by WP4 to initiate interaction with the user.

2.2 Development of repertoire of robot's social signals

The action received from WP4 is then used by the second software package to realise the generation of social signals that the robot can display (e.g., distance with respect to others, voice, body orientation, light patterns etc.) and relevant to the application scenarios targeted in MuMMER. This software interface will be implemented as per the interface defined by WP5.

3 Results

The main result obtained by the deliverable is that we have injected software packages into the MuMMER software system required to support the development of the social signal processing software for the subsequent deliverables for WP3.

The software package in between the WP2 and WP4 software is designed to play host to social signal processing methods focussed on the multimodal fusion of social state estimation (T3.1). The software package in between the WP4 and WP5 software is designed to play host to social signal processing methods used to develop a repertoire of the robot's social signals (T3.2). In combination the software packages are designed to support the design of an entertainment-oriented social strategy based on personality impressions (T3.3).

4 Outputs and Future Directions

The outputs for D3.1 consist of the two software packages developed. They provide a framework for seamlessly developing and deploying social signal processing methods, software, and techniques into the MuMMER software system.

Instructions for accessing the software packages are available through the MuMMER website at the following location: http://www.mummer-project.eu/outputs/software/

Subsequent research will focus on populating the framework described with the required social signal processing methods, to be developed as software components, as described in the workplan, milestones, and tasks for WP3. The software package in between WP2 and WP4 will play host to social signal processing methods focussed on the multimodal fusion of social state estimation (T3.1), while the software package in between WP4 and WP5 will play host to social signal processing methods used to develop a repertoire of the robot's social signals (T3.2). Combined the software packages will support the design of an entertainment-oriented social strategy based on personality impressions (T3.3).

5 Deviations

There have been no deviations from the workplan.

Perceptual Control Theory for Engagement and Disengagement of Users in Public Spaces

Amol Deshmukh, Mary Ellen Foster, Alessandro Vinciarelli
School of Computing Science
University of Glasgow, Glasgow, United Kingdom
{Amol.Deshmukh, MaryEllen.Foster, Alessandro.Vinciarelli}@glasgow.ac.uk

ABSTRACT

This paper presents Perceptual Control Theory—a model that explains behaviour as an attempt to keep sensory inputs in a desired range—and demonstrates that it can be used to develop an approach designed to make robots capable of human interaction. In particular, we present an approach that embodies the most salient features of the theory through a feedback loop. This approach has been implemented on a Pepper robot, and a preliminary experiment has been performed by deploying the robot in the entrance hall of a university building. The results show that the robot effectively engages and disengages the attention of people in 43% and 39% of cases, respectively. This result has been obtained in a fully natural setting where people were unaware of being involved in an experiment and therefore behaved spontaneously.

CCS Concepts

•**Human-centered computing** → **User models;** •**Computing methodologies** → **Computational control theory;** •**Computer systems organization** → **Robotic autonomy;**

Keywords

Perceptual control theory, social signals, social robotics

1. INTRODUCTION

This paper presents an approach—based on Perceptual Control Theory (PCT) [1]—designed to make a robot capable of interacting with people in public spaces. In particular, we present an experiment where a Pepper robot is expected to use a system based on PCT to engage people passing through a public space, and then to disengage them after delivering a short message (less than five seconds). The reason for focusing on engagement and disengagement is that these are the two fundamental steps of any interaction, so it is important to ensure that a social robot is at least capable of engaging and disengaging its users as a prerequisite for implementing more complex interactive behaviour. This work is inspired by the attention of the linguistic community to opening and closing of conversations, the two indispensable steps of any human-human interaction [2].

The key idea of PCT is that agents do not directly control their behaviour; rather, they control their perceptions of the world [3]. In other words, agents behave in such a way that sensory inputs, or at least the subset of these that is relevant to a particular setting, remain in a predefined desirable range. In this respect, PCT has two main advantages when it comes to the development of approaches aimed at controlling a robot. The first is that the most salient features of PCT can be implemented through feedback loops, processes that have been thoroughly studied and modelled from a mathematical point of view. The second is that PCT is fully materialistic: that is, it takes into account only observable, measurable and machine-detectable phenomena. This appears particularly suitable for robots that, being machines, cannot take into account non-observable inputs.

We have therefore implemented a model based on PCT to control the engagement and disengagement behaviour of a humanoid robot. We have then tested the model by deploying the robot in the entrance hall of a university building and recording whether it is successfully able to engage and disengage with passersby. The results, to be considered preliminary, show that engagement and disengagement attempts were successful 43% and 39% of the times, respectively.

The rest of this paper is organised as follows: Section 2 presents the proposed approach, Section 3 shows experiments and results and Section 4 draws some conclusions and discusses future work.

2. THE APPROACH

According to the PCT, the behaviour of agents is aimed at "maintaining some variable at or near a specified fixed value or pattern of values despite disturbances" [3]. Such a process can be modeled with a feedback loop [4] where the **system** is the robot, the **sensor** is the sensing apparatus of the robot and the **controller** is the approach adopted to verify whether the sensory input is in the desirable range (see Figure 1).

The system is the robot that displays behavioural cues. The sensor is the RGB-D camera that detects the head pose of the user. The controller checks whether the head pose belongs to a cluster corresponding to a desired behavioural response (meaning that is in the Voronoi region of its centroid). The robot behaviour is determined by the output of the controller.
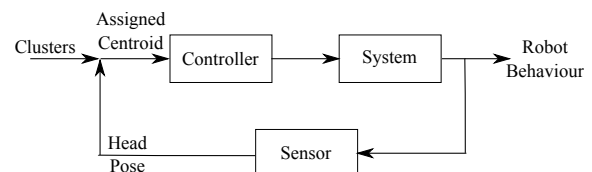

Figure 1: Feedback loop and Perceptual Control Theory.

Figure 2: Left: Pepper robot with feedback system, Right: Pepper robot with users

In the current implementation, the sensory input is a vector $\vec{x} \in \mathbb{R}^3$ where the three components account for pitch, yaw and roll of a users' head. In other words, $\vec{x}$ represents the head pose of a user (detected with a RGB-D camera mounted on the robot). During a training phase (before the actual experiments), head poses were recorded and then clustered using a k-means algorithm with $k = 3$ [5]. One of the clusters corresponds to the head poses of the people that look at the robot. Thus, during the actual experiments, it is possible to determine which cluster the head pose of a person belongs to and, hence, to check whether a person is looking at the robot or not. Such a task corresponds to the control step of the feedback loop (see Figure 1).

The robot behaviour is implemented as follows: the robot utters an expression—"hello" or "goodbye" depending on whether it is trying to engage or disengage the user—and then checks whether $\vec{x}$ belongs to the desired cluster(s): the "looking at the robot" cluster in the engagement attempt, and the other two clusters in the disengagement attempts. Whenever the head pose is in the right cluster, the attempt is considered successful. If an attempt is not successful, the robot escalates by using longer expressions (e.g., "hello there") uttered more loudly. After three attempts, the robot breaks the feedback loop and looks for another person.

3. EXPERIMENTS AND RESULTS

The above model was implemented on a Pepper robot, and a deployment study was performed in the hall of a university building. Note that the people that interacted with the robots were not aware of being involved in an experiment, and therefore acted in a fully spontaneous way and not according to a predefined script or protocol. In this respect, the experimental setup simulates real conditions in which the robot is deployed in a public space where people move and gather regardless of the robot's presence.

To avoid parasitic attention effects, the LED lights and the tablet on the chest of the Pepper robot were switched off. When a user was detected and not paying attention, the robot would try to get user's attention in 3 attempts, each attempt with progressive louder volume saying *"Hello"* (50%, 70%, and 95% of the maximum volume respectively). The robot would give up after making 3 attempts. After a user was determined to be paying attention, the robot would deliver a short welcome message (less than 5 seconds). After the message the robot tried to disengage the user by saying *"Good bye"* again in 3 attempts, using the same volume sequence. The experimental set-up is illustrated in Figure 2.

Overall, the robot was deployed for 4 hours over 2 consecutive days and made 100 attempts to engage with people with a success rate of 43% (success was determined by measuring whether people started to look at the robot in the 200 ms following the greeting

utterance). For the users that were engaged, the disengagement attempts were successful 39% of the times (i.e., people looked away). Preliminary analysis of the results show that the main reason for failure was inability in head pose detection because a person walked too fast or in a direction perpendicular to the robot.

4. CONCLUSIONS AND FUTURE WORK

This paper has presented an approach based on Perceptual Control Theory aimed at making a robot capable to engage and disengage with people walking through a public space. The approach has been implemented as a feedback loop where the system is a Pepper robot, the sensor is a RGB-D camera and the controller checks whether the head pose of a person corresponds to a desired behavioural response: looking at the robot for engagement, and looking away for disengagement. The control step of the feedback loop embodies the key idea of PCT, namely the process of keeping sensory input in a desired range.

The robot has been deployed in a real-world setting, and the preliminary results of this deployment are encouraging, and also illustrate the limitations of the approach. In particular, it is necessary to make the detection of the head pose faster in the case of people that walk with higher speed. Furthermore, it is necessary to adopt alternative strategies in the case of people that walk in a direction perpendicular to the robot. This will be the subject of future work. In addition, we will explore other robot engagement behaviours such as body gestures and the use of the tablet and/or LEDs. More generally, we plan to integrate aspects of this model into a larger-scale system designed to support autonomous and engaging human-robot interaction in the context of a public shopping mall.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] William Treval Powers. *Behavior: The control of perception*. Aldine Chicago, 1973.

[2] E.A. Schegloff and H. Sacks. Opening up closings. *Semiotica*, 8(4):289–327, 1973.

[3] G. Cziko. *The things we do*. MIT Press, 2000.

[4] S. Skogestad and I. Postlethwaite. *Multivariable feedback control: analysis and design*. Wiley and Sons, 2007.

[5] C.M. Bishop. *Pattern recognition*. Springer Verlag, 2006.