ICSR 2013

international conference on social robotics / 27-29 October 2013, Bristol, UK

COMPANIONSHIP

Workshop

On

Embodied Communication of Goals and Intentions

Organisers: Katrin Lohan and Konstantinos Theofilis

Workshop on Embodied Communication of Goals and Intentions

This workshop aims to bring together researchers from different fields working on robots that communicate with humans. The focus is on human-robot interaction and on embodied communication of goals and intentions. It is assumed that there is a strong influence of action performance, gazing behaviour, spatial arrangement and spatial flow of action to infer goals and intentions from humans. Learning and understanding in a social context should not be considered as an one-sided process. Thus, it is interesting to study situations from the perspective of both the learner's and the teacher's perspective. In this workshop, the intention is to investigate the challenges posed by such complex interaction systems from different research perspectives. Therefore, we host talks given by researchers with different backgrounds. The aim is to report on the state-of-the-art and promote the exchange of ideas on how to enable a robot to interact with a human in a more natural way so that it can directly learn from human instruction.

The organisers:
Katrin Lohan
Konstantinos Theofilis

Invited Talk

Planning for Human-Robot Interaction: Representing Time and Intention

Frank Broz, University of Plymouth

In many social tasks, it is important to reason about the intentions of others in order to coordinate behaviour when goals are shared or resolve conflicts when differing goals could lead to them. In this talk, I will describe a modelling approach that represents human-robot interactions as partially observable Markov decision processes (POMDPs) where the intention of the human is represented as an unobservable part of the state space and the robots own intentions are expressed through the rewards. The state space and transition structure for these models are designed to represent time-dependence in action outcomes and changes in the environment, which is necessary to successfully coordinate behaviour in many domains. I will present results comparing the performance of policies from these models to the performance of humans interacting with other humans for an interaction task in a simulated environment. I will also discuss potential applications this approach to other domains in human-robot interaction, including face-to-face interaction.

Bio:
Dr. Frank Broz is a research fellow at the Adaptive Behaviour and Cognition Lab at Plymouth University. His current research involves the design and evaluation of multimodal robot interfaces for eldercare as part of the Robot-Era project. His research interests also include planning for HRI and the role of mutual gaze in face-to-face interaction. He received his PhD in robotics from Carnegie Mellon University in 2008.

Invited Talk

Interaction with socially interactive robot companions
Kheng Lee Koay, University of Hertfordshire

The talk will discuss the role of embodied communication and interaction in human-robot interaction scenarios in an assistive context. Examples of research on robot companions will be presented, i.e., home companion robots meant to assist people in their own homes. The emphasis of the talk will be on modes and modalities of interaction in order to create engaging scenarios.

Bio:
Kheng Lee Koay received his B.Sc. degree in robotics and automated systems and Ph.D. degree from the University of Plymouth, U.K. in 1997 and 2003, respectively. He is currently a Senior Research Fellow at the Adaptive Systems Research Group at the University of Hertfordshire, U.K. His research interests include Mobile Robotics, Social Robotics, Robotic Home Companion, Human-Robot Interaction and Agent Migration. He was involved in several European projects Cogniron (Cognitive Robot Companion), LIREC (Living with Robots and Interactive Companions) and is currently working in the FP7 European project ACCOMPANY (Acceptable Robotics Companions for Ageing Years) and the EPSRC project Trustworthy Robotic Assistants.

Invited Talk

Communicating by Moving - Anecdotes about and Insight into Human-Robot Spatial Interaction
Marc Hanheide, University of Lincoln

Enabling a robot to move among humans is not only a question of safety, but also of non-verbal communication of intentions and goals. The spatially interacting partners (humans and robots) continuously monitor and signal these by means of motion trajectories, body orientation, facial expressions, and gaze. In my talk, I will present our research in this area covering the understanding of communicative signals, qualitative reasoning about trajectories and our ideas on long-term adaptation of navigation patterns in human-robot spatial interaction.

Bio:
Marc Hanheide is a senior lecturer in the School of Computer Science at the University of Lincoln, UK. He received the Diploma in computer science from Bielefeld University, Germany, in 2001 and the Ph.D. degree (Dr.-Ing.), also in computer science, from Bielefeld University in 2006. In 2001, he joined the Applied Informatics Group at the Technical Faculty of Bielefeld University where he worked in the European Union IST project VAMPIRE. From 2006 to 2009, he held a position as a senior researcher in the Applied Computer Science Group as a PI in the EU cognitive robotics project COGNIRON. From 2009 until 2011 he was a research fellow at the School of Computer Science at the University of Birmingham, UK, working in the EU cognitive robotics project CogX. Marc Hanheide is also a PI in a number of projects funded by the CoR-Lab and the Cluster of Excellence CITEC, Bielefeld. In all his work, he researches on autonomous robots, human-robot interaction, interaction-enabling technologies, and architectures for cognitive systems.

International Conference on Social Robotics
27-29 October 2013, Bristol, UK

Workshop 2
Embodied Communication of Goals and Intentions

Programme

9.00am
Invited Talk: Analysis of Mutual Gaze and Speech Using Automated Methods
Frank Broz, University of Plymouth

10.00am
Communicating Simulated Emotional States of Robots by Expressive Movements
Sara Baber Sial and Aleksandar Zivanovic

10.20am
Artificial Emotions to Assist Social Coordination in HRI
Jekaterina Novikova and Leon Watts

10.40am - 11.10am: Coffee Break

11.10am
WoZ Pilot Experiment for Empathic Robotic Tutors - Opportunities and Challenges[1]
Amol Deshmukh, Srinivasan Janarthanam, Helen Hastie, Shweta Bhargava and Ruth Aylett

11.25am
Train Robots - A Dataset for Natural Language Human-Robot Spatial Interaction through Verbal Commands
Kais Dukes

11.25am
Phenomena in Conveying Information during Oral Task Descriptions
Stephanie Schreitter and Brigitte Krenn

11.45am
The Development of Robot-specific Behaviour for Tour Guide Robots[2]
Daphne Karreman, Betsy van Dijk and Vanessa Evers

12.20pm - 1.30pm: Lunch Break

1.30pm
Invited Talk: Interaction with Socially Interactive Robot Companions
Kheng Lee Koay, University of Hertfordshire

---

[1] Video submission
[2] Short or position paper

2.30pm
The Use of Social Robot Ono in Robot-assisted Therapy[2]
Cesar Vandevelde, Jelle Saldien, Maria-Cristina Ciocci and Bram Vanderborght

3.00pm
Social Task Engagement - Striking a Balance between the Robot and the Task
Lee J. Corrigan, Christopher Peters and Ginevra Castellano

3.15pm - 3.30pm: Coffee Break

3.30pm
Invited Talk: Communicating by Moving - Anecdotes about and Hindsights into Human-Robot Spatial Interaction
Marc Hanheide, University of Lincoln

4.30pm
Towards Legible Robot Navigation - How to Increase the Intend Expressiveness of Robot Navigation Behaviour
Christina Lichtenhäler and Alexandra Kirsch

4.50pm
Social Navigation - Context Is King[2]
David V. Lu and William D. Smart

[1]Video submission
[2]Short or position paper

Acknowledgements

We would like to thank the support given by our supervisors (Kerstin Dautenhahn, Chrystopher Nehaniv and Giorgio Metta). Also, the ICSR organisers for their flawless cooperation. Futhermore, we would like to thank for the financial support the ICSR consortium, the University of Hetfordshire, the Italian Institute of Technology, the RobotDoc Project and the Heriot-Watt University.

Finally, we would like to thank our reviewers for their good work:

Frank Broz
Christian Dondrup
Kerstin Fischer
Sascha S. Griffiths
Marc Hanheide
Stefan Krueger
Oliver Lemon
Katrin S. Lohan
Lorenzo Natale
Iris Nomikou
Francesco Nori
Ali Paikan
Maha Salem
Alessandra Sciutti
Dag S. Syrdal
Konstantinos Theofilis
Nick Wilkinson

Gratefully,
Katrin S. Lohan and Konstantinos Theofilis

International Conference on Social Robotics
27-29 October 2013, Bristol, UK

Workshop 2
Embodied Communication of Goals and Intentions

Programme

9.00am
Invited Talk: Analysis of Mutual Gaze and Speech Using Automated Methods
Frank Broz, University of Plymouth

10.00am
Communicating Simulated Emotional States of Robots by Expressive Movements
Sara Baber Sial and Aleksandar Zivanovic

10.20am
Artificial Emotions to Assist Social Coordination in HRI
Jekaterina Novikova and Leon Watts

10.40am
WoZ Pilot Experiment for Empathic Robotic Tutors - Opportunities and Challenges[1]
Amol Deshmukh, Srinivasan Janarthanam, Helen Hastie, Shweta Bhargava and Ruth Aylett

10.50am - 11.15am: Coffee Break

11.15am
Train Robots - A Dataset for Natural Language Human-Robot Spatial Interaction through Verbal Commands
Kais Dukes

11.35am
Phenomena in Conveying Information during Oral Task Descriptions
Stephanie Schreitter and Brigitte Krenn

11.55am
The Development of Robot-specific Behaviour for Tour Guide Robots[2]
Daphne Karreman, Betsy van Dijk and Vanessa Evers

12.30pm - 1.30pm: Lunch Break

1.30pm
Invited Talk: Interaction with Socially Interactive Robot Companions
Kheng Lee Koay, University of Hertfordshire

2.30pm
The Use of Social Robot Ono in Robot-assisted Therapy[2]
Cesar Vandevelde, Jelle Saldien, Maria-Cristina Ciocci and Bram Vanderborght

---
[1] Video submission
[2] Short or position paper

3.00pm
Social Task Engagement - Striking a Balance between the Robot and the Task
Lee J. Corrigan, Christopher Peters and Ginevra Castellano

3.15pm - 3.30pm: Coffee Break

3.30pm
Invited Talk: Communicating by Moving - Anecdotes about and Hindsights into Human-Robot Spatial Interaction
Marc Hanheide, University of Lincoln

4.30pm
Towards Legible Robot Navigation - How to Increase the Intend Expressiveness of Robot Navigation Behaviour
Christina Lichtenhäler and Alexandra Kirsch

4.50pm
Social Navigation - Context Is King[2]
David V. Lu and William D. Smart

[2] Short or position paper

Communicating Simulated Emotional States of Robots by Expressive Movements

Sara Baber Sial and Aleksandar Zivanovic

Middlesex University, The Burroughs, London, NW4 4BT, United Kingdom
{s.sial, a.zivanovic}@mdx.ac.uk

Abstract. This research focuses on the non-verbal communication of non-android robots by comparing the results produced by three different emotional models: Russell's circumplex model of affect, Tellegen-Watson-Clark model and PAD scale. The relationship between the motion of the robot and the perceived emotion is developed. The motion parameters such as velocity and acceleration are changed systematically to observe the change in the perception of affect. The embodiment is programmed to adopt the smooth human motion profile of the robot in contrast to the traditional trapezoidal velocity profile. From the results produced it can be concluded that the emotions perceived by the user is the same on all three scales, validating the reliability of all the three emotional scale models and also of the emotions perceived by the user. Moreover the selected motion parameters of velocity and acceleration are linked with the change of perceived emotions.

Keywords: Nonverbal communication, Human-robot interaction, Perceived emotion, Smooth spline motion, Affective robots, Social robotics

1 Introduction

Increasingly, research is focusing on techniques whereby robots can work together with humans in order to carry out tasks [1]. The safety and effectiveness of cooperation may be enhanced by the human understanding the robot's behavior and being able to anticipate what the robot will do next. It is natural for people to perceive motion in terms of emotional behavior [2]. Nonverbal communication through motion itself contains a lot of information about the physical state and the intentions of robots [3]. The central focus of this research is to develop low level programming of movement trajectories to represent a simulated emotional state of the robot. Three gestures were programmed and were compared using three different emotional models that are discussed in this paper. The key aspect of the experiment is that the robotic embodiment used does not have significant anthropomorphic or zoomorphic features. The embodiment is programmed to adopt a smooth motion profile based on human movement characteristics, unlike the traditional trapezoidal motion used in industry which looks "unnatural" [4].

2 Modeling Machine Emotions

The perception of emotions is very subjective. There are many different models that are available for categorizing the emotions of the machine by the user. The detailed study for these models can be found in [5] and [6]. Three different models of affective emotional experiences used for visualizing the emotions are: Russell's circumplex model of affect, the Tellegen-Watson-Clark model and the PAD scale. The first two models represent emotions in 2D space. The last scale used added the third dimension for measuring the perceived emotion [7]. The perceived emotion for the same robotic motion is measured by all the three scales to approve the validity of the user's perception of emotions as well as testing the reliability of all the scales used.

3 Robot Platform and Software

The robotic embodiment used for this research is a 5 degree of freedom arm made by IGUS©, as shown in Fig. 1. This platform was chosen because it allowed low-level control of the motor trajectory which was implemented using National Instruments LabVIEW running on a PC and a real-time hardware platform, cRIO 9074 with 5 stepper drive modules (9501).

4 Selection of Gestures and Features that Affect Emotions

Three different gestures were used for observing the emotional states of the robot: basic point-to-point motion, waving of the robotic arm and "bowing down to welcome". The motion parameters used to show the change in emotional state were velocity and acceleration. Changing these parameters, the robot changed its speed, trajectory, time taken to reach same point and curvature [3]. Thus these two features appeared to be relevant for the perception of the emotional state. The choice for the values of velocity and acceleration were a subjective decision bearing in mind the limitations of the robot hardware. The set of values for acceleration and velocity in arbitrary units corresponding to three different gestures are shown below in Fig. 2.

Fig. 1. IGUS© Robotic arm used for expressing emotions

Gesture: 1 Point-point motion	
Velocity	Acceleration
250	10
800	50
2000	300

Gesture:2 Waving of robotic arm	
Velocity	Acceleration
100	15
100	5
100	1.5

Gesture:3 Bowing of robotic arm	
Velocity	Acceleration
30	30
50	50
100	100

Fig. 2. Set of values used for all three gestures (arbitrary units)

5 Recognition of Emotions

Russell's and Tellegen-Watson-Clark model for measuring emotions are divided into four quadrants based upon the range of emotions. The third scale PAD is divided into three ranges to measures the overall perceived effect of emotion. Moreover pleasure, arousal and dominance are also measured individually at three different levels of low, medium and high.

6 Methodology

The 18 participants who took part in the study were shown the three variations of each of the three motions and were asked to mark on each of the model graphs the characteristic they perceived. Examples are show in Figure 3.

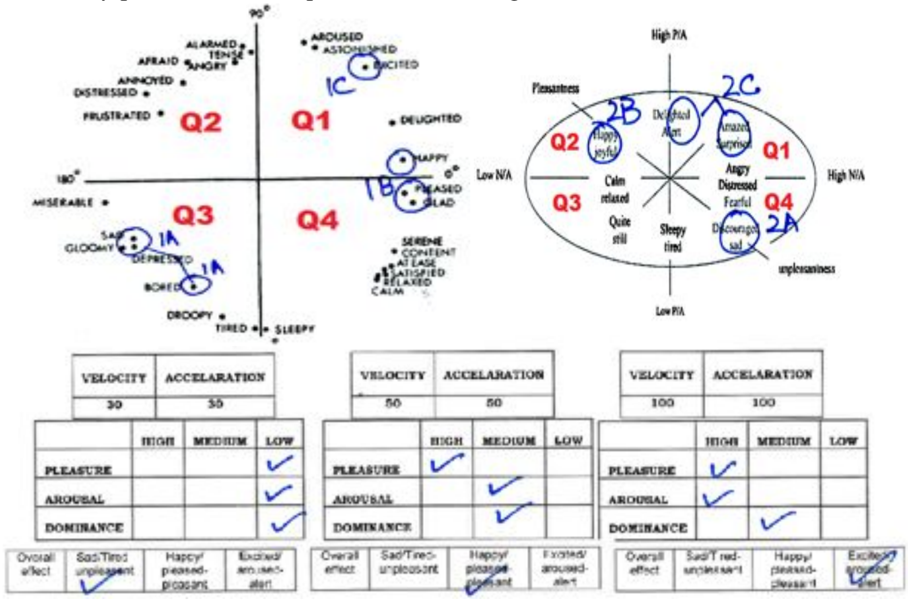


Fig. 3. Sample of marked questionnaires by the participants for all three scales

7 Results

7.1 Low velocity and low acceleration

For point-to-point motion at low values of acceleration and velocity, 95%, 83% and 100% of people perceived the emotion as sad, unhappy or tired according to Russell's, Tellegen-Watson-Clark and PAD model respectively. Similarly for waving of robot 83%, 83% and 95% of participants marked the motion as same for the three scales respectively. The same holds true for bowing gesture i.e. 83%, 72% and 100% of people marked same emotions for all three scales respectively.

7.2 Medium velocity and medium acceleration

At medium values of acceleration and velocity, the majority of subjects observing the point-to-point motion i.e. 39%, 61% and 61% of people according to Russell's, Tellegen-Watson-Clark and PAD model respectively, marked the emotions as happy, pleased or calm. Similarly for the waving gesture, 39%, 61% and 78% of participants marked the same perceived emotion for all the three scales respectively. The results were similar for the bowing gesture i.e. 83%, 72% and 100% of people marked same emotions for all three scales respectively.

7.3 High velocity and high acceleration

When the values of acceleration and velocity were raised, it was found that for point-to-point motion 67%, 50% and 67% of participants according to Russell's, Tellegen-Watson-Clark and PAD model respectively marked the emotions as excited, alert, aroused or surprised. Similarly for waving of robot 72%, 67% and 78% of participants marked the motion as same for the three scales respectively. The results are similar for the bowing gesture i.e. 61% of people according to all three scales marked the same perceived emotion.

8 Discussion of results

From the results, it can be seen that the selected motion parameters of velocity and acceleration are linked with the change in the perceived emotions. According to each model for all the three gestures i.e. point-to-point motion, waving of the robotic arm and bowing down to welcome, the majority of the participants perceived the motion to be sad, unhappy or tired if the velocity and acceleration were low. As the acceleration and velocity increased, the perceived emotion changed to happy and then excited.

9 Conclusion

A link was developed between the change in user perception of emotions and the variation of the motion parameters of velocity and acceleration [8]. Moreover, a robotic embodiment without any android features was capable of conveying emotions to the user. The robot's velocity profile at five different joints closely resembled the human arm trajectory profile as shown by curves obtained by LabVIEW in Fig.4, Fig.5 and Fig.6. For sad/unpleasant emotions the spline is distributed on the graph, for pleasant/happy emotion it is contracted with an increase in amplitude and for excited/alert emotion, amplitude is high and the splines have shown contraction on the graph**.** This profile can be compared with the velocity profile trajectories produced by human arm [10].
It was observed that the audible noise produced by the robot changed with change of emotional behavior. Research (e.g. [9]) has shown that sound is connected to emotional expression. When the robot was perceived to be sad or unhappy the noise associated with it was very low. However as the perceived emotion changed from sad to happy and then to excited as the noise associated with robotic embodiment increased.

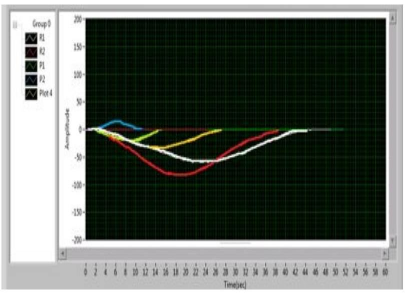


Fig.4. Splines at V=25

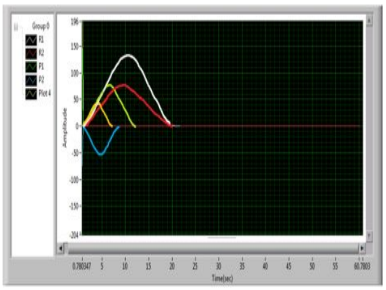


Fig.5. Splines at V=800

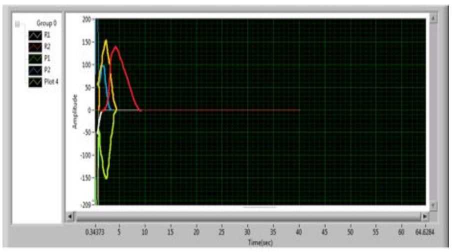


Fig.6. Splines at V=2000

10 Limitations and Future Work

It is important to highlight that the gestures were deliberately selected to be expressive. However it was important from the aspect of developing a movement that should be expressive and communicative to the user [11]. This might have had an effect on the results found in this research.

Another potential bias associated with this robotic embodiment is the noise that it makes during its motion (the sound from the stepper motors). This might help the user to identify the perceived emotions.

The experiments performed for this research were preliminary. Clearly, more detailed and varied experiments should be carried out. For instance, comparing the spline motion with the traditional trapezoidal trajectory generation or by repeating the experiments with more gestures and emotions. Moreover, this research could be performed on an anthropomorphic robot to see if the perception of emotions differs according to the type of robot.

This research into the emotional behavior of the robot gives rise to several questions that remain to be answered e.g. in the field of care and medication, is slow motion perceived as a sad gesture or a careful gesture by the patient? For industrial purposes can these emotional robots have the same efficiency and productivity rate as the ones used now?

References

1. TUM Cognition for Technical Systems (CoTeSys) - Munich Center for NeuroSciences - Brain and Mind - LMU Munich http://www.mcn.uni-muenchen.de/affil_research/centers/cotesys/index.html
2. Heider, F., Simmel, M., An Experimental Study of Apparent Behaviour. The American Journal of Psychology 57, 243 (1944)
3. Saerbeck, M., Bartneck, C., Perception of affect elicited by robot motion. ACM Press, p. 53 (2010)
4. Wu , W., Zhu, S. & Liu, S., 2009. Smooth joint trajectory planning for humanoid robots based on B-splines. In IEEE International Conference on Robotics and Biomimetics (ROBIO). Guilin, pp. 475 – 479.
5. Lewis, M., Jones, J., Barrett, L., Handbook of emotions, 3rd ed. 2008 The Guilford Press (2008)
6. Barrett, L.F., Mesquita, B., Ochsner, K.N., Gross, J.J., The Experience of Emotion. Annu Rev Psychol 58, 373–403 (2007)
7. Lang, P.., Bradley, M., Cuthbert, B., International Affective Picture System (IAPS): Technical Manual and Affective Ratings (1997)
8. Ian , B., Debayan , G. & Fei , H., 2005. Effects of low level changes in motion on human perception of robots. In. 4th International Conference on Development and Learning. pp. 1–5.
9. Eun, S.J. et al., 2009. Sound Production for the Emotional Expression of Socially Interactive Robots, Advances in Human-Robot Interaction, pp. 342
10. Atkeson, C.G., Hollerbach, J.M., Kinematic features of unrestrained vertical arm movements. J. Neurosci. 5, 2318–2330 (1985)
11. Beck, A., Hiolle, A. & Cañamero, L., 2013. Using Perlin Noise to Generate Emotional Expressions in a Robot. CogSci 2013, pp.1845–1850.

Artificial Emotions to Assist Social Coordination in HRI

Jekaterina Novikova, Leon Watts

Department of Computer Science
University of Bath
Bath, BA2 7AY
United Kingdom
j.novikova@bath.ac.uk

Abstract. Human-Robot Interaction requires coordination strategies that allow human and artificial agencies to interpret and interleave their actions. In this paper we consider the potential of artificial emotions to serve as coordination devices in human-robot teams. We propose an approach for modelling action selection based on artificial emotions and signalling a robot's internal state to human team member. We describe an architecture that drives the display of artificial emotional gestures with a model of latched internal emotional states. We also present preliminary data on human recognition rates for a candidate set of artificial emotional expressions in a Lego robot.

Keywords: artificial emotions, action selection, human-robot collaboration

1 Introduction

Robots could act as members of a human team by assisting people who share a given physical workspace, by performing action relevant to their joint goals. Research on human-robot interaction (HRI) must address a number of challenges to make coordinated action possible. Robots must act in a way that is understandable to the people with whom they are working, through the way they move and interact with objects in the shared space. In this paper we consider the potential of artificial robot emotions to serve as coordination devices in human-robot teams. In order to benefit coordination, artificial emotions should first of all be clearly expressed in a way comprehensible for humans. At the same time, for robot emotion to function effectively in human interactions, it is necessary to consider the internal relevance of the emotional state for the robot's operation so that intelligible mappings can be made to a set of signals for the robot's human partner. Without this step emotion is unlikely to serve interactions well. The general aim of our work is to try to find a general way of communicating internal robot's state to humans in a way, which is both meaningful and natural for humans and thus insure a successful social coordination between human and robot. In this paper we propose a model for incorporating artificial emotions and emotional expressions into an emotionally-triggered action selection mechanism based on Behaviour Oriented Design.

Visual cues such as facial expressions are important in human-human coordination because they assist people to make inferences about one another's task-relevant state. For example, a grimace might indicate difficulty or a smile may suggest some success. Knowledge of this kind can help co-workers to bring their actions together at particular points, or to reschedule or reallocate work in case of difficulty. Research on emotion recognition, expression, and emotionally enriched communication is of great potential importance to HRI and has been the subject of significant research effort since the mid-1990s [1],[2],[4],[5],[8]. Most of the existing work in social and humanoid robotics focuses on the recognition of human emotions or mimicking their expression [1],[5]. However, from an interaction perspective, understanding of social cues and a social context should not be considered as a one-sided process. In addition to understanding human emotions, more work should be done on the role of artificial emotions in influencing human behaviour in human-robot teams and their impact on interaction.

Most of the state of the art research is focusing either specifically on the expression of artificial emotions or on emotionally-based action selection. Our approach, however, focuses on coupling a subsystem that generates robot's internal affective state with a social signalling subsystem. We argue that this approach can assist in communicating the intent of a robot to a human during interaction by creating meaningful expectations of actions associated with specific emotional expressions, like in human-human coordination.

This paper is organized as follows. As a starting point we present the framework for modelling artificial robotic emotions. It is followed by the description of the performed study aimed to understand how a non-humanoid robot should express artificial emotions in the way understandable for a human. Details of the study are given together with the results and analysis. Finally, we conclude with a discussion of the results and suggest both implications for HRI and directions for further work.

2 Method

Artificial emotions in this study are represented as a factor for dynamic action selection. A mechanism for generating an internal emotional state is used in conjunction with a selection mechanism for visual cues that are intended to communicating the current emotional state to a human. The artificial emotion system is designed to run concurrently with other robot subsystems, such as planning, learning and signal processing. Emotional states are thus continuously computed and can drive the production of emotional signals before and during the execution of actions [6].

2.1 Modelling Artificial Emotions

The framework for modelling artificial robotic emotions is presented in Figure 1. The first phase of the emotional action selection includes detecting specific internal and/or external conditions (C1, C2 and C3 in the Figure 1), following Breazeal [1]: presence of an undesired stimulus, presence of a desired stimulus, a sudden stimulus, delay in achieving goal. For determining an appropriate emotional state we use a simple valence-arousal representation for modelling basic emotional states, in a manner analogous to Russell's approach [8]. Here, *arousal* (EA in Figure 1) represents the strength of a stimulus, and the *valence* (EV in Figure 1) shows a positive/negative value of a stimulus.

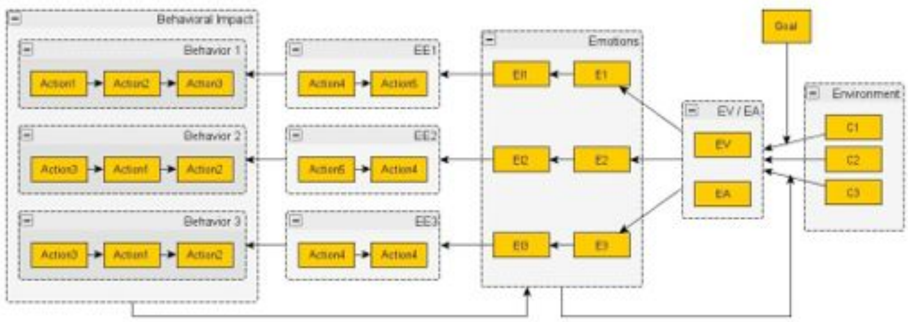


Fig. 1. The framework for modelling artificial emotions in robot.

All the detected conditions influence both valence and arousal values and thereby a robots emotive response (E1, E2 and E3 in Figure 1). We also use intensity (EI1, EI2 and EI3 in Figure 1) as an additional property of an emotion. Emotional intensity in this model is an internal state of an agent, which is changed dynamically while robot is experiencing an emotion. Intensity depends on time, number of detected stimuli, and an impact factor of an executed behaviour. Each emotion calls a specific behaviour of a dynamic plan. We use an impact factor (Behavioural impact in Figure 1) as a property of a behaviour that depresses the intensity of the emotion this behaviour was triggered by. While the selected behaviour is being executed it inhibits the intensity of the emotion it was triggered by, i.e. intensity of an emotion is a function of a behavioural impact over time. 'Feeling' an emotion is modelled as a latched process [7], during which an intensity of the emotion is increasing over time from zero value until the maximum threshold of 100, and is reducing back to zero after the executing behaviour inhibits it.

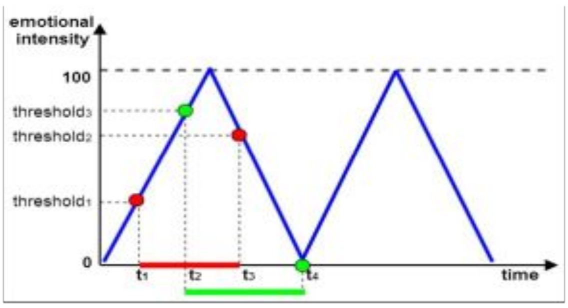


Fig. 2. Latched process of 'feeling' an emotion.

The expression of emotion starts after an increasing intensity of the emotion reaches the specified level *threshold1* and stops when the specified level *threshold2* is reached while the intensity is decreasing, as shown in Figure 2. The red line shows the time period while the emotion is being expressed. The execution of the selected behaviour starts when the intensity of an emotion reaches *threshold3*. The execution of behaviour, if not interrupted, stops when intensity of the emotion is zero. The green line indicates the period of time while the selected behaviour is being executed.

The execution of the selected behaviour starts when the intensity of an emotion reaches a specific level which is above the level of the start of expressing the emotion and below the maximal intensity level. The execution of behaviour, if not interrupted, stops when intensity of the emotion is zero. For managing interruptions, the following model is used: if interruption happens when emotion intensity is below the *threshold1* level the behaviour stops, otherwise the behaviour is resumed. A latched process of emotional intensity helps the system not to get "stuck" swapping rapidly back and forth between two emotions, thus solving a common problem in other behaviour-based architectures. There is always a delay between the expression of an artificial emotion and the initiation of a behaviour it selects. Such a delay serves two important purposes: 1) this presents a co-worker with the opportunity to infer its state and potential next action in relation to their own actions, and to adjust their work accordingly, 2) it has a role of an emotional 'memory' and makes the system more robust.

2.2 Expressing Artificial Emotions

We are planning a series of studies to better understand whether a non-humanoid robot can express artificial emotions in the way understandable for a human. As a first step, we have been experimenting with a Lego robot using two basic movements of its body: moving the 'neck' forward/backward, and raising/lowering 'eyebrows'. The inspiration for this simple scheme is drawn from our basic arousal-valence underlying model, with approach and avoidance of the neck as a metaphor for valence and eyebrows reflecting the arousal concept. We programmed six combinations of these two movements and then photographed them from two angles – front and ¾ views, as shown in Figure 3:

Fig. 3. Lego robot, expressing artificial emotions using a combination of two basic movements of its body: moving the 'neck' forward/backward, and raising/lowering 'eyebrows'.

The six pairs of pictures were used to construct a questionnaire. 27 people (14 females and 13 males) agreed to participate in a study to determine whether our simple set of valence-arousal robotic gestures could be interpreted as emotional signals. 18 had no previous experience with any kind of robots, 4 considered themselves as roboticists, and the rest had some previous interaction experience with robots. 18 were over 40 years old, 3 were between 30 and 39 years old, and six were between 20 and 29 years old.

For each pair of images, participants were asked to select the most appropriate emotional term from a set of possible responses: *sadness, happiness, anger, surprise, excitement, fear, other, no specific emotion* and *don't know***.** They were also asked to use a five-point Likert scale (*very easy, easy*, *neutral*, *difficult* and *very difficult*) to rate their degree of confidence making that judgement.

3 Preliminary Results

Our preliminary data suggest that our simple two-dimensional robot movements can be interpreted by people as expression of several basic emotions – sadness, happiness, anger, surprise and excitement. However, judgements of sadness and surprise were made most consistently participants, with a few alternative interpretations. For example, the robot captured in the picture No3 was interpreted as being surprised by 14 participants, while none of alternative interpretations collected more than 6 votes (Figure 4). Most participants rated their confidence in judging the emotional meaning of the robot images as *Easy* (Figure 5), with no clear indication that some were harder to interpret than others.

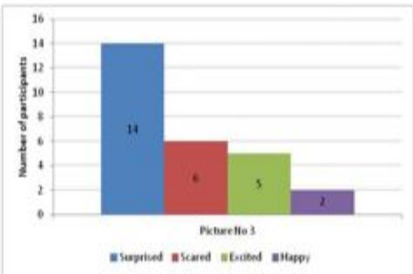


Fig. 4. Expression No3 was interpreted as 'surprise' by most users.

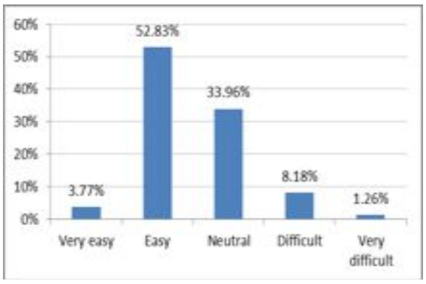


Fig. 5. Confidence of interpreting emotions, expressed by a Lego robot.

4 Discussion and Future Research Direction

The robot signals, treated as artificial emotional expressions, were recognized quite easily by most participants. However, although the questions where users were asked to evaluate the certainty of detected emotions were mixed with other questions in the questionnaire, some of their answers were influenced by a central tendency bias, i.e.

users avoided extreme ratings. The questionnaire is going to be improved for the future studies in order to avoid this problem.

The preliminary data shows that people find it easy to ascribe emotional states to the robot they see in the picture. Several pictures of the robot were interpreted with less noise and these were the pictures where users associated the robot with the emotions of *surprise, sadness* and *fear*. Some of the results correspond to previous research in human facial expressions recognition, such as e.g. raised “eyebrows” expressing surprise or lowered “eyebrows” expressing sadness [3]. We can suggest that high level of arousal (as in *surprise* and *fear*) or high level of valence (as in *sadness*) helps people to interpret artificial emotions expressed by robot more easily. However, more studies are required to check this suggestion, especially in a dynamic context.

The architecture proposed earlier has been implemented and is currently being validated by tests. We plan to further investigate our model of emotional action selection for a human-robot collaborative task where the robot must solicit assistance to achieve its goal. We intend to investigate joint action as a function of three conditions: emotional action selection with and without expression and emotion free action selection. Our observations will help us to understand whether emotional communication will empower the robot to influence a human’s behaviour, and how social coordination in HRI may be helped or hindered as a consequence.

5 References

1. Breazeal, C.: Emotion and sociable humanoid robots Int. J. Hum.-Comput. Stud. In Applications of Affective Computing in Human-Computer Interaction, Vol. 59, No. 1-2, pp. 119-155 (2003)
2. Bryson, J., Tanguy, E.: Simplifying the Design of Human-Like Behaviour: Emotions as Durative Dynamic State for Action Selection, in The International Journal of Synthetic Emotions,1(1):30-50 (2010)
3. P. Ekman. Facial expression and emotion. *American Psychologist*, 48:384–392, 1993.
4. Gonsior B., Sosnowski S., Buss M., Kuhmlenz K.: An Emotional Adaption Approach to increase Helpfulness towards a Robot, Proceeding of International Conference on Intelligent Robots and Systems, Vilamoura, Algarve, Portugal (2012)
5. Kawamura K., Wilkes D., Pack T., Bishay M., Barile J.: Humanoids: future robots for home and factory. In: Proceedings of the First International Symposium on Humanoid Robots (HURO96). Tokyo, Japan, pp. 53-62. (1996)
6. Novikova J., Gaudl D., Bryson J.: Emotionally Driven Robot Control Architecture for Human-Robot Interaction, Proceedings of Towards Autonomous Robotic Systems (TAROS-13), Oxford, UK (Accepted, 2013)
7. Rohlfshagen, P, Bryson, J.: Flexible Latching: A Biologically-Inspired Mechanism for Improving the Management of Homeostatic Goals. In: Cognitive Computation, Vol. 2, No. 3, pp. 230-241 (2010)
8. Russell, J.: Reading emotions from and into faces: resurrecting a dimensional contextual perspective. In: Russell, J., Fernandez-Dols, J. (Eds.), The Psychology of Facial Expression. Cambridge University Press, Cambridge, UK, pp. 295-320. (1997)

WoZ Pilot Experiment for Empathic Robotic Tutors: Opportunities and Challenges

Amol Deshmukh, Srinivasan Janarthanam, Helen Hastie, Shweta Bhargava, and Ruth Aylett

School of Mathematical and Computer Sciences
Heriot-Watt University, Edinburgh UK
a.deshmukh@hw.ac.uk

Abstract. We discuss the challenges and opportunities in building empathic robotic tutors based on a preliminary Wizard-of-Oz (WoZ) pilot study. From the data collected in this study, we identify situations where empathy in a robotic tutor could have helped the conversation between the learner and the tutor. The video presented with this paper captures these situations where two children participants are interacting with a map application and a robot tutor operated by a wizard.

1 Introduction

Wizard-of-Oz frameworks have been used in several studies [1] in order to collect human-computer dialogue data to help design dialogue systems. WoZ systems have been used to collect data to learn [2] and evaluate dialogue management policies [3]. The main objective of this pilot WoZ experiment was to collect multimodal data namely video, audio, user-wizard interaction to help understand the requirements for building an artificial embodied intelligent tutoring system to engage in *empathic* interactions.

The WoZ setup described in [4] comprised of the wizard interface, interactive touch table with map application, cameras and the robot. The participants aged 8-10 had to solve a treasure hunt map-reading activity and follow the tutor's instructions in a step-by-step manner. In this paper[1], we give a preliminary qualitative analysis of the pilot data gathered to inform requirements for an empathic tutor.

2 Opportunities and Challenges

There are clear challenges involved in such a WoZ data gathering experiment. In this section, we describe these challenges and discuss lessons learned and requirements going forward. Firstly, when the map application faltered (i.e. zoomed in,

[1] This work was partially supported by the European Commission (EC) and was funded by the EU FP7 ICT-317923 project EMOTE. The authors are solely responsible for the content of this paper and video

became non-responsive, etc), users looked frustrated. They seemed helpless and did not know how to proceed until there was some other form of intervention to reset the application.

Secondly, on occasion there were circumstances when a human tutor could have easily pointed out where some of the map features are (e.g. when finding the train station, etc) or directions when users are confused. However, it is challenging for a robot to do so using its arms. This presents us with an opportunity to utilise multi-modal outputs through the application running on the touch-table, for example pointing out map features by overlaying shapes such as circles, bounding boxes and arrows on top of the map.

Thirdly, response times of the robot (i.e. wizard) were perceived as too long as evidenced by the children's "blank" expressions after giving an answer and waiting for a response. It is important to intervene quickly when the user is about to make poor choices (such as walking in the wrong direction or looking in a totally different zone for answers). This presents us with the challenge of effective turn management wherein the tutoring system needs to decide how to stall during diagnosis (for examples using backchannels or encouragement), which dialogue move to select and when to intervene by continuously monitoring the state of the map application.

3 Conclusion and Future work

The scenarios described above present the tutor with opportunities to be empathic and help the learner to handle difficult situations while staying inside the zone of proximal development. It has been shown through this initial study, evidenced in the video, that key to this empathic behaviour is responsiveness and expressivity of the robot tutor.

Future work includes a full WoZ experiment whereby the data will be used to understand how human tutors, through a robotic interface, adapt to learners' emotions and cognitive states in tutorial tasks. The intention is then to use these data to learn appropriate pedagogical moves and dialogues strategies for an autonomous empathic agent.

References

1. Fraser, N.M., Gilbert, G.: Simulating speech systems. Computer Speech and Language **5**(1) (1991) 81 – 99
2. Strauss, P.M., Hoffmann, H., Scherer, S.: Evaluation and user acceptance of a dialogue system using Wizard-of-Oz recordings. In: 3rd IET International Conference on Intelligent Environments. (2007) 521–524
3. Cuayahuitl, H., Kruijff-Korbayova, I.: An Interactive Humanoid Robot Exhibiting Flexible Sub-Dialogues. In: Proceedings of The 2012 Conference of the NAACL-HTL: Demo Session. (June 2012)
4. Bhargava, S., Janarthanam, S., Hastie, H., Deshmukh, A., Aylett, R., Corrigan, L., Castellano, G.: Demonstration of the Emote Wizard of Oz Interface for Empathic Robotic Tutors. In: Proceedings of SIGdial. (2013)

Train Robots: A Dataset for Natural Language Human-Robot Spatial Interaction through Verbal Commands

Kais Dukes

Institute for Artificial Intelligence, School of Computing
University of Leeds
sckd@leeds.ac.uk

Abstract. Developing intelligent robotic systems that can robustly understand verbal spatial commands requires large datasets for accurate training. Previous work has been limited by using examples of only several hundred utterances. This paper presents Train Robots (http://www.TrainRobots.com), a new online game with a purpose that has collected over 5,000 commands through crowd-sourcing. In the game, players are shown before and after images of a board with blocks and a simulated robotic arm. Participants are asked to enter commands to instruct a hypothetical robot to rearrange blocks to match subsequent images. To promote high-quality data, players vote for each others' commands. We describe the design of the collaborative game and compare the different methods used by players to instruct the robot to manipulate its spatial environment.

Keywords: Human-robot interaction, natural language, spatial reasoning, robot commands, machine learning dataset, crowdsourcing, game with a purpose.

1 Introduction

For a robotic system to naturally and effectively interact with humans, understanding commands given in natural language is essential. Over the past decade, state-of-the-art natural language processing components such as statistical parsers [1, 2] and semantic taggers [3] have been applied to robotic systems with increasingly promising results. However, most of these experiments have been performed on a small scale with limited amounts of data. In contrast, the datasets used to construct machine learning systems for classical problems in computational linguistics such as statistical parsing are much larger, with some treebanks at over a million words in size used for training [4]. However, no datasets of a comparable size exist for training integrated robotic systems that utilize natural language processing components.

This paper proposes the construction of a new large-scale dataset to address this challenge, focusing on a spatial task. The dataset consists of 5,000 commands (77.3K words) with corresponding images. Due to its nature as a free game, it is expected that the resource will grow significantly over time. Once completed, the resulting data will be made freely available to researchers, to encourage the development of intelligent systems that can perform spatial understanding with comparable accuracy to humans.

International Conference on Social Robots – Embodied Communication of Goals and Intentions Workshop

2 Previous Work

Data for building language understanding systems has been collected for a variety of robotic systems, such as the RoboCup corpus, consisting of 300 examples of coaching advice given to embodied football robots [1]. However, although used to train natural language understanding systems, these sentences were annotated in English after the event, based on instructions given to robots in an alternative machine representation. In more recent work, Amazon Mechanical Turk has been used to collect commands for a simulated forklift robot by annotating actions in video data. This dataset is also not large-scale, consisting of only a few hundred example sentences [3].

In contrast, much larger datasets have been successfully constructed by human annotators for non-robotic linguistic work. Phrase detectives is one such effort [5]. Presented as a game with a purpose, over one million words of text have been annotated by volunteer players for pronoun anaphora resolution. Players are rewarded through a points system, and compete to obtain the highest scores. Train Robots is inspired by this and other games with a purpose such as Google Image Labeler [6]. In contrast to paid annotation projects, collecting data through free online gaming has the benefit that it is open-ended and highly scalable, with the potential to accrue much larger volumes of data over time without incurring on-going development costs.

3 Game Design

The eventual aim of the Train Robots dataset is to construct a robot able to robustly understand spatial commands. It is envisioned that this robot will learn from the dataset once this reaches a significant size. For this reason, the game has been designed to allow for automatic evaluation of the accuracy of robotic commands, by using a discrete simulated spatial environment (Figure 1).

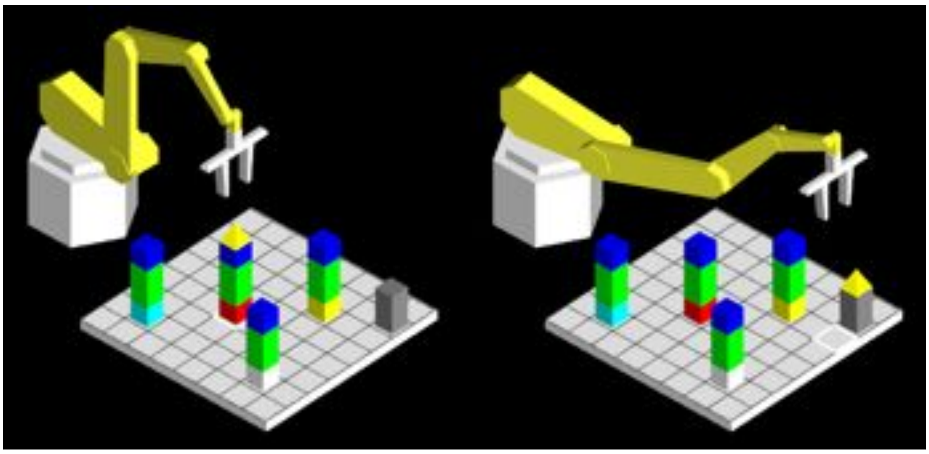


Fig. 1. An example board shown to players consisting of before (left) and after (right) images.

3.1 Environment

The robot's simulated environment consists of an 8 x 8 board that can hold prisms and cubes. These blocks occur in eight different colors, chosen for their contrast (red, green, blue, yellow, cyan, magenta, gray and white). The game also has a basic set of implicit physical rules. Blocks can either be supported by the board, the robot arm or by other blocks, so that a block left midair will fall. However, prisms cannot support other blocks so that these must be placed on the board, or on top of cubes. The robot's gripper can move to any discrete position within an 8 x 8 x 8 space above the board. The lengths and angles of the arm's joint segments have been designed to allow the gripper to pick up blocks in nearly all configurations without collision. The images of the environment shown to players were rendered offline using Java with OpenGL, with the arm based on a 3D model developed for the game using the free version of Google SketchUp. The game utilizes a pool of 1,000 image pairs that were manually developed offline. These board configurations are not random. Instead they have been purposefully designed to include a variety of examples of spatial ambiguity, with subtle differences in layout that are challenging to describe verbally.

3.2 Game Flow

To uniquely identify participants, new players are asked to register before joining by providing an email address and a game password. As well as sign-in credentials, timings of actions within the game and IP addresses are recorded to detect duplicate logins. This feature was added after a small minority of players initially logged in to the game using the same account on multiple different machines.

Once users have signed in, their previous score is displayed and the game indefinitely repeats a three-round cycle. For incentive, as well as to provide examples of accurate commands, the first two rounds of each cycle are voting rounds, in which players rate previous commands from 1 to 5. Points are awarded for accurate voting compared to the majority. Similarly, players who enter high-quality commands gain bonus points when their command is voted as accurate. In the third round of each cycle, players enter new commands for a random image pair selected from the 1,000 pre-constructed board configurations. The cycle repeats until the player logs off, with voting scores and commands saved to a database.

4 Crowdsourcing Accuracy

Data accuracy is a common problem for crowdsourcing projects. For Train Robots, an accurate command is one that is of 'gold standard' quality and of potential use for robot training. Accuracy can be measured either by using the in-game player voting metric, or by comparing to an expert player. Although crowd voting is useful for incentivizing players within the game, it is subject to a similar set of errors as crowd annotation when only a few players vote. This section focuses on the second approach. An expert reviewed a random sample of 1,000 commands, and judged each command as accurate only if it correctly described the move between the two boards.

Before the game was made available to the general public, five annotators were paid to initially seed the database, each playing for 900 rounds and entering 300 commands (a total of 1,500 commands). This section evaluates the paid commands together with over 3,500 further commands entered by 92 volunteers. Table 1 below presents the results of the evaluation process, as judged by an expert player. Only 69% commands on average were considered valid for the corresponding pair of images. However, the average time taken to enter commands as well as command accuracy was found to be consistent between the two groups of players, demonstrating that unpaid volunteers playing for enjoyment perform as effectively as paid annotators.

A manual review across both groups found the main reasons for invalid commands included: confusing the two images (68% of all invalid commands), ambiguous commands (12%), incorrect perspective (7%) and typos or missing words (5%). Surprisingly, many players confused the before and after images. For example, for Figure 1, a common mistake is a command similar to 'Move the yellow pyramid on the gray tower to the tower nearest the center of the board'. Although the game has each image clearly labelled as before or after, once players identify image differences they typically enter their command quickly and move to the next round, possibly leading to this confusion. A future improvement could be to replace static images with videos. Ambiguous commands are another challenge to data quality, such as 'Move the yellow block to another tower' for Figure 1. Over-generalized commands such has these are difficult for automatic evaluation, as a command with several possible outcomes may require a human to decide if a robot carried out the task correctly. The third most frequent source of error was incorrect perspective, in which a small minority of players used left and right to mean their perspective and not the robot's.

The evaluation also found examples of typos judged too inaccurate for gold-standard data. Interestingly, nearly all commands entered by players lack punctuation or capitalization, as players prefer to play as rapidly as possible. In addition, determiners are often dropped, such as in 'Put yellow block on gray block'. Abbreviated commands are judged as accurate, as a robust robot would be expected to understand these. However, despite brevity, the average command length was 15.47 words, as most players provided detailed spatial descriptions to the hypothetical robot to convey the intended meaning of the move between boards. Other mistakes included spelling errors such as 'reb' for 'red', as well as more serious errors such as missing words, leading to commands that were not understandable even to expert players.

Metric	**Commands by paid players**		**Commands by volunteers**	
	Valid	Invalid	Valid	Invalid
Accuracy	69.5%	30.5%	69.2%	30.8%
Average time	29.2 seconds	28.8 seconds	29.8 seconds	29.6 seconds

Table 1. Accuracy measured by an expert and average time taken for entering commands.

5 Resolving Spatial Ambiguity

Two common methods used by players to resolve spatial ambiguity within the game are referring to spatial context and assuming a shared knowledge of game physics. For the image pair in Figure 2 below, players are required to instruct the robot to move a specific green prism. The method employed consistently by all players for this scene was to ask the robot to move the prism laying *between* the cyan and magenta blocks (although often with alternative color names). For this scene, the obvious spatial description strategy is to use landmarks as spatial context, a technique commonly used to ground objects in natural language [7]. Players of the game assumed the hypothetical robot to have a similar level of spatial awareness.

For Figure 3, all players assumed a command such as 'move the gray block to the nearest blue block' would be unambiguous, as the robot would understand that choosing a gray block from within a stack would be unphysical in a single move. Because the game is a simplified simulation of a real three-dimensional scene, players assume that the rules of the game are based on real physics. It would be expected that humans describing their intentions to a embodied version of this robot would also assume it to have an understanding of the physical rules of its environment.

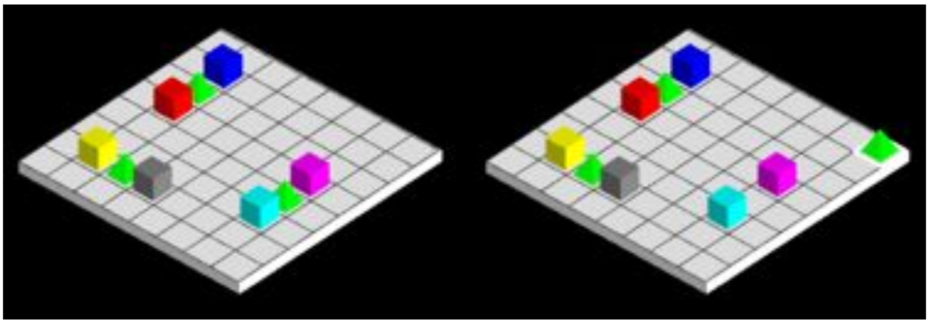


Fig. 2. Using spatial context to move one of three possible green prisms.

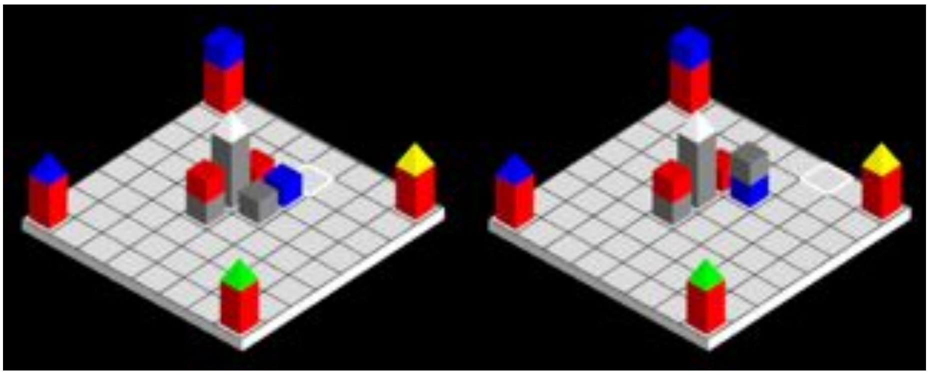


Fig. 3. Knowledge of game physics for choosing a gray block.

6 Conclusions and Future Work

This paper presented Train Robots, a new game with a purpose for annotating robot commands. At present, the dataset consists of 5,000 commands (77.3K words), of which an estimated 70% is considered to be of sufficiently high-accuracy to be suitable for training machine learning systems. Future work is planned to include improving the data collection task as well as using the data to train a robotic system.

A more interesting version of the data collection task would be interactive turns via human-human interaction, so that two players incrementally annotate a sequence of moves. It is likely that an incremental version would involve shorter commands that are refined by players in later turns, as well as provide better examples of more complex spatial language [8]. It is also planned to relate the resource to other spatial corpora [9], by re-annotating related spatial scenes in the context of robotic commands.

Finally, it is planned to use the dataset to develop a spatial planner and a semantic parser. It is envisioned that these two computational components would form the core of a natural language understanding system. A planner would execute commands specified in a formal robot control language (RCL), while a semantic parser would translate natural language (NL) into RCL statements. Planned work towards this involves annotating the NL commands in the dataset with a formal semantic representation, to enable machine learning for joint spatial grounding and semantic parsing.

References

1. G. Kuhlmann, P. Stone, R. Mooney and J. Shavlik, "Guiding a Reinforcement Learner with Natural Language Advice: Initial Results in RoboCup Soccer," in *AAAI-2004 Workshop on Supervisory Control of Learning and Adaptive Systems*, 2004.
2. C. Matuszek, E. Herbst, L. Zettlemoyer and D. Fox, "Learning to Parse Natural Language Commands to a Robot Control System," in *Proceedings of the International Symposium on Experimental Robotics (ISER)*, 2012.
3. S. Tellex, K. Thomas, D. Dickerson, W. Matthew, A. Banerjee, S. Teller and N. Roy, "Understanding Natural Language Commands for Robotic Navigation and Mobile Manipulation," in *AAAI*, 2011.
4. M. Marcus, M. Marcinkiewicz, and B. Santorini, "Building a Large Annotated Corpus of English: The Penn Treebank," in, *Computational linguistics (19.2)*, 1993.
5. J. Chamberlain, M. Poesio, and U. Kruschwitz, "A Demonstration of Human Computation using the Phrase Detectives Annotation Game," in *Proceedings of the ACM SIGKDD Workshop on Human Computation*, 2009.
6. L. Ahn and L. Dabbish, "Labeling Images with a Computer Game,", in *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, 2004.
7. T. Regier and L. Carlson. "Grounding Spatial Language in Perception: An Empirical and Computational Investigation," in *Journal of Experimental Psychology (130.2)*, 2001.
8. A. Herskovits. "Schematization" in *Representation and Processing of Spatial Expressions (149–162)*, 1998.
9. M. Barclay and A. Galton. "A Scene Corpus for Training and Testing Spatial Communication Systems", in *Proceedings of the AISB Convention Communication, Interaction and Social Intelligence*, 2006.

Phenomena in conveying information during oral task descriptions

Stephanie Schreitter and Brigitte Krenn

Austrian Research Institute for Artificial Intelligence, 1010 Vienna, Austria,
firstname.lastname@ofai.at

Abstract. A robot has to deal with a broad variety of information conveyed via verbal and non-verbal channels to be able to observe and listen to a task presented by a human teacher. We have collected a small corpus of human-human dyads to investigate how information is presented through verbal and/or visual channels. Apart from the characteristics of spoken language, the qualitative analysis of the data shows: (i) broad variation in wording regarding objects and actions, as well as omissions of lexical referents, (ii) patterns of use of verbal references and/or communicative gestures for directing the attention of the learner, (iii) a temporal structuring of the task by verbal means for all teachers, and (iv) the use of generic "you" for most of the teachers.

Keywords: task descriptions, embodied language processing, oral communication

1 Background

In face-to-face communication, people do not only use speech but a multitude of non-verbal behaviours such as nods, communicative gestures, gazes, object manipulation gestures, etc. The vocal and the gestural acts together comprise the information necessary for the observer/learner to understand. Findings from embodied cognition have shown the importance of action and perception during language comprehension in humans [5, 6, 12, 13]. If robots are to interact with humans in natural ways in the future, a number of serious issues in multimodal communication must be tackled. With the present contribution, we aim at illustrating the problem and state (minimal) requirements for system functionality.

Imagine a robot that can analyse, interpret, and learn from task oriented presentations where a human teacher shows some activity to the robot learner and explains what she/he is doing by means of task accompanying speech. In the present paper, we investigate which kinds of communicative signals and their variations a robot should be able to deal with when it is presented with a task. We recorded human-human dyads to see which information is typically conveyed by which channels.

Clark and Krych [4], for instance, argue that human-human dialogue is a bilateral, opportunistic, and multimodal process where common ground is continuously updated. The authors emphasize that in dialogue, participants use vocal and gestural modalities in parallel and that the visual modality is faster and more secure than the auditory modality for certain types of communication. Gestures are an integral part of language, synchronous and co-expressive with speech, cf. [9, 1]. In a study by Lozano and Tversky [8], communicators explained how to assemble a simple object using either speech with gestures or only gestures. In the "gestures only"-condition, the assembly task was learned better and fewer assembly errors were made than in the "speech with gesture"-condition.

It is a challenge to equip robots with system components necessary to understand multimodal natural human communication. In a task description context, system components and the robot's architecture must (i) allow for robust incremental processing of natural speech and of multimodal communicative signals, (ii) include visual perception of the objects in the scene and the ongoing activity, and (iii) integrate all this in multimodal representations and the robot's episodic memory. Recent attempts have been made to address task-based natural language understanding on robots. Scheutz et al. [10] developed the robot architecture DIARC aiming at more natural human-robot interactions. The architecture includes mechanisms for natural language processing, intentional behaviours, and monitoring mechanisms to detect faults and recover from them. Kopp et al. [7] propose a model of how meaning can be organized and coordinated in speech and gesture. Their model is based on spreading activation within dynamically shaped multimodal memories.

The aim of the present work is to present and discuss an inventory of phenomena characteristic for showing and explaining to a learner what she/he should do. In the following sections, the corpus is presented and the communicative phenomena present in the corpus are discussed together with implications they have regarding specific components necessary for a robotic learner.

2 Corpus: human-human task descriptions

A corpus comprising 19 German recordings (video plus audio) was created where one person (the teacher or actor) showed another person (the observer or learner) how to mount a tube in a box with holdings, see Figure 1. [1] Two markers differing in colour had to be put in two different holdings. The teacher performed the task and verbally explained what had to be done. Thus, the task descriptions contain language mirroring the human perception and structuring of the task. The observer was told to carefully watch and listen to the explanations to be able to pass the information on to a new learner. The utterances were recorded as well as a frontal video of the setting including arms, hands, and torso of each teacher and learner. Although head and shoulders are not visible in the recordings, the transmission of non-verbal cues is already extensive. The manipulation task is borrowed from a robotic setting. Letting humans do the same task, and in addition let them explain it, helps to better understand what a robot would have to deal with when it were in the learner's position.

[1] The subjects were German students from the Technical University Munich (16 male, 3 female).

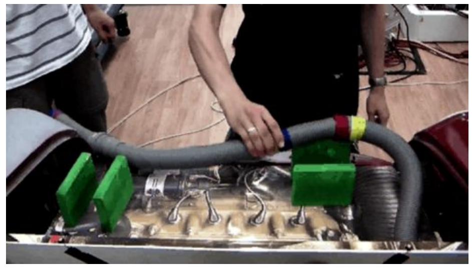


Fig. 1. A picture of the setting. A teacher is mounting a tube in a box with holdings.

3 Phenomena: how information is conveyed

The task to be described is quite simple: containing a grasp for the tube at a coloured marker, adjusting the tube between two green holdings, then grasping the tube at another coloured marker and putting it between another pair of green holdings. On average, the task duration was 21 seconds (12-34s). Although the task was quite simple and the learners had the assignment to listen carefully and forward the information to a new learner, there was quite some variation in how teachers presented the task. In the following, prevalent phenomena are presented and requirements for respective system components are briefly discussed.

Characteristics of spoken language Several properties typical for spoken language are present in the data: wrong word substitutions – 'holdings' (*Hindernis*)[2] instead of 'marker' (*Markierung*); repairs – 'red eh blue and yellow marker' (*rot äh blau-gelben Marker*); insertions – 'äh'; contractions – 'through the' (*durchs*, 'durch das'); errors – *habst* for 'have' ('hast').

These phenomena call for robust incremental language processing, e.g. [11], in addition to standard language technology tools such as automatic speech recognition, tokenization, part-of-speech, morphological analysis, phrase chunking, dependency parsing, and the such.

Variations in wording Objects in the task are the tube, two pairs of holdings, and three markers. For tube, all teachers used the same German word *Schlauch* (tube), except for three who did not verbally refer to the object at all. For 'marker', two teachers used the Anglicism *Marker*, and two used either 'point' (*Punkt*) and 'gripping point' (*Greifpunkt*) or 'endpoint' (*Punkt* / *Endpunkt*). The other 15 teachers used 'marker' (*Markierung*). For the holdings, there was a wide variation in naming: 'obstacle' (*Hindernis*), 'thing' (*Ding*), 'block' (*Block*), 'beam' (*Balken*), 'rail' (*Schiene*), 'marker' (*Markierung*), 'log' (*Klotz*), 'opening' (*Öffnung*). Again, there was one teacher who did not verbally refer to the holdings. The actions 'grasping the tube' and 'mounting the tube in the box

[2] For better readability, the English translation is in the main text and the actual German word choice is in brackets.

with the holdings' also showed some variance. For grasping, 'grasp' (*greifen*), 'have' (*haben*), 'take' (*nehmen*), 'span' (*umfassen*), 'change grip' (*umgreifen*) were used, and for 'putting the tube between the holdings': 'put' (*legen*), 'insert' (*führen / einführen / einspannen / einlegen / einsetzen / einfügen*), 'put inside' (*reinstellen / reinlegen*), 'clamp' (*klemmen*), and 'thread' (*einfädeln*).

Taking the above into account, the learner – may it be a human or a robot – has to infer objects and actions by listening and observing. The action is still the same, although 11 different verbs were used (up to two per teacher for the same action). Multimodal knowledge representations are a necessary prerequisite for dealing with lexical variation and omitted verbal references for objects and actions. The robot has to be able to resolve the connection of an abstract entity to an entity in the world, cf. [2], e.g. the words *Block*, *Klotz*, and *Hindernis* are all three referring to the green holdings. A comparison of what is visually perceived and what is uttered reveals how the same actions and objects are verbally expressed. In addition, the unspoken needs to be grounded in the scene. As Clark and Krych put it: "when the workspace is visible, the partners ground what they say not only with speech, but with gestures and other actions" [4], p.69. Thus, even though some teachers did not mention important elements of the task, the observers were able to understand.

Time markers Three teachers verbally signalled their respective learner that the task will now start, e.g. 'it is about' [...] (*es geht darum* [...]), 'the goal is' [...] (*Ziel ist* [...]), 10 told their learners when the task was done, e.g. [...] 'that was it' ([...] *das wars*), [...] 'that's all' ([...] *das ist alles*). All teachers used lexical time markers, such as 'first' (*zuerst*), 'then' (*dann*), 'subsequently' (*anschließend*) to signal the sequencing of the sub-tasks.

Therefore, as a technical basis, a (simple) model of before, after, and concurrency along a common timeline is required together with mechanisms to identify and interpret cues for temporal structuring. These may be lexical (as above), grammatical (tense) or determined by the course of multimodal action.

The teachers' perspective 13 teachers used 2nd person singular when explaining while carrying out the task by themselves, e.g. 'you grasp the tube with the right hand' (*du greifst den Schlauch mit der rechten Hand*). One participant interpreted the 'you' (*du*) as referential "you", and made a step forward to conduct the task himself. When the teacher continued explaining, he stepped back again to watch and listen. Another three teachers used imperative 'you have to [...]' (*du musst* [...]). Elliptic form – 'first to grasp here' (*zuerst hier greifen*), 1st person plural – 'we have to insert the tube here' (*wir müssen den Schlauch hier einfädeln*), and 3rd person singular – 'Muriel has to...' (*Muriel muss...*) were used by one person each. One teacher who started with 2nd person singular and the teacher who used 1st person plural switched to the elliptical form during explanation.

For a robot to be able to deal with these varieties, the following capabilities are required: (i) the ability to distinguish between the perception of self and other, (ii) a robust interpretation of the perspective from which the action accompanying utterance is issued, and (iii) a model for taking initiative, i.e. for the observer to understand when to just go on observing and when to step in the actor's position.

Verbal and gestural references to visual perception 13 teachers verbally referred to objects, actions or locations, e.g. 'here' (*hier*), 'like this' (*so*), 'this obstacle' (*dieses Hindernis*). The most frequent kind of gestures during task explanations were deictic gestures and holds during object manipulation to refer to objects or actions in the visual scene. Both gestures serve as indicators for directing the attention of the listener to certain objects or actions. Three teachers used verbal references and communicative gestures simultaneously (e.g. 'here' (*hier*) + deictic gesture). One teacher neither used communicative gestures nor verbal references to the visual scene. He only mentioned the grasping of the marker and did not mention that the tube has to be mounted in the box with the holdings. This could only be inferred by the learner from the visual scene. In this respect, Herbert Clark argues that "placing things just in the right manner" ([3], p.243) is an indicative act in which an object is moved into the addressee's attention.

For gestural and verbal references to visual perception, the robot has to be able to deal with (i) object recognition, (ii) feature recognition, and (iii) gesture recognition. In addition to visual gesture recognition and the recognition of verbal reference to visual perception such as 'here' (*hier*), 'like this' (*so*), (iv) an attention model is required to enable the robot to detect and interpret the attention directing signals issued by the teacher.

Verbal backchannels 10 learners signalled their understanding via verbal backchannels to their respective teacher, e.g. *ok*, *mhm*. Non-verbal backchannels such as head nods etc. were not visible in the present videos. The interplay of verbal and non-verbal backchannels in joint activity (speaking and listening together form a joint activity, cf. [4]) will be topic of further investigations.

4 Conclusion and future work

In this paper, we discussed phenomena occurring in a corpus of 19 simple task descriptions (action plus speech) of how to mount a tube in a box with holdings. They include characteristics of spoken language, variations in wording, verbal time marking, variation of teacher's perspective, and verbal and gestural references to the scene. These results highlight the importance of multimodal signal processing in human-robot interaction.

Depending on the phenomena and their functional challenge, there exist none up to a variety of proposed technical solutions for system functionalities. The interplay of the components and the requirement for real-time processing are still far from being reached. More research and integration work is needed on the way toward human-like task-based natural language processing for robots.

A second corpus has already been collected which is a follow-up and extension to the presented corpus. The new corpus includes 3 video streams – one of the teacher, the listener and the setup respectively – an audio stream, motion data, and force data when collaboratively manipulating an object. The interplay of head movement, eye gaze, gesture, facial expression, verbal and non-verbal backchannel feedback, body posture etc. will be further analysed based on the new data and the experiences from the initial corpus.

5 Acknowledgements

The first author is a recipient of a DOC Fellowship of the Austrian Academy of Sciences at the Austrian Research Institute for Artificial Intelligence. The authors would also like to thank the Institute for Information Oriented Control (ITR) at Technical University of Munich and the Cluster of Excellence Cognition for Technical Systems (CoTeSys) for their support in recording the data.

References

1. Bergmann K.: The production of co-speech iconic gestures: empirical study and computational simulation with virtual agents. PhD Thesis. Bielefeld, University (2012)
2. Cantrell, R., Scheutz, M., Schermerhorn, P., Wu X.: Robust spoken instruction understanding for HRI. Proceedings of the 2010 HRI Conference. (2010)
3. Clark, H. H.: Pointing and placing. In S. Kita (ed.) Pointing. Where language, culture, and cognition meet. Hillsdale, NJ, Erlbaum (2003) 243–268
4. Clark, H. H., Krych, M. A.: Speaking while monitoring addressees for understanding. Journal of Memory and Language, **50(1)** (2004) 62–81
5. Horton, W. S., Rapp, N. D.: Occlusion and the accessibility of information in narrative comprehension. Psychonomic Bulletin & Review, **10** (2002) 104–109
6. Kiefer, M., Barsalou, L.: Grounding the human conceptual system in perception, action, and introspection. In Tutorials in action science. MIT Press (2011)
7. Kopp, S., Bergmann, K., Kahl, S.: A spreading-activation model of the semantic coordination of speech and gesture. In Proceedings of the 35th Annual Meeting of the Cognitive Science Society (2013)
8. Lozano, S., Tversky, B.: Communicative gestures facilitate problem solving for both communicators and recipients. Journal of Memory and Language. **55(1)** (2006) 47–63
9. McNeill, D.: Gesture and thought. University of Chicago Press, Chicago (2005)
10. Scheutz, M., Briggs, G., Cantrell, R., Krause, E., Williams, T., Veale, R.: Novel mechanisms for natural human-robot interactions in the DIARC architecture. Proceedings of AAAI workshop on intelligent robotic systems (2013)
11. Scheutz, M., Cantrell, R., Schermerhorn, P. Toward humanlike task-based dialogue processing for human robot interaction. AI magazine. **32(4)** (2011) 77–84
12. Stanfield, R. A., Zwaan, R. A.: The effect of orientation derived from verbal context on picture recognition. Psychological Science, **12** (2001) 153–156
13. Zwaan, R. A., Taylor, L.: Seeing, acting, understanding: motor resonance in language comprehension. Journal of Experimental Psychology, **135** (2006) 1–11

Position paper: the development of robot specific behavior for tour guide robots

Daphne Karreman, Betsy van Dijk, Vanessa Evers

Human Media Interaction, University of Twente
Enschede, The Netherlands
{d.e.karreman, e.m.a.g.vandijk, v.evers}@utwente.nl

1 The FROG-project

The EU FP7 project FROG (Fun Robotic Outdoor Guide) proposes to develop a guide robot with a convincing personality and behaviors that will engage tourists in a fun exploration of outdoor attractions. The project work encompasses innovation in the areas of vision-based detection, robot design, navigation, human-robot interaction, affective computing, intelligent agent architecture and dependable autonomous outdoor robot operation. This paper focusses on the design of the behaviors of the FROG-robot. FROG's behaviors will be designed based on the findings of a combination of an iterative and user-centered design approach, social behavioral studies and exploratory field studies on human-robot interaction. The intelligent agent architecture is a platform that will allow the integration of low-level guidance and communication controls with high-level interaction generation including affective computing algorithms and contextual recognition. This approach will lead to the creation of a new generation of highly sophisticated autonomous outdoor robotic guide services.

Keywords: Tour guide robot, User-centered design, anthropomorphism, specific robot behavior, design.

2 Effective robot specific behavior

More and more robots become available for public and private spaces. About these robots Fong et al. state: "it (*the robot*) must establish appropriate social expectations, it must regulate social interaction (using dialogue and action), and it must follow social convention and norms" [1]. Most behavior of robots is copied from humans (anthropomorphism), or animals (zoomorphism), because people tend to understand human- and animal-like cues best. However, are human- or animal-like communication cues and behavior the only possible behaviors for robots that people understand?

2.1 Presumption

Robots are built to help people with various tasks; therefore, robots should be designed to be able to perform these tasks. Even though current robots often have anthropomorphic features or have a human-like appearance, for the purposes of the task the robot needs to perform (fetch, carry, clean etc.) a human-like shape may not be necessary. For instance, the design of a robot that has to swim will more likely resemble a fish rather than a person. Robot bodies should be designed in a way that is optimal for the task at hand. We argue that this also holds for the design of robot behavior. Designers should not just copy human behavior and communication cues one-on-one to robots. Instead, they need to identify robot specific communication cues people will understand intuitively and experience as natural.

2.2 User-centered design approach

We adopted a user-centered design approaches taken from Industrial/Product Design. We would like to incorporate these methods more in the field of human-robot interaction (HRI) as the first author has a product/interaction design background. In user-centered design approaches, designers develop products to effectively perform a function while continuously keeping the user (and their needs and requirements) in mind during all phases of design. With a user-centered design approach, we can design the robot for people, but not by definition as a person. Most important is that the robot will support a person in an effective and intuitive way with a task or a series of tasks.

2.3 Anthropomorphism

The first computers were not designed to resemble humans. Nevertheless many people tended to treat the computer as a social actor; they projected human social behavior norms to a computer [2]. This is even more the case with robots, now robots enter everyday social environments. People tend to anthropomorphize objects or robots they see. They tend to assign human-like attributes to objects/robots so they can apply mental models they have already learned [3].

2.4 How to design specific robot behavior

It is important that a robot's behavior, personality and appearance match [3]. The robot can have the perfect nice and gentle behavior, but if the shell of the robot looks aggressive, many people would judge from the shell that the robot will not be nice to interact with in the first place. For the design of the robot behavior, we argue that iterative design and continuous user testing will help find the best solutions. Anthropomorphic form or behavioral patterns can be a starting point in designing robot behaviors. However, we argue that the (intended) effects and outcomes of the human behavior should be studied and robot behavior should be designed to evoke the same communication goals.

In the next section, we describe how we used a user-centered design approach and iterative, explorative sessions to analyze human tour guide behavior and guide robot behavior. Results from these studies lead to design guidelines for guide robot behavior, with a focus on behavior that is robot specific and intuitively understandable.

3 User-centered design in the FROG-project

3.1 Studies to effective guide behavior

For the FROG-project we started with gaining an understanding of visitors' likes and dislikes when visiting tourist sites with a participatory design approach. We found that visitors really liked the structure provided in guided tours, however, they did not like the rush of tours and long duration [4]. Also, we observed human tour guide behaviors, and we found that human tour guides use many (non-verbal) strategies to gain and keep the visitor's attention, to direct their attention and to balance the information given [5]. In these observations, we focused on the intended effects and outcomes of the guide's actions and interactions with visitors.

In May 2013, we conducted an iterative and exploratory study with a machine-like robot in a real world cultural heritage setting (the Royal Alcazar in Seville, Spain). We did not directly copy the robot behaviors from human tour guides, but we developed robot behaviors based on the intended interactional outcomes. Hence, in iterative sessions, we tested several different robot behaviors for its orientation and for its utterances. Preliminary results of this study indicate three zones of proximity: 1) visitors stand very close, almost touching the robot, 2) they stand more than three meters away, 3) they stand somewhere in between. Also, we found that gaze direction of the robot influenced where visitors looked at. Only well-designed text can direct the gaze of visitors to a point the robot did not gaze at. Finally, we found that visitors who left the robot did not necessarily influence the other people in the group.

3.2 Preliminary design guidelines for guide robot behavior

Previous mentioned studies formed the basis for developing design guidelines for guide robot behavior. Note that the robot will differ from human tour guides: it will give short tours, based on the interests of the visitors. Also, it will only guide in some places of the tourist site, so visitors still get a chance to explore the site on their own as well. In addition, the robot can carry devices that a human tour guide will not carry, such as a projector or a screen. These devices can make the robot tour more lively and interesting; also, it forces the design of the robot to differ from the human body.

From the results of the participatory design study and the iterative exploratory study, we deduced design guidelines for the robot behavior that resemble human tour guide behavior. First, the robot should use some specific strategies human tour guides uses. For example, give curiosities to capture and keep the visitors' attention. Second, the perceived gaze direction of the robot can steer the visitors' gaze direction. If the point of interest is somewhere else than the robot seems to gaze to, the robot should give sufficient information in text about where to look.

On the other hand, visitors tend to show interest in the robot only (not in its guiding function), which is very different from their reaction to human tour guides. Therefore, some guidelines are very different from human tour guide behavior. First, as long as visitors pay attention to the story, the robot should go on giving information. However, when people only show interest in the robot itself, the robot should display that it is aware of these visitors. Then after a while, the robot can try to bend this more playful interaction into a guide-visitors interaction. Second, the robot does not only catch the attention of people that are standing close, also visitors who stay at a distance show interest. Therefore, the robotic tour guide should scan the surroundings occasionally. It should continue the tour when it detects visitors, who probably stand further away, but show an orientation towards the robot and stay there during the whole story. Reacting to these visitors becomes important if a visitor close by loses interest and walks away. In that case, the robot should not stop or interrupt the story and keep focused on the visitor that left, but instead the robot should focus on visitors that still show interest. If the robot does not show interest in these visitors anymore, they are likely to lose their interest in the robot as well. Last, the robot should not solely rely on its detection of visitors by gaze (cameras directed to the front-side) to continue or stop the explanations because in some situations visitors tend to stand next to or behind the robot, while they still show interest in the story of the robot.

4 Conclusion and future work; a tour guide robot in the field

Observations of human tour guides and iterative design of robot behavior show that robot behavior can partly be copied from human behavior; however, as people tend to react to the robot itself, the robot should show some specifically designed robot behavior as well. In this paper, we gave some preliminary guidelines for robot tour guide behavior. In near future, we will analyze all video material of the exploratory field study. We will use a specially developed "HRI analysis tool" that helps to speed up and standardize the analysis of large HRI datasets; however, the development of the HRI analysis tool is still ongoing work. The results finally will lead to a set of design guidelines for tour guide robots.

5 Acknowledgement

The research leading to these results received funding from the European Community's 7th Framework Programme under Grant agreement 288235 (http://www.frogrobot.eu/).

6 References

1. T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A Survey of Socially Interactive Robots : Concepts , Design , and Applications Terrence Fon," no. November, 2002.
2. C. Nass, B. Fogg, and Y. Moon, "Can computers be teammates?," *International Journal of Human-Computer …*, vol. 45, no. 6, pp. 669–678, 1996.
3. B. R. Duffy, "Anthropomorphism and the social robot," *Robotics and Autonomous Systems*, vol. 42, no. 3–4, pp. 177–190, Mar. 2003.
4. D. E. Karreman, E. M. A. G. Van Dijk, and V. Evers, "Using the visitor experiences for mapping the possibilities of implementing a robotic guide in outdoor sites," in *RO-MAN 2012: The 21st IEEE International Symposium on Robot and Human Interactive Communication.*, 2012, pp. 1059–1065.
5. D. E. Karreman, E. M. A. G. van Dijk, and V. Evers, "Contextual Analysis of Human Non-verbal Guide Behaviors to Inform the Development of FROG, the Fun Robotic Outdoor Guide," *Human Behavior Understanding*, vol. LNCS 7559, pp. 113–124, 2012.

The Use of Social Robot Ono in Robot Assisted Therapy

Cesar Vandevelde[1], Jelle Saldien[1], Maria-Cristina Ciocci[1], Bram Vanderborght[2]

[1]Ghent University, Dept. of Industrial Systems and Product Design, Campus Kortrijk, Belgium
{cesar.vandevelde, jelle.saldien, cristina.ciocci@ugent.be
[2]Vrije Universiteit Brussel, Dept. of Mechanical Engineering, Brussels, Belgium
bram.vanderborght@vub.ac.be

1 Introduction

Ono is a low cost DIY reproducible social robot designed to make large scale human-robot interaction (HRI) studies more feasible financially, as well as to make social robotics accessible to hobbyist experimenters. While many HRI platforms exist, most are not suitable for the aforementioned scenarios because they are either too expensive or too hard to modify. Examples of existing platforms are Kobian [1], HRP-4C [2], WE-4RII [3], iCub [4], Kismet [5], Probo [6], Nao [7]. Low-cost options also exist, such as My Keepon [8] and Kaspar [9]. In our opinion, many existing platforms are hard to modify because hardware/software source files are not made available or because the platform relies on high-end components and manufacturing techniques. Our social robot, named Ono was developed with the following goals in mind to address these issues:

Open source.
The aim for Ono is to distribute both open hardware and open software. By allowing unrestricted access to hardware and software, other researchers have the opportunity to easily extend the capabilities of the robot, enabling them to adapt the robot to their specific needs. The source files of the robot can be found in a public Github repository [10], however full assembly instructions are not available yet.

Do-It-Yourself.
Our goal is that Ono can be built without the aid of paid experts or professionals.

Modular.
By dividing the robot into smaller functional subunits, repairs can be made quicker and more easily, modules can be reused in other projects and more specialized can be developed, allowing for a degree of customizability.

Reproducible.
Ono is constructed from standardized components and readily available materials. Custom components can be produced using low volume manufacturing techniques, most notably laser cutting. With this approach, we aim to make it possible to replicate

this robot anywhere in the world, without the need for high-end components or manufacturing machines.

Social Expressiveness.

Ono's face contains 13 degrees of freedom (DOF), allowing the robot to gaze and to express facial expressions. The DOFs are based on the Action Units (AU) defined by the Facial Action Coding System (FACS) [11] as well as our experiences with the Probo social robot [6]. A mapping algorithm translates a valence and arousal parameter to a set of positions for all DOFs using Russel's circumplex model of affect [12], allowing for a smooth transition between emotions.

The goal of this paper is to provide a brief overview of the construction of Ono, as well as to present the results of our first study with children.

2 Construction

Fig. 1. Child interacting with Ono.

Ono was developed as a social robot for children; this had several consequences for the design of the robot. The entire robot is covered in a soft foam and textile skin to attain a soft and inviting appearance for the children, as well as to protect the internal components from damage. The robot has a disproportionally large head to make its facial expressions more noticeable and is posed in a sitting position to improve stability. The main components of the robot are:

Skeletal frame.

The frame of the robot consists of a series of interlocking cross-sections that together form a sturdy structure onto which modules can be attached.

Modules.

Sets of related sensors and/or actuators are grouped into modules. The current prototype has 3 types of modules: 2 eye modules, 2 eyebrow modules and 1 mouth module. Modules are attached to the main frame using snap connectors, making it easy to replace them.

Foam and skin covering.

The outer layers of Ono consist of a polyurethane foam shell covered with stretchable textile. The foam shell is made from multiple pieces that were cut from a flat sheet of foam and were subsequently sewn together over the frame.

Electronics and interface.

The robot is currently controlled from a separate control box with joystick interface. The control box also contains the robot's power supply. Power and data are sent to the robot using the same cable.

3 Pilot study

A pilot study to evaluate the use of Ono in Robot-Assisted Therapy (RAT) was performed in Romania. The robot was tested with 5 children with autism spectrum disorder (ASD), aged 3 to 10 years old. Children were asked to identify the emotion expressed by the robot, they were then asked to mimic the robot's facial expressions and were finally allowed some time to freely interact with Ono. Table 1 shows the interaction rates during the pilot study. Imitation is measured as the number of times that the child had the same facial expression as Ono. Touching is measured as the number of times the child touches the robot. Verbal initiation is measured as the number of times the child talks to the robot. Table 2 shows the recognition rates of happiness, anger, sadness and surprise. The emotions happiness, anger, sadness and surprise were shown in random order, but each emotion was shown 4 times. Because one child did not want to participate in this part of the study, only 16 measurements were obtained for each emotion. The children could easily identify happiness and sadness, anger was often confused with being scared or sad and surprise was often confused with happiness or sadness. During the free play phase of the study, most children continued to show interest in the robot. One child asked to play with Ono after the study ended; he even controlled the robot himself using the joystick interface. Another child played a musical instrument to the robot, and a third child tried to feed Ono. The tests suggest that Ono has an overall inviting appearance that elicits interaction, however some emotions are difficult for the children to identify and should be improved.

Table 1. Interaction rates

	Imitation	Touching	Verbal initiation	Engagement
Child I	11	70	25	3.08s / 4.16s
Child II	0	4	15	2.10s / 3.46s
Child III	0	9	5	2.43s / 9.43s
Child IV	7	13	3	2.09s / 1.29s
Child V	0	0	40	0.23s / 2.55s

Table 2. Emotion recognition rates

	Correct	Incorrect	Don't know
Happiness	15	1	0
Anger	3	10	3
Sadness	15	1	0
Surprise	6	7	3

4 Next steps

Our first results suggest that a low cost social robot can become a valuable tool in RAT and HRI studies. Even though Ono does not possess the same level of capabilities and features as other social robots, we believe there is a place for this type of robots. Low cost social robots such as Ono make it possible to perform large scale experiments, which has not been practical in the past because of the high costs and because often only one prototype of the robot exists.

Our next steps for Ono include fixing the problems discovered during the pilot study, such as eliminating the control box interface, improving the appearance of the facial expressions (anger and surprise in particular) and designing a new, more robust eye module. Additionally, we would like to try new degrees of freedom, such as arm movement and pan and tilt movement of the head. Our goal is to then evaluate different DOF configurations in a large study, to determine the optimal degrees of freedom for applications such as robot-assisted therapy.

We hope that the reproducible design of Ono means social robotics can become accessible to researchers and hobbyists around the world, and that the robot can be used as a therapeutic tool to help children with autism.

5 References

1. Zecca, M., Endo, N., Momoki, S., Itoh, K., Takanishi, A.: Design of the humanoid robot KOBIAN-preliminary analysis of facial and whole body emotion expression capabilities. In: IEEE-RAS International Conference on Humanoid Robots (Humanoids 2008), pp. 487-492. (Year)
2. Kaneko, K., Kanehiro, F., Morisawa, M., Miura, K., Nakaoka, S., Kajita, S.: Cybernetic Human Hrp-4c. IEEE-RAS International Conference on Humanoid Robots (Humanoids 2009) 7-14 (2009)
3. Miwa, H., Itoh, K., Matsumoto, M., Zecca, M., Takanobu, H., Roccella, S., Carrozza, M.C., Dario, P., Takanishi, A.: Effective emotional expressions with emotion expression humanoid robot WE-4RII. In: IEEE/RSJ International Conference on Intelligent RObots and Systems (IROS 2004), pp. 2203-2208. (Year)
4. Tsagarakis, N.G., Metta, G., Sandini, G., Vernon, D., Beira, R., Becchi, F., Righetti, L., Santos-Victor, J., Ijspeert, A.J., Carrozza, M.C., others: iCub: the design and realization of an open humanoid platform for cognitive and neuroscience research. Advanced Robotics 21, 1151--1175 (2007)
5. Breazeal, C.: Toward sociable robots. Robotics and Autonomous Systems 42, 167--175 (2003)
6. Goris, K., Saldien, J., Vanderborght, B., Lefeber, D.: Mechanical design of the huggable robot Probo. International Journal of Humanoid Robotics 8, 481 (2011)
7. Gouaillier, D., Hugel, V., Blazevic, P., Kilner, C., Monceaux, J., Lafourcade, P., Marnier, B., Serre, J., Maisonnier, B.: The NAO humanoid: a combination of performance and affordability. Arxiv preprint arXiv:0807.3223 (2008)
8. Kozima, H., Michalowski, M.P., Nakagawa, C.: Keepon. International Journal of Social Robotics 1, 3-18 (2009)
9. Dautenhahn, K., Nehaniv, C.L., Walters, M.L., Robins, B., Kose-Bagci, H., Mirza, N.A., Blow, M.: KASPAR--a minimally expressive humanoid robot for human--robot interaction research. Applied Bionics and Biomechanics 6, 369--397 (2009)
10. https://github.com/cesarvandevelde/Ono
11. Ekman, P., Friesen, W.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press (1978)
12. Russell, J.A.: A circumplex model of affect. Journal of personality and social psychology 39, 1161 (1980)

Social-Task Engagement : Striking a Balance between the Robot and the Task

Lee J. Corrigan[1], Christopher Peters[2], and Ginevra Castellano[1]

[1] School of Electronic, Electrical and Computer Engineering, University of Birmingham, United Kingdom
[2] School of Computer Science and Communication, Royal Institute of Technology (KTH), Sweden

Abstract. In this paper, we consider engagement in a triadic human-robot-task interaction. More specifically, we discuss why we need to perform 'online' differentiation and balancing of task and social engagement during human-robot interactions. The results of this work will help us to progress towards uncovering novel ways to design personalised human-robot interaction experiences. We start by defining the type of engagement that we are interested in, then we explain the methodology we are using to explore our hypothesis.

1 Introduction

At present, engagement is a broadly used term in human-robot interaction (HRI), typically characterised by an elements of concentration, enjoyment and flow [1] [2] [3] [4]. However, this umbrella-like definition is often used to explain aspects of engagement which are individually distinguishable as owing either to the task being performed or to the robot being interacted with [1] [5].

In this paper we begin with three clear and distinct definitions of engagement which are relevant to social HRI. We propose these definitions in an attempt to bring clarity and meaning to the exact type of engagement being considered in our work.

Imagine a scenario where you are asked to perform some task in a dyadic human-task relationship (Figure 1-b), such as that found in the typical one-player game scenario. The task could involve physical objects which you can manipulate, such as a board game or a block building task, or a virtually represented task hosted on a computer, tablet or phone. The task is considered to be an explicit task in which your input and the corresponding output caused by performing that task are intrinsically linked to one another. Now, lets say that you find yourself becoming immersed in the task, you are enjoying and concentrating on your inclusion in that task. This is considered to be 'task engagement'. Likewise, you can become disengaged from the task, but this is still considered to be 'task engagement', albeit, at its lowest extreme.

Furthermore, imagine another scenario where you are interacting with a socially capable robot where there is no task involved (Figure 1-a). An example might be a form of entertainment robot which is capable of sociable and friendly interaction. If during this scenario you become engaged with the robot, you are socially engaged. This is 'social engagement'. Again, at its lowest extremes you would become disengaged from the robot.

Now to extend on this further, imagine another scenario where you are interacting with a socially capable robot, where both you and the robot work together to perform an explicit task (Figure 1-c). An example of this could be a collaborative task where both robot and human work together to build something. The question here is, if you become engaged in what you are doing, are you engaged with the task or are you engaged with the robot? It would be far too ambiguous to simply call this engagement, so we will need to define this phenomenon as 'social-task engagement'. Furthermore, stating that one is simply engaged does nothing to help distinguish the proportion of engagement attributable to different aspects of the interaction. For example, lets say at some point during this scenario you become more engaged with the robot or less productive in the task. Was it the task or the robot which caused that to happen?

With this in mind, we hypothesize that engagement with the task must be separable from engagement with the social robot. Further still, 'online' differentiation and balancing of social and task engagement (i.e. updating both the task and the robot throughout the interaction) will lead to a more personalised and productive experience for both the robot and the user.

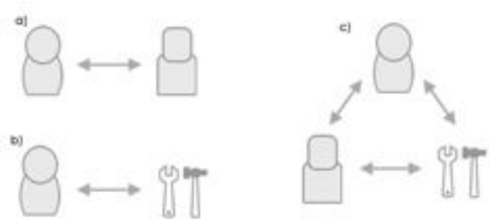


Fig. 1. a) Social Engagement, b) Task Engagement c) Social-Task Engagement

2 Engagement

Engagement is a much talked about phenomenon in HRI, but what is engagement really? A definition taken straight from a dictionary states "the act of engaging or the state of being engaged", but this does not help us to explain what engagement is. Digging deeper we find more functional terms related to engagement which might help us to characterise this phenomenon, such as participation, commitment, concentration, involution and immersion.

Engagement shares many of the same characteristics involved in flow [5], in education settings it has been found that the more challenging assignments lead to flow, whereas in the workplace having a clear concept of the goal and having immediate feedback was more effective. In terms of causality, the first thing that comes to mind is that engagement is the effect of an internal state, a low level desire or a state of being, such as curiosity, intrigue, interest, amazement, wonder or concern. It could be that these internal states act as incentives for becoming engaged. In fact, further studies involving flow have found that a "need for achievement" is a personal characteristic which fosters flow [6] [7].

In addition to this, engagement could also cause more affective aspects of conciousness, states of enjoyment or even provide some other form of arousal which is beneficial or at least pleasing to the recipient, therefore, warranting the initial investment of becoming engaged. One might hypothesize that engagement is driven by ones underlying motives for wanting to satisfy their own goals and desires.

3 Related Work

Further to our previous work, where we consider the measurement of task engagement during human-robot interactions [8], we have become aware of the need to perform 'online' balancing of social and task engagement during experiments. This has shown us that situations exist where the engagement the user experiences in a triadic human-robot-task relationship is associated with either the task, the robot or combinations of both. The amount of engagement experienced is scalar as oppose to being present or not.

Whilst engagement is often associated with learning performance [9] [10], and efforts have been made to explore social [11] and task engagement [12], very little work has been done to differentiate task from social engagement during a human-robot interaction involving a social-task. At present, social engagement is defined as "the process by which two (or more) participants establish, maintain and end their perceived connection during interactions which they jointly undertake" [1], and "the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and continuing the interaction" [13]. Task engagement is derived from studies involving flow experience, characterised by elements of attention, concentration and enjoyment with the learning task [5].

Context is an important aspect in human-robot interactions, [12] consider its relevance during child and robot interaction involving a chess game, and [14] use task state information to classify interest of children interacting with a game.

Engagement is far more than a binary concept (i.e. engaged or disengaged), [15] considered the 'level of engagement' which details how much the user was looking at relevant objects at appropriate times, and the 'quality of engagement' where users were considered as being engaged, superficially engaged or uninterested in the scene/action space. Here, we intend to learn from and extend upon that concept by evaluating the interaction in terms of the task and social elements of the interaction.

Recent unpublished work by [16] showed that when a social robot interacting with a child in a shared physical space struggled to adjust the screen, the child without hesitation notices the problem and immediately moves the screen for the robot. This leads us to believe that the child was highly engaged with the robot causing him to do something of which he was not expected. At the same time the child was also performing well in the task.

Currently, we are unable to detect and differentiate between the level and quality of task and social engagement during such an interaction, but with the advances we intend to make during this project we will be able to look at social-task interactions in a completely different light.

4 Methodology

4.1 Pilot Study

Our first experiment is a pilot study involving adults. We have consulted with psychologists in an attempt to design the experimental conditions which will help us to identify the most pertinent indicators of both task and social-task engagement. The experimental set-up comprises of a large touch screen to run interactive tasks, several cameras detecting valence and affective display from facial expressions, an Affectiva Q Sensor[3] detecting arousal from galvanic skin responses and a Microsoft Kinect[4] for reading lean position and posture [17] through depth perception. In addition to this gaze direction will be clamped to either the task, robot or elsewhere using data derived from the users' head direction.

Interactive Tasks We are using three tasks and each one has been designed to elicit different states of engagement. The first is based on a simple Whack-A-Mole style game and is considered to be an engaging task which requires much effort and concentration. The second is a simple sequence following block tapping task, designed to be far less engaging. In the third task we use a memory game involving cards to observe social-task engagement during a novel human-robot interaction scenario.

Experimental Conditions Participants are divided into two groups, representing the two conditions in the study i.e. engaging and non-engaging. Participants from both groups are then divided again into two further groups, here, half perform task one followed by task two, and the other half perform the tasks in the opposite order. This ensures the data we collect is not biased by the ordering of the tasks. Furthermore, the user is not introduced to the robot until the third and final task involving the human-robot-task experiment, this is to prevent biasing the social relationship with the robot.

[3] http://www.qsensortech.com/, Affectiva Q sensor, Last accessed 25-4-2013
[4] http://www.microsoft.com/en-us/kinectforwindows/, Microsoft Kinect, Last accessed 01-09-2013

Robot Behaviours The engaging group experience a robot which is friendly, helpful and instructive, the robots behaviours are designed to be personable, pulling on the empathic strings of the participant. The robot describes why 'they' need to work together to build 'their' battery, looking directly at the participant and addressing them by their first name. In contrast, the non-engaging group experience a neutral and partially helpful robot which although provides some help is far less personable, refrains from mutual gaze and does not address the participant by name.

4.2 Wizard-of-Oz Study

Our second experiment is a Wizard-of-Oz study involving children aged between 11 and 13. The task is grounded in geography, more specifically map reading. The robot will be semi-autonomous and capable of social, empathic and pedagogical intervention. During the interaction we intend to experiment with different levels of task difficulty and various robot behaviours. Here we will utilise the same experimental design as the pilot study to collect a corpus of interaction data, yet we will have remote control of the robot, with the goal of giving the perception of realistic social intelligence as well as both task and situational awareness.

5 Conclusion

At present it is common to bundle all elements of engagement during human-robot interactions into a single classification, but without further research in this area we will be unable to design interactions that can be balanced and personalised towards the individual user.

In this paper we have described how we intend to explore engagement owing to differing aspects of the interaction. The results of this work will enable us to move forward and further explore both situational and contextual indicators of task and social-task engagement, helping us to progress towards uncovering novel ways to design personalised human-robot interaction experiences.

6 Acknowledgements

We would like to thank Arvid Kappas, Christina Basedow, Dennis Küster and Pasquale Dente from Jacobs University Bremen for helping to formulate and refine the experiments and conditions required for the pilot study.

This work was partially supported by the European Commission (EC) and was funded by the EU FP7 ICT-317923 project EMOTE (EMbOdied-perceptive Tutors for Empathy-based learning). The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

References

[1] C. L. Sidner, C. D. Kidd, C. Lee, and N. Lesh, "Where to look: a study of human-robot engagement," in *Proceedings of the 9th international conference on Intelligent user interfaces*, pp. 78–84, ACM, 2004.

[2] K. Pitsch, H. Kuzuoka, Y. Suzuki, L. Sussenbach, P. Luff, and C. Heath, "The first five seconds : Contingent stepwise entry into an interaction as a means to secure sustained engagement in hri," in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, 2009.

[3] M. Michalowski, S. Sabanovic, and R. Simmons, "A spatial model of engagement for a social robot," in *Advanced Motion Control, 2006. 9th IEEE International Workshop on*, pp. 762–767, 0-0.

[4] C. Rich, B. Ponsleur, A. Holroyd, and C. L. Sidner, "Recognizing engagement in human-robot interaction," in *5th ACM/IEEE International Conference on Human-Robot Interaction, HRI 2010*, pp. 375–382, 2010.

[5] M. Csikszentmihalyi, *Flow : The Psychology of Optimal Experience*. New York: Harper Perennial, 1990.

[6] S. Engeser and F. Rheinberg, "Flow, performance and moderators of challenge-skill balance," *Motivation and Emotion*, vol. 32, no. 3, pp. 158–172, 2008.

[7] R. Eisenberger, J. R. Jones, F. Stinglhamber, L. Shanock, and A. T. Randall, "Flow experiences at work: For high need achievers alone?," *Journal of Organizational Behavior*, vol. 26, no. 7, pp. 755–775, 2005.

[8] L. J. Corrigan, C. Peters, and G. Castellano, "Identifying task engagement: Towards personalised interactions with educational robots," in press.

[9] J. T. Guthrie and K. E. Cox, "Classroom conditions for motivation and engagement in reading," *Educational Psychology Review*, vol. 13, pp. 283–302, 2001.

[10] J. D. Finn, "Withdrawing from school," *Review of educational research*, vol. 59, no. 2, pp. 117–142, 1989.

[11] A. L. Thomaz and C. Breazeal, "Experiments in socially guided exploration: lessons learned in building robots that learn with and without human teachers," *Connection Science*, vol. 20, no. 2-3, pp. 91–110, 2008.

[12] G. Castellano, I. Leite, A. Pereira, C. Martinho, A. Paiva, and P. W. McOwan, "Detecting engagement in hri: An exploration of social and task-based context," in *2012 ASE/IEEE International Conference on Social Computing and 2012*, pp. 421–428, 2012.

[13] I. Poggi, *Mind, Hands, Face and Body: A Goal and Belief View of Multimodal Communication*. Weidler Buchverlag Berlin, 2007.

[14] A. Kapoor and R. W. Picard, "Multimodal affect recognition in learning environments," in *Proceedings of the 13th annual ACM international conference on Multimedia*, pp. 677–682, ACM, 2005.

[15] C. Peters, S. Asteriadis, and K. Karpouzis, "Investigating shared attention with a virtual agent using a gaze-based interface," *Journal on Multimodal User Interfaces*, vol. 3, no. 1, pp. 119–130, 2010.

[16] O. Blanson Henkemans, B. Bierman, J. Janssen, M. Neerincx, R. Looije, and J. A. Van der Bosch, H.and Van der Giessen, "Using a robot to personalise health education for children with diabetes type 1: A pilot study.," in *Patient education and counseling 2013 (volume 92 issue 2 Pages 174-181)*.

[17] J. Sanghvi, G. Castellano, I. Leite, A. Pereira, P. W. McOwan, and A. Paiva, "Automatic analysis of affective postures and body motion to detect engagement with a game companion," in *Proceedings of the 6th international conference on Human-robot interaction*, pp. 305–312, ACM, 2011.

Towards Legible Robot Navigation - How to Increase the Intend Expressiveness of Robot Navigation Behavior

Christina Lichtenthäler[1], Alexandra Kirsch[2]

[1] Institute for Advanced Study, Technische Universität München, Garching, Germany
[2] Department of Computer Science, Tübingen University, Tübingen, Germany

Abstract. The work at hand addresses the question: How can we achieve legible robot navigation? To this end, we investigate current state-of-the-art assumptions and methods regarding legible robot navigation in order to propose key factors for the development of a legible robot navigation. We reviewed 18 articles regarding legible robot behavior and present the conclusions from our own research. We found three important factors for legible robot navigation: straight lines, stereotypical motions and the use of additional gestures.

Keywords: human-aware navigation, legibility, intend expressive navigation

1 Introduction

Robots will increasingly become part of the habitats and work spaces of humans. Wherever they are located, in the factories as co-workers, in nursing homes or hospitals as care assistants, as guides in supermarkets, or as household-robots in our homes, one crucial behavior, which they all have in common, is navigation. A robot has to move through spaces where humans live and as Althaus et al. [2] already stated *"The quality of the movements influences strongly the perceived intelligence of the robotic system."*. The way a robot moves affects not only the perceived intelligence, but also the perceived safety, comfort and predictability [7, 18]. One important finding of a study Dautenhahn et al. [8] conducted to explore peoples' perception towards the future use of robot companions was that the behavior of a robot has to be predictable.

From all this we can conclude that (1) navigation is a crucial behavior, (2) a motion is not only an instrument to reach a goal position, it is also a way to communicate, and (3) predictability is important for robot behavior. The question now is how to generate predictable motions. Before going into details we want to point out our definition of predictable robot behavior. We use the term **legibility** to describe robot behavior that is (1) intent expressive, meaning that a human is able to infer the next actions, goals and intentions of a robot and (2) the robot behavior fulfills the expectations of a human interaction partner [17, 16]. This definition of legible robot behavior is in line with determinations of legibility in [1, 3, 5, 14, 15, 28, 29, 24, 30].

Research Question With the work at hand we want to answer the question how to generate legible robot navigation. What are the key factors for a legible robot navigation? To this end, we reviewed current literature regarding legibility in robot behavior and combined the thus collected insights with our own findings in order to propose factors that have to be considered in order to generate legible robot navigation.

2 Results of Literature Research

We will start with a literature survey regarding legibility of robot behavior. We systematically reviewed 18 articles [1, 3, 5–7, 9, 11–14, 22–27, 29, 30] published in the primary HRI publication venues from 2005 - 2013. In order to find all relevant papers we used the search terms: legibility/legible, readability/readable, and predictability/predictable in combination with motion/behavior and robot. For the work at hand we considered only 18 articles from the initial set of papers (32) claiming assumptions and/or approaches to generate legible robot behavior.

One very obvious assumption regarding legible robot behavior is that **human-like** behavior would be perceived as legible [5, 14, 11], because human behavior is well-known for humans. Therefore, the development of methods imitating human motions is very common in the HRI community.

Furthermore, Beetz et al. [5] claimed that a **stereotypical motion** is predictable, and thus legible. This assumption is supported by results from Bortot et al. [6].

In [23] the authors claim that the use of **complementary gestures** made by the robot could achieve legibility. Therefore, using their proposed gesture classification can improve the legibility of the robot. Similarly, Sisbot et al. [24] integrated complementary gestures in order to make the motion more intend expressive. This *complementary gesture assumption* is also supported by results from Basili et al. [3]. They were able to show that gaze behavior increases the ability to predict where someone is heading to.

Takayama et al. [30] claim that the **use of animation principles** makes the robot behavior more legible. They implemented additional gestures in order to let the robot show forethought and the results of their conducted study supported their assumption.

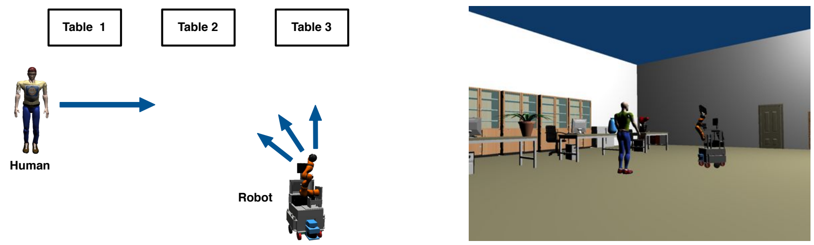
Another assumption is to **take into account social constraints, human preferences and abilities** [13, 1]. Following this Kirsch et al. [12, 13] proposed an approach to achieve legible task execution behavior. They suggest to learn human preferences and abilities in order to integrate this knowledge into a high level task planner.

Several authors like Takayama et al. [30], Guzzi et al. [11] and Kruse et al. [14] claim that **efficiency** is also one factor of legibility. Humans expect a robot to interact in an effective manner. If the robot behavior is non-legible, the human is not able to predict goals and intentions resulting in less efficient interactions.

Visibility is not only a prerequisite for legibility, because a human is not able to anticipate anything from a hidden motion, it is also a very important factor for generating legible motions. Sisbot et al. and Dehais et al. claim that a legible robot motion must be as visible as possible [9, 26]. This assumption is implemented in the Human Aware Motion Planner [22, 24, 27, 29] as well as in the Human Aware Navigation Planner [28, 25]. The visibility assumption is based on results from Dautenhahn et al. [7].

To conclude, in order to generate legible robot behavior, the following assumptions were proposed in the reviewed articles:

- model human-like behavior [5, 14, 11]
- generate stereotypical motions [5, 6]
- generate efficient motions [11, 14, 30]
- add complementary motions (gestures) in order to clarify intentions (e.g. gaze, pointing, use animation principles) [23, 24, 3, 30]
- take into account social constraints, human abilities, and preferences [13, 12, 1]
- robot motion must be as visible as possible [9, 26].



(a) Experimental Setup (b) Simulator view

Fig. 1. In the experiment we showed the participants simulated videos (b) of a robot and a human crossing the robots path. The robot is heading towards one of the tables (a).

3 Lessons Learned from Our Research

In the following we present our own findings regarding legible robot navigation. In our work we concentrate on human-robot path crossing scenarios.

3.1 Investigate Legibility of Existing Navigation Methods

In our first experiments we evaluated the legibility of existing navigation methods [16]. In a simulator-based experiment we showed the participants different videos of a robot and a human crossing the robot's path in an office environment (see Fig. 1). We measured legibility by asking the participants to (1) predict the goal after seeing a part of the video and after seeing the complete video (2) to rate how much the actual robot behavior matched their expectations and (3) how surprisingly the behavior was perceived.

From our results we concluded that existing state-of-the-art navigation algorithms fail in the presence of humans regarding legibility. In situations were a human was crossing the robots path the tested algorithms produced strange movements like small curves, or the robot was spinning around, or in the worst case the robot crashed into the human. All these strange movements were rated as non-legible although straight movements towards the goal and decreasing the velocity when approaching the human revealed higher legibility ratings. This fact is also confirmed by results from our second experiment [19] were we showed the participants first-person perspective videos of a similar setup.

3.2 Measuring Human Expectations

For further investigations towards a legible robot navigation we designed and conducted a study in order to find how a human would expect a robot to move when a human is crossing its way [21, 20]. To this end, we let participants steer a robot in a real-world scenario in which an instructed confederate was crossing the robot's path (see Fig. 2). We captured the movements of the robot and the confederate in order to find (1) the expected behavior in a path-crossing scenario and (2) to find the spatial coherence between robot, confederate and the behavioral reaction of the robot. We found out that a good strategy is to drive straight towards the goal and only react (stop) to a crossing human when the spatial relationship predicts to stop, otherwise drive on towards the goal.

(a) Steering BIRON (b) Laboratory (c) Study Setup

Fig. 2. In the study we let participants steer the robot (a) in our Laboratory (b). An instructed confederate crosses the robots path by chance as depicted in (c).

4 Conclusion

In the following we conclude the aforementioned findings regarding legible robot behavior and draw the key factors for generating legible robot navigation behavior.

Straight Towards the Goal The first, and from our opinion the main factor for a legible robot navigation is that a robot should always move as far as is possible straight towards its goal and react as smoothly as possible to a human. Bortot et al. [6] showed in their experiment that a straight and stereotypical motion leads to higher human performance and well-being. Moreover, this fact was formerly stated by Beetz et al. [5]. Furthermore, in our own aforementioned study [21, 20], where we let participants steer the robot in order to find their preferred robot behavior, we also found that a straight way towards the goal is preferred. Straight lines towards the goal are also fulfilling the efficiency criteria that we mentioned earlier in our review as one factor of legible motion. In addition, our simulator based experiment [19] showed that driving curves or spinning around leads to lower legibility. Participants told us, that they were confused by the strange movements the robot was performing in some trials. In addition, straight lines are also in line with the claim for human like behavior. In a human-human path crossing experiment Basili et al. [4] found that humans do not swerve. Decreasing the velocity in order to avoid a collision was the observed behavior.

We know that this approach is contrary to the results Dragan et al. [10] observed in their experiment. They could show that a sweeping arm motion towards a goal is more legible in terms of goal predictiveness. Nonetheless, we think there is an important difference between legible arm motions without any human interaction and a navigation where a human might cross a robot's path. This is one point that has to be further investigated in the future.

Stereotypical Behavior As suggested by Beetz et al. [5] and validated by Bortot et al. [6] a robot that behaves the same way in similar situations is way more legible than a somehow optimized motion with permanently varying trajectories. Therefore, another very important key factor for a legible navigation algorithm is to produce consistent trajectories.

Additional Gestures Another factor to increase the legibility of robot navigation is the use of additional gestures or motions like gaze, pointing, or torso direction in order to communicate goals and intentions. In navigation especially the direction the robot is heading to could be communicated with gaze cues. This is in line with results of an experiment conducted by Basili et al. [3]. Also Sisbot et al. [24] used additional gaze and head motions in order to increase the legibility of a motion. Furthermore, Takayama et al. [30] suggested to use animation principles in order to make robot behavior more legible. Some participants suggested in a debriefing of a study that one can use winkers at a robot to indicate directional changes. To conclude, every motion or gesture that indicates the robot's direction or goal increases legibility.

The aforementioned key factors are a first step towards a legible robot navigation. This list is not intended to be exhaustive. Further research is necessary to investigate the factors for a legible navigation and a more important issue is the implementation of these factors into a navigation method.

References

1. Rachid Alami, Aurélie Clodic, Vincent Montreuil, Emrah Akin Sisbot, and Raja Chatila. Toward Human-Aware Robot Task Planning. In *AAAI Spring Symposium*, 2006.
2. Philipp Althaus, H. Ishiguro, T. Kanda, T. Miyashita, and H.I. Christensen. Navigation for Human-Robot Interaction Tasks. In *IEEE ICRA*, 2004.
3. Patrizia Basili, Markus Huber, Omiros Kourakos, Tamara Lorenz, Thomas Brandt, Sandra Hirche, and Stefan Glasauer. Inferring the Goal of an Approaching Agent: a Human-Robot Study. In *IEEE RO-MAN*, 2012.
4. Patrizia Basili, Murat Sag, Thibault Kruse, Markus Huber, Alexandra Kirsch, and Stefan Glasauer. Strategies of Locomotor Collision Avoidance. *Gait & Posture*, 2012.
5. Michael Beetz, Freek Stulp, Piotr Esden-Tempski, Andreas Fedrizzi, Ulrich Klank, Ingo Kresse, Alexis Maldonado, and Federico Ruiz. Generality and Legibility in Mobile Manipulation. *Autonomous Robots*, 28(1), 2010.
6. Dino Bortot, Maximilian Born, and Klaus Bengler. Directly or on detours? How should industrial robots approximate humans? In *ACM/IEEE HRI*, 2013.
7. K. Dautenhahn, M. Walters, S. Woods, K.L. Koay, C.L. Nehaniv, A. Sisbot, R. Alami, and T. Siméon. How May I Serve You? A Robot Companion Approaching a Seated Person in a Helping Context. In *ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, 2006.
8. Kerstin Dautenhahn, Sarah Woods, Christina Kaouri, Michael L Walters, Kheng Lee Koay, and Iain Werry. What is a Robot Companion-Friend, Assistant or Butler? In *IEEE/RSJ IROS*, 2005.
9. Frédéric Dehais, Emrah Akin Sisbot, Rachid Alami, and Mickaël Causse. Physiological and Subjective Evaluation of a Human-Robot Object Hand-Over Task. *Applied Ergonomics*, 2011.
10. Anca D Dragan, Kenton CT Lee, and Siddhartha S Srinivasa. Legibility and Predictability of Robot Motion. In *ACM/IEEE HRI*, 2013.
11. Jérôme Guzzi, Alessandro Giusti, Luca M Gambardella, Guy Theraulaz, and Gianni A. Di Caro. Human-friendly Robot Navigation in Dynamic Environments.
12. Alexandra Kirsch, Thibault Kruse, and Lorenz Mösenlechner. An Integrated Planning and Learning Framework for Human-Robot Interaction. In *Workshop on Planning and Plan Execution for Real-World Systems*, 2009.

13. Alexandra Kirsch, Thibault Kruse, E.Akin Sisbot, Rachid Alami, Martin Lawitzky, Dražen Brščić, Sandra Hirche, Patrizia Basili, and Stefan Glasauer. Plan-Based Control of Joint Human-Robot Activities. *KI - Künstliche Intelligenz*, 24(3), 2010.
14. Thibault Kruse, Patrizia Basili, Stefan Glasauer, and Alexandra Kirsch. Legible Robot Navigation in the Proximity of Moving Humans. In *IEEE ARSO*, 2012.
15. Thibault Kruse, Alexandra Kirsch, Emrah Akin Sisbot, and Rachid Alami. Exploiting Human Cooperation in Human-Centered Robot Navigation. In *IEEE RO-MAN*, 2010.
16. Christina Lichtenthäler, Tamara Lorenz, Michael Karg, and Alexandra Kirsch. Increasing Perceived Value Between Human and Robots—Measuring Legibility in Human Aware Navigation. In *IEEE ARSO*, 2012.
17. Christina Lichtenthäler, Tamara Lorenz, and Alexandra Kirsch. Towards a Legibility Metric: How to Measure the Perceived Value of a Robot. *ICSR*, 2011.
18. Christina Lichtenthäler, Tamara Lorenz, and Alexandra Kirsch. Influence of Legibility on Perceived Safety in a Virtual Human-Robot Path Crossing Task. In *IEEE International Symposium on Robot and Human Interactive Communication*, 2012.
19. Christina Lichtenthäler, Tamara Lorenz, and Alexandra Kirsch. Influence of Legibility on Perceived Safety in a Virtual Human-Robot Path Crossing Task. In *IEEE RO-MAN*, 2012.
20. Christina Lichtenthäler, Annika Peters, Sascha Griffiths, and Alexandra Kirsch. Social Navigation - Identifying Robot Navigation Patterns in a Path Crossing Scenario. In *ICSR*. Springer, 2013.
21. Christina Lichtenthäler, Annika Peters, Sascha Griffiths, and Alexandra Kirsch. Be a Robot! Robot Navigation Patterns in a Path Crossing Scenario. In *ACM/IEEE HRI*, 2013.
22. Jim Mainprice, E Akin Sisbot, Léonard Jaillet, Juan Cortés, Rachid Alami, and Thierry Siméon. Planning Human-Aware Motions Using a Sampling-Based Costmap Planner. In *IEEE ICRA*, pages 5012–5017, 2011.
23. Chrystopher L Nehaniv, Kerstin Dautenhahn, Jens Kubacki, Martin Haegele, Christopher Parlitz, and Rachid Alami. A Methodological Approach Relating the Classification of Gesture to Identification of Human Intent in the Context of Human-Robot Interaction. In *IEEE RO-MAN*, 2005.
24. Emrah Akin Sisbot and Rachid Alami. A Human-Aware Manipulation Planner. *IEEE Transactions onRobotics*, 28(5), 2012.
25. Emrah Akin Sisbot, Rachid Alami, Thierry Siméon, Kerstin Dautenhahn, M Walters, and S Woods. Navigation in the Presence of Humans. In *IEEE-RAS International Conference on Humanoid Robots*, 2005.
26. Emrah Akin Sisbot, Aurélie Clodic, Rachid Alami, and Maxime Ransan. Supervision and Motion Planning for a Mobile Manipulator Interacting with Humans. In *ACM/IEEE HRI*, pages 327–334, 2008.
27. Emrah Akin Sisbot, Luis F Marin, and Rachid Alami. Spatial Reasoning for Human Robot Interaction. In *IEEE/RSJ IROS*, 2007.
28. Emrah Akin Sisbot, Luis F Marin-Urias, Rachid Alami, and Thierry Simeon. A Human Aware Mobile Robot Motion Planner. *IEEE Transactions on Robotics*, 23, 2007.
29. Emrah Akin Sisbot, Luis F Marin-Urias, Xavier Broquere, Daniel Sidobre, and Rachid Alami. Synthesizing Robot Motions Adapted to Human Presence. *International Journal of Social Robotics*, 2(3), 2010.
30. Leila Takayama, Doug Dooley, and Wendy Ju. Expressing Thought: Improving Robot Readability with Animation Principles. In *ACM/IEEE HRI*, 2011.

Social Navigation: Context is King
A Play in Three Acts

David V. Lu[1] and William D. Smart[2]

[1] Washington University in St. Louis, St. Louis, MO 63130, USA,
davidlu@wustl.edu,
[2] Oregon State University, Corvallis, OR 97331, USA

Abstract. While many approaches exist to giving context and structure to interactions between humans and robots, one that has gained increasing support in recent years is to view the interactions as theatre. This has the benefits of being strongly communicative in nature as well as giving us the capacity to view the interaction as a complete action that changes over time in a structured way. In this paper, we examine robot navigation in this capacity and discuss how theatrical frameworks can guide the behavior.

1 Introduction

When some people think about how robots move and behave, the characteristics that come to mind are not always flattering, especially given certain media portrayals. Their moves are jerky and have sharp accelerations, as though they were dancing "The Robot." They are unresponsive or slow to respond to the environment around them. They act with very narrow scope, and anything outside of that scope "does not compute." This collection of attributes has come to be represented by the term "robotic" even in non-robotic scenarios. These qualities can be reinforced by the actual behavior of certain types of robots. If the only goal is efficiency, the result is robots programmed to perform specific tasks in efficient ways, resulting in many of the qualities that people have come to expect.

What such work lacks is an appreciation for *context*, i.e. the circumstances of the robot's current situation. The easiest solution is often to only consider the contexts that directly lead to an efficient solution, e.g. what objects are in the way, where is the goal located. With the increasing use of proper human-robot interaction techniques, additional contexts are added to the optimization, such as the effects on nearby people and their impressions of the robot. Taking all contexts into account is likely intractable. To circumvent that, we turn to a multidisciplinary approach to help us sort through a number of different contexts. We use principles derived from the theatre to motivate our work in human-robot interaction. In theatre, all of the physical actions on stage are motivated by contexts, whether in relation to characters' objectives or to the larger structure of the narrative arc.

In this paper we aim to examine the problem of social robot navigation (see Kruse et al. [2] for an excellent survey) and apply the principles of theatre (first discussed in Lu and Smart [3]). In particular, we will show how the navigation task can be broken up to mirror traditional dramatic structures, namely, the three-act play.

2 Three Act Structure

One particular context that is not taken into account in navigation is the robot's progress on its task. Consider the navigation task of moving from one location to another. In order for a robot to navigate in dynamic environments with uncertain elements, navigation algorithms are written to allow the robot to continuously re-plan and change its high-level behavior (the global path) almost instantly. This creates relatively robust behavior, but does not take into account how people viewing the robot's behavior may interpret it. This behavior implements an implicit Markov assumption, in that the robot only takes into account its current state, rather than also including its progress through the action.

One way to explore this context is to confer every physical action the robot performs with a narrative arc, specifically that of the archetypal three act structure. The first act is the exposition, in which the characters and setting are established. Second is the rising action, the main course of action in which the protagonist faces multiple obstacles as they move toward their goal. Then finally, the protagonist will arrive at a point where there is only one possible scenario: the climax in which we see the protagonist either achieve or fail to achieve their goal.

To properly use this structure with a navigation task, we must first define what is meant by "goal". For navigation, it may refer to simply the goal pose of the robot, or the goal pose with some constraints on how to get there (no collisions, minimal path length). In a theatrical context, the term for goal is objective, the motivation behind every action the character does. In a play, if a character crosses the stage, it might be to move away from someone to make someone feel isolated. These objectives are always posed in relation to others and not in isolation. Hence, the robot's objective cannot just be to move from place to place, but to move from place to place in relation to others around it. (Note, in this discussion we will use "objective" to refer to the motivation, and "goal" to refer to the desired pose.)

In the exposition, we need to establish the robot's "character," i.e. what it is capable of and likely to do. In most scenarios, it is impossible and impracticable to endow the robot with as much character as traditional dramatic characters like Austen's Mr. Darcy. Instead, the aim is merely to introduce static qualities of the robot that will be present throughout the action as a way of providing information to help people predict what the robot will do. This could mean exploring the different modalities of the robot (i.e. the different ways the robot can move/act). Establishing this is important, even if the modalities are not functionally necessary, so that if/when the robot employs these behaviors later, they do not come as a surprise to the audience. Furthermore, establishing the type of movements the robot will perform can also be beneficial. Consider the difference between a robot that starts moving in a straight line to its goal, and one that moves more erratically. An observer may think the latter may need more attention or that the former is more deserving of trust. Not only does introducing these qualities early on have the benefit of helping predict future behavior, but also molds an observer's vital first impression.

The first act is also where the robot will begin to move toward its goal, which may require some preparation. The start of the action must be done in a way that is consistent with the robot's objective. For many robot navigators, the objective is simply to move towards the goal. However, in social navigation, the objective includes moving toward the goal in a way that does not disturb the people around it or cause them to be uncomfortable. For some large robots, the simple act of them starting to move their bulk toward the goal can be unquieting. One way around that is to use the additional modalities of the robot besides its mobile base to indicate that the robot is about to move. This could entail moving the head around to ensure the area is clear or a slight raise in the torso to indicate imminent action. This sort of anticipatory gesture is also suggested by Van Breemen [7] and Mead and Matarić [4].

During the middle act, i.e. the bulk of the movement toward the goal, the robot's objective must be to move to the goal, deal with unforeseen obstacles it encounters along its planned path, and to make people aware of those activities in a way that makes them continue to be comfortable. Importantly, the robot should react to the obstacles during this middle act in a way appropriate to the context of the action as a whole. The robot should not stop completely and act as though it were planning a brand new motion from the beginning again. It should react in a way that indicates that it is still pursuing the same goal while taking into account new information about the obstacles. Similarly, the scale of the reaction to unforeseen obstacles needs to be adjusted based on when it happens. One would not expect a robot to react the same way to a change in plans at the beginning of an action than at the end when it is almost at its goal.

The relationship between the robot and the people in the environment is centered around the idea of legibility, i.e. making the robot's actions clear and readable[1, 5]. Legibility is particularly important in this middle act for ensuring a smooth transition between the robot's initial goal and the ultimate outcome of that goal, since illegible behavior could be read as not acting toward that goal. Certain people in the environment require extra consideration since the robot is actively moving in the same space as them. As a result, the robot must take into account contexts related to them, their objectives, and the robot's relationship to those objectives. A person should be treated differently than other mere obstacles, in that they are often mobile and have personal space which it is better not to enter. The person may be particularly sociable and want explicit interaction with the robot. On the other hand, if the person's sole objective is to get to their destination as quickly as possible, then the robot may need to adjust its behaviors for that. If the robot's only objective is its own goal, then it can ignore the other person's objective and move in a way that my detrimental to their goal. However, if the robot's objective does include other people's objectives, then it should move in a way that enables both to complete their objectives.

In the final act, it becomes clear whether the robot has achieved its goal or not. If a robot just stops, it is difficult to determine whether the robot is actually at its goal position or whether there is a problem, especially if the goal is unknown to the people observing the robot. [3] Thus, adding cues like looking at the goal or changing the robot pose in some manner will help indicate the final outcome, making its behavior more legible. One particularly useful approach that has not seen common-place usage in navigation tasks is including success and failure animations to explicitly mark the outcome of the action[6].

3 Discussion

Thinking of the navigation tasks as a three act play as the first step toward a context-sensitive implementation of behavior can help to alleviate the perception of robots as robotic (as defined in the introduction). Robots should ease into the initial motion, to avoid the perception of jerky motion. They should react appropriately when encountering unforeseen obstacles during the middle of their actions. Finally, they should demonstrate their success or failure in order to acknowledge the additional context of the entire action that has come before. We argue that adding in these additional layers of context will make robot navigation behavior more legible and more naturally understood by people.

Additional contextual data will improve behavior in specific contexts. This approach contrasts with the usual aim of creating universally applicable behavior. Instead of pursuing behaviors that work adequate in most situations, we should attempt to create interactions which work particularly well in given situations with given contexts. The failure to do so will lead to human robot interactions that are both homogeneous and mediocre, a phenomenon we term, the heat death of robotics.

Acknowledgements

Many thanks to Annamaria Pileggi for her insights on all things theatrical and to Doug Dooley for his animated ideas.

References

[1] Anca Dragan and Siddhartha Srinivasa. Generating legible motion. In *Robotics: Science and Systems*, June 2013.

[3] This is especially true while testing new iterations of navigation algorithms. Spoken from experience.

[2] Thibault Kruse, Amit Kumar Pandey, Rachid Alami, and Alexandra Kirsch. Human-aware robot navigation: A survey. *Robotics and Autonomous Systems*, 2013.

[3] David V. Lu and William D. Smart. Human-Robot Interactions as Theatre. In *RO-MAN 2011*, pages 473–478. IEEE, 2011. ISBN 978-1-4577-1572-3.

[4] Ross Mead and Maja J Matarić. Automated caricature of robot expressions in socially assistive human-robot interaction. In *The 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI2010) Workshop on What Do Collaborations with the Arts Have to Say about HRI*, 2010.

[5] E.A. Sisbot, L.F. Marin-Urias, R. Alami, and T. Simeon. A human aware mobile robot motion planner. *IEEE Transactions on Robotics*, 23(5):874–883, 2007.

[6] L. Takayama, D. Dooley, and W. Ju. Expressing thought: improving robot readability with animation principles. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 69–76. ACM, 2011.

[7] AJN Van Breemen. Bringing robots to life: Applying principles of animation to robots. In *Proceedings of Shaping Human-Robot Interaction workshop held at CHI*. Citeseer, 2004.