Towards Intuitive Verbal and Non-Verbal Communication for Incidental Robot-Human Encounters in Clinic Hallways

Ivana Kruijff-Korbayová, Johannes Hackbarth, Caspar Jacob, Bernd Kiefer, Matthias Schmitt, Tanja Schneeberger, Tim Schwartz
<first.last>@dfki.de
DFKI, Germany

Hanns-Peter Horn
horn@human-factors-consult.de
HFC Human-Factors-Consult, Germany

Karsten Bohlmann
karsten.bohlmann@ek-automation.com
E&K Automation GmbH, Germany

ABSTRACT

Robots autonomously navigating in public spaces need to use appropriate nonverbal and verbal behaviours to signal their intentions during incidental encounters with bystanders and passersby. We introduce our initial system design concepts regarding social navigation, verbal communication and an avatar face and present our initial experimental observations for robots handling incidental encounters in the environment of a clinic.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; *HCI theory, concepts and models*; *Interaction paradigms.*

KEYWORDS

human-robot interaction; incidental encounters; Wizard-of-Oz; verbal and non-verbal; social navigation; assistance robots

ACM Reference Format:
Ivana Kruijff-Korbayová, Johannes Hackbarth, Caspar Jacob, Bernd Kiefer, Matthias Schmitt, Tanja Schneeberger, Tim Schwartz, Hanns-Peter Horn, and Karsten Bohlmann. 2020. Towards Intuitive Verbal and Non-Verbal Communication for Incidental Robot-Human Encounters in Clinic Hallways. In *Cambridge 2020: 15th Annual ACM/IEEE International Conference on Human Robot Interaction, March 23–26, 2020, Cambridge, UK*. ACM, New York, NY, USA, 4 pages. https://doi.org/???

1 INTRODUCTION

Robots have been used in real world public spaces before, including clinics. However, autonomous operation within an active environment such as a clinic hallway still poses challenges. A robot will encounter clinic personnel, clients and visitors. Particularly the latter two are likely unfamiliar with the robot and thus first/incidental encounter situations are to be expected. We aim to improve the predictability of robot intentions during incidental encounters by investigating and introducing concepts of a hybrid verbal and non-verbal communication towards intuitive interaction.

Figure 1: Robot platforms used in the INTUITIV project

This paper briefly describes our project and use cases, introduces the initial design concepts of our system and presents observations from two preliminary studies addressing incidental encounters.

2 PROJECT AND USE CASES

In our project[1] we study intuitive-nonverbal and informative-verbal robot-human communication. The overall aim is that the robots exhibit behaviours that help to establish understanding and trust in their actions. We study the need for linguistic communication and its appropriate realization in various situations, together with the influence of iconic information in the form of sounds, displays, simulated eye movements, etc. on interpretability and legibility. Our goal is a situation-adapted selection of an appropriate interaction possibility and combination of realization means.

The application scenario is a rehabilitation clinic environment. Mobile robots are to accompany and support the clients and staff in various situations, such as guiding the clients to their room while transporting their luggage or guiding them to their therapy appointments. A stationary robot arm is to hand objects to clients or staff. Fig.1 shows our robots: a motorized walking aid, an omnidirectional transport platform and the arm.

The use cases for the mobile robots include *establishing and breaking off contact* when picking up and departing from a client; *giving indoor route directions* when guiding a client to a location; and *handling incidental encounters* with bystanders and passersby when returning to base (the reception) after leaving a client at a target location. In this paper we focus on the incidental encounters.

Cambridge 2020, March 23–26, 2020, Cambridge, UK

ACM ISBN ???…$??.??
https://doi.org/???

---

[1] INTUITIV: https://www-cps.hb.dfki.de/research/projects/INTUITIV

3 SYSTEM CONCEPT

3.1 Overall architecture

The overall system architecture is designed to be portable between the three robot platforms. The system is distributed and multi-layered with cognitive nodes being decoupled from the actuator controlling modules. For perception and semantic understanding of the environment the robots use Intel Realsense D435 cameras as principal sensors. To process the 3D-pointcloud data we plan to integrate GPU-accelerated Single-Board-Computers like Nvidia Xavier in the next design step. Full person safety under all circumstances is a central requirement for all robots in a populated environment, especially in healthcare. For the omnidirectional transport robot ISO 13482-compliant safety is realized using certified LIDARs and a diverse-redundant supervisor with continuous plausibility checks and a PL/d rated voting.

3.2 Social Navigation

An autonomous robot in a public space needs to comply with implicit social conventions (e.g., driving on the right-hand side), respecting the proxemics of a person [5] and respecting social relations between people [13]. Such behaviour shall ensure more predictable trajectories for people who encounter the robot in a hallway and has already been used in a variety of use cases [2][8][12].

In the clinic environment the robot encounters staff, clients and visitors. The clients often use a mobility aid. Their dependence level may vary from functional independence, when they use their aids for a sense of safety, to functional dependence, when they are in real need of their mobility aid [11]. People with mobility aids have to be robustly recognised as such by the robot when driving autonomously [7] and treated with special care to ensure their safety and comfort. An exemplary behaviour might be the passing of a walking aid user in a hallway at a greater distance and a smaller speed, in order to reinforce that no harm is to be expected from the robot. A constraint for the navigation could be, that users of a walking aid should not be overtaken, because we cannot yet classify their functional dependence on the aid.

A social navigation framework must account for these implicit social conventions to ensure non-verbal communication of the robot's motion intentions. This can be realised for example via additional layers in a costmap [9] already used for navigation. Mobility aid users can be incorporated into the navigation either by increasing the costs in a greater area around this person in the costmap or directly in the planner with an extra cost parameter.

In case of a navigation failure, e.g. when the robot encounters a group of people that it cannot avoid, a recovery behaviour which makes use of speech interaction, may be preferable. Simply turning around and navigating an alternative route or brute-forcing the way through this group could be perceived as irritating and non-intuitive behaviour by bystanders.

3.3 Speech Interaction

The speech interaction component is driven by a reactive dialogue manager implemented using the open-source VOnDA framework [6]. VOnDA was designed specifically for applications such as virtual or robotic assistive agents. The dialogue capabilities are driven by the robot's information state, which contains all the knowledge it has about past interactions with the user, aggregated sensor data as far as it is relevant for the current task, and the user's personal preferences. This knowledge is stored in an RDF/OWL storage layer, which supports traditional reasoning, but also streaming reasoning, which is important for the integration of robot sensor data.

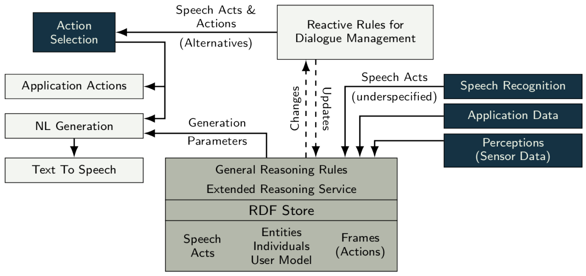

Figure 2: VOnDA Architecture

Changes in the information state trigger a declarative rule system with statistical selection that attempts to compute the most appropriate reaction to the current situation. For this process, it can take into account everything that is contained in the information state, from the urgency of its task to the difficulty of the local navigation state. This allows for producing appropriate user and situation adapted verbal interactions, with the additional benefit that previous interactions can be used as references to give users a feeling of recognition and familiarity.

For ASR and TTS, we use off-the-shelf solutions, viz., Kaldi (https://github.com/kaldi-asr/kaldi) or Nuance Cloud ASR, and Mary TTS (http://mary.dfki.de/). Language interpretation is handled by a hybrid approach using RASA NLU (https://rasa.com/) machine learned models and hand-crafted grammars using an extended version of the SRGS format, which is parsed by our own open-source parser (https://github.com/bkiefer/srgs2xml).

3.4 Avatar

We give the robot a face in order to make its behaviour and its people awareness more transparent, i.e., show that the person is recognized and taken into account in robot navigation.

The abstract 2D-animated face shown in Fig. 3 displayed on a tablet mounted on the robot, uses the robot's gaze direction as a further modality to emphasize intentions referring to the navigation pathway, detected persons and other informative expressions related to objects in the robot environment. The objects depicting the eyes, i.e. pupils and eyebrows, are movable to simulate gaze direction and the mouth is a horizontally and vertically mirrored bar chart for visualization of the speech interaction. Several factors have been considered in this design. First, the Mona Lisa effect

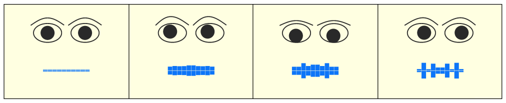

Figure 3: The 2D-animated face model

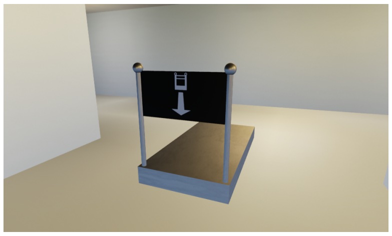

Figure 4: Virtual reality environment for the VR study

[4], i.e., the effect that the Mona Lisa always seems to look at the observer, regardless of where they are standing. If the robot system does not allow the use of head shaped masks as proposed in [1, 3], then indicating target person selection by the avatar's gaze or indicating direction of movement becomes problematic with groups of several passersby. The visualization of a 3D head model on a monitor is a projection of the depth information onto a plane, so when rotating the plane, the observer perceives the projection with a changing scale instead of being shown a rotation of the head model itself. Due to this scaling of the projection, the perceived gaze direction of a picture or photograph works fairly accurately when the observer is standing at the picture's station point but highly decreases when the picture is slanted (i.e. the observer is standing outside the station point), as two studies described in [14] indicate. It seems therefore crucial that the eye-movement of the 2D-animated face is accompanied by a rotation of the robot base, such that the screen showing the 2D face minimizes slanting in respect to the person we are trying to (non-verbally) communicate with or that this effect is being minimized by an alternative approach. We are currently setting up an experiment to investigate how these effects can be overcome.

4 OBSERVATIONS FROM PRE-STUDIES

To inform the early design stages, we gathered information on how humans react to an incidental encounter with a robot in a (simulated) clinic hallway, and consequently, how the robots should behave during these. Doing so at an early stage was crucial, as it allowed for shaping the entire (behavioral) design process. Our observations complement the results of [10], a seminal ethnographic study which examined the integration of an autonomous delivery robot into the workflow of different hospital units from the staff's perspective over long-term. Specifically regarding the use of the physical environment their robot was perceived as taking precedence over people in high traffic and/or cluttered hallways. In our studies we explored various human-passing behaviours.

4.1 Explorative Study

We conducted a preliminary study (12 student participants) to explore interaction patterns and collect a first set of interaction data to shape the multimodal communication module. A room was prepared with barriers to create a corridor of about 6m length and 3m width. In every run of the experiment, one participant enacted the ideal fully autonomous walking aid, pushing the device and communicating as s/he found most appropriate, while three other participants enacted the passersby meeting the robot. All participants were instructed to pretend that the robot could handle any type of interaction and behave in the most appropriate way.

We defined the following situations: (1) pick up a person to guide to another location; (2) robot and person coming from different directions meet (2a) on different sides of the corridor: no conflict; (2b) on the same side: one has to sidestep; (3) robot and person walk in the same direction and (3a) encounter an obstacle on robot's side; (3b) encounter an obstacle on person's side; (4) robot and single person walk in opposite directions and encounter an obstacle; (5) robot approaches group of persons from behind and (5a) there is enough space to overtake them (5a.1) robot's task is not urgent; (5a.2) robot's task is urgent; (5b) there is not enough space (5b.1) robot's task is not urgent; (5b.2) robot's task is urgent; (6) robot approaches a group of people blocking the corridor.

After each situation, every participant filled a questionnaire, stating how comfortable s/he was with the situation, and as how natural and how predictable the robot's action was perceived. In addition, the participant could add suggestions on how the robot's behaviour could be improved. One group session ended with an informal interview to collect further comments and general suggestions concerning the interaction strategy of the robotic walker. The whole procedure took around one hour.

From the questionnaires, the interviews and the video material, we extracted some general guidelines for the behaviour of the robot, as well as verbal interaction patterns that will be used for the dialogue processing pipeline. The following list of requirements was supported by almost all participants: Keep enough distance when passing or overtaking a person; Use a distinct sound when approaching from behind to avoid surprise; Use verbal interaction only when needed, e.g., in unclear situations; Use clearly recognizable default strategies, e.g., always run on the same side of a corridor, always give way to people; Let the robot appear alive, e.g., using a face which always performs little movements.

Opinions were divided on to which extent the robot should react to a person during an encounter considering path planning and social navigation. While some subjects expressed to expect the robot to not abruptly change its path while driving to make room for passersby, others mentioned that the robot should always prioritize a passerby's pathway over its own. There seems to be a difference between subjects in expectations about the intelligence of the robot and in the understanding of an autonomous technical system as a servant. This issue is subject to further investigation during our experiments at the clinic.

4.2 Virtual Reality Study

We also conducted a virtual reality study, in which a sample of 30 participants repeatedly traversed a virtual hallway while a simple transportation robot with omnidirectional moving ability (see Fig. 4) was heading towards them and eventually had to evade them to pass them by. The sample was deliberately chosen to include people from a broad demographic and age spectrum to reflect the variety of different people that could incidentally become "stakeholders" by encountering such a robot in a public space. Participants moved through the virtual hallway by walking through the spacious VR lab with their movement being tracked and translated into the virtual world, therefore both increasing their sense of immersion and at the same time eliminating the common problem of VR sickness.

The robot they encountered in every trial (with the exception of some baseline trials without a robot) was moving according to one of two strategies. Strategy A (“early reaction, full speed") meant that the robot started its evasive movements earlier (approx. 8m away from the participant), while maintaining most of its speed (2.8 km/h) while passing by the participant. Strategy B (“late reaction, careful approach") made the robot swerve comparatively later (approx. 4m), but also involved the platform reducing its speed down to approx. 1 km/h during its approach. A major problem in implementing mobile robots in clinic settings is that oftentimes there is only sparse room for maneuvering. One of the goals thus was to assess the effect and cost of implementing the latter, less spacious navigation strategy. In addition, the robot either signaled its movement intentions via an arrow on its display, via yellow lights resembling turn indicators, or did not signal its intentions at all. Dependent variables measured included, among others, the time to cross the hallway (efficiency), distance chosen by the participant (objective measure for sense of safety), and the amount of trust the participants showed regarding the different robot variations. The study had a repeated-measures design. The participants first completed one block of trials with one of the two strategies where all three signaling methods were applied. After a short break and the completion of a questionnaire, they concluded the study with the second block of trials with the other evasion strategy and again with all three signaling types.

Among the two evasion strategies, Strategy A turned out to produce a significantly shorter time to traverse the hallway, which means that on average participants were more disturbed by the “Strategy B - robots”. However, the difference between the robots, while statistically significant, was quite small (8.06s for Strategy A vs. 8.4s for Strategy B), as was expected due to the overall short completion time of the task. More interestingly, there was no significant difference and not even a recognizable trend in trust ratings for both robot variations. Furthermore, participants accepted a smaller distance from the Strategy B robot (an average of 1.25m vs. 1.35m for Strategy A), which is a sign that the comparatively more aggressive late-evasion of Strategy B could successfully be compensated by the reduced movement speed.

Regarding supportive intention signaling, the turn indicator-lights were superior to both the display arrows (smaller traversal time, overall approval) and especially to the non-signaling robots (traversal time, trust, overall approval), even though most of those differences were rather small in size and overall approval for the robots was rather high in most cases. Two additional details emerged: Firstly, even though the display for the “signaling by arrow” condition was explicitly designed to avoid confusion whether an arrow indicates a movement intention or an request towards the participant to move in the specified direction (the arrow was pointed towards the participant and at its base was a small depiction of the robot itself), several participants were still confused regarding the matter. This is a sign of how difficult a non-ambiguous design for this signaling type is, further emphasizing the benefits of using turn-indicator-like lights which rely on knowledge readily available to most people (i.e., the meaning of an automotive turn indicator blinking). Secondly, even though the study was conducted in VR, we consider the data to be quite valid and applicable to the real world, as all participants reported high or highest immersion while in the simulation, with several participants, against better knowledge, actively trying to physically interact with the robot (e.g., trying to ride on its cargo platform or asking the examiner seriously whether they would be allowed to do so).

5 CONCLUSION AND OUTLOOK

We presented our initial considerations and concept for a system that will handle incidental encounters with bystanders and passersby in a clinic environment. At the time of publication of this paper we have just completed a Wizard-of-Oz study in the clinic.

For the social navigation framework we expect to learn, what velocity and distance constraints and what behaviour while overtaking are appropriate to ensure the comfort of bystanders and passersby during incidental encounters. For the different gaze strategies of the 2D face we investigate how they improve the communication and how the gaze in each situation is perceived by the encountered persons. We also compare the acceptance of overtaking maneuvers with and without a verbal warning.

ACKNOWLEDGMENTS

This work was funded by the German Federal Ministry of Education and Research under grant number 16SV7979 (INTUITIV). We would like to thank all our project collaborators.

REFERENCES

[1] Samer Al Moubayed, Jonas Beskow, Jens Edlund, Björn Granström, and David House. 2011. Animated Faces for Robotic Heads: Gaze and Beyond. In *Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues*, Vol. 6800. Springer, Berlin, Heidelberg, 19–35.

[2] Yu Fan Chen, Michael Everett, Miao Liu, and Jonathan P How. 2017. Socially aware motion planning with deep reinforcement learning. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 1343–1350.

[3] Frédéric Delaunay, Joachim Greeff, and Tony Belpaeme. 2010. A study of a retro-projected robotic face and its effectiveness for gaze reading by humans. In *International Conference on Human-Robot Interaction (HRI 2010)*. 39–44.

[4] Jens Edlund, Samer Al Moubayed, and Jonas Beskow. 2013. Co-present or not? In *Eye Gaze in Intelligent User Interfaces: Gaze-based Analyses, Models and Applications*. Springer, London, 185–203.

[5] Edward T Hall. 1963. A System for the notation of proxemic behavior 1. *American anthropologist* 65, 5 (1963), 1003–1026.

[6] Bernd Kiefer, Anna Welker, and Christophe Biwer. 2019. VOnDA: A Framework for Ontology-Based Dialogue Management. In *International Workshop on Spoken Dialogue Systems Technology (IWSDS)*. Springer, 12.

[7] Marina Kollmitz, Andreas Eitel, Andres Vasquez, and Wolfram Burgard. 2019. Deep 3D perception of people and their mobility aids. *Robotics and Autonomous Systems* 114 (2019), 29–40.

[8] Timm Linder, Stefan Breuers, Bastian Leibe, and Kai O Arras. 2016. On multi-modal people tracking from mobile platforms in very crowded and dynamic environments. In *International Conference on Robotics and Automation (ICRA'16)*. IEEE, 5512–5519.

[9] David V Lu, Dave Hershberger, and William D Smart. 2014. Layered costmaps for context-sensitive navigation. In *International Conference on Intelligent Robots and Systems*. IEEE, 709–715.

[10] Bilge Mutlu and Jodi Forlizzi. 2008. Robots in organizations: The role of workflow, social, and environmental factors in human-robot interaction. *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (2008), 287–294.

[11] Sarah Louise Pampel. 2017. No-go, Slow-go und Go-go sind Geschichte. *Info Diabetologie* 11, 6 (2017), 48–49.

[12] Kun Qian, Xudong Ma, Xianzhong Dai, and Fang Fang. 2010. Socially acceptable pre-collision safety strategies for human-compliant navigation of service robots. *Advanced Robotics* 24, 13 (2010), 1813–1840.

[13] Jorge Rios-Martinez, Anne Spalanzani, and Christian Laugier. 2015. From proxemics theory to socially-aware navigation: A survey. *International Journal of Social Robotics* 7, 2 (2015), 137–153.

[14] Sheena Rogers, Melanie Lunsford, Lars Strother, and Michael Kubovy. 2003. The Mona Lisa effect: Perception of gaze direction in real and pictured faces. In *Studies in Perception and Action VII – Proceedings of the International Conference on Perception and Action*. Lawrence Erlbaum Associates, Publishers, 19–24.