Socially Compliant Human-Robot Interaction for Autonomous Scanning Tasks in Supermarket Environments

Benjamin Lewandowski[1], Tim Wengefeld[1], Sabine Müller[2], Mathias Jenny[2], Sebastian Glende[2], Christof Schröter[3], Andreas Bley[3] and Horst-Michael Gross[1]

***Abstract*— In this paper, we present a system for socially aware robot navigation for a wide range of service tasks in supermarkets. It comprises modules for real-time person detection and tracking to gain situation awareness, modules to react to situations, and means for human-robot communication. The technical performance of the situation awareness was evaluated in a shelf out-of-stock (SOOS) detection scenario under real-world conditions in a supermarket in Germany. Furthermore, in order to investigate whether and to what extent our social navigation strategy can improve the acceptance and application of a mobile service robot in a supermarket, we have conducted surveys with $N = 60$ participants and usability tests with $N = 8$ participants during a three-day field test. We can show that a robot for SOOS detection operating in a supermarket during the opening hours is generally accepted by customers and that the integration of a real-time person perception is crucial, especially for keeping appropriate distances to persons and for improving user-centered communication. Furthermore, our results indicate that various communication channels (e.g. speech, a video projector, and LED lights) are beneficial in order to address a wider user group in the targeted supermarket setting.**

I. INTRODUCTION

An increasingly attractive area for the use of mobile autonomous robots is the retail. Supermarkets in particular offer potential for various robotic service tasks, like guiding [1], cleaning [2], or shelf out-of-stock (SOOS) detection. Since supermarkets in Germany typically have long (up to 10m) and narrow aisles (partly under 1.60m width) equipped with shelves, one challenge is to realize a robot behavior that does not annoy the customers when operating during the opening hours. Such behavior, called human-aware behavior, has to comprise an appropriate situation awareness and suitable responses to situations, including polite waiting and appropriate means for human-robot communication (Fig. 1).

In this paper, we present a system for a socially aware robot behavior in supermarkets. In particular, we introduce our robotic platform (see Fig. 3 right) and describe how it detects situations that occur when performing service tasks within supermarket aisles. Our example scenario is embedded in the ROTATOR project (th**r**ee-dimensional **o**ut-of-stock de**t**ection using **a**u**t**onomous m**o**bile **r**obots, duration: 11/2016 – 10/2019), where a robot performs a SOOS detection as service task. We further present, how the robot reacts to detected situations by adopting its navigation behavior appropriately and by using various complementary kinds of actuators for human-robot communication, like an LED-omnilight, a video projector, an LCD display and speakers. We show results of a technical evaluation of our system for situation awareness conducted under real-world conditions. The development of the robot followed a user-centered-design approach (e.g. like [3]) in order to ensure a high customer acceptance. To proof this concept, we conducted usability tests with $N = 8$ participants and surveys with $N = 60$ participants during an extensive three-day field test in a supermarket in Ilmenau (Germany). With these tests, we aimed to answer the following research questions:

- RQ1: Does a robot for shelf out-of-stock detection require the ability of a socially aware behavior at all?
- RQ2: Does human-robot communication increase the feeling of safety for customers and the acceptance of the robot?

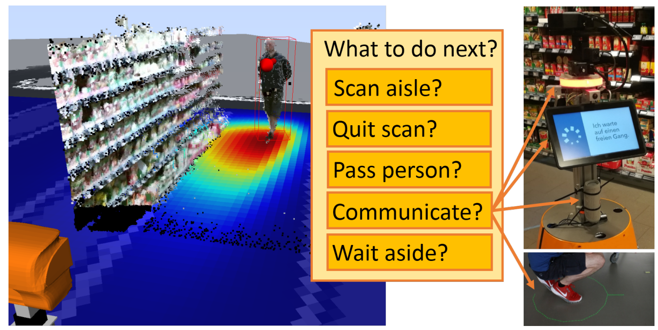


Fig. 1: Left: Supermarket scene (shown as 3D point cloud), a detected person, his/her estimated upper body orientation (red arrow) and a visualization of the asymmetrical personal space. Right: Means of communication on our platform including the omnilight (top) and video projections (buttom).

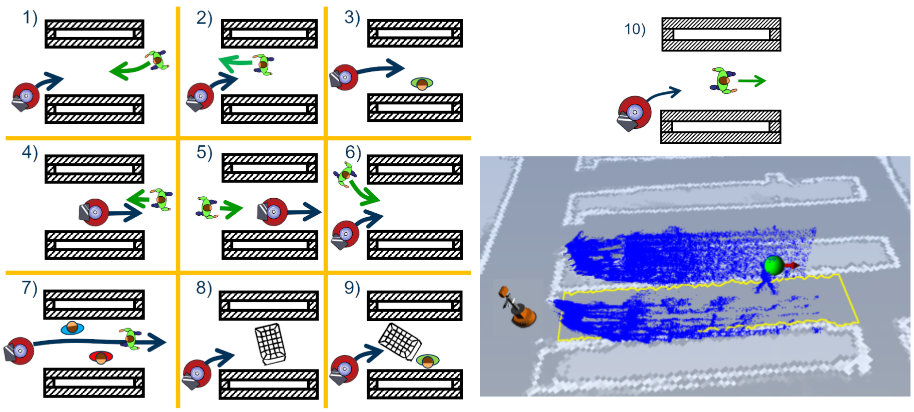
II. RELATED WORK

Mobile robots have already been deployed in a wide range of applications, ranging from public [4] to domestic [5] environments. In supermarkets, robots have been used either as shopping assistant [1], [6] or autonomous cleaning devices [2]. In this domain, a relatively new field for robotic applications is the so-called shelf out-of-stock (SOOS) detection. Large annual losses caused by empty

This work has received funding from the German Federal Ministry of Education and Research (BMBF) to the project ROTATOR (grant agreement no. 03ZZ0437D) in the program Zwanzig20 and to the project FRAME (grant agreement no. 16SV7829K).

[1,2,3]Authors are with Neuroinformatics and Cognitive Robotics Lab, Technische Universität Ilmenau, 98694 Ilmenau (Germany), fg-nikr@tu-ilmenau.de (1); YOUSE GmbH, 13187 Berlin (Germany), info@youse.de (2) and MetraLabs GmbH, 98693 Ilmenau (Germany), robotics@MetraLabs.com (3).

Int. Symp. on Robot and Human Interactive Communication (RO-MAN), Naples, Italy, pp. 363-370, IEEE 2020



	situation	reaction
1)	robot wants to scan, person enters aisle on opposite side	wait & observe
2)	robot wants to scan, person is moving to same side	wait
3)	robot wants to scan, person looks at goods in aisle	wait & observe
4)	robot is scanning, person appears in front	wait on side
5)	robot is scanning, person appears from behind	go on scanning
6)	robot wants to scan, person enters on same side	wait & observe
7)	robot wants to scan, multiple persons in aisle	scan next aisle
8)	robot wants to scan, non-human obstacle blocks scan trajectory	scan next aisle
9)	robot wants to scan, obstacle blocks scan & person in aisle	scan next aisle
10)	robot wants to scan, person is moving to opposite side	start scanning

Fig. 2: Examples for situations that can occur during shelf scanning (1-10). The picture in the middle visualizes situation 10 from the robot's point of view. The point cloud from a Kinect2 sensor is shown in blue, the green ellipsoid represents a tracked person, and the red arrow marks the person's upper body orientation. The free space of the aisle within the occupancy grid map (grey) is framed with a yellow polygon derived from the grid map. The table on the right describes the situations and the corresponding reactions of the robot. Please note that the situation of a free and ready-to-scan aisle (class 0 in the functional tests of Sec. IV) is not specifically listed.

and missing stocks, as reported in [7], should be reduced using mobile robots that autonomously detect these empty stocks to initiate a quick refilling by the supermarket staff. The advantages of such mobile systems [8], [9] are obvious. In comparison to static SOOS systems [10], mobile robots require just a fraction of the amount of sensors a static system needs to monitor large scale super- or hypermarkets. In addition, maintenance is reduced to a single device, which may lead to lower operational costs. Such mobile systems even have made it to purchasable products [11], [12], [13], [14]. However, to be effective, the SOOS detection process needs to be performed during opening hours, where the customer acceptance is a critical factor that cannot be neglected. Socially compliant robots have already been deployed in various scenarios. A robot for the long-term application in an office building and an elder-care facility was developed in [15] that navigates adaptively near humans. In [16] a robot for assistance, information, and guidance of passengers at airports considers human social behavior.

One widely used concept to ensure the fit between robots and humans is usability (e.g. [17], [18]). It originated in the field of human-computer interaction (HCI) and has also been applied in the field of human-robot interaction (HRI). Usability refers to the ease of using a product. The ISO 9241 [19] defines it as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use". For the robotic application described in this paper, the main dimension of usability is satisfaction as the efficiency and effectiveness dimensions need specific tasks to be performed. From a customer's point of view the goal of the robot is to navigate through aisles of a supermarket politely without bothering customers and, thus, we focus on the dimension of satisfaction. Others have proposed frameworks and guidelines for good human-robot interaction specifically for robotic applications. One concept is the USUS Evaluation Framework for HRI [20]. It outlines several factors that determine good and appropriate HRI: usability, social acceptance, user experience, and societal impact with different indicators of each factor. For the application described in this paper, again, because the interaction between customers and this type of service robots in the supermarket should ideally be minimized, and there are no user-relevant tasks to be performed with the application, we consider the factors emotion, feeling of security, and human-oriented perception as the most relevant for a good HRI in our context of use. According to [20], emotion hereby refers to the fact that people tend to interact with computers (and robots) socially, human-oriented perception requires the robot to track human features, and invoking a feeling of security should inform the design of the robotic applications.

In our project, we implemented and evaluated methods to increase the customers' acceptance with approaches for HRI, described in the following. For the context of our application scenario, we specify HRI as the combination of a socially aware robot navigation and human-robot communication.

III. SOCIALLY AND SITUATIONAL AWARE ROBOT NAVIGATION

Typical German supermarkets consist of a series of long and narrow aisles, which normally do not offer much space for large evasive maneuvers. Hence, a deadlock situation can easily occur. Depending on the current task and location of the service robot within the market (inside a narrow aisle or outside), the reaction of the robot to such a deadlock should vary. For a normal *drive-to-task* it is appropriate to plan an alternative route to avoid aisles that are occupied by persons, shopping carts, or cardboard boxes. In contrast, the behavior has to be changed when the robot has to drive in a specific aisle for a particular service task, like the SOOS scanning. Then, the robot should only move to another aisle if it is predictable that the current one will not be passable in the near future. Otherwise, the better option may be to wait in a non-disturbing position or to queue behind persons walking in the same direction (Fig. 2).

For this reason, we implemented two polite navigation strategies that differ depending on the task. While we rely on classic navigation with the dynamic window approach [21] considering the personal space for *drive-to-tasks*, the robot explicitly classifies the current aisle situation for *scanning-tasks* into one of eleven classes and reacts accordingly. All situations, specified in a requirements analysis, are shown in Fig. 2. In the following, we describe all hard- and software modules that are required in order to enable a situation awareness based on this classification of situations. We also show how the robot reacts to detected situations, and how HRI components are applied in this context.

A. System Overview

The developed system has a layered architecture (see Fig. 3 left) similar to that presented in [22]. At the top, the application layer manages transitions between different states, the robot can be in, during SOOS scanning. The states themselves are defined in the behavior layer. There, the robot classifies the current situation and decides which reaction is appropriate (Fig. 2 left). Behaviors make use of modules for person perception and navigation from the underlying skill layer. These skills receive their inputs from the sensor layer at the bottom. Fig. 3 (right) shows the locations of the sensors and actuators used for human-robot interaction on our robot.

B. Detection of HRI-Situations

In Fig. 2, situations that can occur during the SOOS detection differ with respect to obstacles which block the aisle, persons in the vicinity, their upper body orientation within the aisle, and their walking velocity. Hence, the robot has to be able to recognize all these aspects in order to achieve situation awareness.

Person detection: Since persons may occur at far distances to the robot, like on the other entrance of an aisle, as well as in close range, we rely on a combination of three types of sensors and detectors for person detection. Thereby, it is particularly important that the detection runs in real time on mobile hardware in order to react appropriately fast to situational changes. For larger distances, we rely on a Kinect2 mounted on a Pan-Tilt-Unit (Fig. 3). We apply a fast 3D point cloud detector [23] and a skeleton estimator [24] that increases the robustness of the person perception. Both approaches are well suited for detecting persons with occlusions and in body postures that are typical for customers moving in a supermarket, such as squatting or bending down in front of a shelf. The skeleton estimator [24] is also applied on three fisheye cameras to enable a robust 360° close range perception. All skeleton estimators run on NVIDIA Jetson TX2 modules on the robot. As a third type of sensor, we use two planar SICK laser range finders and apply leg detectors [25], [26]. Since these detectors are prone to a high amount of false positives and tend to miss persons occluded by objects, like cardboard boxes, we use them primarily to increase the accuracy of spatial position estimation and the temporal update rate of the succeeding tracker.

Person tracker: To combine all detections into a common representation in world coordinates, we apply a probabilistic person tracker [27]. Tracking improves the robustness of the person perception due to a temporal filtering. Since this involves a data association step, it also fuses different detections of the same person. For each hypothesis, the tracker outputs a location, shown as green ellipsoid in Fig. 2, and a certainty to be a person rather than a false detection. Furthermore, differences in time and location of a trajectory are used to calculate a 2D velocity vector per person.

Upper body orientation estimation: We explicitly compute upper body orientations on point cloud detections with the fast and precise DeepOrientation approach [28] which relies on a deep neural network for a regression of orientation angles. We combine the estimated angles and their uncertainties with tracked velocity vectors in a separate tracking module [29]. The velocity further enables an approximation of orientations in walking direction for all detected persons and is not restricted to the Kinect2 like the explicit approach.

Obstacle detection: For performance reasons, we do not explicitly classify other objects than persons. Instead, obstacles, like cardboard boxes or shopping carts, are detected in general by using the local 2D and 3D obstacle mapping based on 2D laser scanners and RGB-D cameras respectively. All obstacles that do not correspond to a person are treated as objects in the aisles and represent potential blockages.

Classification of aisle situations: Situations are classified using a rule-based system that has been integrated into the state machine of the application. Input is the information described above in combination with information of the aisles to scan and a global 2D occupancy map of the environment. When a certain aisle shall be scanned, a polygon is spanned over the free space in that aisle. Together with information about obstacles as well as the position, velocity, and orientation of persons within and outside of that polygon, the current situation is classified into one of the defined classes shown in Fig. 2 using threshold operations. All rules and thresholds were derived manually by an expert from the specified situations. The free space polygon is calculated with ray tracing on the global 2D map. To this end, the end points of the rays along all normal vectors of the scan-trajectory are determined to span up the polygon. An example for such a free space polygon is shown in the middle of Fig. 2.

C. Reacting to HRI-Situations

While in *scanning mode*, the robot has to react appropriately to detected situations. It has to decide whether to enter the aisle in order to start its scanning task, wait until the aisle becomes free on a non-disturbing position in front of the aisle where it still can observe the aisle, or put the current aisle back into the queue and try to scan the next one instead, e.g. due to blockages by nonhuman objects. If persons appear during a running shelf-scan, the robot immediately stops, turns its omnilight red (Sec. III-D), and drives to the side to

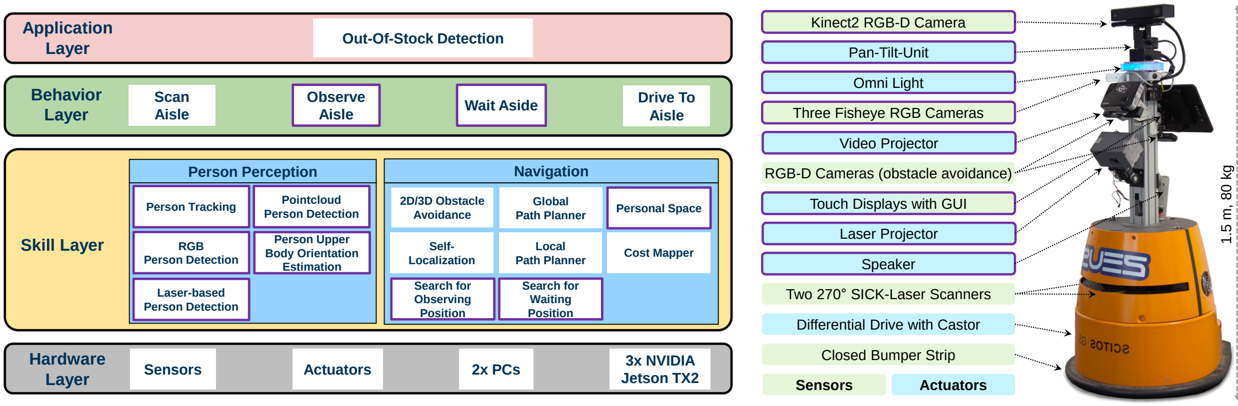


Fig. 3: Left: System overview with submodules. Right: Robot and hardware configuration. In the experimental section, we compare an application configuration with HRI (all software and hardware modules enabled) with a configuration without HRI (just modules without purple boxes enabled). Please note, that in the final application the robot is equipped with additional cameras for SOOS detection (not shown in the picture).

let the person pass. When the remaining (not yet scanned) part of the aisle becomes free again, the scanning is resumed. All reactions are briefly summarized in Fig. 2 (right). Note that a certain reaction can lead to one of the other situations. If so, the situation is classified again and the corresponding reaction is made. To avoid endless loops, a timeout is used that triggers the current aisle to be queued again, e.g. if a person is present in the aisle for a longer period of time. Afterwards the next aisle is processed.

To find suitable waiting positions, we rely on the approach based on a particle swarm optimization (PSO) presented in [30]. As optimization criteria for the PSO, we use, among others, the distance to the aisle, the observability of the aisle as intersection between sensor and aisle polygons, as well as the proximity to the shelves. This should ensure that access to the aisle is not blocked and that customers can get past the robot as easily as possible.

In order to increase the persons' feeling of safety, we rely on the following channels for human-robot communication.

D. Human-Robot Communication

For reacting to situations close to persons as described above, we implemented various complementary cues for human-robot communication on our platform in order to increase the feeling of being perceived by the robot and, thus, to increase the feeling of safety. In addition to short but concise display and male voice outputs, like "I have detected a person, I'm waiting on the side until the aisle becomes free", we also use a mobile video projector and a self-developed LED omnilight for this purpose.

The projector projects the robot's planned trajectory as an arrow to show driving intentions. Furthermore, to signal a person that he/she is perceived by the robot it also projects detected and tracked persons with their upper body orientation as oriented circles into the scene (directly onto the persons' positions), as shown in Fig. 1. The LED omnilight consists of 53 RGB LEDs and is attached to the "neck" of the robot. It can be used for a huge variety of visualizations. For example, the neck can light up in any direction in which a person is tracked, and the color can be selected by the person's tracking ID. However, previous expert tests have shown that such complex visualizations are not clearly comprehensible by persons without technical background. Hence, we have limited its application to simple and constant traffic light colors on all LEDs: green - no person influences the robot's movement, yellow - person is close to the robot but does not influence its movement, red - person in front of the robot has been detected, and the robot stops.

E. Driving with an Asymmetrical Social Space

To make the navigation generally more polite whenever the robot is moving, we have integrated an asymmetric personal space cost function into the driving behavior. The personal space is based on the theory of proxemics [31] and should ensure that the robot keeps a certain distance to persons. It is particularly important that the intimate (up to 0.46m) and personal (up to 1.22m) zones are not entered by the robot if possible. We implemented the personal space according to [32], realizing it as density function that results from the superposition of several Gaussian functions for each detected person. This approach considers the person's position and orientation to weight the areas in front and behind a person differently. In supermarkets, it is helpful to adopt a smaller space behind persons in order to avoid crossing the customers' view on a shelf and, thus, produce less disturbances. Similar to [33], we additionally scale the front space by the person's walking velocity. This should ensure that evasion maneuvers are initiated earlier when persons are moving fast. Furthermore, inspired by [34], we integrated a second asymmetry into the cost function by weighting the right side higher than the left side (Fig. 1 left). Thus, the robot applies a right-hand drive behavior in accordance with the rules of German pedestrian traffic.

IV. FUNCTIONAL EVALUATION

We conducted a functional evaluation of our system for situation awareness with data of a real supermarket (in

Thuringia, Germany). The robot was configured to alternately scan two aisles for noodles and canned food while non-instructed customers were shopping. In total, our system classified 463 real-world situations. Ground truth labels were manually noted by an observer. Results are shown in a confusion matrix in the left of Fig. 4. While most situations were classified correctly, the robot had problems with situations 7 (*multiple persons in aisle* mostly misclassified as *just one person in aisle*), 8 (*obstacle in aisle* mostly misclassified as *aisle is free*), and 9 (*person with shopping cart in aisle* mostly misclassified as *just one person in aisle*). The two major problems were occluded persons and far and empty shopping carts. Due to the long and narrow aisles, people occlude each other to more than 70% across multiple frames and thus, the robot was not able to detect all of them from its point of view, neither with vision nor laser-based detectors. However, since often at least one person was detected the robot's behavior was still socially acceptable most of the times, since it was waiting in front of the aisle according to situation 3, avoiding disturbances. The problem with shopping carts are their thin metal struts, which were barely perceived by our laser and depth-based sensors for obstacle detection. Especially empty shopping carts are difficult to detect. With increasing distance, the problem becomes greater, as the struts are less likely perceived by the sensors. An explicit vision-based detection of shopping carts could provide a remedy here. However, for most cases of situation 9 the robot could at least detect the person (corresponds to situation 3) and thus, it could at least be avoided that the robot enters the aisle.

Since some situations lead to the same reaction, we also evaluated our system with respect to the chosen action. For this purpose, we have summarized the robot's actions as follows: 0 $\widehat{=}$ *start scanning*, 1 $\widehat{=}$ *wait & observe*, 2 $\widehat{=}$ *wait on side*, 3 $\widehat{=}$ *go on scanning*, and 4 $\widehat{=}$ *scan next aisle*. The results are shown in the right of Fig. 4. While most actions were made correctly, it is obvious that often the action *wait & observe* was taken instead of directly driving to the next aisle. This can again be explained by the difficulty of perceiving shopping carts. However, although the robot did not act as intended, it still waited politely in front of the aisle.

V. FIELD TEST EVALUATION

The human-robot interaction was developed following a user-centered design approach, e.g. like in [3]. Prior to the field tests in the supermarket, the robot platform enabled with social navigation was tested in an expert evaluation according to [35]. Findings, like simplifications of voice outputs and visualizations, were then implemented on the platform as part of the iterative development. Furthermore, the maximum speed of the robot was limited to 0.4 m/s, which corresponds to results presented in [36] for approaching persons. The modified platform was tested with users in a supermarket afterwards and compared to a robot without any kind of social navigation or human-robot communication. The goal of the field tests was to evaluate the system and to ensure higher acceptance of the robot among customers in a final product. In surveys, customers ($N = 60$) were asked about

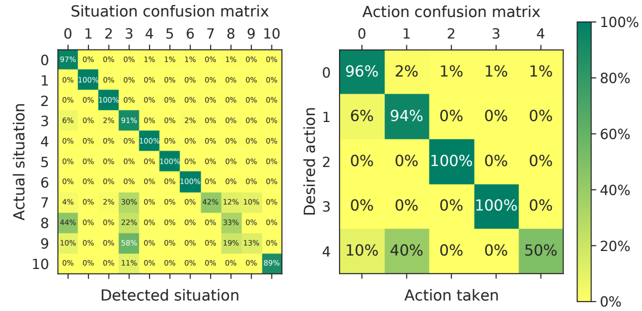


Fig. 4: Results of the functional evaluation with 463 real-world situations. Left: Confusion matrix for actual and detected situations according to Fig. 2. Class 0 represents the situation of a free aisle, ready to scan. Right: Confusion matrix for the chosen action.

their impressions during their observation of the robots. In usability tests, customers ($N = 8$) interacted with the robots.

A. Methods

Design: During the three-day supermarket field test, extensive surveys were carried out which provided information about HRI. The platform presented in Fig. 3, which offered the possibility of social navigation and human-robot communication (HRI) and a similar platform that was not equipped with the corresponding skills (NO_HRI) were tested. While both can perceive obstacles and, therefore, avoid collisions, only the HRI platform is capable of explicitly distinguishing between humans and other objects, which enables person-aware behavior. Concrete differences between both test configurations are highlighted in Fig. 3.

In addition to the robot's onboard safety precautions (i.e. bumpers, an emergency stop switch, and a reactive navigation component), the tests were constantly monitored by a technical staff member who was equipped with a remote control to increase safety during the experiments.

Customer survey: For each of the two robots (NO_HRI and HRI) 30 persons were asked to fill in a survey based on external observations of the robot's behavior. The survey consisted of seven identical items for both robots covering general reactions of the use of such robots in the supermarket. The dimensions covered were:

- perceived distance between human and robot
- anxiety towards the robot
- speed of the robot
- impact on buying behavior

All dimensions were measured on self-developed items (e.g. "The robot distracts me while shopping") and measured on five point likert scales (from 1 = completely disagree to 5 = completely agree). Moreover, for the robot with HRI, additional items have asked about the shoppers' impressions of voice output and how information is communicated.

AttrakDiff2-mini: The general impression of the robots was measured using the AttrakDiff2-mini questionnaire [37]. It covers three dimensions, pragmatic and hedonic quality as well as perceived attractiveness of the robot and is widely used in different fields and also in robotics (e.g. [38]). The pragmatic quality (PQ) essentially corresponds to the task-related usability and describes how certain goals can be achieved effectively, efficiently, and satisfactorily [39]. The hedonic quality (HQ) refers to the user himself and evaluates to what extent the user can identify with the product, i.e. the robot. The attractiveness (ATT) is evaluated based on the perception of a product. The questionnaire consists of ten antonym pairs rated on a 7-point scale.

Usability test: While we wanted to obtain quantitative customer feedback with the surveys based on external observations, we also aimed for qualitative feedback in the sense of a mixed methods approach with longer usability tests ($\approx$ 45 minutes). In these usability tests, participants were asked to go through specific reference situations with both robots by themselves. The aim was to assess how customers react when a robot passes through the aisles during their shopping and does not interact directly with them. In order to keep the test scope manageable, we have selected three out of all reference situations from Fig. 2 that were evaluated:

3) robot wants to scan, person looks at goods in aisle - reaction: wait and observe (HRI), start scan (NO_HRI)
4) robot is scanning, person appears in front - reaction: wait on side (HRI), stop if path blocked (NO_HRI)
5) robot is scanning, person appears from behind - go on scanning (HRI & NO_HRI)

We have chosen these situations because they occur frequently and reflect well the spectrum of possible HRI reactions. Since a customer does not always notice when the robot politely waits in front of an aisle (depending on the distance), only situation 3 with this behavior was added.

The participants were informed about the experiments before they were conducted and could stop them at any time. The order of the robot that was experienced first was randomized between participants. During a test sessions, participants were asked to think aloud.

Participants: In the surveys, $N = 60$ persons were interviewed. The participants who evaluated the robot with HRI were about 49 years old on average ($M_{\text{HRI}} = 49.22; SD_{\text{HRI}} = 22.12$). The participants who evaluated the platform without HRI were about 41 years old on average ($M_{\text{NO\_HRI}} = 41.52; SD_{\text{NO\_HRI}} = 21.14$). The usability tests were conducted with eight persons ($N = 8$) at an average age of 26 years ($M = 26.25; SD = 6.92$). Differences in age were caused by the real-world setting, where older customers were less likely willing to take part in the longer sessions. However, choosing participants randomly from the test environment ensures more unbiased impressions.

B. Results

General reactions: If the ratings from the customer survey of both platforms are considered together, the following results are evident: The respondents could imagine shopping in a supermarket where such a robot is used ($M_{\text{HRI}} = 3.77; SD_{\text{HRI}} = 1.55$), ($M_{\text{NO\_HRI}} = 4.14; SD_{\text{NO\_HRI}} = 1.30$) and do not feel disturbed ($M_{\text{HRI}} = 1.55; SD_{\text{HRI}} = 0.87$), ($M_{\text{NO\_HRI}} = 1.34; SD_{\text{NO\_HRI}} = 0.55$). The robots do not invoke a feeling of fear ($M_{\text{HRI}} = 1.34; SD_{\text{HRI}} = 0.90$), ($M_{\text{NO\_HRI}} = 1.28; SD_{\text{NO\_HRI}} = 0.59$). The speed of the robots was not found to be too fast ($M_{\text{HRI}} = 1.64; SD_{\text{HRI}} = 1.19$); ($M_{\text{NO\_HRI}} = 1.28; SD_{\text{NO\_HRI}} = 0.45$). All of these items were not statistically different between the two platforms HRI and NO_HRI ($p > .05$).

AttrakDiff2-mini: The results of AttrakDiff2-mini show that the robot with HRI was rated with a significantly higher perceived HQ ($M_{\text{HRI}} = 5.05, M_{\text{NO\_HRI}} = 4.47$), ($t(49.92) = -2.17, p = .035$). The pragmatic quality (PQ) was not significantly different between the two robots ($M_{\text{HRI}} = 5.07; M_{\text{NO\_HRI}} = 4.77, p > .05$). The attractiveness (ATT) of the robots was also not rated significantly differently ($M_{\text{HRI}} = 5.04; M_{\text{NO\_HRI}} = 4.96, p > .05$).

Human-robot communication: The descriptive data and the results of the usability test indicate that customers feel slightly more distracted by the HRI ($M_{\text{HRI}} = 3.24; SD_{\text{HRI}} = 1.27$) than by the NO_HRI ($M_{\text{NO\_HRI}} = 3.03; SD_{\text{NO\_HRI}} = 1.12$) although the differences are not statistically significant. However, the robot with social navigation ($M_{\text{HRI}} = 2.30, SD_{\text{HRI}} = 1.37$) was found to be perceived as significantly more intrusive than the NO_HRI robot ($M_{\text{NO\_HRI}} = 1.70, SD_{\text{NO\_HRI}} = 0.67$), ($t(43.05) = -2.12, p = .039$).

The customers were asked about the possibilities of human-robot communication with the HRI robot (Sec. III-D). The voice output was found useful by the customers ($M_{\text{HRI}} = 3.67; SD_{\text{HRI}} = 1.21$), and the information was perceived as clear and unambiguous ($M_{\text{HRI}} = 3.96, SD_{\text{HRI}} = 0.93$). The male voice was rated as pleasant ($M_{\text{HRI}} = 3.96, SD_{\text{HRI}} = 1.31$). The information displayed on the monitor was rated as useful ($M_{\text{HRI}} = 3.85; SD_{\text{HRI}} = 1.13$). Most participants of the usability tests found it helpful that the robot communicates its tasks and actions via a voice output and a display. The consideration of accessible communication, which enables visually impaired and blind as well as hearing impaired people to get all necessary information about the robot, was positively evaluated by the participants.

The voice output of the HRI robot helped to understand the robot's actions. However, it was questioned whether all customers understood the auditory cues. "*I heard the robot was saying something, but I did not fully understand whether this had anything to do with me*" (TP07). Most test persons were disturbed by the fact that the robot without social navigation did not communicate what its task is. Due to this lack of communication, the customers did not know how to interact with the robot. "*I have no idea what it [the robot] is*" (TP05). The HRI-based robot was preferred by seven out of eight test persons, presumably because of its communication.

The test persons also commented on other forms of communication described in Sec. III-D. Many people recognized the importance of the colors red, yellow, green of the signal lamp "*He switched to red as soon as he recognized me*" (TP01). The importance of the traffic light colors of the signal lamp was recognized by many. The detection of persons, which was graphically beamed on the floor, was clear to most and was perceived very positively. For some

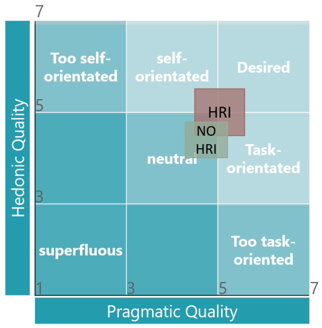


Fig. 5: Ratings of the AttrakDiff2-mini questionnaire [37]. Boxes are 95% confidence intervals.

testers, the contrast to the floor was too weak. For this reason, we developed an alternative to the video projector solution based on the deflection of a laser beam [40]. An evaluation of it in the application environment remains future work.

Social distance: The test of the social distance, which is achieved by combining the social space (Sec. III-E) with the anticipatory reaction to certain situations (Sec. III-C), showed that the distance of the robot with social navigation ($M_{\text{HRI}} = 4.15; SD_{\text{HRI}} = 0.99$) was reported significantly better than the one without social navigation $((M_{\text{NO\_HRI}} = 3.43, SD_{\text{NO\_HRI}} = 1.48), t(47.31) = -2.13, p = .038)$. Most of them noticed that the HRI-based robot recognized the own person and waited at the end of the aisle. "*He works, he sees me, he waits. The robot notices something*" (TP01). While the HRI robot scanned the shelves, most of the test subjects had enough space to shop. "*I had enough space and he didn't bother me*" (TP03). "*He is not threatening. He is not so big too*" (TP08). The distance that the robot kept to the customers was rated positively by nearly everyone. "*What I find even better is that it drives to the side and does not stay in the way (...) you can still walk past it*" (TP06). "*Very far away, [.] that doesn't bother me at all. Very respectful, so to speak*" (TP07). However, almost all participants noticed that the robot NO_HRI did not keep enough distance and came too close to them. "*He is already getting quite close to me. Which can be frightening in any case, if you don't know much about it yet.*" (TP01). "*So a little bit more distance, maybe an arm length, I would find quite appropriate*" (TP07).

C. Discussion

Overall our results indicate that a mobile service robot is generally accepted by customers of almost all age groups, regardless of a social navigation behavior. The maximum robot speed of 0.4 m/s is considered appropriate as it is slow enough so that customers do not feel insecure, but still fast enough to ensure its service tasks in a reasonable time.

In the following, we give answers to the research questions we asked at the beginning.

RQ1: *Does a robot for shelf out-of-stock detection require the ability of a socially aware behavior at all?*
According to the results of our surveys and usability tests, this question can be answered with yes. A robot needs a socially aware communication and navigation behavior in order to keep an appropriate distance to people as far as possible. The robot without social navigation (NO_HRI) came too close to the customers, making them feel uncomfortable. Increasing the minimum distance to obstacles in general is no solution for that problem because the aisles in a supermarket are too narrow and, thus, the robot would not be able to navigate at all. Hence, an explicit person detection and its integration into the navigation is required. Our HRI robot therefore keeps larger distances to humans, if possible, by considering a social space and especially by reacting to situations according to Fig. 2 (right).

However, social navigation has no effect on the pragmatic quality, i.e. on the usability, when using the robot for out-of-stock detection only. This is probably due to the lack of direct interaction of the customers with the robot. It informs the customers audio-visually, but reacts only implicitly to them by recognizing situational changes (see Fig. 2).

It should also be considered that a socially aware behavior including person perception may require extra hardware, depending on the sensor and hardware configuration already used for detecting out-of-stocks.

RQ2: *Does human-robot communication increase the feeling of safety for customers and the acceptance of the robot?*
Also our second research question can be answered with yes. The results indicate that people felt more comfortable in the presence of the HRI robot and they could identify themselves better with it. That means that the HRI robot has a higher hedonic quality than the NO_HRI robot (Fig. 5). At the beginning and without communication, this robot was a quite big, autonomously driving machine to the customers. They did not know its functions, if it was a danger or restricts them during shopping. In contrast, the HRI robot informed people about its actions and, thus, they knew what it was doing at all times. However, the HRI robot was also evaluated as more intrusive. From this, it can be concluded that communication is very important. However, it must be balanced and should not disturb the customers too much while shopping.

In order to reach a large user group, i.e. increasing accessibility also for people with a disability, several easily understandable communication channels are required. Combining speech, mobile video projector, and display outputs enables persons to experience the robots actions, that it recognizes them and that it drives to the side in order to wait in narrow aisles. The simple LED omnilight with adapted traffic light colors (green, yellow, red) as additional visual signal further improves the robot's communication, especially from larger distances without being too intrusive.

VI. Conclusion

We have presented a system for social navigation and human-robot communication for mobile robots in supermarkets, using the example of a shelf out-of-stock (SOOS) detection. It classifies situations that occur during scanning in order to react politely to customers. Anticipatory waiting and an asymmetrical social space ensure that the robot keeps as much distance to persons as possible. The use of speech and display outputs, a video projector, and an LED omnilight allows to communicate information to a large user group and takes accessibility to people with disabilities into account.

Functional tests demonstrated that the situation classification generally works well, but the person detection has sometimes problems with shopping carts or occlusions. Surveys and usability tests we conducted during a three-day field test have shown that customers generally accept robots for scanning tasks. The results indicate that robust person perception is crucial for keeping social distances and to enable user-centered communication. Multiple and easy to understand communication channels are important, but the communication must be balanced and should not be too intrusive in order to disturb customers as little as possible. It remains future work to investigate in long-term experiments which combination of information and communication channels are less disturbing and intrusive, or whether customers get used to them over time.

Although our system was developed and evaluated for the task of SOOS detection, we believe that it can be easily transferred to other tasks in a supermarket-like environment like for cleaning, stocktaking, or guiding purposes. However, it remains future work to further examine such scenarios.

ACKNOWLEDGMENT

The authors would like to thank the Sander family, who made their Edeka supermarket in Ilmenau available for our robot field tests.

REFERENCES

[1] H.-M. Gross, *et al.*, "TOOMAS: Interactive Shopping Guide Robots in Everyday Use – Final Implementation and Experiences from Long-Term Field Trials," in *Proc. of IROS*, 2009, pp. 2005–2012.

[2] H. Endres, *et al.*, "Field test of a navigation system: autonomous cleaning in supermarkets," in *Proc. of ICRA*, 1998, pp. 1779–1781.

[3] J. J. Garrett, *Elements of user experience, the: user-centered design for the web and beyond*. Pearson Education, 2010.

[4] H.-M. Gross, *et al.*, "Mobile Robot Companion for Walking Training of Stroke Patients in Clinical Post-stroke Rehabilitation," in *Proc. of ICRA*, 2017, pp. 1028–1035.

[5] H. M. Gross, *et al.*, "Living with a Mobile Companion Robot in your Own Apartment - Final Implementation and Results of a 20-Weeks Field Study with 20 Seniors," in *Proc. of ICRA*, 2019, pp. 2253–2259.

[6] A. Marin-Hernandez, *et al.*, "Conception and Implementation of a Supermarket Shopping Assistant System," in *11th Mexican Int. Conf. on Artificial Intelligence*, 2012, pp. 26–31.

[7] T. W. Gruen, *et al.*, *Retail out-of-stocks: A worldwide examination of extent, causes and consumer responses*. Grocery Manufacturers of America Washington, DC, 2002.

[8] M. Paolanti, *et al.*, "Mobile robot for retail surveying and inventory using visual and textual analysis of monocular pictures based on deep learning," in *Proc. of ECMR*, 2017, pp. 1–6.

[9] S. Kumar, *et al.*, "Remote retail monitoring and stock assessment using mobile robots," in *Int. Conf. on Technologies for Practical Robot Applications (TePRA)*, 2014, pp. 1–6.

[10] E. Frontoni, *et al.*, "Design and test of a real-time shelf out-of-stock detector system," *Microsystem Technologies*, vol. 24, no. 3, pp. 1369–1377, 2018.

[11] Zebra Technologies Corp. (2020) Zebra Smartsight. [Online]. Available: https://connect.zebra.com/smartsight

[12] Bossa Nova Robotics. (2020) Bossa Nova out-of-stock. [Online]. Available: https://www.bossanova.com/out-of-stocks

[13] MetraLabs GmbH. (2020) Tory Inventory Robot. [Online]. Available: https://www.metralabs.com/en/tory-shelf-automated-shelf-scanning-inventory/

[14] Simbe Robotics, Inc. (2020) Simbe Robotics out-of-stock management. [Online]. Available: https://www.simberobotics.com/solutions/out-of-stock-management/

[15] N. Hawes, *et al.*, "The STRANDS Project: Long-Term Autonomy in Everyday Environments," *IEEE Robotics Automation Magazine*, vol. 24, no. 3, pp. 146–156, 2017.

[16] R. Triebel, *et al.*, "Spencer: A socially aware service robot for passenger guidance and help in busy airports," in *Field and service robotics*. Springer, 2016, pp. 607–622.

[17] O. M.-E. Sucar, *et al.*, "From HCI to HRI - usability inspection in multimodal human - robot interactions," in *12th Int. Workshop on Robot and Human Interactive Communication, ROMAN*, 2003, pp. 37–41.

[18] N. Doering, *et al.*, "User-Centered Design and Evaluation of a Mobile Shopping Robot," *Int. Journal on Social Robotics (JSR)*, vol. 7, pp. 203–225, 2015.

[19] *ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability*. International Organization for Standardization, 1998.

[20] A. Weiss, *et al.*, "The USUS evaluation framework for human-robot interaction," in *Symp. on new frontiers in human-robot interaction*, vol. 4, no. 1, 2009, pp. 11–26.

[21] D. Fox, *et al.*, "The dynamic window approach to collision avoidance," *IEEE Robotics & Automation Magazine*, vol. 4, no. 1, pp. 23–33, March 1997.

[22] H.-M. Gross, *et al.*, "ROREAS - Robot Coach for Walking and Orientation Training in Clinical Post-Stroke Rehabilitation: Prototype Implementation and Evaluation in Fields Trials," *Autonomous Robots (AR)*, vol. 41, no. 3, pp. 679–698, 2017.

[23] B. Lewandowski, *et al.*, "A Fast and Robust 3D Person Detector and Posture Estimator for Mobile Robotic Application," in *Proc. of ICRA*, 2019, pp. 4869–4875.

[24] Z. Cao, *et al.*, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," in *Proc. of CVPR*, 2017, pp. 7291–7299.

[25] L. Spinello *et al.*, "Human detection using multimodal and multidimensional features," in *Proc. of ICRA*, 2008, pp. 3264–3269.

[26] Ch. Weinrich, *et al.*, "People Detection and Distinction of their Walking Aids in 2D Laser Range Data based on Generic Distance-Invariant Features," in *Proc. of RO-MAN*, 2014, pp. 767–773.

[27] T. Wengefeld, *et al.*, "A Multi Modal People Tracker for Real Time Human Robot Interaction," in *Proc. of RO-MAN*, 2019, pp. 1–8.

[28] B. Lewandowski, *et al.*, "Deep Orientation: Fast and Robust Upper Body Orientation Estimation for Mobile Robotic Applications," in *Proc. of IROS*, 2019, pp. 441–448.

[29] S. Müller, *et al.*, "A Multi-Modal Person Perception Framework for Socially Interactive Mobile Service Robots," *Sensors*, vol. 20, no. 3, p. 722, 2020.

[30] T. Q. Trinh, *et al.*, ""Go Ahead, Please": Recognition and Resolution of Conflict Situations in Narrow Passages for Polite Mobile Robot Navigation," in *Social Robotics*. Springer International Publishing, 2015, pp. 643–653.

[31] E. T. Hall, *The hidden dimension*. Garden City, NY: Doubleday, 1966, vol. 609.

[32] A. Vega-Magro, *et al.*, "Socially acceptable robot navigation over groups of people," in *Proc. of RO-MAN*, 2017, pp. 1182–1187.

[33] P. Papadakis, *et al.*, "Social mapping of human-populated environments by implicit function learning," in *Proc. of IROS*, 2013, pp. 1701–1706.

[34] R. Kirby, *et al.*, "COMPANION: A Constraint-Optimizing Method for Person-Acceptable Navigation," in *Proc. of RO-MAN*, 2009, pp. 607–612.

[35] E. Clarkson *et al.*, "Applying Heuristic Evaluation to Human-Robot Interaction Systems." in *Flairs Conference*, 2007, pp. 44–49.

[36] M. Buss, *et al.*, "Towards proactive human-robot interaction in human environments," in *Int. Conf. on Cognitive Infocommunications (CogInfoCom)*, 2011, pp. 1–6.

[37] M. Hassenzahl *et al.*, "The Inference of Perceived Usability From Beauty," *HumanComputer Interaction*, vol. 25, no. 3, pp. 235–260, 2010.

[38] A. D. Frederiks, *et al.*, "Towards Participatory Design of Social Robots," in *IFIP Conf. on Human-Computer Interaction*. Springer, 2019, pp. 527–535.

[39] S. Diefenbach *et al.*, "Handbuch zur fun-ni toolbox," *User Experience Evaluation auf drei Ebenen. Retrieved May*, vol. 27, p. 2019, 2010.

[40] T. Wengefeld, *et al.*, "A Laser Projection System for Robot Intention Communication and Human Robot Interaction," in *Proc. of RO-MAN*, 2020.