



# **Towards Socially Intelligent Robots In Real World Applications: Challenges and Intricacies**

Melanie Jouaiti, Sera Buyukgoz, Eloïse Zehnder, Steven Lawrence, Amir Aly,  
Kerstin Dautenhahn

## **► To cite this version:**

Melanie Jouaiti, Sera Buyukgoz, Eloïse Zehnder, Steven Lawrence, Amir Aly, et al. (Dir.). Towards Socially Intelligent Robots In Real World Applications: Challenges and Intricacies. 2022. hal-03763780

**HAL Id: hal-03763780**

**<https://hal.science/hal-03763780>**

Submitted on 29 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Proceedings of the Workshop

Towards Socially Intelligent  
Robots In Real World  
Applications: Challenges And  
Intricacies

in Conjunction with the 31st IEEE  
International Conference on Robot &  
Human Interactive Communication  
(Ro-Man 2022)

Melanie Jouaiti, Sera Buyukgoz, Eloise  
Zehnder, Steven Lawrence, Amir Aly,  
and Kerstin Dautenhahn

# Social Robot Navigation with Adaptive Proxemics Based on Emotions

Baris Bilen<sup>1</sup>, Hasan Kivrak<sup>2</sup>, Pinar Uluer<sup>3</sup>, Hatice Kose<sup>1</sup>

**Abstract**—The primary aim of this paper is to investigate the integration of emotions into the social navigation framework to analyse its effect on both navigation and human physiological safety and comfort. The proposed framework uses leg detection to find the whereabouts of people and computes adaptive proxemic zones based on their emotional state. We designed several case studies in a simulated environment and examined 3 different emotions; positive (happy), neutral and negative (angry). A survey study was conducted with 70 participants to explore their impressions about the navigation of the robot and compare the human safety and comfort measurements results. Both survey and simulation results showed that integrating emotions into proxemic zones has a significant effect on the physical safety of a human. The results revealed that when a person is angry, the robot is expected to navigate further than the standard distance to support his/her physiological comfort and safety. The results also showed that reducing the navigation distance is not preferred when a person is happy.

**Index Terms**—Social Robots, Socially Aware Robot Navigation, Emotions, Affective Robot

## I. INTRODUCTION

Socially aware navigation is an active and expanding topic of study that brings together human-robot interaction, perception, and motion planning topics. It is crucial for mobile robots to be able to navigate in a social environment without endangering humans' physical safety. In order to perform and navigate in human-inhabited areas, a mobile robot needs to consider humans not only as an obstacle but as social beings. For this reason, a socially-aware navigation approach that integrates social norms and cues is needed. Most of the state-of-the-art robot navigation algorithms perform inherently in human-populated environments while assuming that all humans are in the same state of mind. However, emotions have a significant impact on our behaviours in our daily lives. Different emotions changes our boundaries in the physical movement space [1]. Therefore, emotion-aware navigation strategies might play an important role for mobile robots to improve their behaviours while considering human safety and comfort.

In this paper, we designed and developed a social navigation framework for mobile robots to navigate in a human-

inhabited environment while interpreting human emotions and integrating them into the decision mechanism of human proxemics. We evaluated the effect of this system and also a survey is conducted with 70 participants. The results show the significant effect of integrating emotions into the decision mechanism of robot during its navigation.

## II. RELATED WORK

Robot navigation with social awareness is beneficent for assisting mobile robots in generating socially acceptable behaviours in social environments. To create socially acceptable behaviours, one has to understand the behaviours of humans in a social environment. Edward T. Hall, et al. proposed proxemic criteria for humans defining 4 zones: (i) intimate, (ii) personal, (iii) social and (iv) public [2]. For a robot to have socially acceptable behaviours, it has to avoid entering the personal space of a human. Robot navigation methods that allow a mobile robot to approach people have been thoroughly researched in past years. There are algorithms that widely used for path planning on a static environment such as Dijkstra, A\*, D\* [3], D\* Lite [4] and RRT\* [5]. Also, the aforementioned methods are studied for the static and dynamic environment but none of them manages to achieve social navigation. A mobile robot may approach a person who is standing, moving, or sitting using conventional human-approaching frameworks. The dynamic window approach (DWA) [6] offers a collision avoidance mechanism for mobile robots in dynamic environments. The DWA selects the best way to approach or pass humans by deriving an approximation of the robots' and humans' trajectories. The researchers investigated the preference of people on how to be approached while sitting and presented a cost-based path planning strategy to mimic these preferences [7]. In another study, the authors proposed a framework that allows a mobile robot to use spatial information to establish and sustain a dialogue with standing individuals [8]. These frameworks are mostly designed to approach a single person. Although they can be used to avoid obstacles and create collision-free trajectories in human-inhabited environments, they don't offer a physical safety for humans.

One other way to approach humans is using proxemics theory-based costmaps. Robot operating system (ROS) [9] offers layered costmap plugins and allow users to work with separate layers such as static map, obstacle, inflation, and other user-specified layers (e.g., socially costmap, range sensor). This plugin can subscribe to objects and people and then alters the costmap by adding Gaussian costs around the subscribed objects and people. By the use of these additional costs, robots

<sup>1</sup>Dept. of AI and Data Engineering, Istanbul Technical University, TURKEY {bilenb20, hatice.kose}@itu.edu.tr

<sup>2</sup>Dept. of Computer Engineering, Karabuk University, TURKEY kivrakh@karabuk.edu.tr

<sup>3</sup>Dept. of Computer Engineering, Galatasaray University, TURKEY puluer@gsu.edu.tr

During this study, Baris Bilen was supported by Turkcell-Istanbul Technical University researcher funding program. This study is partially supported by Roborehab TUBITAK project under the contract number 118E214.

can navigate around people in collision-free trajectories by respecting their personal zones [10]. The social force model (SFM) [11] is another method to describe the motions of pedestrians. SFM is not just used for foreseeing the behaviours of pedestrians but also can be used as a local planner [12].

One other approach proposes a learning scheme that can learn the navigational behaviours of people by observing them. This approach has the benefit of adapting to different conditions and increasing its efficiency over time simply by learning on observations. This approach attempts to use human behaviour model on the local planner. Deep reinforcement learning is used in [13] for a mobile robot to learn human navigational behaviour using DQN. [14] uses actor-critic method for reinforcement learning to create a model for mobile robot. [15] extends the time elastic band approach by predicting human behaviours and using this information to create a social navigation planner. Furthermore, people's mood and emotional status can affect their behaviours. For example, in a negative mood, baseline walking speed of pedestrians is likely to decrease [16]. This can also affect their preferred distance in a proxemic zone depending on which mood they are on, i.e., personal space of a person might decrease or increase depending on their mood [1]. There are several studies involving the use of human emotions into the robot's proxemics in social navigation literature. One of them [17] uses emotion estimation from both faces and trajectories of pedestrians to create a trajectory for a humanoid robot. [18] proposes an adjustable proxemic zone depending on people's mood. Although this is a similar work to our work, we also conducted a more comprehensive survey to investigate the adaptive proxemic zones depending on emotions between human-human and human-robot interaction, as well as the expectations for children and the elderly in the same situations.

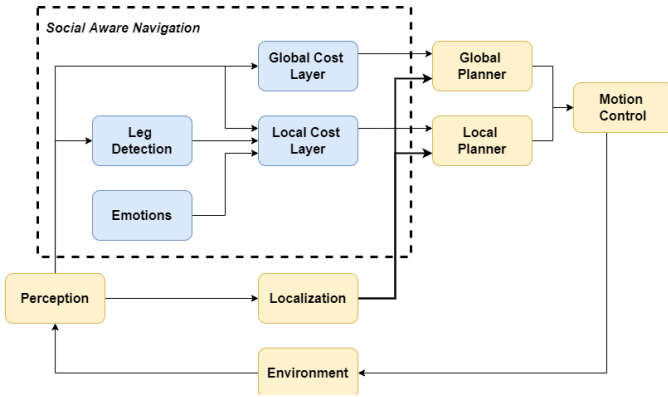


Fig. 1. Navigation framework scheme for mobile robots. Consists of 2 parts; 1) Social aware navigation framework and 2) conventional navigation framework.

### III. FRAMEWORK

This paper proposes a framework that can integrate adaptive social proxemics and emotions into the navigation framework to see the effects of emotions on a mobile robot's movements and also to analyse its effects on physical safety of humans.

For this to be achieved, the robot needs to locate humans in the environment. Figure 1 shows the architecture of the proposed navigation system. The system consists of two parts: 1) conventional navigation framework and 2) socially aware navigation framework. In the first part, the framework consists of perception, localization, planners (global and local) and motion control function blocks. The second part consists of leg detection, emotion and costmap layer (global and local) function blocks.

#### A. Leg Detection

In this paper, a leg detection system [19] is used. The basic idea of this system is to use laser scan info to determine if there is a human leg nearby. The algorithm processes the laser scan data with machine learning methods to inform about possible legs.

#### B. Emotions

We use emotions to increase or reduce the diameter of social proxemic zone depending on the people's emotions. 3 different emotions has been used: 1) Happy, 2) Angry and 3) Neutral. First two emotions are selected from the far edges of valence axis having high arousal components, to be able to observe the clear effects that emotions have on the proxemics and human safety. Selected personal proxemic zone distances for emotions are as follows: for a person with happy emotions, it is 0.5 m, for a person with neutral emotions, it is 1 m [2] and for a person with angry emotions, it is 1.5 m.

#### C. Costmap Layers

In this study, 4 different layers have been used to generate a costmap: 1) Static layer, 2) Obstacle layer, 3) Inflation layer, and 4) Social layer. The static layer represents the unchanged parts of the costmap. The obstacle layer is used to determine and mark the obstacles as read by laser scan. The inflation layer optimizes or adds costs to important or valuable objects to represent a more realistic costmap for the robot to use. And finally, the social layer tracks the whereabouts of the pedestrians as a layer. This layer also represents the personal space specifications of pedestrians i.e., the boundaries of the proxemic zone that the robot should not cross over.

#### D. Global and Local Costmap

Global costmap uses static, obstacle, and inflation layers to create a global costmap for the robot to navigate in. This costmap is used by the global planner. Local costmap uses all 4 layers to create a local costmap. This costmap is generated in the immediate vicinity of the robot and used by the local planner.

#### E. Global and Local Planner

Global planner is used to generating a global path for the robot to track. The global planner uses Dijkstra's algorithm to find the shortest path to the goal while considering the global costmap values. Although the robot can follow the global planner and reach its goal, the local planner ensures that the movements of the robot are within the range of its motors and

wheels to prevent any collision or breakdown. Also, it reacts to the new obstacles that might not be there when the global planner generates the path. A good example would be while the robot is following the global path, a new person is detected on the path or near the path and the robot can react to this situation with the help of local planner.

#### F. Human Safety and Comfort Indices

To validate our work, we used the social individual index (SII) [20]. SII is based on Hall's [2] proxemics. SII is being used to measure the physical safety of each individual. According to SII, physical safety of humans gets in danger if the distance between a robot and a human is smaller than the sum of the area they occupy in the space. Physiological safety is violated if the distance between human and robot is lower than the personal space distance.

### IV. EXPERIMENTAL STUDIES AND RESULTS

In this paper, the proposed framework is simulated using ROS [9] and Gazebo [21] simulation environments.

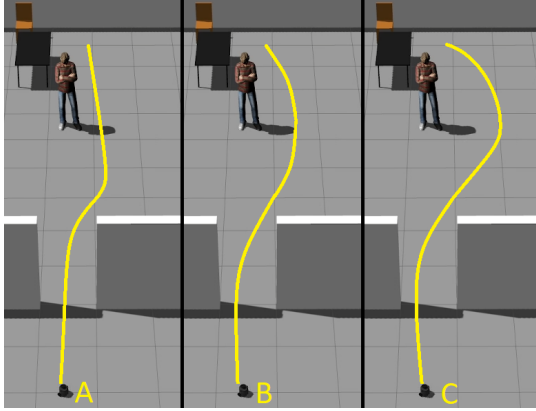


Fig. 2. From left to right movement paths of the robot when the person is Happy (A), Neutral (B) or Angry (C)

#### A. Simulation Study and Results

In this study, we aim to examine the effect created by knowing a person's emotions and adapting proxemics accordingly. A total of 4 different simulations were performed for 2 different emotions. 2 of the simulations were conducted with a person that has known and unknown happy emotions (see Figure 2 Robot A and B). Other 2 simulations were conducted with a person that has known and unknown angry emotions (see Figure 2 Robot C and B). SII metric that is used to evaluate these different simulated actions the robot takes according to the emotions of people, in terms of the people's safety and comfort. We also show the visuals of these simulated actions to the participants and ask them to evaluate the people's safety and comfort subjectively, as reported in the Section IV-B2.

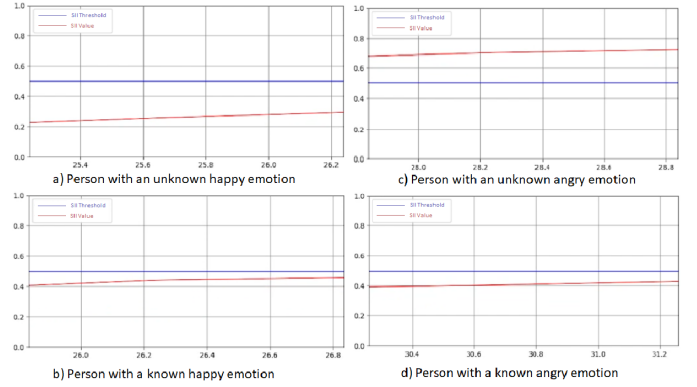


Fig. 3. These plots show the physiological safety of a person within different known and unknown emotions. Blue lines represent SII threshold value and red lines represents measured SII value.

1) *Happy Emotion Simulations:* In happy emotion simulations, we investigated the path created by the robot with known and unknown happy emotion. In the first simulation (See Figure 3.a) we decrease the personal proxemic zones diameter of the pedestrian to 0.5 m to simulate happy emotion but did not allow the robot to adjust its proxemic zone accordingly. We observed that the robot passed through the person without exceeding physiological safety. In the second simulation (See Figure 3.b) we let the robot adjust its proxemic zone according to the emotion of the person and observed that the robot still did not exceed physiological safety. In the first simulation the total path travelled by the robot was 6 m. and in the second simulation the total path travelled by the robot was 5.5 m. We observed that the detection of the emotion of happiness has a positive effect on the path travelled by the robot.

2) *Angry Emotion Simulations:* In angry emotion simulations, we investigated the approach distance of a robot passing by a known and an unknown angry person and its effects on the person's physiological safety. To accomplish this, in the first simulation (See Figure 3.c) we increase the personal proxemic zone diameter of the pedestrian to simulate angry emotion but did not allow the robot to adjust its proxemics and in the second simulation (See Figure 3.d) we let the robot adjust its proxemic zone according to a known angry person. After the simulations, we observed that the detection of angry emotion makes the robot navigate better in terms of physiological safety of humans.

#### B. NARS and Post-Test Questionnaire Results

The survey study was conducted with 70 people (34 female, 36 male). The age distribution of the test participants as follows, 12-18 is 4.2%, 18-30 is 47.9%, 30-45 is 14.1%, 45-65 is 25.4%, 65+ is 8.5%. Also, 95.8% of the people has a bachelor or a higher education degree.

1) *Negative Attitudes Toward Robots Scale (NARS):* We used NARS to evaluate participants' attitudes toward robots, i.e. to find out if they have any prejudice towards robots [22]. NARS is commonly used to evaluate the attitudes of participants in human-robot interaction and to explain the

TABLE I  
THE RESULTS OF ONE SAMPLE T-TEST BASED ON NARS

No.	Subscale	Mean (SD)	t(69)	p values
Q1	S2	3.36 (0.99)	3.01	0.003
Q3	S3	3.10 (1.02)	0.81	0.416
Q4	S1	2.07 (0.95)	-8.02	<0.001
Q5	S3	3.26 (1.04)	2.05	0.043
Q6	S3	3.17 (0.97)	1.46	0.146
Q8	S1	2.51 (1.04)	-3.89	<0.001
Q10	S1	2.98 (1.16)	-0.10	0.918
Q11	S2	2.53 (1.02)	-3.71	<0.001
Q12	S1	2.21 (0.97)	-6.61	<0.001
Q13	S2	2.26 (1.00)	-6.08	<0.001
Q14	S2	2.68 (1.06)	-2.45	0.016

behavioural disparities between them. NARS has 14 items, categorized in 3 sub-scales: S1 (6 items), S2 (5 items), S3 (3 items). S1 is to measure the negative attitude towards interaction with robots, S2 is to measure the negative attitude toward the social influence of robots and S3 is to measure the negative attitude toward emotional interactions with robots. Each item is scored based on a 5-point semantic differential scale (1 being strongly disagree and 5 being strongly agree).

A one-sample t-test is performed for all items shown in Table I based on the participants' answers. Item numbers are written to be equivalent to the item numbers specified in the NARS test. The results showed that participants did not have any priors about the robots, we conducted a one-tailed t-test for each item to see the difference in the corresponding means. Table I shows the items with their statistics.

The t-test results showed that all items from S1 subscale were smaller means and all items from S3 were higher means which was the expected outcome and showed us that participants has positive tendency towards robots. The results showed most of the S2 sub-scale questions means were smaller which was also expected. Even though the average of the Q1 questions had higher means, in another question ("Do you think our emotions affect interactions with robots?") 62% of the participants stated that their emotions are effective in robot interaction.

2) *Post-Test Form*: It is a basic form that includes demographic questions and additional 7 inquiries (Table II) scored on a same semantic differential scale with NARS. Simulated robot behaviours in Figure 2 are also used in the questionnaire. First 2 inquires (I1 and I2) are asked to see that whether the theory presented was also valid among humans. The one sample t-test showed (see Table III) that the theory presented is valid for I2. Then, simulation videos and robot behaviours in Figure 2 are shown to the participants and they were asked to answer the remaining inquiries according to this video and picture. We also conducted a one sample t-test to the rest of the inquiries, which is also shown in Table III. The results of I3 showed that people find the movement of the robot acceptable when a person is in neutral emotion. The I4 and I6 inquiries results showed that participants prefer robots to pass further away when they are angry but interestingly, there is uncertainty among the participants for robots passing closer when a person

TABLE II  
POST-TEST FORM QUESTIONS

No.	Inquiries
I1	It doesn't bother me when people I don't know stand closer to me than usual when I'm happy.
I2	I prefer people to stay at a further distance from me than usual when I am angry.
I3	The distance between the human and the robot B was sufficient as the robot passed the human.
I4	In the picture, the human is disturbed by the distance the robot A has passed.
I5	Robot A's behavior would be correct if the person in the picture was elder or a child.
I6	In the picture, the human is disturbed by the distance traveled by the robot C.
I7	Robot C's behavior would be correct if the person in the picture was elder or a child

TABLE III  
THE RESULTS OF ONE SAMPLE T-TEST BASED ON POST-TEST FORM

No.	Mean (SD)	t(69)	p values
I1	3.14 (1.05)	1.13	0.260
I2	3.98 (0.75)	10.81	<0.001
I3	3.76 (0.80)	7.90	<0.001
I4	3.02 (1.10)	0.21	0.829
I5	2.44 (1.00)	-4.53	<0.001
I6	2.18 (0.68)	-9.75	<0.001
I7	3.42 (0.84)	4.12	<0.001

is happy. Another point, even though the NARS results showed that the participants did not have any negative attitudes towards robot, the results for the inquiries I5 and I7 showed that if the person to be passed is a child or an old person, they preferred the robot to pass from a further distance.

## V. CONCLUSION

This paper presents a socially aware navigation framework that can adjust its proxemic distance depending on a person's emotions to maintain people's physical safety. In this work, emotions of the pedestrians are used to change their personal space during the social navigation task of the robot. This paper investigates the effects of integrating emotions in robot's social navigation on the comfort and safety of humans. The SII metric and questionnaires based on several simulated robot actions are used to evaluate these effects. The results showed that detecting emotions and adjusting personal space depending on the human emotion can create a more comfortable and safer environment for humans.

Although the proposed framework can navigate in a human-inhabited environment while respecting human physical safety, a few changes can be made to make the navigation process more efficient, such as employing cameras to detect the pedestrians and their emotions as well as the laser scan data currently being in use.

This work is aimed at safe and comfortable navigation of Pepper robot among children with hearing impairments in hospitals and audiology centres. In the future, we will further improve the proposed framework by adding a real-time emotion detection module, and employ the whole framework on a physical Pepper humanoid robot to be used in real world human inhabited environments.

## REFERENCES

- [1] Ma, Yingliang, Helena M. Paterson, and Frank E. Pollick. "A motion capture library for the study of identity, gender, and emotion perception from biological motion." *Behavior research methods* 38.1 (2006): 134-141.

- [2] Hall, Edmund T., and Edward Twitchell Hall. The hidden dimension. Vol. 609. Anchor, (1966): 113-129.
- [3] Stentz, Anthony. "The focussed d\* algorithm for real-time replanning." IJCAI. Vol. 95. (1995): 1652-1659.
- [4] Koenig, Sven, and Maxim Likhachev. "D\* lite." Aaai/iaai 15 (2002): 476-483.
- [5] Otte, Michael, and Emilio Frazzoli. "RRT\*: Real-Time Motion Planning/Replanning for Environments with Unpredictable Obstacles." Algorithmic Foundations of Robotics XI. Springer, Cham, (2015). 461-478.
- [6] Fox, Dieter, Wolfram Burgard, and Sebastian Thrun. "The dynamic window approach to collision avoidance." IEEE Robotics and Automation Magazine 4.1 (1997): 23-33.
- [7] Dautenhahn, Kerstin, et al. "How may I serve you? A robot companion approaching a seated person in a helping context." Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction. (2006): 172-179.
- [8] Shi, Chao, et al. "Spatial formation model for initiating conversation." Proceedings of robotics: Science and systems VII (2011): 305-313.
- [9] Robot Operating System, <https://www.ros.org>, 02.08.2022
- [10] Lu, David V., Dave Hershberger, and William D. Smart. "Layered costmaps for context-sensitive navigation." 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, (2014): 709-715.
- [11] Helbing, Dirk, and Peter Molnar. "Social force model for pedestrian dynamics." Physical review E 51.5 (1995): 4282.
- [12] Shiomi, Masahiro, et al. "Towards a socially acceptable collision avoidance for a mobile robot navigating among pedestrians using a pedestrian model." International Journal of Social Robotics 6.3 (2014): 443-455.
- [13] Do, Nam Thang, et al. "Deep reinforcement learning based socially aware mobile robot navigation framework." 2020 7th NAFOSTED Conference on Information and Computer Science (NICS). IEEE, (2020): 226-231.
- [14] Tung, Truong Xuan, and Trung Dung Ngo. "Socially aware robot navigation using deep reinforcement learning." 2018 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). IEEE, (2018): 1-5.
- [15] Khambhaita, Harmish, and Rachid Alami. "A human-robot cooperative navigation planner." Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction. (2017): 161-162.
- [16] Ostir, Glenn V., et al. "Emotional well-being predicts subsequent functional independence and survival." Journal of the American Geriatrics Society 48.5 (2000): 473-478.
- [17] Bera, Aniket, et al. "The emotionally intelligent robot: Improving social navigation in crowded environments." arXiv preprint arXiv:1903.03217 (2019).
- [18] Ginés, Jonatan, et al. "Social navigation in a cognitive architecture using dynamic proxemic zones." Sensors 19.23 (2019): 5189.
- [19] Leg Detection, [http://wiki.ros.org/leg\\_detector](http://wiki.ros.org/leg_detector), 02.08.2022
- [20] Truong, Xuan-Tung, and Trung-Dung Ngo. "'To approach humans?': A unified framework for approaching pose prediction and socially aware robot navigation." IEEE Transactions on Cognitive and Developmental Systems 10.3 (2017): 557-572.
- [21] Gazebo Environment, <https://gazebo.org/home>, 02.08.2022
- [22] Nomura, Tatsuya, et al. "Psychology in human-robot communication: An attempt through investigation of negative attitudes and anxiety toward robots." RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication (IEEE catalog No. 04TH8759). IEEE, (2004): 35-40.

# The Magni Human Motion Dataset: Accurate, Complex, Multi-Modal, Natural, Semantically-Rich and Contextualized

Tim Schreiter<sup>1</sup>, Tiago Rodrigues de Almeida<sup>1</sup>, Yufei Zhu<sup>1</sup>, Eduardo Gutierrez Maestro<sup>1</sup>, Lucas Morillo-Mendez<sup>1</sup>, Andrey Rudenko<sup>2</sup>, Tomasz P. Kucner<sup>3</sup>, Oscar Martinez Mozos<sup>1</sup>, Martin Magnusson<sup>1</sup>, Luigi Palmieri<sup>2</sup>, Kai O. Arras<sup>2</sup>, and Achim J. Lilienthal<sup>1</sup>

**Abstract**—Rapid development of social robots stimulates active research in human motion modeling, interpretation and prediction, proactive collision avoidance, human-robot interaction and co-habitation in shared spaces. Modern approaches to this end require high quality datasets for training and evaluation. However, the majority of available datasets suffers from either inaccurate tracking data or unnatural, scripted behavior of the tracked people. This paper attempts to fill this gap by providing high quality tracking information from motion capture, eye-gaze trackers and on-board robot sensors in a semantically-rich environment. To induce natural behavior of the recorded participants, we utilise loosely scripted task assignment, which induces the participants navigate through the dynamic laboratory environment in a natural and purposeful way. The motion dataset, presented in this paper, sets a high quality standard, as the realistic and accurate data is enhanced with semantic information, enabling development of new algorithms which rely not only on the tracking information but also on contextual cues of the moving agents, static and dynamic environment.

## I. INTRODUCTION

In recent years, the topics of human motion prediction and human-robot interaction have been rapidly growing, driven by the human-aware robotics research and industry interests. Most approaches require plentiful motion data recorded in diverse environments and settings to train on, as well as for the evaluation [1]. Among the growing number of human trajectory datasets, most focus on capturing interactions between the moving agents in indoor [2], outdoor [3] and automated driving [4] settings. These datasets are designed to study the geometric and velocity aspects of human motion.

Human motion is influenced by a large amount of contextual cues, which include semantic attributes of the static and dynamic environment, space topology and its activity patterns, social roles, relations and preferences of the target agents. Studies of these contextual aspects of human motion are gaining traction, creating the need for new datasets containing relevant cues.

<sup>1</sup> Örebro University, Sweden {tim.schreiter, tiago.almeida, eduardo.gutierrez-maestro, yufei.zhu, lucas.morillo, oscar.mozos, martin.magnusson, achim.lilienthal}@oru.se

<sup>2</sup> Robert Bosch GmbH, Corporate Research, Stuttgart, Germany {andrey.rudenko, luigi.palmieri, kaioliver.arras}@de.bosch.com

<sup>3</sup> Mobile Robotics Group, Department of Electrical Engineering and Automation, Aalto University, Finland tomasz.kucner@aalto.fi

This work was supported by the European Union's Horizon 2020 research and innovation program under grant agreement No. 101017274 (DARKO) and the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.



Fig. 1. Laboratory room layout including the floor markings (1) in Scenario 1B. The environment contains various static obstacles, including a narrow corridor (2) in the right with entry limited by a no-entry sign. The table displays the motion capture helmets (3).

In this work, we follow on and further develop the protocol for human motion data collection introduced in [5]. There we proposed a weakly-scripted indoor scenario for generating diverse, natural, and goal-driven human motion in crowded social spaces with static obstacles and a moving robot. The THÖR dataset<sup>1</sup>, recorded according to the proposed procedure, includes 9 participants, moving alone and in groups, whose positions and head orientations are tracked with a motion capture system<sup>2</sup>. The THÖR dataset also includes first-person gaze information for a subset of participants. To diversify the recorded motion patterns, participants in THÖR move between fixed goal positions in the environment, receiving at each goal a random card with the next target. The recording features over 60 minutes of motion and over 600 individual and group trajectories. THÖR is gaining attention in the scientific community, for instance in robotics [6], [7] and predictive motion modeling [8], and serves as a building block for the Atlas motion prediction benchmark [9].

In this paper, we extend THÖR in many aspects. The new recording includes 160 minutes of motion on 4 acquisition days with a total of 30 unique participants. In addition to the static obstacles in the room, we augment the environment with semantic context, such as one-way passages and yellow tape markings for caution areas, which are non-geometric cues that influence motion patterns.

The introduction of the semantic context further enriches the recorded data. Moreover, capturing semantic features enables explainability of motion flow models [10] or enhances

<sup>1</sup><http://thor.oru.se/>

<sup>2</sup><https://www.qualisys.com/>

the downstream tasks with semantics [11]. To further diversify the recorded motion patterns, in addition to cards indicating the next motion goal of the participants, we introduce remote instructions via voice command (using *Discord* [12]). In addition to the gaze directions in the 2D eye-tracker image plane, we also provide 3D gaze vectors in the environment reference frame. In addition to the motion capture and eye-gaze data, we record on-board robot sensors (LiDAR, RGB fish-eye, and RGB-D cameras). Lastly, we propose two variations in the teleoperated robot motion, namely the “differential drive” and “omnidirectional” motion, which enables the study of human-robot collision avoidance under varying conditions.

This paper presents the data collection procedure, describes sensors, scenarios, and the participants’ priming (Sec. II), as well as highlights a portion of the recorded data (Sec. III). We will make the full dataset available in the near future. Once the post-processing is complete, we will systematically describe the recorded data and analyze its application in HRI research.

## II. DATA COLLECTION

### A. Scenarios Description

We designed three scenarios for data collection, which differ in the room layout, motion mode of the robot and the tasks executed by the participants. In all scenarios, participants navigate in the room between 7 goal positions (see Fig. 2).

Scenario 1 is designed as a baseline to capture “regular” social behavior of walking people in a static environment. It has two variations: 1A which only includes static obstacles, and 1B which additionally includes floor markings and stop signs in a one-way corridor. Scenario 1B is designed with the focus on Maps of Dynamics (MoDs) [10]. MoDs are maps that encode dynamics as a feature of the environment, containing information about motion patterns in an environment. MoDs can provide information for planning and navigation in populated environments. The Scenario 1B provides motion data affected by invisible obstacles (safety zones) and flow controlling signs (one-way passages).

Scenario 2 features the same room layout as Scenario 1A (i.e., without semantics). In addition to the basic goal-driven navigation, this scenario introduces people performing different tasks. These tasks aim to emulate regular activities performed in industrial contexts, such as transporting stacks of different objects between various goal locations. Therefore, one participant carries small objects (i.e., a bucket), while another participant carries medium objects (i.e., a box) between two different goal points. Finally, a group of two people moves a large object (i.e., a poster stand) instructed over *Discord* [12].

In Scenario 3, the robot (which remained stationary in the previous scenarios, see its position on the left in Fig. 2) navigates in the room. Scenario 3 has two variations: 3A and 3B. In Scenario 3A, the teleoperated robot moves as a regular differential drive robot. In Scenario 3B, the robot moves in an omnidirectional way.

In all scenarios, we randomly divided the participants into individuals or groups of two or three people who share the

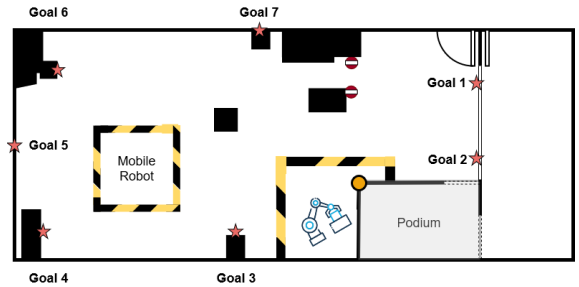


Fig. 2. Layout for Scenario 1B with the focus on Maps of Dynamics. For the other scenarios, we remove the black-yellow striped lane markings. Also, in Scenario 3, the robot on the left becomes a moving obstacle.

Scenario	Description	Robot	Duration
1A	Baseline motion	Static Obstacle	8 minutes
1B	Semantic features	Static Obstacle	8 minutes
2	People with tasks	Static Obstacle	8 minutes
3A	People with tasks	Directional	8 minutes
3B	People with tasks	Omni-directional	8 minutes

TABLE I

SHORT DESCRIPTION OF THE CONDUCTED SCENARIOS, THE MOTION MODE OF THE ROBOT, AND THE DURATION OF RECORDINGS IN ONE DAY

navigation goal. Except for the group carrying the larger object, every group reaching the goal points takes a random card, indicating the number of their next goal. Each group takes one card at a time.

### B. Room Setup

The room for data collection is the robot lab at Örebro University – the same as in the THÖR dataset [5], which creates continuity between the recordings, while allowing to study human motion in the presence of varying contextual factors and obstacle layouts. Fig. 2 depicts the room layout. Seven goal positions are placed specifically to drive purposeful navigation through the room, generating frequent interactions between groups in the center. Several static obstacles (robotic manipulators and tables) are placed in the room to prevent walking between goals in a straight path.

Apart from static obstacles, two robots are placed in the room. One is a static robotic arm placed near the podium, as shown on the right in Fig. 2. The other one is on the left in Fig. 2: an omnidirectional mobile robot with a robotic arm on top (DARKO Robot). The mobile robot is static in scenarios 1A, 1B and 2. In Scenario 3A, the mobile robot navigates as a differential drive robot, while in Scenario 3B, it navigates as an omnidirectional robot. In both cases, an operator drives the mobile robot using a remote controller. In Scenario 3A and 3B, the mobile robot is also used for data collection. The robot base is RB-Kairos+ and the arm is the Collaborative Robot Panda from Franka Emika. The robot base dimensions are 760×665×690 mm. The maximum reach height of the robot arm is 855 mm. The robot has one Ouster OS0-128 LiDAR, two Azure Kinect RGB-D cameras (one used in these recordings), two Basler fish-eye RGB cameras, and two Sick MicroScan 2D safety LiDARs. The Azure Kinect camera has a 75-degree horizontal field of view and a tracking range of up to 5 m.

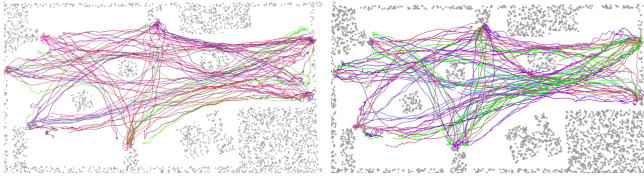


Fig. 3. Recorded trajectories for one run in Scenario 1A (left) and Scenario 1B (right), which includes the environment semantics. In both cases, the room contains various static obstacles, including a narrow corridor in the top right area. Trajectories show that most people would instinctively avoid the “dangerous areas” around the robots, marked with yellow tape (see the layout in Fig. 2).

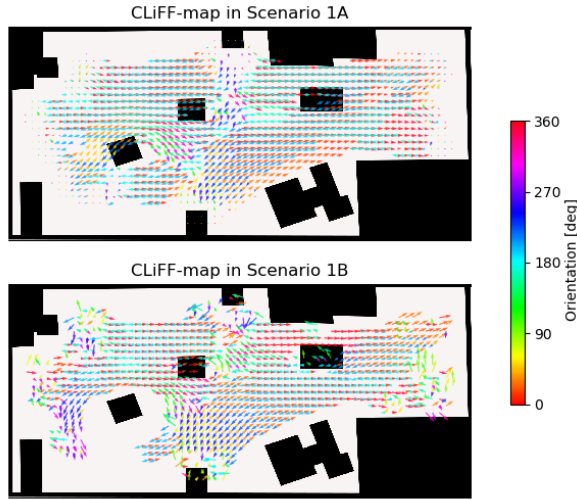


Fig. 4. Maps of dynamics created from Acquisition I - IV (40 minutes) in Scenario 1A (top) and Scenario 1B (down). CLIFF-map [10] is used to represent statistical information about flow patterns.

In Scenario 1B, floor markings and stop signs are added. With black and yellow warning tapes, floor markings are placed around the mobile robot and the robot arm, showing the border of the dangerous zone. Two stop signs are placed near the right permanent obstacle, indicating that the passage from right to left is blocked.

### C. Recording Procedure and Participants' Priming

At the beginning of each session, participants filled out a demographic questionnaire. We recorded in total three different scenarios per session. Scenario 2 had one variation, and Scenario 1 and 3 had two. For each variation, we recorded two runs with a length of 4 minutes each. A summary of the scenarios and duration is given in Tab. I. We always started from Scenario 1B to avoid biasing the participants' motion by letting them observe how the lane markings and the stop signs are prepared. After the two runs of this scenario, we followed with Scenario 1A and Scenario 2. Finally, we proceed with each variation from Scenario 3 in no particular order across the recording days.

After each run, participants fill the Raw version of the NASA-Task Load Index (RTLX) [13], [14]. The scale consists of a 21-point set of sub-scales [1=Low; 21=High], each of which assesses the mental demand, physical demand, temporal demand, and frustration produced by the task as reported by the participants, as well as their self-perceived

performance and frustration. By the end of the session, after the last run of Scenario 3, participants fill out two extra questionnaires with regard to the mobile robot. First, the Godspeed Questionnaire Series [15], a semantic differential set of subscales [5-points] that measures the participants' perception of the robot in terms of anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety, respectively. Second, a 5-point likert scale [1=Strongly disagree; 5=Strongly agree] to evaluate trust towards the robot in industrial human-robot collaborations [16]. The participants filled out all the questionnaires on paper.

Before recording each run, an instructor calibrates the three eye-trackers (Tobii Glasses 2 and 3) and adjusts the gazes for the Pupil Invisible Glasses. The instructor then returns to the stage and sets a 4-minute alarm. We check with the participants if everyone is ready to begin the measurements. If so, we start the recordings of the motion capture system and the eye trackers simultaneously as the instructor counts down to three to signal the participants the start of a run. Additionally, we record rosbag files including sensor data from the robot platform, like the image feed of its onboard RGB and RGB-D cameras and the point cloud recorded by the LiDAR, as well as topics regarding people tracking. After 4 minutes, we simultaneously stop all recordings and the ringing of the alarm signalizes the participants the end of a run.

Between each run, while the participants fill out the questionnaires, we prepare the next run; i.e., we remove the floor markings (after the last run of 1A), set up a phone for the *Discord* voice chat (before 2 and 3), check on the batteries of the eye trackers and potentially change them and finally prepare the robots for Scenario 3. As the participants finish filling out the questionnaire, we shuffle the roles in Scenario 2 and 3 and always assign new groups consisting out of one to three participants for the next run, hereby we always follow the rule, that for groups there can only be one participant with an eye tracker. We assign each group a new goal point to start from at the next run. For the scenarios with roles (2 and 3) we also give a short recap on the task connected with each role, if that participant has not been assigned this role before.

## III. RECORDED DATA

We recorded data on 4 acquisition days for a total of 30 unique participants (9 on Day I, 7 on Days II-IV). As described in Sec. II-A, each acquisition day consists of three different scenarios, and two of them have two different variants. Furthermore, we recorded two 4-minute runs per scenario. Therefore, each acquisition comprises ten runs comprehending all scenarios and yielding 40 minutes of multi-modal data: 3D motion patterns, eye-gaze data from 3 eye trackers, and robot sensor data.

Fig. 3 shows 2D motion trajectories, collected during one 4-minute run in Scenario 1A (left) and Scenario 1B (right). It shows the difference between the two scenarios in areas delimited by the lane markings (see Fig. 2 for the layout reference). Specifically, participants in Scenario 1B tended to navigate farther from the delimited static objects than in Scenario 1A. In addition, Fig. 4 shows the maps of

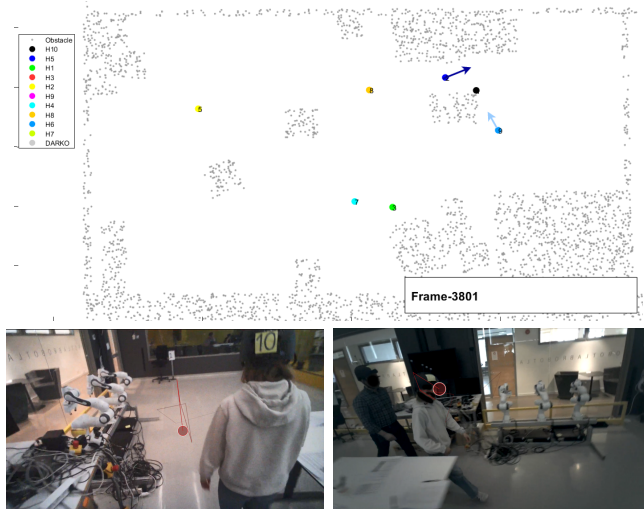


Fig. 5. Eye-gaze vectors, recorded for the participants wearing eye-tracking glasses. **Top:** gaze-vectors mapped into the 3D global map frame for participants 2 (dark blue arrow) and 9 (light blue arrow). **Bottom:** corresponding first-person views for participants 2 and 9. Red line displays the gaze history in the past 2 seconds, and the red circle shows the gaze point in the current frame.

dynamics [10] generated from the collected trajectories from all runs in Scenario 1A and 1B. It shows that in Scenario 1B the flow is less intensive near the “dangerous areas” around the robots. Also, one-way passage flow pattern in the top right corner from Scenario 1B is clearly visible.

Fig. 5 shows the example eye-gazes, recorded for two participants wearing the tracking glasses in the same frame. The 2D gaze direction is provided in the first-person video frame, furthermore we calculate the 3D gaze coordinates in global map frame. Finally, Fig. 6 provides an example of the data recorded with the on-board robot sensors (LiDAR, RGB and fish-eye cameras), displayed in RViz.

#### IV. CONCLUSION AND FUTURE WORK

In this paper we present a new contextually-rich recording of human-robot co-navigation in an indoor environment. The multi-modal data on human motion, collected from the motion capture system, eye-gaze trackers and the on-board sensors of a moving robot, aims to supply the research on human motion prediction, obstacle avoidance, maps of dynamics and human-robot interaction.

In future work we plan to extend the co-navigation scenarios with explicit forms of human-robot communication, for instance by signalling the robot’s intentions, and collaboration, for instance in loading, transporting and unloading the boxes.

#### REFERENCES

- [1] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrilu, and K. O. Arras, “Human motion trajectory prediction: A survey,” *Int. J. of Robotics Research*, vol. 39, no. 8, pp. 895–935, 2020.
- [2] D. Bršćić, T. Kanda, T. Ikeda, and T. Miyashita, “Person tracking in large public spaces using 3-d range sensors,” *IEEE Trans. on Human-Machine Systems*, vol. 43, no. 6, pp. 522–534, 2013.
- [3] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, “Learning social etiquette: Human trajectory understanding in crowded scenes,” in *Proc. of the Europ. Conf. on Comp. Vision (ECCV)*. Springer, 2016, pp. 549–565.

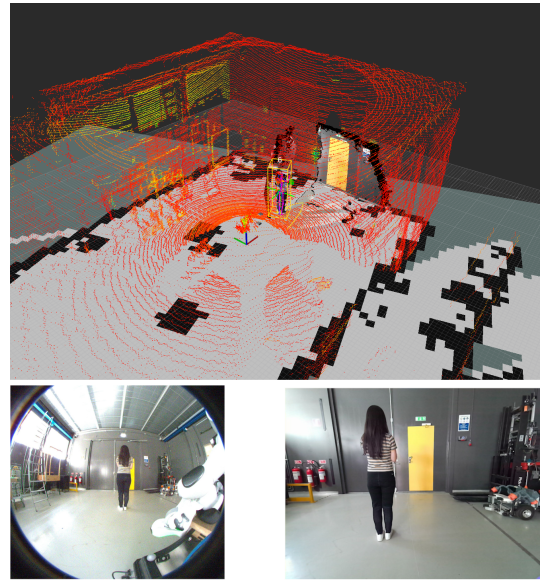


Fig. 6. Data collected from the moving robot. **Top:** 3D visualization of sensor data in RViz: Ouster point cloud shown as red/yellow points, an occupancy grid map for the laboratory and the output of the people tracking module (yellow bounding box) [17]. **Bottom left:** Fish-eye RGB camera image. **Bottom Right:** Azure Kinect RGB-D camera image.

- [4] J. Bock, R. Krajewski, T. Moers, S. Runde, L. Vater, and L. Eckstein, “The ind dataset: A drone dataset of naturalistic road user trajectories at german intersections,” *IEEE*, 2020, pp. 1929–1934.
- [5] A. Rudenko, T. P. Kucner, C. S. Swaminathan, R. T. Chadalavada, K. O. Arras, and A. J. Lilienthal, “THÖR: Human-robot navigation data collection and accurate motion trajectories dataset,” *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 676–682, 2020.
- [6] W. Zhi, T. Lai, L. Ott, and F. Ramos, “Anticipatory navigation in crowds by probabilistic prediction of pedestrian future movements,” in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2021.
- [7] H. Yoon and S. Sankaranarayanan, “Predictive runtime monitoring for mobile robots using logic-based bayesian intent inference,” in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2021.
- [8] W. Zhi, L. Ott, and F. Ramos, “Probabilistic trajectory prediction with structural constraints,” in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 9849–9856.
- [9] A. Rudenko, W. Huang, L. Palmieri, K. O. Arras, and A. J. Lilienthal, “Atlas: a benchmarking tool for human motion prediction algorithms,” in *Proc. of the Robotics: Science and Systems (RSS) Workshop on Social Robot Navigation*, 2021.
- [10] T. P. Kucner, A. J. Lilienthal, M. Magnusson, L. Palmieri, and C. S. Swaminathan, *Probabilistic mapping of spatial motion patterns for mobile robots*. Springer, 2020.
- [11] A. Rudenko, L. Palmieri, J. Doellinger, A. J. Lilienthal, and K. O. Arras, “Learning occupancy priors of human motion from semantic maps of urban environments,” *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3248–3255, 2021.
- [12] Discord Inc., “Discord.” [Online]. Available: <https://discord.com>
- [13] S. G. Hart and L. E. Staveland, *Adv. Psychol.*, vol. 52, pp. 139–183, 1988.
- [14] “Nasa-task load index (nasa-tlx); 20 years later,” 2006, pp. 904–908.
- [15] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, “Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots,” *Int. J. Soc. Robot.*, vol. 1, no. 1, pp. 71–81, 2009.
- [16] G. Charalambous, S. Fletcher, and P. Webb, “The development of a scale to evaluate trust in industrial human-robot collaboration,” *International Journal of Social Robotics*, vol. 8, pp. 193–209, 4 2016.
- [17] T. Linder, K. Y. Pfeiffer, N. Vaskevicius, R. Schirmer, and K. O. Arras, “Accurate detection and 3d localization of humans using a novel yolo-based rgb-d fusion approach and synthetic training data,” in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2020.

# Experimental Design of An Approach for Active Data Collection Using a Mobile Social Robotic Platform

Pedro Custodio, Heitor Carodoso, Rui Pimentel de Figueiredo, Plinio Moreno and Alexandre Bernardino

**Abstract**—The collection of patient reported measures is extremely useful for both patients and health workers since the feedback given by these data saves time, money and lives. For these reasons this collection of data is being adapted to modern technologies. Most of the times this is achieved through mobile apps that although useful for a big portion of the population leave the elderly at a disadvantage. Personal robots offer a more convenient solution for this problem since they are capable of performing these questionnaires actively and directly with the patients. This work presents an active health data collection approach via a mobile robot that autonomously seeks and recognizes patients to perform health data collection through voice and touch-screen based questionnaires.

## I. INTRODUCTION

Consistent health data collection of patient reported measures is crucial for a high quality and close monitorization of a patient's condition or their stage of rehabilitation after a treatment or surgery, and its use is becoming more and more common [1]. These types of data are PROMs (Patient Reported Outcome Measures) and PREMs (Patient Reported Experience Measures) [2]. As the names indicate, the first set of data represents the patient's perception of their own health condition while the second represents the patient's perception of some experience they have been through, be it a surgery, some sort of treatment or even a consultation. This consistent monitorization allows the health professionals to follow a patient closely without the inconvenience of constant check-ups and consultations. Furthermore, it can be fundamental in catching irregularities early on a patient's state and thus prevent future health problems. The 2017 study by Kingsley and Patel [3], shows how these types of data are obtained and their practical importance. The constant collation, interpretation and utilisation of these types of data as a source of feedback also helps reaching an important medium to long term improvement of patient care and general medical services [4].

PROMs and PREMs are usually collected via questionnaires made directly to the patients. Simple statements about the person's health state are made and a set of defined multiple answers is given. The patient, then, has to choose the answer which they think better describes their current condition. In the past, these questionnaires were normally conducted face-to-face by nurses, caretakers or other health professionals. Nowadays, in the digital era, other options have surfaced, and, more and more, this collection of data is done through mobile apps that frequently remind the patients to answer their questionnaires and register their answers for evaluation. These new options have very positive aspects but they still

lack some user friendliness. In this project another solution is presented. With the current evolution in robotics, social robots can be very helpful in this data collection. The main goal is to have a personal robot capable of conducting these questionnaires with a high level of social awareness in order to make the interaction with the patients as smooth and comfortable as possible. Through the rest of this article we present our approaches for active collection of health data using a social robot.

## II. RELATED WORK

Collection of PROMs and PREMs has a wide variety of applications in the medical area, not only for screening and monitorization of patients but also for evaluations and improvements of procedures. One study from 2013 [5] about hernia surgery even shows how these measures can be used to determine the cost-effectiveness of specific procedures and treatments. The application of modern technology in health data acquisition has been explored in many different fronts for a number of years. In the specific case of PROM data collection, most examples of that occur in the form of mobile applications that take advantage of the technological capabilities of smart phones and smart watches to closely monitor patients. In 2016, a randomized study by Johnston et al. [6] showed how a smartphone support App on drug adherence for Myocardial Infarction patients significantly improved their self reported drug adherence and was even associated to an improved cardiovascular lifestyle when compared to the ones not using the app. In a similar study, Seto et al. [7], also proved how the quality of life of patients with continuous monitoring of heart functions through their mobile phones was significantly greater than the one presented by the control group. Also, in 2017 Rudin et al. [8] conducted a study with asthma patients using the mHealthgm App to efficiently identify core components for an mHealth-based asthma symptom-monitoring intervention using patient-reported outcomes. Although this same mHealth App together with a speech component was then studied in 2019 by Welbie et al. [9]. That study was an Usability Assessment and the results revealed that the level of usability still needed improvements before its public launch and that participants less used to using these types of technology had to be taken into account more in the tests. As they were the ones still showing difficulties in using the App. These were normally older adults, i.e. people above 70 years old. For that reason, this is the target population of this project. Not only are they less knowledgeable of these technologies but also, often times, struggle with memory loss problems and

general forgetfulness [10], which lessens the effectiveness of these applications.

Although social robotics is still a relatively novel topic, some examples of projects in this area are emerging as social robots are a very popular tool to use in the medical area. In 2018, the Institute for Systems and Robotics in Lisbon [11], experimented Vizzy, a Humanoid designed specifically for assistive robotics [12], as an exercise coach for the elderly with positive results. Also in 2018, Bauer et al. [13], studied the idea of using a service robot, for the detection of falls in older adults. Thus the application of social robots to the collection of PROMs, specially when dealing with an older demographic, becomes the logical next step.

### III. METHODOLOGIES

This work proposes a solution in which the robot is aware of which PROM questionnaires the user should answer and when they should be answered. When that moment arrives the robot should take the initiative to seek out the user, identify and approach them to initiate the dialog (see Figure 1).

#### A. Robot Platform: Temi

The robot chosen for this work is Temi robot (see Figure 2a). Temi can be controlled through a tablet with a HD touch screen placed in the face area of the robot. This tablet runs java applications on an Android Operating System. The manufacturer provides an SDK to allow developers to control Temi and also to give access to some extra services in the cloud. In terms of sensors, Temi uses a 360° LIDAR, an IMU sensor, 6 time of flight linear sensors for accurate mapping and navigation. And has 3 different cameras. One of this cameras is a high resolution 13MP camera and the others are a wide view camera used for remote navigation and a time of flight depth camera up to 5 meters with a field of view of 90 degrees. Temi has a very advanced voice interaction system.

#### B. Person Localization

All PROM questionnaires have a determined frequency with which they should be answered, thus every time the user has to answer a questionnaire, the robot has to be able to localize and approach the user. For this to happen, three steps are needed, first the robot has to be able to find and approach people, then it has to perform a simple user availability check, which is described in the next section of this chapter, and third of all it should use a face recognition system in order to confirm that the right person is being approached. In case the wrong person is approached, the robot should then go back to the original state of looking for the user in question.

A solution like the one proposed by Song et al. [14] represents a very interesting option for the problem at hands. Although, it also presents some limitations. In this solution the service provided by the robot is requested via a mobile App and the indoor localization is made possible through Bluetooth signals between the robot and the smart phone.

Temi's ability to detect, recognise and follow specific faces works particularly well for person localization. Furthermore,

Temi maps its user's home, and saves the different rooms in a database. Then, when the time to find the user comes, Temi follows a certain path visiting each of the rooms looking for the user. Once in the same room as the user Temi should be able to detect its face and make the final approach.

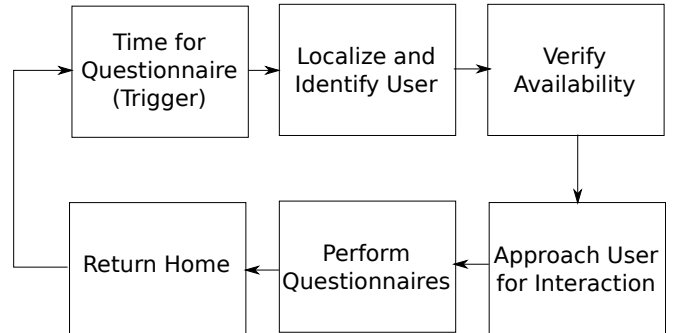


Fig. 1: Block Diagram - Proposed Solution

#### C. Voice interaction

Voice interaction represents the final and most important method of our active data collection solution as it is in this part of the process that, not only the questions are answered and the data is collected but also because this represents the moment in which the Robot's ability to interact with a human is put to the test. A careful presence and approach in the communication might prove to be crucial in the level of acceptance by the users [15].

When first approaching the user, the robot has to greet the user and confirm the availability for answering the questionnaire. Then, the robot should proceed to the questionnaire in need of answering. The PROM questionnaires consist of simple statements to which the user has to attribute a multiple choice answer that better describes their feelings towards the given statement. It is also believed that the Robot should sometimes use reassuring words during these questionnaires according to the answers given by the user.

The target population of this proposal are people in an older demographic. People of this age group tend to be more susceptible to hearing loss and have reduced comprehension ability [16]. Thus it is imperative to present visual aids during the questionnaires, i.e., both the questions and the answers should be visually displayed.

### IV. EXPERIMENTAL METHODS

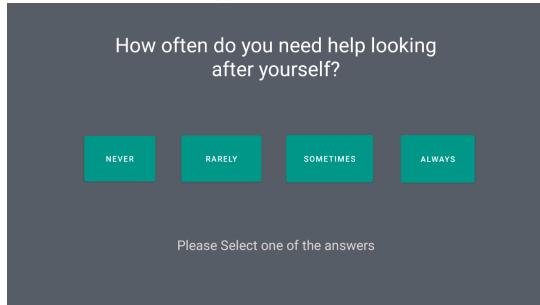
In order to test Temi's abilities within the context of this work, a test application, called temi test dialog, was developed (see Figure 2b).

#### A. Study Design

At first, Temi is trained to recognize the face of a new user. The volunteers are asked to stay in a zone of the laboratory floor of their choosing. After that, Temi is placed in a different zone and asked to look around and find the volunteer in order to start the interaction. The time this process takes and the problems and mistakes made by the robot are closely



(a) Temi Robot



(b) Questionnaire prompt example

Fig. 2: Temi health assistant.

monitored. Once the robot finds and identifies the volunteer, it begins the interaction, proceeding to perform a simple example of a PROM-type questionnaire. Two versions of this App were developed. A more complex one, in which the interaction between user and robot is all done by voice and a simpler one in which the robot will use voice to ask the questions but the user will answer selecting answers in the touch screen. The experiments will take place in two phases, in which both versions are tested and a comparative study between the two is done. Currently, our application provides a simple dialog app, in which, prompts with questions appear on the screen accompanied by a voice dialog, however, we intend to gradually update the application with new functionalities to test.

For now, the questionnaire used is just a simple example used in the development and testing of the app. In the future the focus will be on using real PROM questionnaires. Specifically, the EQ-5D [17], a relatively general PROM questionnaire which can be applied to a variety of illnesses and conditions, and the INTERMED [18], a self-assessment questionnaire, more focused on in-house patients.

## B. Data Acquisition and Analysis

After the questionnaire is finalized and the process terminated, the volunteer then has to answer a quality and usability assessing questionnaire. For this the Post-Study System Usability (PSSU) questionnaire is used alongside some additional questions in order to assess the user acceptance to this solution and to understand which version of the interaction is preferred by users. After the usability assessing experiments are done, an experiment in a more real scenario should also be done. Here the robot is taken to users' homes, preferably users over 70 years old living in their own personal home, as this represents the target audience for this project. Temi is trained to recognize this user and a mapping of the home is also done and then the robot stays in this user's home for a number of days. Temi is programmed to perform a number of PROM questionnaires during each day, some of these questionnaires are taken all in one sitting and for others Temi would let the user take a break and resume the questions later. Every day of the experiment the user is contacted and interviewed about their experience with Temi. This would serve not only to evaluate the general satisfaction of the user with the system but also to understand how their opinion about Temi and the system would develop throughout the experiment (see Table I). The following videos illustrate the robot behaviour when searching for the patients, and also when it begins the interaction to collect the patient answers to the questionnaires (see <https://youtu.be/nXnx8s4TO8I>, <https://youtu.be/wJtlUPpsEyk>).

## V. CONCLUSIONS

In this work we have proposed to use a personal robot assistant for health data collection via PROM and PREM questionnaires, to make this interaction as comfortable and smooth as possible, especially considering that the target population of our active health data collection are elderly people, a series of methods to allow for this interaction were developed, and properly tested in order to guarantee the system is usable and useful. Future work will be focused on developing data collection and analysis tools in order to continuously actively monitor the health of different populations.

## ACKNOWLEDGMENT

The authors would like to acknowledge the financial support of the HAVATAR project (PTDC/EEI-ROB/1155/2020), and the Laboratory of Robotics and Engineering Systems (LARSyS), FCT Project (UIDB/50009/2020).

## REFERENCES

- [1] K. A. Meadows, "Patient-reported outcome measures: an overview," *British journal of community nursing*, vol. 16, no. 3, pp. 146–151, 2011.
- [2] E. M. van der Willik, C. B. Terwee, W. J. W. Bos, M. H. Hemmelder, K. J. Jager, C. Zoccali, F. W. Dekker, and Y. Meuleman, "Patient-reported outcome measures (proms): making sense of individual prom scores and changes in prom scores over time," *Nephrology*, vol. 26, no. 5, pp. 391–399, 2021.
- [3] C. Kingsley and S. Patel, "Patient-reported outcome measures and patient-reported experience measures," *BJA Education*, 2017.

Questions	Answers			
	Never	Rarely	Sometimes	Always
1. How often do you need help looking after yourself?				
2. How about when performing household tasks? Do you often need help?				
3. Do you often feel challenged when getting around your home or community?				
4. Do you ever feel that your health affects negatively your relationships with friends and family?				
5. Regarding your vision, how often do you struggle with seeing clearly?				
6. Do you ever feel any difficulty hearing clearly?				
7. How often do you have a hard time communicating with others?				
8. Do you ever feel difficulty when trying to sleep?				
9. How often do you feel anxious, worried or depressed?				
10. How often do you experience pain or discomfort?				

TABLE I: Example questionnaire used in the experiments

- [4] J. Greenhalgh, R. Pawson, J. Wright, N. Black, J. M. Valderas, D. Meads, E. Gibbons, L. Wood, C. Wood, C. Mills, *et al.*, “Functionality and feedback: a protocol for a realist synthesis of the collation, interpretation and utilisation of proms data to improve patient care,” *BMJ open*, vol. 4, no. 7, 2014.
- [5] S. Coronini-Cronberg, J. Appleby, and J. Thompson, “Application of patient-reported outcome measures (proms) data to estimate cost-effectiveness of hernia surgery in england,” *Journal of the Royal Society of Medicine*, vol. 106, no. 7, pp. 278–287, 2013.
- [6] N. Johnston, J. Bodegard, S. Jerström, J. Åkesson, H. Brorsson, J. Alfredsson, P. A. Albertsson, J.-E. Karlsson, and C. Varenhorst, “Effects of interactive patient smartphone support app on drug adherence and lifestyle changes in myocardial infarction patients: a randomized study,” *American heart journal*, vol. 178, pp. 85–94, 2016.
- [7] E. Seto, K. J. Leonard, J. A. Cafazzo, J. Barnsley, C. Masino, and H. J. Ross, “Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial,” *Journal of medical Internet research*, vol. 14, no. 1, p. e31, 2012.
- [8] R. S. Rudin, C. H. Fanta, Z. Predmore, K. Kron, M. O. Edelen, A. B. Landman, E. Zimlichman, and D. W. Bates, “Core components for a clinically integrated mhealth app for asthma symptom monitoring,” *Applied clinical informatics*, vol. 8, no. 4, p. 1031, 2017.
- [9] M. Welbie, H. Wittink, M. J. Westerman, I. Topper, J. Snoei, and W. L. Devillé, “A mobile patient-reported outcome measure app with talking touchscreen: usability assessment,” *JMIR formative research*, vol. 1, no. 1, p. e1, 2017.
- [10] J. Ballard, “Forgetfulness and older adults: concept analysis,” *Journal of Advanced Nursing*, vol. 66, no. 6, pp. 1409–1419, 2010.
- [11] J. Avelino, H. Simão, R. Ribeiro, P. Moreno, R. Figueiredo, N. Duarte, R. Nunes, A. Bernardino, M. Čaić, D. Mahr, *et al.*, “Experiments with vizzy as a coach for elderly exercise,” in *Proc. Workshop Pers. Robots Exercising Coaching-HRI Conf.(PREC)*, pp. 1–6, 2018.
- [12] P. Moreno, R. Nunes, R. Figueiredo, R. Ferreira, A. Bernardino, J. Santos-Victor, R. Beira, L. Vargas, D. Aragão, and M. Aragão, “Vizzy: A humanoid on wheels for assistive robotics,” in *Robot 2015: Second Iberian Robotics Conference*, pp. 17–28, Springer, 2016.
- [13] J. Bauer, L. Gründel, J. Seßner, M. Meiners, M. Lieret, T. Lechler, C. Konrad, and J. Franke, “Camera-based fall detection system with the service robot sanbot elf,” in *Smart Public Building 2018 Conference Proceedings, Stuttgart, DE*, pp. 15–28, 2018.
- [14] K.-T. Song, P.-C. Lu, and S.-H. Song, “Human-robot interaction design based on specific person finding and localization of a mobile robot,” in *2019 International Automatic Control Conference (CACCS)*, pp. 1–6, IEEE, 2019.
- [15] J. Aydelott, R. Leech, and J. Crinion, “Normal adult aging and the contextual influences affecting speech and meaningful sound perception,” *Trends in amplification*, vol. 14, no. 4, pp. 218–232, 2010.
- [16] N. J. Devlin and R. Brooks, “Eq-5d and the euroqol group: past, present and future,” *Applied health economics and health policy*, vol. 15, no. 2, pp. 127–137, 2017.
- [17] S. Ferrari, A. Van Reedt Dortland, and A. Boenink, “Intermed-self assessment (imsa): Validity and preliminary applications in research,” 2017.
- [18] S. Ferrari, A. Van Reedt Dortland, and A. Boenink, “Intermed-self assessment (imsa): Validity and preliminary applications in research,” vol. 3, no. 3, p. e11617, 2019.