TU Delft

Delft University of Technology

PARSNiP

A Novel Dataset for Better Perceived Appropriateness Detection in Robot Social Navigation with Emotional and Attentional Features

Zhou, Yunzhong; Vroon, Jered; Rusák, Zoltán; Kortuem, Gerd

DOI
10.1007/s12369-025-01266-x

Publication date
2025

Document Version
Final published version

Published in
International Journal of Social Robotics

Citation (APA)
Zhou, Y., Vroon, J., Rusák, Z., & Kortuem, G. (2025). PARSNiP: A Novel Dataset for Better Perceived Appropriateness Detection in Robot Social Navigation with Emotional and Attentional Features. *International Journal of Social Robotics*, *17*(10), 2245-2257. https://doi.org/10.1007/s12369-025-01266-x

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

International Journal of Social Robotics
https://doi.org/10.1007/s12369-025-01266-x

PARSNiP: A Novel Dataset for Better Perceived Appropriateness Detection in Robot Social Navigation with Emotional and Attentional Features

Yunzhong Zhou[1] · **Jered Vroon**[1] · **Zoltán Rusák**[1] · **Gerd Kortuem**[1]

Accepted: 5 May 2025
© The Author(s) 2025

Abstract
Despite advancements in socially aware navigation, robots still often behave inappropriately in social environments. To ensure successful application, robots must detect the human perceived appropriateness of their navigation behaviors. This paper presents a novel dataset covering a complete range of perceived appropriateness and uniquely incorporates human emotion and attention to facilitate the detection of perceived appropriateness of robot social navigation in pathways (PARSNiP). It is created based on a series of human-robot interaction experiments with 30 participants and a mobile robot. Several typical machine learning models are utilized to evaluate the dataset and analyze the contributions of different features in detecting perceived appropriateness. The results indicate that incorporating emotional and attentional features can significantly improve the accuracy of perceived appropriateness detection. There was an increase from 63% to 68% using algorithm-predicted emotional and attentional features, and a further increase to 79% with the emotion and attention data reported by the participants. With the dataset, researchers could train machine learning models to enable robots to detect perceived appropriateness accurately, fostering adaptations that improve their responsiveness and accuracy in social interactions. The dataset is available for download at https://github.com/duibcuiegiosahxois/PARSNiP.git, and videos will be shared upon request by contacting Y.Zhou-13@tudelft.nl.

Keywords Socially aware navigation · HRI · Social errors · Perceived appropriateness

1 Introduction

Mobile robots are increasingly operating in social environments such as offices, hospitals, and public spaces, where they encounter and interact with humans. These robots undertake different tasks, including assisting in nursing and patient rehabilitation, delivering parcels to homes and offices, stocking shelves in warehouses, and cleaning floors in shopping malls [1]. A fundamental prerequisite for the successful application of mobile robots is their ability to navigate around humans [2]. Robots that fail to respect humans and behave inappropriately have been restricted or banned in social environments [3].

Socially aware navigation, also referred to as human-aware navigation [4], socially compliant navigation [5], socially acceptable navigation [6], or socially competent navigation [7], aims to enable robots to navigate social environments safely, effectively, and in a socially acceptable manner. Its primary objective is to integrate social norms into robot navigation behaviors [8], making them, for example, respect personal space [9], minimize path interference, and prioritize human behavior [10]. However, pre-defined social norms lack flexibility and fail to consider contextual factors and individual differences adequately. More sophisticated algorithms have been developed to enable robots to

✉ Yunzhong Zhou
y.zhou-13@tudelft.nl

Jered Vroon
j.h.vroon@tudelft.nl

Zoltán Rusák
z.rusak@tudelft.nl

Gerd Kortuem
g.w.kortuem@tudelft.nl

[1] Faculty of Industrial Design Engineering, Delft University of Technology, Mekelweg 5, Delft, South Holland 2628 CE, The Netherlands

Published online: 28 May 2025

detect and respond to social dynamics, including human physical activities [11], and social signals such as emotion [12, 13], intention [14, 15], and dominance [16] in varied settings including crowded scenes [17], narrow spaces [18], and urban environments [19].

These studies primarily focus on robots perceiving and responding to humans [1, 20], yet they often overlook how humans perceive robots, especially real-time perception. Just as humans pick up on social cues and adapt to others' feedback, robots should also detect whether their behavior is perceived as appropriate by nearby individuals. Humans use a rich set of social cues to indicate if they perceive a behavior as inappropriate, annoying, or unsafe, such as facial expressions, hand gestures, and evasive motions. Yet due to the limited studies and datasets on the perceived appropriateness (PA) of robot navigation behavior, current robots are unable to detect the PA of their navigation from these cues.

This study introduces the PARSNiP dataset, which contains PA labels and uniquely includes emotion and attention to improve robots' ability to detect PA. Collected from human-robot interaction experiments that involved 30 participants and a robot, PARSNiP offers a complete range of PA levels and is enriched with features crucial for PA detection, such as the motion features of both humans and the robot, as well as human emotion and attention. Several typical machine learning models are employed to analyze the dataset, and findings reveal that incorporating emotional and attentional features markedly improves PA detection performance. This dataset can be utilized to develop machine learning models, which could then be applied in robots to enhance their ability to accurately detect PA and adapt to human behaviors more effectively.

The remainder of this paper is structured as follows: Section 2 reviews existing datasets and their limitations, which motivates the creation of PARSNiP. Section 3 describes the dataset creation, including collection and processing methods. Section 4 presents the testing of the dataset on multiple machine learning models, with an analysis and discussion of the experimental results in detecting PA. Finally, Section 5 concludes the study and outlines future research directions.

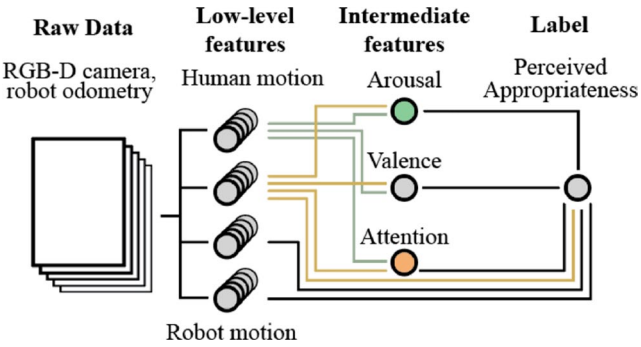


Fig. 1 Assumption for PA detection

2 Related Work

Many datasets have been developed to improve and evaluate socially aware navigation systems. These datasets primarily focus on gathering human motion data [21], covering a variety of environments, both indoors [22–28] and outdoors [28–36]. These datasets are used to train data-driven methods, such as reinforcement learning [37] and deep learning techniques [38], for simulating robot social navigation behaviors. However, they do not take into account the decision-making processes and psychological states of humans.

Additional datasets have integrated affective features, extracted from posture and movement cues [39], to enable the detection of human emotion [40]. Although these datasets have enhanced robots' ability to navigate socially in human environments, they often fail to consider feedback regarding humans' perceived appropriateness of the robots' navigation behavior. In response, recent datasets have shifted their focus to social errors or mistakes, enabling robots to recognize and correct their inappropriate behaviors [41, 42]. While these studies are not within the navigation context, they have yielded significant insights. They demonstrate that social cues play a crucial role in detecting the perceived appropriateness of robot behavior [43, 44], and indicate that integrating emotion and attention could further enhance PA detection performance [41]. In robot social navigation, only 1 dataset has considered and enabled PA detection, albeit with limitations such as focusing solely on robot inappropriate positioning behavior and including only low-level motion features [45].

In this study, the PARSNiP dataset is created to offer a complete range of PA levels and uniquely incorporate emotion and attention to enhance PA detection.

3 Dataset Creation

This study is based on the assumption that human and robot motion features, together with human emotion (through valence and arousal, where valence refers to how positive or negative an event is, and arousal reflects whether an event is exciting or calming [46]) and attention, contribute to the detection of perceived appropriateness (PA) of robot navigation behavior, as shown in Fig. 1. Emotion and attention are intermediate features that can be predicted from low-level human and robot motion features [47, 48].

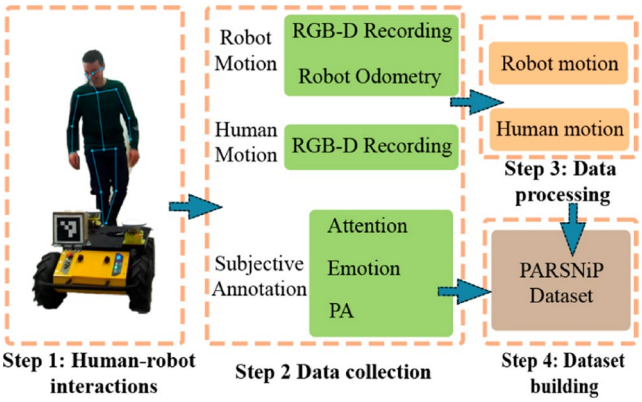


Fig. 2 The workflow for creating PARSNiP

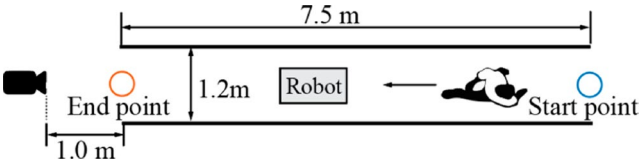


Fig. 3 Overview of the setting. Each participant walks from the start point to the end point for each interaction

Figure 2 shows the workflow to create the PARSNiP dataset. It consists of 4 major steps: Step 1 involves the human-robot interactions, with the participants interacting with a robot 8 times. Step 2 is the data collection, including RGB-D recordings and robot odometry to capture human and robot motion, along with questionnaires to gather participants' subjective annotations of emotion, attention, and PA. In Step 3, the data processing, human and robot motion features are extracted and computed from the raw data. Step 4 is the dataset building, which integrates the computed human and robot features with attention, emotion, and PA to build the PARSNiP. The study was approved by the ethical committee of the university.

3.1 Pilot Study

A pilot study involving both field observations and controlled human-robot interactions has been conducted to determine the setting and types of interactions that contribute to triggering a complete range of PA.

The field observations involved a mobile robot (Clearpath Husky) navigating in diverse outdoor public spaces: narrow pathways, open areas, cross-shaped roads, and street corners. A robot operator fully controlled the robot's navigation behavior to observe the robot's behavioral impact on humans in a natural but safe way. Narrow spaces consistently elicited richer and more nuanced reactions and PA of the robot [49], consistent with the literature indicating a higher incidence of human-robot conflicts in narrow spaces [50]. A subsequent controlled study involved 8 participants interacting with the robot in pathways of different widths (1.0 m, 1.2 m, and 1.4 m). Based on observations and questionnaires, a pathway width of 1.2 m was selected, striking a balance between the richness of PA and human safety. Importantly, the richness of PA persisted even when the pathway was merely marked by floor indicators. During the controlled study, interactions that were highly effective in eliciting diverse human reactions and PA were identified, such as the robot making sudden changes in trajectory and velocity near humans and unconventional human-robot spatial relationships like blocking or squeezing paths.

3.2 Step1: Human-Robot Interactions

3.2.1 Setting

The setting is shown in Fig. 3, which is a pathway delineated by tape markers on the floor. An RGB-D ZED2 camera, positioned on a table 0.75 m high and 1 m away from the end of the pathway, is used to capture the participants' full-body motion (1080p, 30fps). This setup ensures the collection of clean, comprehensive data on human-robot interactions, avoiding the limitations of an onboard camera that has a restricted field of view or experiences data occlusion when humans are in close proximity.

3.2.2 Designed Interactions

8 interactions are designed to trigger a complete range of PA of robot navigation behavior. This is built upon the study of Koay et al., which identified specific interactions causing human discomfort [51], as well as insights from the pilot study. Husky A200 is used as the robot platform to interact with the participants (dimensions of $0.90 \times 0.67 \times 0.39\,\mathrm{m}$) [52]. Each interaction with the robot and human start and end points is visualized in Fig. 4 and detailed below:

- **Interaction 1: Block.** Initiating the interaction from 7.5 meters away—outside the social space [53]-the robot gradually moves toward its endpoint. Midway, it blocks the participant's path at close proximity until the participant passes.
- **Interactions 2, 3: Change Direction.** Starting at a distance of 7.5 meters, the robot moves toward its endpoint while changing direction to approach the participant from different angles at a close distance, until the participant passes.
- **Interaction 4: Squeeze.** As it proceeds toward its endpoint, the robot narrows the available pathway for the participant until the participant passes.
- **Interactions 5, 6: Stop.** Starting at a distance of 7.5 meters from the participant, the robot moves toward its

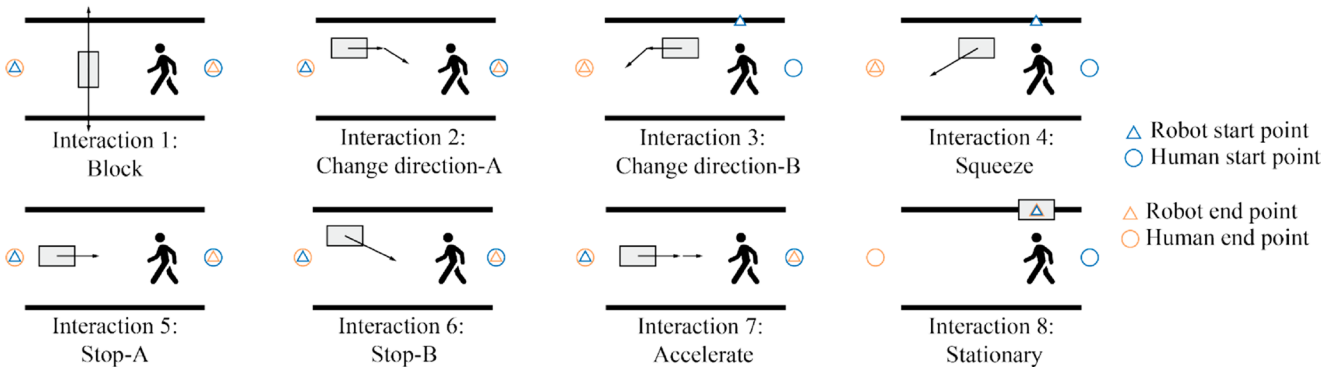


Fig. 4 The sketch of robot-human directional relationships for 8 interactions. During each interaction, the PA of the robot's behavior may vary based on factors such as distance, speed, acceleration rate, etc

goal and then stops when in close proximity to the participant until being passed.

- **Interaction 7: Accelerate.** Starting 7.5 meters away, the robot moves toward its endpoint and then unexpectedly accelerates, continuing to accelerate even after passing the participant.
- **Interaction 8: Stationary.** The robot remains stationary within the environment, serving as a control condition that is expected to be socially acceptable and minimally intrusive the entire time.

These interactions span up to 26 seconds, enabling natural and continuous engagement between the participant and the robot. Each interaction is comprised of a sequence of behaviors, carefully designed to elicit a wide span of PA ratings. While some interactions, such as "Change Direction" or "Squeeze," involve behaviors that might be perceived as inappropriate, they are often initiated with the robot at a distance respecting human social space norms [53]. Furthermore, within each interaction, there are periods where the robot maintains a safe distance, adjusts its trajectory to avoid the participant, or simply passes by, which are expected to be appropriate. For instance, "Interaction 1: Block" begins with the robot 7.5 meters away from the participant, and the blocking behavior only occurs midway through the interaction. "Interaction 5: Stop" also starts with a considerable distance between the robot and participant, and includes moments where the robot is simply moving towards its goal before making any sudden stops. These interactions, therefore, encompass a diverse range of behaviors, from clearly appropriate to potentially inappropriate, allowing for the collection of a comprehensive and balanced dataset of PA ratings. Each 2-second clip of these interactions is treated as an individual data point, capturing the dynamic nature of PA within each interaction.

3.2.3 Experiment

The experimental procedure consists of the following steps:

Preparation: Participants provides informed consent and familiarize themselves with the setting through an initial walkthrough. The experimental setup is explained, including the 7.5 m pathway and interaction protocol.

Robot Positioning: For most interactions, the robot is positioned at the pathway endpoint. interactions 3,4, and 8, the robot is positioned 1.5 m ahead of the participant. All relevant positions are clearly marked to ensure that both the participant and the robot start from the same location in each trial, ensuring consistency across trials.

Interaction: Each participant completes 8 interactions in randomized order to prevent learning effects (see Fig. 4). For each interaction: (1) The participant starts at the designated start point. (2) The participant traverses the pathway to the end point. (3) An experienced operator is trained to manually control the robot, utilizing the Wizard-of–Oz protocol [54]. (4) Both the robot and the participants could step outside the pathway for safety or other concerns, similar to what might occur in real-life pathways.

Figure 5 depicts real-life human-robot interactions, with the detected skeletal points visualized.

3.2.4 Demographics

30 participants are recruited through onsite convenience sampling, including 11 females and 19 males, with no exclusions. Most participants (27) are young adults between 18 and 34 years old, while 3 are adults aged 35 to 54 years. The majority (26 out of 30) are Dutch. Regarding robot experience, most participants have limited interactions with robots: 23 have no prior interactions with robots, 6 have some interactions, and only 1 has frequent interactions.

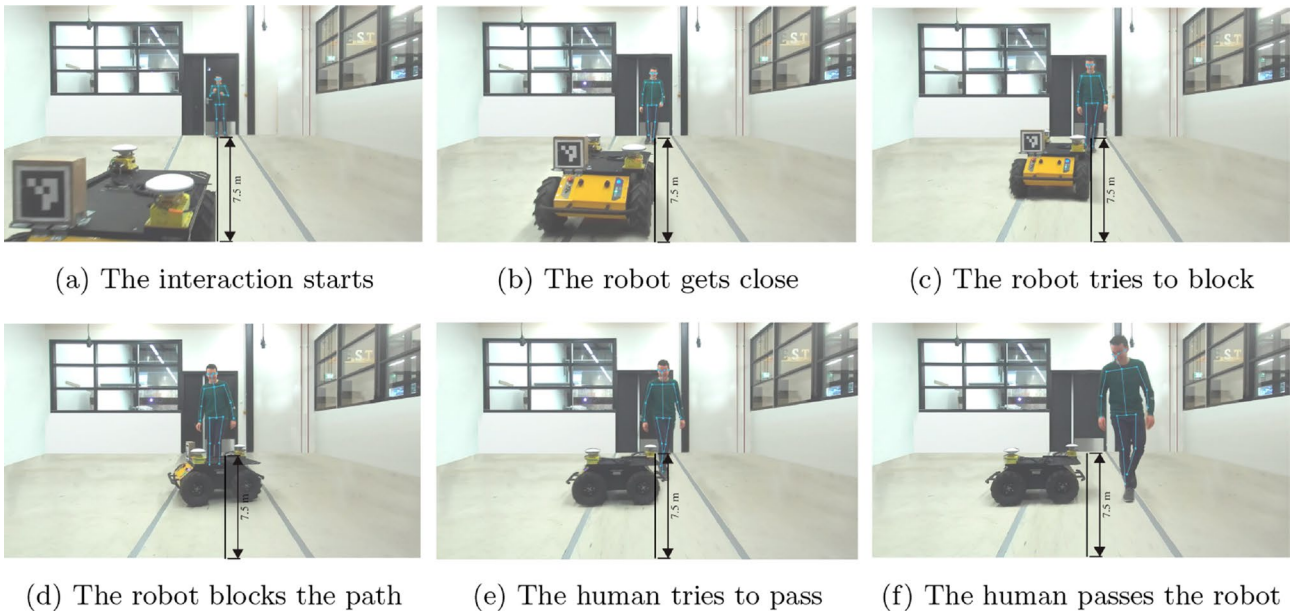


Fig. 5 Human-robot interactions, with the human's 18 skeletal points captured by the ZED SDK

3.3 Step 2: Data Collection

After completing all 8 interactions, the participants are given definitions of valence and arousal [46]. Afterwards, they are given questionnaires to indicate their current levels of valence and arousal. They are also requested to provide explanations for their responses to ensure that they understand how to assess and report their valence and arousal levels accurately. Meanwhile, the recorded interactions from the camera are divided into 2-second clips, resulting in 25–42 clips [55, 56]. These clips are presented to participants with online questionnaires to self-report PA, emotion, and attention, as data points of the dataset. PA is assessed on a 7-point scale ranging from −3 to 3, with participants asked to indicate the appropriateness of the robot's behavior (7-point scale: "The robot behaved completely inappropriately"–"Not sure/Neutral"–"The robot behaved completely appropriately") [45]. Emotion is evaluated in terms of valence and arousal; both are also measured on a 7-point scale from −3 to 3. Attention is rated on a 6-point scale ranging from 0% to 100%. Additionally, participants are asked multiple-choice questions aimed at identifying perceptions of robot behaviors concerning PA value below 3, selecting from a list of factors identified in the pilot study. These factors include moving too fast or slow, taking sharp turns, getting too close, moving unnaturally, being too noisy, blocking the path, appearing threatening, squeezing the walkable area, making sudden changes, and an "other" option.

3.4 Step 3 &4: Data Processing and Building

After the interactions and the questionnaires, all data points are collected. For each data point, the raw data from recorded interactions and robot odometry undergo processing to extract human and robot motion features. The ZED Software Development Kit (SDK) is utilized to extract human motion data, comprising 3D positions of 18 skeletal points and the body center for each frame. A 95% detection confidence threshold is established using the "Accurate" detection model, which fills in missing keypoints using historical data and human kinematics. The robot motion data are captured by the robot's odometer and an Aruco Marker, with the former providing consistent tracking and the latter ensuring reliable position data. The robot's data is transformed into the same coordinate system as the human data to ensure alignment for analysis. Using the Genetic Algorithm (GA) Toolkit in Matlab, this transformation minimizes the mean squared error compared to ground truth measurements and ensures consistent robot positional data for feature computation.

Human and robot motion features are computed based on processed robot and motion data. Both the robot and human motion features are characterized by 5 statistical measures—minimum, maximum, mean, standard deviation, and min-to-max ratio—across each data point. Each point includes 30 robot motion features, including robot speed, acceleration, jerk, robot-human distance, robot-human direction, and robot turning angle [45]. Specifically, the robot-human approach angle is computed as the angle between the robot's current heading and the straight-line vector that

connects the robot's current position to the human's position. These features could provide insights into the dynamics of human-robot interactions. For example, features such as robot-human distance and robot-human approach angle can reveal whether the robot is heading toward a person or will merely pass by. A small approach angle, combined with a low min-to-max distance ratio and a short mean distance, generally indicates that the robot is heading directly toward a person. Conversely, the same small approach angle, when paired with a high min-to-max distance ratio and a longer mean distance, might suggest that the robot, initially aimed toward the person, is moving away or adjusting its path. Additionally, a large approach angle with a low min-to-max distance ratio and a short mean distance could imply that, despite not facing the person directly, the robot is moving toward them.

Each data point also includes 360 human motion features, with 11 groups encompassing kinematic features, curvature, quantity of motion, bounding volume, displacement of joints, motion features, effort component of the Laban Movement Analysis, spatial extension, symmetry of the movement, body balance, and gaze [47, 58, 59]. Specifically, jerk derived directly from position data can be quite noisy. To address the inherent noise issues associated with this calculation, a low-pass filter is applied to the acceleration data [57]. The effort of the Laban Movement Analysis is computed according to the following functions [47, 58]:

$$E(t_i) = \sum_{k \in K} E_k(t_i) = \sum_{k \in K} \alpha_k v^k(t_i)^2 \tag{1}$$

$$\text{Weight}(t_i) = \max_{i \in [1,T]} E(t_i), \quad i = 1, 2, 3, \ldots, N \tag{2}$$

$$\text{Time}^k(t_i) = \frac{1}{T} \sum_{i=1}^{T} \alpha^k(t_i) \tag{3}$$

$$\text{Flow}^k(t_i) = \frac{1}{T} \sum_{i=1}^{T} j^k(t_i) \tag{4}$$

By organizing human and robot motion features and human emotion (valence and arousal), attention, and PA for each data point, the dataset is thus built. A complete overview of the dataset is presented in Table 1.

4 Dataset Analysis

4.1 Dataset Distributions

Table 2 provides a comparison between the PARSNiP dataset and existing datasets in robot social navigation, highlighting its unique attributes and contributions to the field. Unlike its predecessors, PARSNiP encompasses a broad spectrum of human and robot features. It includes human PA, head, body, and hand motions, trajectories, emotion,

Table 1 Overview of the PARSNiP dataset. The dataset includes 922 data points collected from 30 participants, each interacting 8 times with the robot

Data type	Data (source)	Measurement (per 2-second interval)
Low-level features	Human motion (3D position of 18 skeletal points from RGB-D recordings)	11 categories of 360 human motion features, described by the min, max, mean, standard deviation, and min-to-max ratio: •Kinematic features (velocity, acceleration, and jerk (smoothed by low-pass filtering [57]) for head, wrists, and ankles) [47, 58], •Curvature for head, wrists, and ankles [47], •Quantity of motion (aggregated speed over a set of joints) for the arm and head region, upper body, lower body [47], •Bounding volume for the arm region, head region, upper body, lower body [47], •Displacement of joints for the head, wrists, and ankles [47], •Motion features (verticality, extension, left and right elbow flexion, left and right arm shape, hand relationship, and feet relationship) [47, 58], •Effort component of the Laban Movement Analysis (time, weight, flow) for the arm region, head region, upper body, lower body, and whole body [47, 58], •Spatial extension [47, 59], •Symmetry of the movement (horizontal, vertical, and bounding triangle-related symmetry) [47, 59], •Body balance (balance, center of mass for the arm region, head region, upper body, lower body, and whole body) [47], •Human gaze [60]
	Robot motion (Robot odometry, RGB-D recordings)	30 Robot motion features, described by the min, max, mean, standard deviation, and min-to-max ratio: robot speed, robot acceleration, robot jerk, robot-human distance, robot-human approach angle, robot turning angle [45]
Intermediate features	Emotion and attention (Self-report)	Attention (0% to 100%, 6-point scale), valence and arousal (−3 to 3, 7-point scale)
Label	PA (Self-report)	PA (from −3 to 3, 7-point scale)

Table 2 Datasets for socially aware robot navigation

Dataset	Year	PA	Head motion	Body motion	Hand motion	Trajectories	Emotion	Attention
UCY [32]	2007	×	×	×	×	√	×	√
ETH [29]	2009	×	×	×	×	√	×	×
KITTI [33]	2012	×	×	×	×	√	×	×
Edinburg [30]	2009	×	×	×	×	√	×	×
Town center [34]	2011	×	√	√	×	√	×	×
CFF [35]	2014	×	×	×	×	√	×	×
WildTrack [36]	2018	×	×	×	×	√	×	×
SCAND[31]	2022	×	×	×	×	√	×	×
VIRAT [23]	2011	×	×	×	×	√	×	×
ATC [22]	2013	×	√	√	×	√	×	×
L-CAS [26]	2017	×	×	×	×	√	×	×
KTH [24]	2015	×	√	×	×	√	×	×
THOR [25]	2020	×	√	√	×	√	√	×
EgoMotion [27]	2016	×	√	√	√	√	×	×
JackRabbot [28]	2019	×	×	×	×	√	×	×
UNC [39]	2019	×	×	×	×	√	×	×
Vroon [45]	2019	Position	√	√	×	√	×	×
PARSNiP	2023	√	√	√	√	√	√	√

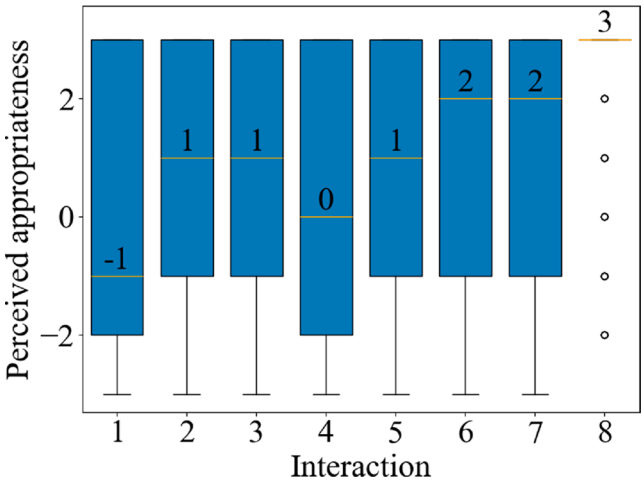


Fig. 6 The distribution of different PA levels

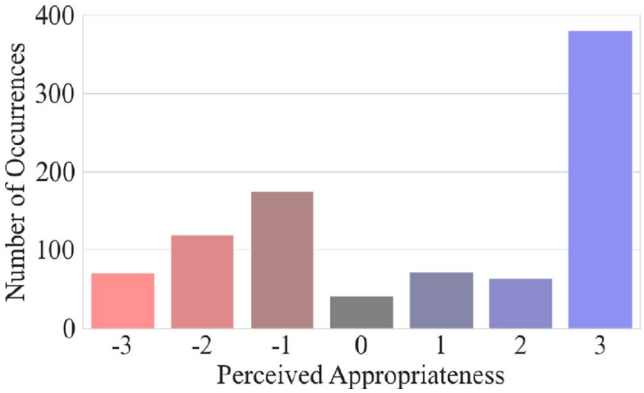


Fig. 7 Interactions and PA levels

and attention. The inclusion of emotion and attention marks a significant advancement, offering insights into better PA detection.

4.1.1 PA Distribution

Participants exhibited variations in their responses to the robot: 3 participants showed signs of disorientation and discomfort in the presence of the robot, even when it remained stationary. Conversely, 1 participant maintained a positive attitude towards the robot, showing high tolerance regardless of the robot's behavior. The range of observed human responses to specific robot behaviors varied significantly, especially during potential conflicts with the robot: Some participants paused to assess the situation before proceeding, while others subtly changed their path or quickened their pace to avoid the robot. A few participants turned their bodies to navigate past the robot.

Figure 6 shows the distributions of different PA levels, illustrating a complete range of PA, from −3 (completely inappropriate) to 3 (completely appropriate). The data show that the participants largely report the robot's behavior as completely appropriate, which highlights the need for further processing when using machine learning techniques.

Figure 7 illustrates the distribution of PA values across the 8 interaction types. Many interactions received a score of 0 to 2, indicating a neutral to slightly positive perception of the robot. However, distinct deviations are observed in interactions 1, 6, 7, 8. Interaction 1 shows a slightly negative value in median PA, while interaction 8 receive a positive

peak, indicating higher PA. These variations in PA scores highlight how different types of interactions are perceived in terms of appropriateness, providing valuable insights into the contextual factors that may influence perceptions and the richness of PA levels.

The dataset includes 11 types of PA, detailed in Fig. 8. Notably, "Block" with 173 instances, "Squeeze" with 112 instances, and "Sudden" with 110 instances are significant factors affecting the human perception of robot behaviors. In contrast, perceptions of "Threat", "Unnatural," and "Fast" are reported less frequently. Different interactions reveal variations in PA, providing insights into human-robot interaction dynamics. Interestingly, although interaction 8 generally receives positive evaluations, previous inappropriate interactions could potentially influence the perceptions of the robot. Overall, these 8 interactions contribute to a complete range of PA levels and richness of PA types, thereby enabling PA detection.

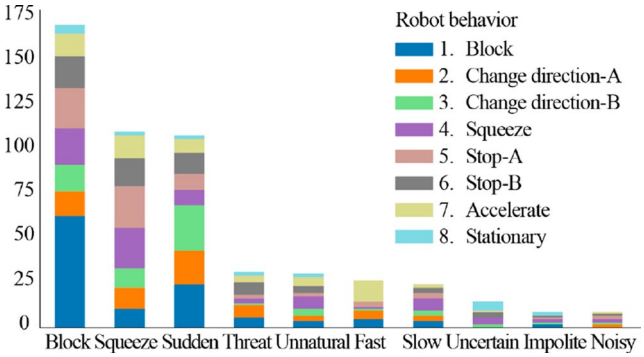


Fig. 8 Interactions and PA types

4.1.2 Emotion and Attention Distribution

Figure 9 illustrates the emotional responses and attention levels across 8 different interactions.

Figure 9a presents the distribution of emotional valence across 8 interaction types. The median valence for interactions 1 to 6 remains consistently neutral, underscoring a uniform emotional response among these categories. Interactions 7 and 8 receive a higher median valence, indicating a more positive emotional reaction. Specifically, interaction 8 shows a narrower interquartile range, suggesting less variability and a more consistent response among participants.

Figure 9b illustrates the arousal levels across the 8 types of interactions. Interactions 2 to 8 display medians at zero, indicating no substantial deviation from a baseline arousal state, while interaction 1 shows a higher median arousal level.

Figure 9c shows that the attention metrics vary significantly across different types of interactions. Interactions 1, 2, and 4 achieve the highest median attention scores of 100%, reflecting heightened cognitive engagement or focus. Interaction 8 notably receives the lowest median and an expanded range, indicating fluctuating levels of participant attention and potentially lower engagement.

The impact of a robot's presence within the participant's field of view on human attention and emotion (valence and arousal) is also examined. Given the dynamic nature of human-robot interactions captured in 2-second intervals, direct comparisons are challenging. Factors such as the robot's movements into or out of view or a transient presence complicate the analysis. Therefore, we differentiated the data points based on two scenarios: A. the robot is entirely outside the participant's view the whole time, and B. the robot has been within view.

A Mann-Whitney U test was conducted comparing the differences between the two scenarios [61]. The statistical analysis revealed significant differences in attention and emotional responses. For Scenario A, the average attention was 82.7%, compared to 58.0% for Scenario B, with a significant p-value of 1.26×10^{-29}. This suggests more human attention when the robot is in view. Emotional responses also varied; valence averaged 0.73 in Scenario A and 0.18 in Scenario B, with a p-value of 5.5×10^{-6}, indicating more negative emotions when the robot is in view. Arousal was -0.44 for Scenario A and 0.25 for Scenario B, with a p-value of 6.0×10^{-8}, indicating that the presence

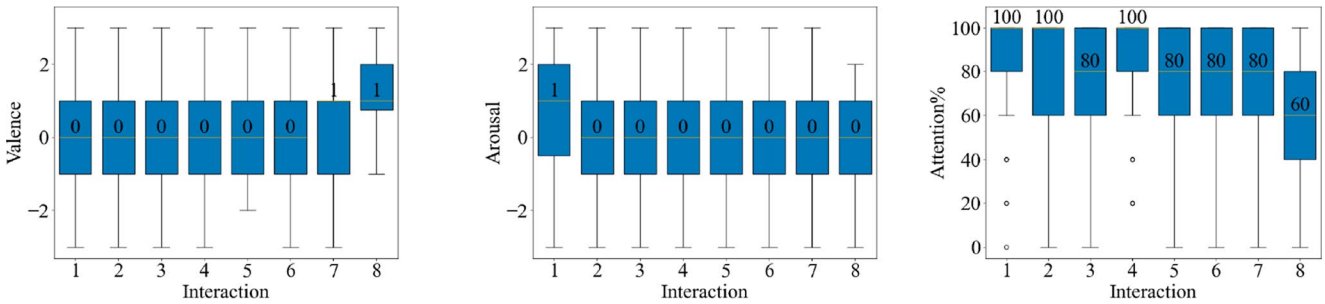
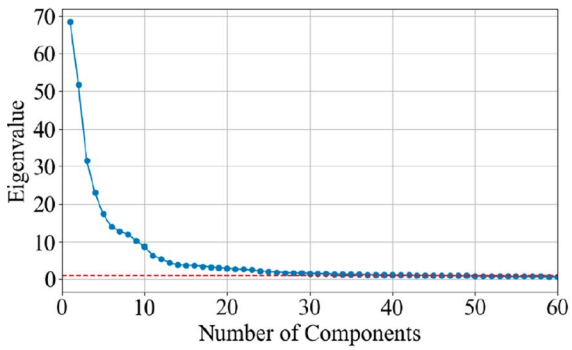
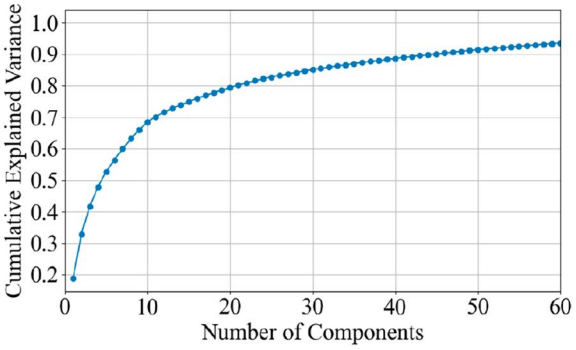


Fig. 9 Valence, arousal, and attention distributions under different interactions



(a) Number of components and eigenvalues



(b) Number of components and explained variance

Fig. 10 PCA results: eigenvalues and explained variance

Table 3 Performance metrics over strategies (Str.) and models (Mod.), with data grouped by clusters (Clus.). The metrics include accuracy (Acc.), precision (Prec.), recall (Rec.), and the F1 score (F1)

Str.	Mod.	Clus.	Acc.	Prec.	Rec.	F1
1	RF	1	0.78	0.64	0.78	0.70
		2	0.49	0.67	0.49	0.57
	GBDT	1	0.74	0.65	0.74	0.69
		2	0.55	0.65	0.55	0.59
	AdaB	1	0.70	0.62	0.70	0.66
		2	0.51	0.60	0.51	0.55
	FNN	1	0.74	0.65	0.74	0.69
		2	0.55	0.65	0.55	0.59
	XGB	1	0.78	0.64	0.78	0.70
		2	0.51	0.67	0.51	0.58
2	RF	1	0.74	0.60	0.74	0.66
		2	0.61	0.75	0.61	0.67
	GBDT	1	0.70	0.60	0.70	0.65
		2	0.63	0.73	0.63	0.68
	AdaB	1	0.63	0.55	0.63	0.59
		2	0.59	0.67	0.59	0.63
	FNN	1	0.70	0.61	0.70	0.65
		2	0.65	0.73	0.65	0.68
	XGB	1	0.70	0.58	0.70	0.63
		2	0.60	0.72	0.60	0.65

of the robot within the view was more emotionally exciting for the participants.

4.2 Dataset Analysis

4.2.1 Clustering and Algorithms

A principal component analysis (PCA) is conducted to reduce the number of dimensions and mitigate noise. This step entails calculating eigenvalues and the cumulative explained variances, as depicted in Fig. 10. The study selects components with eigenvalues exceeding 1 [62], with 45 principal components that account for around 90% of the total variance.

The PA levels are binarized for further analysis, ensuring that the smaller group constitutes at least 40% of the data to enable balanced clusters. This leads to 2 distinct clustering strategies: Strategy 1 assigns PA levels −3 to 0 to Cluster 1 and 1 to 3 to Cluster 2. Strategy 2 groups PA levels −3 to 1 in Cluster 1 and 2 to 3 in Cluster 2.

The dataset is tested using various typical machine learning models, employing a 5-fold cross-validation to ensure the reliability of the results. The models include Random Forest (RF), Gradient Boosting Decision Trees (GBDT), AdaBoost with decision trees (AdaB), Feedforward Neural Network (FNN), and Extreme Gradient Boosting (XGB). The performance was evaluated based on accuracy, precision, recall, and F1 score, as shown in Table 3. Strategy 2 displays more balanced performance metrics between Clusters 1 and 2 across all classifiers, indicating a more uniform data separation, and is selected for further analysis.

4.2.2 Ablation Study

An ablation study is conducted to understand the impact of various feature sets on the performance of PA detection. Based on the assumption that emotional and attentional features may act as intermediate variables that improve PA detection, participants' self-reported emotions and attention are included for comparisons. Since emotional and attentional features are not typically observable in real-world conditions, the study also predicts these features based on human and robot movement data, achieving a valence detection accuracy of 71.0%, arousal detection accuracy of 69.2%, and attention detection accuracy of 72.8%. The study evaluates the effectiveness of the following feature combinations: Robot and human motion (RH), Robot and Human motion + predicted emotion and predicted attention (RHP), Robot and human motion + human reported emotion and attention data (RHEA).

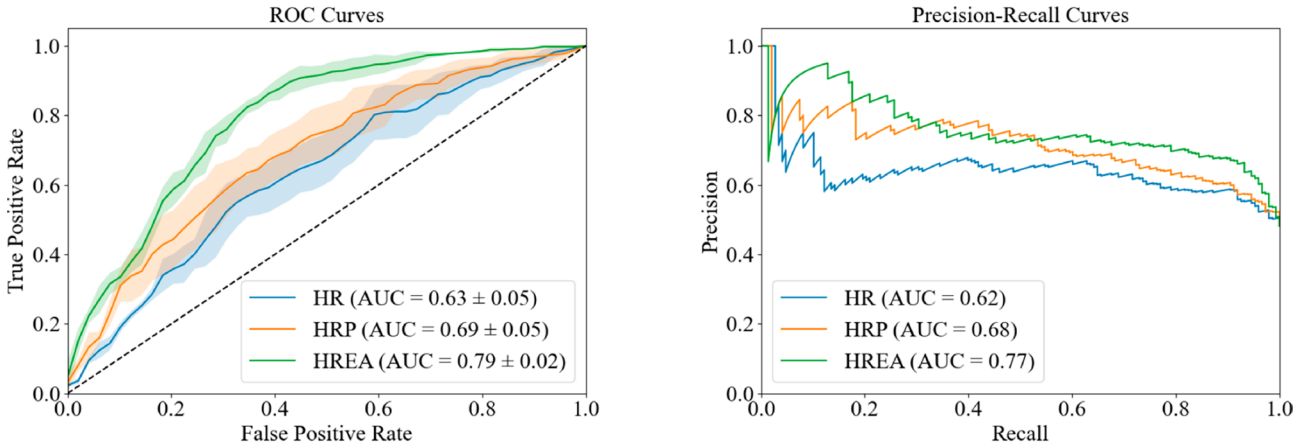
The effectiveness of the XGBoost algorithm across a variety of feature combinations is assessed. Results include the Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves (Fig. 11). The ROC curve (Fig. 11a) illustrates the model's sensitivity (true positive rate) versus its fall-out (false positive rate) at various thresholds. This offers a comprehensive evaluation of the model's discriminative capability, unaffected by class distribution. The PR curve (Fig. 11b), on the other hand, provides insight into the model's precision in relation to its recall, which is especially relevant for imbalanced datasets. The HREA feature set demonstrates the highest performance, achieving the top Area Under the Curve (AUC) scores in both the ROC (0.79 ± 0.02) and PR (0.77) analyses. This indicates a consistent and reliable detection of true positives across different thresholds, thereby enhancing the detection's reliability and robustness. The high AUC values suggest a model that accurately predicts while maintaining a balanced sensitivity and precision. In comparison, the HRP feature set shows significant improvement over the HR set, as indicated by higher AUC values in the ROC (0.69 ± 0.05 vs. 0.63 ± 0.05) and PR (0.68 vs. 0.62) curves. This improvement highlights the benefit of including predicted emotional and attentional features for superior detection capabilities.

The statistical significance of these findings is supported by the hypothesis tests detailed in Table 4. Comparing HREA to HR, and HRP to HR, the analysis reveals significant t-values and small p-values, indicating not just statistical significance but also meaningful effect sizes as measured by Cohen's d. This confirms that the performance differences are statistically significant and practically important. Furthermore, the accuracy and F1 score statistics shown in Table 4 corroborate these results, with HREA outperforming and HRP significantly better than HR. These statistics, together with insights from the ROC and PR curves in Fig. 11, underscore that the inclusion of emotional and attentional features, as represented by HREA, and the predicted emotional and attentional features, as described by HRP, markedly improve the model's detection effectiveness. Specifically, the high PA detection performance with the inclusion of the human self-reported emotion and attention strongly demonstrates an upper bound on the usefulness of emotion and attention to predict PA.

4.3 Limitations and Discussion

The PARSNiP dataset faces several methodological limitations that warrant consideration.

Data Collection and Reliability. The reliance on participants' self-reported data for PA, emotion (valence and arousal), and attention after all interactions may introduce recall biases, potentially compromising data reliability.



(a) Receiver operating characteristic curves

(b) Precision/recall curves

Fig. 11 ROC and Precision/Recall curves showing the PA detection performance with different features

Table 4 Descriptive statistics and hypothesis tests for Accuracy and F1 score with different feature sets using the XGBoost algorithm

(a) Hypothesis tests for PA detection with different sets of features. Since the data used are normal, Cohen's d is used to measure the effect size.

Feature	Accuracy			F1 score		
	$t(4)$	p	d	$t(4)$	p	d
HR vs HREA	12.00	0.004	4.37	12.82	0.003	4.10
HR vs HRP	6.831	0.019	2.00	6.35	0.003	1.56
HRP vs HREA	8.454	0.039	1.53	8.21	0.033	1.96

(b) PA detection statistics for Accuracy and F1 score with different features.

Features	**Accuracy**		**F1**	
	μ	σ	μ	σ
HR	0.63	0.05	0.68	0.05
HRP	0.68	0.05	0.72	0.05
HREA	0.79	0.05		

Despite implementing mitigating measures—such as showing participants recordings of their facial expressions and maintaining brief interactions—recall biases may persist. Future studies should explore real-time annotation methods, including "think-aloud" techniques, which would enable real-time reporting for each interaction.

Temporal Resolution. The study's assumption of consistent human states during 2-second intervals presents another limitation, as it may not fully capture real-world behavioral dynamics. While this interval was chosen to optimize data collection efficiency, real-life applications require a more nuanced approach. We suggest that researchers and engineers implement a sliding-window method for practical applications, using overlapping data windows that shift incrementally (e.g., every 0.5 seconds) [56]. This approach would enable continuous PA detection and allow robots to adapt their behavior dynamically during interactions.

Experimental Setup. While using an onsite camera for data collection is advantageous for controlling experimental conditions, we recognize that the specific setup—positioning the camera on a table—may not fully align with real-world applications. To bridge this gap, several strategies can be employed to adapt the collected data for use with different setups such as onboard cameras. For example, data mapping and transformation techniques can simulate the perspective of an onboard camera, while data augmentation and simulation can create additional datasets that replicate various camera positions and angles. These approaches can improve the practicality of the dataset and deserve further investigation.

5 Conclusion

The PARSNiP dataset represents a significant advancement in enabling robots to navigate social environments more appropriately. Our analysis demonstrates three key findings: First, the integration of emotional and attentional features substantially improves the detection of perceived appropriateness (PA) in robot navigation. The improvement in detection accuracy from 63% with motion features alone to 79% with emotional and attentional features demonstrates the critical role of these psychological factors in human-robot interactions. Second, even predicted emotional and attentional states enhance PA detection capabilities, achieving a 68% accuracy rate. This finding has important practical implications, as it suggests that robots can improve their social navigation even without direct access to human psychological states. Third, the comprehensive range of interaction scenarios captured in PARSNiP provides insights into the appropriateness of robot navigation behaviors, particularly in confined spaces such as pathways. This understanding is crucial for developing more socially acceptable robot navigation strategies. These findings have immediate applications for robotic systems in social environments. Engineers and researchers can implement these insights to develop robots capable of: adapting their navigation behavior based on detected human responses; maintaining appropriate social distance and movement patterns; and responding proactively to potential social discomfort. Looking forward, this work opens new avenues for real-time social navigation systems that can continuously adjust to human preferences and expectations, ultimately leading to more successful integration of mobile robots in human environments.

Declarations

The authors of this paper, hereby declare that we have no financial or non-financial interests that could be perceived as influencing the research work submitted for publication. We confirm that there are no conflicts of interest, such as employment, consultancies, stock ownership, honoraria, or any other competing interests that may have affected the objectivity, integrity, or impartiality of this study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the

source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Kyrarini M, Lygerakis F, Rajavenkatanarayanan A, Sevastopoulos C, Nambiappan HR, Chaitanya KK, Babu AR, Mathew J, Makedon F (2021) A survey of robots in healthcare. Technologies 9(1):8
2. Mavrogiannis C, Baldini F, Wang A, Zhao D, Trautman P, Steinfeld A, Oh J (2023) Core challenges of social robot navigation: a survey. ACM Trans Hum-Robot Interact 12(3):1–39
3. While AH, Marvin S, Kovacic M (2021) Urban robotic experimentation: san francisco, tokyo and dubai. Urban Stud 58(4):769–786
4. Kruse T, Pandey AK, Alami R, Kirsch A (2013) Human-aware robot navigation: a survey. Rob Auton Syst 61(12):1726–1743
5. Gao Y, Huang C-M (2022) Evaluation of socially-aware robot navigation. Front Rob AI 8:721317
6. Shiomi M, Zanlungo F, Hayashi K, Kanda T (2014) Towards a socially acceptable collision avoidance for a mobile robot navigating among pedestrians using a pedestrian model. Int J Soc Robot 6(3):443–455
7. Mavrogiannis C, Baldini F, Wang A, Zhao D, Trautman P, Steinfeld A, Oh J: Core challenges of social robot navigation: a survey. arXiv preprint arXiv:2103.05668 (2021)
8. Johnson C, Kuipers B: Socially-aware navigation using topological maps and social norm learning. In: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, pp. 151–157 (2018)
9. Kamezaki M, Tsuburaya Y, Kanada T, Hirayama M, Sugano S (2022) Reactive, proactive, and inducible proximal crowd robot navigation method based on inducible social force model. IEEE Rob Autom Lett 7(2):3922–3929
10. Che Y, Okamura AM, Sadigh D (2020) Efficient and trustworthy social navigation via explicit and implicit robot–human communication. IEEE Trans Rob 36(3):692–707
11. Ye J, Li X, Zhang X, Zhang Q, Chen W (2020) Deep learning-based human activity real-time recognition for pedestrian navigation. Sensors (Switzerland) 20(9):1–30. https://doi.org/10.3390/s20092574
12. Ma Y, Paterson HM, Pollick FE (2006) A motion capture library for the study of identity, gender, and emotion perception from biological motion. Behav Res Meth 38(1):134–141
13. Bera A, Randhavane T, Manocha D: The emotionally intelligent robot:improving socially-aware human prediction in crowded environments. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2–3 (2019)
14. Rasouli A, Kotseruba I, Kunic T, Tsotsos JKP: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6262–6271 (2019)
15. Narayanan V, Manoghar BM, RV RP, Bera A: Ewarenet: emotion aware human intent prediction and adaptive spatial profile fusion for social robot navigation. arXiv preprint arXiv:2011.09438 (2020)
16. Randhavane T, Bera A, Kubin E, Wang A, Gray K, Manocha D: Pedestrian dominance modeling for socially-aware robot navigation. In: 2019 International Conference on Robotics and Automation (ICRA), pp. 5621–5628 (2019). IEEE
17. Sathyamoorthy AJ, Patel U, Paul M, Kumar NKS, Savle Y, Manocha D (2021) Comet: modeling group cohesion for socially compliant robot navigation in crowded scenes. IEEE Rob Autom Lett 7(2):1008–1015
18. Senft E, Satake S, Kanda T: Would you mind me if i pass by you?: socially-appropriate behaviour for an omni-based social robot in narrow environment. In: 2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 539–547 (2020). IEEE
19. Chen Z, Fan T, Zhao X, Liang J, Shen C, Chen H, Manocha D, Pan J, Zhang W (2021) Autonomous social distancing in urban environments using a quadruped robot. IEEE Access 9:8392–8403
20. Singamaneni PT, Bachiller-Burgos P, Manso LJ, Garrell A, Sanfeliu A, Spalanzani A, Alami R (2024) A survey on socially aware robot navigation: taxonomy and future challenges. Int J Rob Res 02783649241230562
21. Panahandeh G, Mohammadiha N, Leijon A, Händel P (2013) Continuous hidden markov model for pedestrian activity classification and gait analysis. IEEE Trans Instrum Meas 62(5):1073–1083
22. Brščic' D, Kanda T, Ikeda T, Miyashita T (2013) Person tracking in large public spaces using 3-d range sensors. IEEE Trans Human-Mach Syst 43(6):522–534
23. Oh S, Hoogs A, Perera A, Cuntoor N, Chen -C-C, Lee JT, Mukherjee S, Aggarwal J, Lee H, Davis L, et al.: A large-scale benchmark dataset for event recognition in surveillance video. In: CVPR 2011, pp. 3153–3160 (2011). IEEE
24. Dondrup C, Bellotto N, Jovan F, Hanheide M, et al.: Real-time multisensor people tracking for human-robot spatial interaction (2015)
25. Rudenko A, Kucner TP, Swaminathan CS, Chadalavada RT, Arras KO, Lilienthal AJ (2020) Thör: human-robot navigation data collection and accurate motion trajectories dataset. IEEE Rob Autom Lett 5(2):676–682
26. Yan Z, Duckett T, Bellotto N: Online learning for human classification in 3d lidar-based tracking. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 864–871 (2017). IEEE
27. Park HS, Hwang -J-J, Niu Y, Shi J: Egocentric future localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4697–4705 (2016)
28. Martin-Martin R, Patel M, Rezatofighi H, Shenoi A, Gwak J, Frankel E, Sadeghian A, Savarese S: Jrdb: a dataset and benchmark of egocentric robot visual perception of humans in built environments. IEEE transactions on pattern analysis and machine intelligence (2021)
29. Pellegrini S, Ess A, Schindler K, Van Gool L: You'll never walk alone: modeling social behavior for multi-target tracking. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 261–268 (2009). IEEE
30. Majecka B: Statistical models of pedestrian behaviour in the forum. Master's thesis, School of Informatics, University of Edinburgh (2009)
31. Karnan H, Nair A, Xiao X, Warnell G, Pirk S, Toshev A, Hart J, Biswas J, Stone P (2022) Socially compliant navigation dataset (scand): a large-scale dataset of demonstrations for social navigation. IEEE Rob Autom Lett 7(4):11807–11814
32. Lerner A, Chrysanthou Y, Lischinski D (2007) Crowds by example. In: Computer Graphics Forum, vol 26. Wiley Online Library, pp 655–664
33. Geiger A, Lenz P, Urtasun R: Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE Conference

on Computer Vision and Pattern Recognition, pp. 3354–3361 (2012). IEEE
34. Benfold B, Reid I: Stable multi-target tracking in real-time surveillance video. In: CVPR 2011, pp. 3457–3464 (2011). IEEE
35. Alahi A, Ramanathan V, Fei-Fei L: Socially-aware large-scale crowd forecasting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2203–2210 (2014)
36. Chavdarova T, Baqué P, Bouquet S, Maksai A, Jose C, Bagautdinov T, Lettry L, Fua P, Van Gool L, Fleuret F: Wildtrack: a multi-camera hd dataset for dense unscripted pedestrian detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5030–5039 (2018)
37. Li K, Xu Y, Wang J, Meng MQ-H: Sarl: deep reinforcement learning based human-aware navigation for mobile robot in indoor environments. In: 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 688–694 (2019). IEEE
38. Kothari P, Kreiss S, Alahi A (2021) Human trajectory forecasting in crowds: a deep learning perspective. IEEE Trans Intell Transp Syst 23(7):7386–7400
39. Randhavane T, Bhattacharya U, Kapsaskis K, Gray K, Bera A, Manocha D: Identifying emotions from walking using affective and deep features. arXiv preprint arXiv:1906.11884 (2019)
40. Bera A, Randhavane T, Prinja R, Kapsaskis K, Wang A, Gray K, Manocha D: How are you feeling? multimodal emotion learning for socially-assistive robot navigation. In: 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), pp. 644–651 (2020). IEEE
41. Loureiro F, Avelino J, Moreno P, Bernardino A: Self-perception of interaction errors through human non-verbal feedback and robot context. In: International Conference on Social Robotics, pp. 475–487 (2022). Springer
42. Stiber M, Taylor R, Huang C-M: Modeling human response to robot errors for timely error detection. In: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 676–683 (2022). IEEE
43. Mirnig N, Stollnberger G, Miksch M, Stadler S, Giuliani M, Tscheligi M (2017) To err is robot: how humans assess and act toward an erroneous social robot. Front Rob AI 21
44. Giuliani M, Mirnig N, Stollnberger G, Stadler S, Buchner R, Tscheligi M (2015) Systematic analysis of video data from different human–robot interaction studies: a categorization of social signals during error situations. Frontiers in Psychology 6:931
45. Vroon J, Englebienne G, Evers V: Detecting perceived appropriateness of a robot's social positioning behavior from non-verbal cues. In: 2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI), pp. 216–225 (2019). IEEE
46. Kensinger EA (2004) Remembering emotional experiences: the contribution of valence and arousal. Rev Neurosci 15(4):241–252
47. Ahmed F, Bari AH, Gavrilova ML (2019) Emotion recognition from body movement. IEEE Access 8:11761–11781
48. Bera A, Randhavane T, Manocha D: Improving socially-aware multi-channel human emotion prediction for robot navigation. In: CVPR Workshops, pp. 21–27 (2019)
49. Vroon J, Zhou Y, Kortuen G: Factors for robot social navigation among pedestrians. [under review]. (2023)
50. Gross H-M, Scheidig A, Debes K, Einhorn E, Eisenbach M, Mueller S, Schmiedel T, Trinh TQ, Weinrich C, Wengefeld T, et al. (2017) Roreas: robot coach for walking and orientation training in clinical post-stroke rehabilitation—prototype implementation and evaluation in field trials. Auton Rob 41:679–698
51. Koay KL, Dautenhahn K, Woods S, Walters ML: Empirical results from using a comfort level device in human-robot interaction studies. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction, pp. 194–201 (2006)
52. Robotics C: Husky UGV - Outdoor Field Research Robot. https://clearpathrobotics.com/husky-ugv/ Accessed 2024-02-14
53. Hall ET, Birdwhistell RL, Bock B, Bohannan P, Diebold Jr AR, Durbin M, Edmonson MS, Fischer J, Hymes D, Kimball ST, et al. (1968) Proxemics [and comments and replies]. Curr Anthropol 9(2/3):83–108
54. Dahlbäck N, Jönsson A, Ahrenberg L: Wizard of oz studies: why and how. In: Proceedings of the 1st International Conference on Intelligent User Interfaces, pp. 193–200 (1993)
55. Airas M, Alku P (2004) Emotions in short vowel segments: effects of the glottal flow as reflected by the normalized amplitude quotient. In: Tutorial and Research Workshop on Affective Dialogue Systems. Springer, pp 13–24
56. Kim Y, Provost EM: Emotion classification via utterance-level dynamics: a pattern-based approach to characterizing affective expressions. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 3677–3681 (2013). IEEE
57. Zhang L, Diraneyya MM, Ryu J, Haas CT, Abdel-Rahman EM (2019) Jerk as an indicator of physical exertion and fatigue. Autom Constr 104:120–128
58. Larboulette C, Gibet S: A review of computable expressive descriptors of human motion. In: Proceedings of the 2nd International Workshop on Movement and Computing, pp. 21–28 (2015)
59. Glowinski D, Dael N, Camurri A, Volpe G, Mortillaro M, Scherer K (2011) Toward a minimal representation of affective gestures. IEEE Trans Affect Comput 2(2):106–118
60. Hu Z, Yang D, Cheng S, Zhou L, Wu S, Liu J (2022) We know where they are looking at from the rgb-d camera: gaze following in 3d. IEEE Trans Instrum Meas 71:1–14
61. Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann Math Statist 50–60
62. Johnstone IM (2001) On the distribution of the largest eigenvalue in principal components analysis. Ann Stat 29(2):295–327

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.