SSUP-HRI: Social Signaling in Urban Public Human-Robot Interaction dataset

Fanjun Bu
Cornell University, Cornell Tech
New York, USA
fb266@cornell.edu

Wendy Ju
Cornell University, Cornell Tech
New York, USA
wendyju@cornell.edu

Figure 1: Our dataset features hundreds of human-robot interactions in public spaces from the robots' perspective, which showcase rich social signals in a variety of contexts.

ABSTRACT

This paper introduces our dataset featuring human-robot interactions (HRI) in urban public environments. This dataset is rich with social signals that we believe can be modeled to help understand naturalistic human-robot interaction. Our dataset currently comprises approximately 15 hours of video footage recorded from the robots' perspectives, within which we annotated a total of 274 observable interactions featuring a wide range of naturalistic human-robot interactions. The data was collected by two mobile trash barrel robots deployed in Astor Place, New York City, over the course of a week. We invite the HRI community to access and utilize our dataset. To the best of our knowledge, this is the first dataset showcasing robot deployments in a complete public, non-controlled setting involving urban residents.

CCS CONCEPTS

• **Human-centered computing** → **Collaborative and social computing systems and tools**; **Empirical studies in interaction design**; *Empirical studies in collaborative and social computing*.

KEYWORDS

Dataset, human-robot interaction, wizard-of-oz, public interaction, field experiment, urban spaces, behavioral elicitation, social robots

1 INTRODUCTION

While mobile robots are not yet a common feature in public spaces, the emergence of autonomous agents such as self-driving vehicles and unmanned food delivery robots is on the horizon. It will be important to understand how everyday people who encounter these robots might interact with these robots; as part of this, the social signals that they perform in and around the robots will be important for the robots to pick up on and respond to. It is also important to understand how people perceive the movement of the robot itself as a form of social signaling. People inherently attribute social meaning to movement, even of the most basic of shapes [9, 24]. Robots, even those not designed for verbal or social interaction, need to be adept in understanding the signals people make at them, and making sure they respond appropriately in kind. This understanding is crucial to ensure they navigate public spaces safely and respectfully, avoiding causing inconvenience or offense to passersby. Moreover, it is imperative for mobile robots to align their movements with the existing dynamics, norms, and expectations of the public spaces they inhabit.

In this paper, we present a unique dataset that captures interactions between people and two mobile trash barrel robots in a bustling public plaza in New York City. These robots are limited to motion as their sole mode of interaction, and are controlled in a Wizard-Of-Oz (WoZ) method to make sure their movements align with human expectations in the situated context. Our dataset will contribute to the social science understanding of public human-robot interactions and assist in the modeling of social signaling in human-robot interactions in the wild.

2 RELATED WORKS

2.1 HRI in Public Spaces

Previous research on human-robot interactions in semi-public settings like hotels and hospitals has provided specific insights related to particular sites and tasks [3, 8, 20–22]. Broader studies have

concentrated on modeling and predicting human movement to ensure robots can navigate without causing obstructions [1, 2, 6, 25]. Furthermore, significant work in social contexts has delved into proxemic behaviors, developing models to determine the ideal distance between robots and humans during interactions to maintain comfort [11, 12, 18]. However, models for cooperative interaction in public also need to account for the social factors that influence whether people want to interact with a robot or not, such as Koile et al. [16]'s work on "activity zones" in homes, which highlights spatial influences on interactions. Public interaction is similar, but far more complex, with subtle factors like bench arrangement and sunlight direction significantly influencing behavior patterns and social dynamics [27].

2.2 Datasets for HRI

Datasets focusing on physical interactions between humans and robots are scarce because collecting real-time robotics data is inefficient. Consequently, many datasets are centered around human-human interaction data, with the aim for researchers to use the data to create algorithms applicable to robots [4, 17, 19]. Joo et al. [14] released a dataset of nonverbal signals among people in a triadic social interaction scenario, with the hope of teaching machines nonverbal communication skills. In efforts to incorporate robots into data collection, the Wizard-of-Oz method is often employed, wherein robots are remotely controlled by unseen researchers to simulate authentic behaviors [13]. For example, Tian et al. [26] collected a human-robot collaboration dataset featuring handovers with a researcher teleoperating the robot. The MHHRI dataset features triadic interactions between two participants and a robot to study personality and engagement [5]. The robot's dialogue and expressions are also managed through Wizard-of-Oz. Given the complex and context-sensitive nature of human-robot interactions, the Wizard-of-Oz approach is frequently utilized as a reliable method for capturing genuine interaction data.

Different from the aforementioned datasets, which were all collected in controlled lab settings, our dataset features social signaling in a naturalistic environment without explicitly engineered tasks. Here, the signals from passersby are unprompted. Though the range of activities and factors in the dataset are noisier, models built from the data might be anticipated to be more ecologically valid.

3 DATA COLLECTION

In this section, we detailed our study deployment and data collection process. All deployments took place in a public space in New York City and complied with local regulations. Our study was approved by Cornell Tech's IRB protocol #IRB014571.

3.1 Robot Deployments

We deployed two trash barrel robots, one for recycling, and the other for landfill, in Astor Place, New York City. The robots roved the public square and collected trash from people passing by or sitting at the chairs and tables set out in the area. We operated the robots for 2 hours in the early afternoons for 5 days when weather permitted. The hours coincided with high traffic and high occupancy moments in the day.

Methodologically, we used a semi-supervised Wizard-of-Oz protocol [13, 29]. Trained researchers piloted the robot, with the goal of engaging the users as a service robot would. The Wizard-of-Oz method allows us to have the robots respond to social signals in ways that the users were likely to expect; the decisions of the robot operators were as much part of the study as the reactions of the public. For safety reasons, the operators always maintained a line of sight to the robots they were controlling; however, they obscured their role by hiding the controllers or wearing sunglasses so it was less obvious where they were looking. Following people's interactions with the robot, we had another researcher approach to interview willing individuals for their feedback and thoughts on the interaction.

During the deployment, the wizards were encouraged to keep two robots physically close to each other. They were also encouraged to communicate with each other and coordinate the robots' motions. In rare cases where one robot got stuck, the other robot was expected to "bump" the robot in need, instead of direct human intervention.

3.2 Collection Site

To make sure we were collecting naturalistic social signals, we looked for public spaces that were intensively used by residents daily, where the use case of having a mobile robot in the space seemed plausible. Other site selection criteria included foot traffic, availability of public seating, business around the plaza, and ease of research activity operation.

This deployment eventually took place in Astor Place, a public plaza located in the heart of Manhattan, New York City. This public plaza was a trapezium-shaped traffic island with multiple tables and chairs laid out for people to socialize. Originally we intended the robots to be robot chairs and tables, but in our meetings with the business improvement district that operates the plaza, it was determined that robot trash barrels, like those previously deployed by Yang et al. at Stanford [28], would fit in better and gain more interaction. The arrangements of the seating were not fixed day to day. As a result, the wizards sat at different locations for each deployment.

Figure 2: A bird-eye view of our deployment space.

3.3 Recording Apparatus

Each trash barrel robot is a simple mobile platform. Each robot primarily interacts with passersby through its movement. This minimalism helps focus our study and subsequent model on presence and movement, without the confounds that other sounds or visual elements would bring. On the other hand, such design choice also constrains the range of people's signaling towards the robots,

which also largely depends on motion and gesture instead of spoken instructions.

To capture the social signals around them, each trash barrel robot is equipped with an Insta360 One R camera[1] with built-in microphones and an external IMU to record the surrounding and robot motions. The robots are powered by recycled hoverboards; the robust metal chassis provides the structural integrity needed to navigate outdoor urban terrain. Computationally, each robot currently runs on a Raspberry Pi 4 running the Ubuntu 20.04 server and Robot Operating System (ROS 1 Noetic) [23].
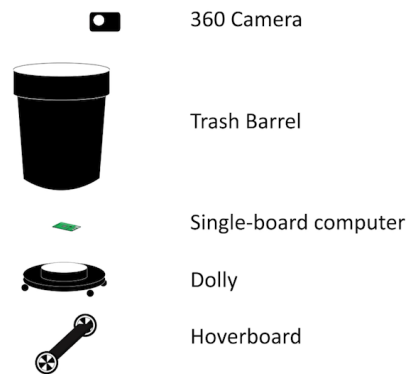
360 Camera

Trash Barrel

Single-board computer

Dolly

Hoverboard

Figure 3: Exploded diagram of robot.

3.4 Data Annotation

To isolate the social signals in the dataset, two researchers from our research group annotated the dataset and labeled every explicit or implicit interaction moment. Between two of the researchers, one of them was involved in the data collection process, where he was on-site for each deployment. The other researcher was not involved during the collection process. The two researchers labeled the dataset individually and then consolidated a spreadsheet of all interaction moments. The spreadsheet will also be publicly available. Based on our interests, our annotation focused on explicit interactions between plaza users and trash barrel robots. For every interaction, each annotator labeled time, which robot was involved in the interaction, and a summary of the scenario.

4 DATASET OVERVIEW

Overall, we observed 274 interactions between the plaza users and trash barrel robots, ranging from normal disposal of trash to people helping the robot to get unstuck. Each interaction consists of unique exchanges of social signals between the users and the robots.

Within the 274 observed interactions, 118 interactions feature people and trash barrel robots negotiated to dispose of trash, for which we provide more detailed annotations focusing on behavioral signaling. Among trash disposal interactions, people interacted with both robots equally (48 interactions with the recycling robot, 54 interactions with the landfill robot, and 16 interactions with both). Over half ($N = 68$) of the interactions feature negotiations between a single user and the robots, while party sizes range from 1 to 5

---

[1]https://www.insta360.com/product/insta360-oner_twin-edition

people. Even though people use different signals to initiate and terminate interactions with the robots, we are able to categorize them into the categories shown in table 1. Lastly, 18 individuals or groups try to have a conversation with the robot during the interaction.

	Signal Type	Count
	Passive: Robot approached	23
	Waving without objects	7
Initiation	Waving/holding objects	47
	Walking towards robots	37
	Standing up	4
	Walking away	47
	Verbal	8
	Turning around	4
	Sitting down	6
Termination	Hand Gesture	16
	Passive: Robots drove away	11
	Looking away	25

Table 1: Different signals people use to initiate and terminate interactions.

5 USE CASES

We anticipate our dataset to significantly impact social science understandings of human-robot interaction in public spaces and aid in advancing social intelligence in machine learning models. Methodologically, we believe our dataset enables bridging qualitative and quantitative practices in the field of human-robot interaction. Pragmatically, our dataset can help establish guidelines to improve the quality of interaction provided by social robots in public.

5.1 Qualitative Analysis

Ethnographic researchers may find our video clips valuable and can discover insights generalizable to other applications. Our dataset features naturalistic exchanges of social signals in the context of social navigation and service robotics. Furthermore, interview transcripts can be used for thematic analysis to understand the mental models and thought processes people go through when encountering robots in uncontrolled environments and deciding ways of signaling. Furthermore, the narration and reflection of the encounter depict how users interpret social signals presented by the robots. Our dataset touches upon topics including, but not limited to, people's sensemaking processes, assumptions, expectations, interaction sequences, and cultural influences on human-robot interaction.

5.2 Quantitative Analysis

Our dataset also opens room for researchers to perform quantitative analysis to provide numerical support for qualitative analysis. From the data, statistics on activities such as people's initiation and termination actions, count party sizes and physical formation, characterize stances (sitting, standing, or walking), and interaction duration are easily observable. Additionally, researchers can profile the robots' actions, based on metrics like approach angles, speed, and distance covered. Our dataset also offers a new playground to

"The image shows a city setting, possibly an outdoor cafe or public space, with two men sitting at a table. One man, wearing a red shirt, is holding a beverage. Possibly a coffee cup, and looking directly at the camera. The other man, wearing a denim shirt, is extending his arm towards the camera, offering or showing a white object, possibly a napkin or a piece of paper. There's also a blue recycling bin nearby. The backdrop shows buildings, pedestrians, and an outdoor vendor. Both men seem aware of the camera's presence, and their body language suggests curiosity or a moment of interaction with the photographer or camera operator."

Figure 4: LLM generated captions with inputs from our field study.

analyze existing theoretical models, such as the social force model and F-Formation [7, 10, 15].

5.3 Machine Learning Test Bed

While our dataset volume is not yet large enough to train a machine learning algorithm from scratch, it provides a test bed for computer scientists to test out their algorithms with realistic real-world data. It is possible both to pick out the social signals generated by passersby who want a robot to approach or stay away and also to model what motions and signals the robots are giving off that people are responding to. Furthermore, given the recent progress in the development of foundational models, we are optimistic that our dataset will play a role in establishing a foundation model for interactive mobile robots navigating social environments in urban public spaces.

6 FUTURE DIRECTION

This project marks the initial phase of establishing a dataset focused on social interactions between humans and mobile robots within urban environments. Our objective is to extend the deployment of these robots to numerous public plazas, with the ultimate aim of covering all five boroughs of New York City. We intend to expand and enrich our dataset as we continually gather new data.

7 CONCLUSION

We present SSUP-HRI, a dataset that features human-robot interaction in public social spaces. The dataset is exploratory in nature, where both the users and the wizards controlling the robots are trying to decipher the social signals presented by each other. Since the data were collected in unconstrained public space, we observed parties of various sizes negotiated with the robots for various purposes under different contexts. Therefore, we believe our dataset will be beneficial to both researchers and practitioners in the understanding of social signals in the context of human-robot interaction in urban spaces.

ACKNOWLEDGMENTS

REFERENCES

[1] Vida Adeli, Mahsa Ehsanpour, Ian Reid, Juan Carlos Niebles, Silvio Savarese, Ehsan Adeli, and Hamid Rezatofighi. 2021. Tripod: Human trajectory and pose dynamics forecasting in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 13390–13400.
[2] Alexandre Alahi, Vignesh Ramanathan, Kratarth Goel, Alexandre Robicquet, Amir A Sadeghian, Li Fei-Fei, and Silvio Savarese. 2017. Learning to predict human behavior in crowded scenes. In *Group and Crowd Behavior for Computer Vision*. Elsevier, 183–207.
[3] Bani Anvari and Helge A Wurdemann. 2020. Modelling social interaction between humans and service robots in large public spaces. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 11189–11196.
[4] Alessandro Carfì, Francesco Foglino, Barbara Bruno, and Fulvio Mastrogiovanni. 2019. A multi-sensor dataset of human-human handover. *Data in brief* 22 (2019), 109–117.
[5] Oya Celiktutan, Efstratios Skordos, and Hatice Gunes. 2017. Multimodal human-human-robot interactions (mhhri) dataset for studying personality and engagement. *IEEE Transactions on Affective Computing* 10, 4 (2017), 484–497.
[6] Pasquale Coscia, Francesco Castaldo, Francesco AN Palmieri, Alexandre Alahi, Silvio Savarese, and Lamberto Ballan. 2018. Long-term path prediction in urban scenarios using circular distributions. *Image and Vision Computing* 69 (2018), 81–91.
[7] Marco Cristani, Loris Bazzani, Giulia Paggetti, Andrea Fossati, Diego Tosato, Alessio Del Bue, Gloria Menegaz, and Vittorio Murino. 2011. Social interaction discovery by statistical analysis of F-formations. In *Proceedings of the British Machine Vision Conference 2011*. British Machine Vision Association.
[8] Marta Díaz-Boladeras, Dennys Paillacho, Cecilio Angulo, Oriol Torres, Jonathan González-Diéguez, and Jordi Albo-Canals. 2015. Evaluating group-robot interaction in crowded public spaces: A week-long exploratory study in the wild with a humanoid robot guiding visitors through a science museum. *International Journal of Humanoid Robotics* 12, 04 (2015), 1550022.
[9] Fritz Heider and Marianne Simmel. 1944. An experimental study of apparent behavior. *The American Journal of Psychology* 57, 2 (1944), 243–259.
[10] Dirk Helbing and Peter Molnar. 1995. Social force model for pedestrian dynamics. *Physical review E* 51, 5 (1995), 4282.
[11] Zachary Henkel, Robin Murphy, Vasant Srinivasan, and Cindy L Bethel. 2012. A proxemic-based HRI testbed. In *Proceedings of the Workshop on Performance Metrics for Intelligent Systems*. 75–81.
[12] Zachary Henkel, Robin R Murphy, and Cindy L Bethel. 2012. Towards a computational method of scaling a robot's behavior via proxemics. In *Proceedings of the 7th annual ACM/IEEE International Conference on Human-Robot Interaction*. 145–146.
[13] Eric Horvitz, Jack Breese, David Heckerman, David Hovel, and Koos Rommelse. 1998. The Lumiére project: Bayesian user modeling for inferring the goals and needs of software users. In *Proceedings of the Fourteenth conference on Uncertainty in Artificial Intelligence*. 256–265.
[14] Hanbyul Joo, Tomas Simon, Mina Cikara, and Yaser Sheikh. 2019. Towards social artificial intelligence: Nonverbal social signal prediction in a triadic interaction. In

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 10873–10883.

[15] Adam Kendon. 1990. Spatial organization in social encounters: The F-formation system. *Conducting interaction: Patterns of behavior in focused encounters* (1990).

[16] Kimberle Koile, Konrad Tollmar, David Demirdjian, Howard Shrobe, and Trevor Darrell. 2003. Activity zones for context-aware computing. In *UbiComp 2003: Ubiquitous Computing: 5th International Conference, Seattle, WA, USA, October 12-15, 2003. Proceedings 5.* Springer, 90–106.

[17] Alap Kshirsagar, Raphael Fortuna, Zhiming Xie, and Guy Hoffman. 2023. Dataset of bimanual human-to-human object handovers. *Data in Brief* (2023), 109277.

[18] Jonathan Mumm and Bilge Mutlu. 2011. Human-robot proxemics: physical and psychological distancing in human-robot interaction. In *Proceedings of the 6th International Conference on Human-robot interaction.* 331–338.

[19] Jan Ondras, Abrar Anwar, Tong Wu, Fanjun Bu, Malte Jung, Jorge Jose Ortiz, and Tapomayukh Bhattacharjee. 2023. Human-robot commensality: Bite timing prediction for robot-assisted feeding in groups. In *Conference on Robot Learning.* PMLR, 921–933.

[20] Eva Blessing Onyeulo and Vaibhav Gandhi. 2020. What makes a social robot good at interacting with humans? *Information* 11, 1 (2020), 43.

[21] Yadong Pan, Haruka Okada, Toshiaki Uchiyama, and Kenji Suzuki. 2013. Direct and indirect social robot interactions in a hotel public space. In *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO).* IEEE, 1881–1886.

[22] Yadong Pan, Haruka Okada, Toshiaki Uchiyama, and Kenji Suzuki. 2015. On the reaction to robot's speech in a hotel public space. *International Journal of Social Robotics* 7 (2015), 911–920.

[23] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, Andrew Y Ng, et al. 2009. ROS: an open-source Robot Operating System. In *ICRA workshop on open source software*, Vol. 3. Kobe, Japan, 5.

[24] Linda Smith and Michael Gasser. 2005. The development of embodied cognition: Six lessons from babies. *Artificial Life* 11, 1-2 (2005), 13–29.

[25] Angelique Taylor, Darren M Chan, and Laurel D Riek. 2020. Robot-centric perception of human groups. *ACM Transactions on Human-Robot Interaction (THRI)* 9, 3 (2020), 1–21.

[26] Leimin Tian, Kerry He, Shiyu Xu, Akansel Cosgun, and Dana Kulic. 2023. Crafting with a Robot Assistant: Use Social Cues to Inform Adaptive Handovers in Human-Robot Collaboration. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction.* 252–260.

[27] William Hollingsworth Whyte. 1980. *The social life of small urban spaces.* Conservation Foundation Washington, DC.

[28] Stephen Yang, Brian Ka-Jun Mok, David Sirkin, Hillary Page Ive, Rohan Maheshwari, Kerstin Fischer, and Wendy Ju. 2015. Experiences developing socially acceptable interactions for a robotic trash barrel. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).* IEEE, 277–284.

[29] J.D. Zamfirescu-Pereira, David Sirkin, David Goedicke, Ray LC, Natalie Friedman, Ilan Mandel, Nikolas Martelaro, and Wendy Ju. 2021. Fake It to Make It: Exploratory Prototyping in HRI. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '21 Companion).* Association for Computing Machinery, New York, NY, USA, 19–28. https://doi.org/10.1145/3434074.3446909