Proceedings Paper:

Wang, Tianyuan, Camara, Fanta, Woolley, Robert et al. (1 more author) (2025) Identifying Public Engagement with Autonomous Art Through Human Pose and Speed Detection. In: 17th International Conference on Social Robotics + AI (ICSR+AI 2025). International Conference on Social Robotics + AI, 10-12 Sep 2025 , ITA.

Identifying Public Engagement with Autonomous Art Through Human Pose and Speed Detection

Tianyuan Wang[1], Fanta Camara[1], Robert Woolley[1], Darren Reed[2]

[1]Institute for Safe Autonomy, University of York, UK.
[2]Department of Sociology, University of York, UK.

Abstract

The successful integration of autonomous systems in public spaces hinges on balancing technical performance with public acceptance and safety concerns. While much research has focused on the technical aspects of autonomous technologies, fewer studies have explored how these systems are perceived and interacted with by the public in artistic contexts. This study addresses this gap by examining how autonomous devices can achieve "social legibility" through a combination of technical and human-centered approaches. The Wheel, an autonomous kinetic sculpture, was used as a tool for studying human-robot interaction in a real-world setting. Deployed at York Festival of Ideas, The Wheel provided an opportunity to observe how the audience members respond to autonomous systems in a public, artistic context. Analysis of audience members' movement and behavior reveals that while The Wheel drew attention and engaged passersby, deeper interactions were less frequent. These findings suggest that creating socially legible autonomous systems requires careful attention to both technical design and public perception. This research contributes to the understanding of how autonomous systems can be better designed to align with human expectations and reduce perceived risks.

Keywords: Autonomous art, Robot performance, Social interactions, Human-robot interaction, Computer vision, AlphaPose

1 INTRODUCTION & BACKGROUND

Autonomous technologies, once used primarily in industrial applications and controlled environments, are increasingly emerging in public spaces, where their integration presents unique challenges and opportunities [1]. This development raises a crucial issue regarding how the integration of such technologies can be optimized to enhance public acceptance and safety. This study, as part of a larger project, explores this question by merging technical innovation with artistic practice to engage the public in meaningful ways. The project leverages The Wheel (shown in Fig. 1), an autonomous kinetic sculpture developed in collaboration with the art practice IOU Theatre, to study and enhance interactions between autonomous systems and the public. This installation not only showcases the potential of robotics in public art but also serves as a research tool to explore the dynamics of human-robot interaction in real-world environments. Public acceptance is often a decisive factor in the successful adoption of new technologies [2]. Despite the growing presence of autonomous systems in public spaces, research on their public interaction, especially through art, remains sparse. For example, [3] investigated the non-verbal behaviors (proxemics) of audience members interacting with a mobile robot manipulator at a festival. [4] described the design of *Fish-Bird*, a kinetic artwork in the form of two wheelchairs that aims to investigate different forms of dialogue between two autonomous robots and their levels of participation in human-robot interaction. [5] developed *Ikit*, an artwork comprised of three robot platforms that autonomously move towards people and make contact with them. [6] used robotic installations as an "artistic medium" to engage with the public. Several large-scale robotic structures and environments were used to induce empathy from audience members towards the mechanistic characters. The present work uses a similar approach to [6], with the aim of collecting data from a multisensor kit and developing a new method to detect audience members' interaction with the kinetic sculpture and its miniature character in a lively festival setting. This dual focus on technical enhancement and public interaction helps advance the technology while ensuring it aligns with societal norms and expectations.

2 DATA COLLECTION

The Wheel is a mobile stage production and large, remote-controlled mechanical sculpture featuring a gleaming, self-propelled, hub-less wheel. The structure, with a diameter of 2.5 meters, has no central axle or spokes, creating a mysterious visual effect as it moves. Inside the wheel, a miniature character walks in sync with the wheel's motion, giving the illusion of an epic journey within its confines. Designed to captivate audience members, The Wheel invites viewers to slow down, observe, and reflect on the scene, fostering a contemplative and interactive experience. The observational data collection for this study was conducted during the York Festival of Ideas, a free public engagement and family-friendly event, where The Wheel served as a central interactive installation. Moving at a speed of 1 meter per minute along a planned route, The Wheel attracted audience members who engaged with the artwork as it progressed. The audience was encouraged to interact by observing the miniature figure walking inside the Wheel and contributing their thoughts or drawings at the accompanying Mobile Gallery, which moved ahead of The Wheel (cf. Fig. 1a). A data acquisition system, referred to as the sensor kit, which is installed on the structure, collected visual and LiDAR data in real time, capturing the movement, behavior, and interactions of

(a) The Wheel.

(b) Sensor kit.

Fig. 1: Pictures of the Wheel at the Festival of Ideas. (**a**) The Wheel with the Mobile Gallery moving ahead. (**b**) The sensor kit used for the data collection.

the audience as they followed and engaged with the installation throughout the event. Ethical approval was obtained from the University of York Ethics Committee.

This sensor kit is composed of a single-board computer (Raspberry Pi 5B with 4GB RAM and a 512GB SSD), two cameras (Raspberry Pi Camera Module V2), and a LiDAR (Unitree L1 PM Lidar). The entire sensor system is powered by an external 12V power supply mounted on The Wheel. The Raspberry Pi accesses the 5V rail via an adapter board (BCRobotic Power + Fan HAT), which then supplies power to the cameras and the SSD, while the LiDAR is powered directly from the 12V supply. All components were assembled using a 3D printed frame, with the Raspberry Pi and power hat positioned on the lower layer, the two cameras placed on both sides of the Raspberry Pi, and the LiDAR mounted facing downward on the upper layer using a trapezoidal bracket, as shown in Fig. 1b. The sensor kit was mounted at the rear of The Wheel at a height of approximately 2 meters above the ground to ensure the audience remains within the LiDAR's sensing range. The data were acquired using the ROS2 (Humble) stack running on Raspberry Pi OS (Bookworm 64-bit). Image and point cloud data were collected by two camera nodes (camera-ros) and one LiDAR node (unitree-lidar-ros2), respectively, with the data being timestamped by ROS2 and stored using the rosbag2 node.

3 METHODS

In this study, the sensors were not directly mounted at the center of the object of interest, which is a relatively large structure. As a result, the viewing angles of the audience often varied significantly. In fact, when the audience gazed at the camera, they were not necessarily interacting with the object of interest but the sensors, as illustrated in Fig. 2. Consequently, gaze detection based on facial features did not consistently provide accurate interaction information. In addition, to include individuals

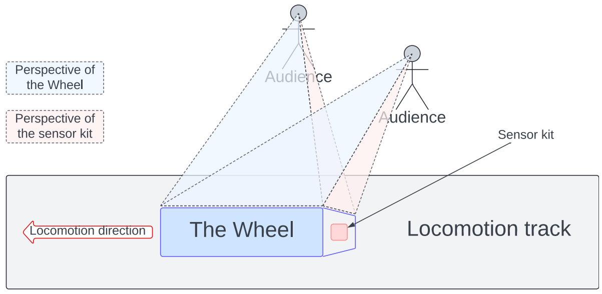


Fig. 2: The perspective deviation when observing The Wheel and the sensor kit, as well as the wide view angle when observing The Wheel.

who may not have approached the artwork for direct interaction but were nevertheless drawn to it, slowing their pace as they passed by, we employed interaction detection based on the audience member's walking speed.

3.1 Position Extraction and Preprocessing

To obtain the position information of audience members, AlphaPose [7] was utilised offline to extract the keypoints data. An example detection output of AlphaPose is shown in Fig. 4a. For the selection of position reference points, to mitigate the offset caused by the bounding box only partially enclosing the body and to account for potential drift when using the nose as a reference point due to head movement, the midpoint between the shoulders was used as the reference point for determining the audience member's position. Various disturbances and noise encountered during data collection were addressed through preprocessing. Initially, data with a confidence score below 0.5 were discarded (confidence or visibility is the uncertainty of the point from the pose detection algorithm), as such data exhibited significant positional jumps, making them unreliable for accurate position analysis. Facial data were not utilized for subject identification in this study.

Video analysis revealed that AlphaPose's tracking of individuals was not consistently reliable, with the algorithm occasionally misidentifying different individuals appearing in consecutive frames as the same person. To eliminate this interference, individuals whose positional data exhibited large jumps between consecutive frames were re-labeled. Additionally, audience members that appeared in the video for less than two seconds were excluded. Furthermore, the sensor kit mounted on the object of interest included a mechanical LiDAR, which inevitably introduced vibrations during scanning, leading to noisy positional reference outputs. To mitigate this, a notch filter was first applied to the raw positional data in both the x and y directions to remove the 11 Hz noise introduced by the LiDAR. Based on the filtered audience member's velocity data, the rolling standard deviation was calculated. This method helps mitigate potential misjudgments caused by perspective-related speed variations among audience members positioned at different distances from the sensor kit. Moreover, the rolling standard deviation provides a more comprehensive representation of the velocity changes over time, avoiding the disproportionate influence of brief, extreme

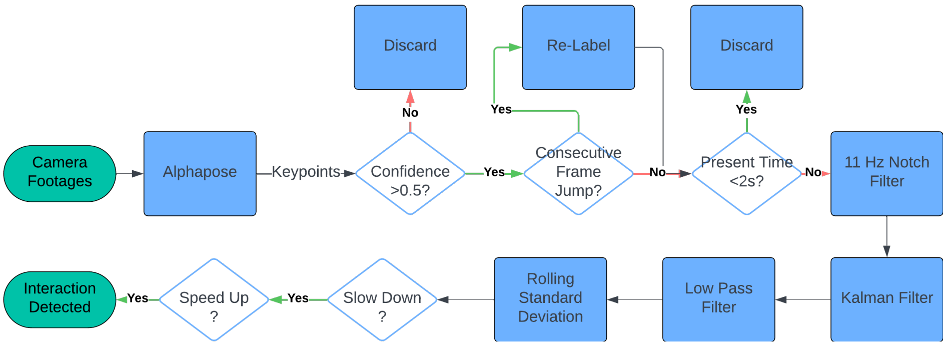


Fig. 3: The steps of the data pre-processing and interaction detection analysis.

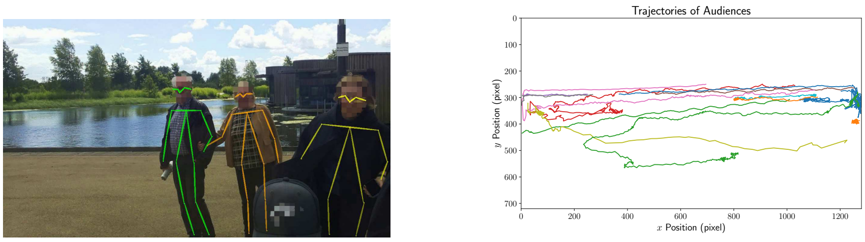
fluctuations in speed. Notably, some data exhibit high initial values that gradually decrease to a lower level. This is typically observed when audience members suddenly appear in close proximity to the sensor kit, leading to significant changes in recorded speed. However, these abrupt changes do not necessarily indicate meaningful interaction, which is why both a decrease in speed and a subsequent departure are used as criteria for assessing interaction.

3.2 Interaction Detection

Subsequently, based on the preprocessed keypoints data, the velocity of each individual was calculated. Given the inherent limitations in the accuracy of the pose detection algorithm, a Kalman filter was first applied to the velocity data to reduce noise in the observed velocities. A low-pass filter was then applied to the velocity data to eliminate the influence of walking gait, resulting in a smooth velocity curve. The data processing workflow is illustrated in Fig. 3. To detect interaction between audience members and the object of interest, the processed velocity curves were analyzed. A key characteristic of audience members who were passing by and became attracted to The Wheel is that they slowed down or stopped their movement. This behavior is reflected in the data as a decrease in walking speed. After concluding the interaction, the audience members would leave the area of interest, indicated by an increase in walking speed. Since the employed pose detection algorithm does not provide three-dimensional positional information, and to address potential issues caused by varying distances and different baseline walking speeds among audience members, the relative rate of speed decrease and subsequent increase was used as the criterion. When an audience member's walking speed exhibited a decrease followed by an increase exceeding the specified threshold, it was inferred that interaction with The Wheel had occurred during the period of reduced speed.

4 RESULTS & DISCUSSION

To facilitate data visualization in this paper, a randomly trimmed 60-second footage was used as a sample and analyzed. The audience members' position information



(a) AlphaPose output.   (b) Post-processed Trajectories.

Fig. 4: (**a**)Snapshot of the video with keypoints generated by AlphaPose labelled in the scene. The faces of the audience members were blurred to comply with ethical requirements. (**b**) The trajectories of the audience members passing by The Wheel after the data pre-processing. The x and y axes of the figure correspond to the width and height of the image, respectively.

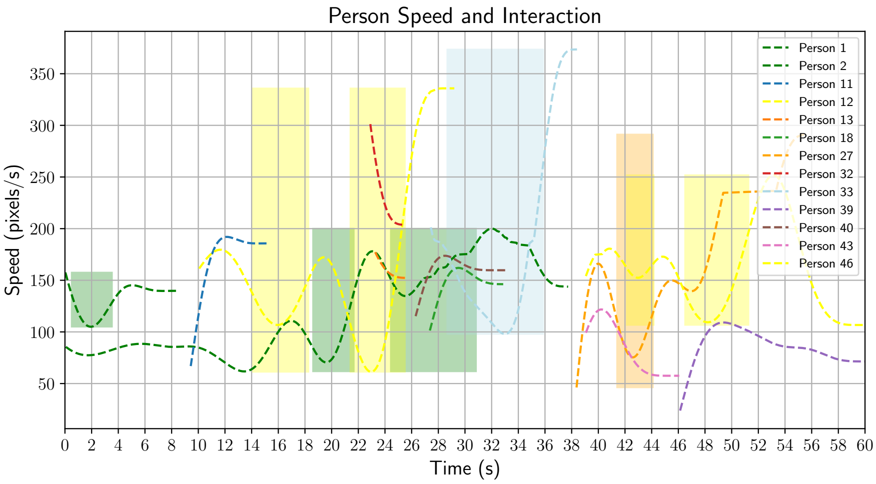


Fig. 5: The detected interaction activities of the audience members.

before and after the pre-processing stage was firstly compared, where their trajectories in the footage is plotted and shown in Fig. 4b. The results indicated a significant improvement in the continuity and smoothness of the audience members' trajectories after data pre-processing. Trajectories displaying large variation along the y axis represent audience members who were more strongly attracted when they passed by and approached closer for interaction.

Fig. 5 depicts the variation in audience members' walking velocity over time, with periods of detected interaction highlighted using rectangles colored corresponding to the relevant data. From the filtered velocity profiles, it is clear that audience members' interactions can be identified with relative clarity. Multiple deceleration events can be observed during an audience member's interaction, likely reflecting instances where audience members paused to examine different parts of The Wheel, a large structure that naturally draws audience members' attention to its various elements. Conversely, some velocity profiles show brief appearances of audience members without any noticeable deceleration, indicating a lack of interaction with The Wheel. Decreased walking velocity is one component of the production of a socially meaningful action of 'showing interest in the Wheel'. Alongside velocity, we might find body direction change and potentially other embodied actions such as gaze adjustment and gesture production (such as a pointing action). These and other meaningful actions are available through an approach to social interaction called Embodied Conversation Analysis (ECA), which identifies sequences of action through video analysis [8, 9]. These actions are socially meaningful and consequential for the progression of collective behaviors. A more detailed social analysis is provided in [10].

A key observation from the analysis is that audience members' interaction with The Wheel can be identified primarily through changes in walking velocity. Audience members who slowed down, stopped, or altered their trajectory were inferred to be interacting with the installation. A manual labelling of the social interactions in the video clip shows that the proposed interaction detection approach is about 60% accurate in predicting them. This method revealed several limitations due to its reliance on pose detection and tracking algorithms, such as AlphaPose employed in this study, which introduced inaccuracies when identifying individuals and their movement patterns. For instance, errors in assigning tracking IDs or misidentifying individuals after overlap events in the video led to instances of missed or false interaction detections. These challenges are inherent in real-world, outdoor environments where occlusion, overlapping individuals, and dynamic lighting conditions can complicate accurate detection and tracking. Participants' demographics was not studied in this work, but it is hoped that including this information in the future could improve the modelling and results.

The study also opens several avenues for improving the methodology used to capture human-robot interactions. First, enhancing the robustness of pose detection and tracking algorithms in dynamic, outdoor settings would lead to more reliable data on audience members' movement and interaction. Future research could explore the use of multiple sensor modalities, such as combining visual data with LiDAR sensors, to improve detection accuracy in cases of occlusion or poor lighting. Additionally, integrating machine learning techniques to adaptively adjust tracking algorithms based on real-time feedback could reduce errors in identifying audience members. Future work will also aim to quantify the proxemic and trust behavioural preferences of audience members towards the Wheel using the models from [11–13].

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to IOU Theatre (https://ioutheatre.org) for supporting this work with Wheel and the data collection.

References

[1] Fayyaz, M., González-González, E., Nogués, S.: Autonomous mobility: A potential

opportunity to reclaim public spaces for people. Sustainability **14**(3), 1568 (2022)

[2] Huijts, N.M.A., Molin, E.J.E., Steg, L.: Psychological factors influencing sustainable energy technology acceptance: A review-based comprehensive framework. Renew. Sustain. Energy Rev. **16**(1), 525–531 (2012)

[3] Krenn, B., Gross, S., Dieber, B., Pichler, H., Meyer, K.: A proxemics game between festival visitors and an industrial robot. arXiv preprint arXiv:2105.13812 (2021)

[4] Velonaki, M., Rye, D.: Designing robots creatively. Robots and Art: Exploring an Unlikely Symbiosis, 379–401 (2016)

[5] Doepner, S., Jurman, U.: Robot partner—are friends electric? Robots and Art: Exploring an Unlikely Symbiosis, 403–423 (2016)

[6] Vorn, B.: I want to believe—empathy and catharsis in robotic art. Robots and Art: Exploring an Unlikely Symbiosis, 365–377 (2016)

[7] Fang, H.-S., Li, J., Tang, H., Xu, C., Zhu, H., Xiu, Y., Li, Y.-L., Lu, C.: Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(6), 7157–7173 (2022)

[8] Reed, D.J., Wooffitt, R., Young, J.: Walking with gail: the local achievement of interactional rhythm and synchrony through footwork. Social Interaction. Video-based Studies of Human Sociality (2023)

[9] Reed, D.J.: Turning heads and making conversation on twitch. Discourse, Context & Media **60**, 100802 (2024)

[10] Reed, D., Camara, F., Wang, T.: Social interaction with autonomous art: Combining social analysis and computer vision. In: International Conference on Social Robotics + AI (ICSR+AI) (2025). (In Press)

[11] Camara, F., Fox, C.: Space invaders: Pedestrian proxemic utility functions and trust zones for autonomous vehicle interactions. International Journal of Social Robotics **13**(8), 1929–1949 (2021)

[12] Camara, F., Fox, C.: Extending quantitative proxemics and trust to hri. In: 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pp. 421–427 (2022). IEEE

[13] Camara, F., Fox, C.: A kinematic model generates non-circular human proxemics zones. Advanced Robotics **37**(24), 1566–1575 (2023)