Proceedings Paper:

Reed, Darren James orcid.org/0000-0001-9018-0145, Camara, Fanta and Wang, Tianyuan (2025) Social Interaction with Autonomous Art:Combining Social Analysis and Computer Vision. In: 17th International Conference on Social Robotics + AI (ICSR+AI 2025). International Conference on Social Robotics + AI, 10-12 Sep 2025 , ITA.

Social Interaction with Autonomous Art: Combining Social Analysis and Computer Vision

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

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

Abstract

The deployment of autonomous robots in public spaces is increasing thanks to advancements in robot localisation, mapping and navigation strategies, but there is a slower development of autonomous robotic systems in artistic contexts. This project introduces The Wheel, an autonomous kinetic sculpture that moves through the crowd at art festivals. Using onboard camera data, this work provides some insight into the social interactions happening between The Wheel and people by combining a form of social analysis, called Embodied Conversational Analysis, with computer vision analysis through the AlphaPose algorithm. The findings include a list of meaningful social actions performed by people interacting with The Wheel, thus offering a promising approach to better understanding social interactions with autonomous robots performing on stage and beyond.

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

1 Introduction

Autonomous technologies, primarily designed for industrial applications in controlled environments, are increasingly emerging in public spaces, where their integration presents unique opportunities and challenges [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 explores this question by merging technical innovation, social behavioral analysis, and artistic practice to explore meaningful public engagement in novel ways. An important difference with robotic or autonomous art pieces is that they are meant to garner attention. A successful piece unsettles ongoing perceptions and practices and encourages engagement and proximity. Arguably, rather

(a) (b)

Fig. 1: Pictures of The Wheel at the Festival of Ideas. (**a**) The Wheel with the sensor kit. (**b**) The Wheel at an art festival.

than increasing the likelihood of collision, such attention-paying forms the basis of human-object spatial coordination. Yet with proximity comes greater danger and only through a detailed understanding of existing human behaviors and their 'projection' can the device anticipate potential dangers.

The present 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. Despite the growing presence of autonomous systems in public spaces [2], research on their public interaction, especially through art, remains sparse. Our project begins to fill this gap by employing The Wheel to gather data on how people perceive and interact with autonomous technologies 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. This work is interdisciplinary in that it brings together researchers from robotics, computer vision and sociology, alongside art practitioners. The paper details the development of a novel combination of existing methods and practices, so as to inform real-time machine learning-based models development for social robots. This work contributes the following:

- a novel approach combining embodied conversation analysis with computer vision analysis via the use of AlphaPose to provide some insight into social interactions with an autonomous robotic art system;
- a summary list of meaningful social actions identified in these social interactions and indicative computer vision techniques needed to identify them autonomously.

2 Related Work

The study of human interactions with autonomous vehicles is a very active field, with topics such as human detection and tracking that have well-established solutions

[3] but human trajectory prediction and interaction behaviors are less understood research areas [4]. Most of these works have focused on on-road vehicles that drive people around or mobile robots that deliver some services to them. But recently, several projects have started demonstrating the potential of integrating robotics with art, to study human-robot interaction. For example, the "Shimon" project [5] developed a robotic marimba player capable of improvising music with human musicians, [6] designed robots capable of drawing portraits of human subjects, [7] and [8] showed how to make robots dance alongside humans, and [9] proposed a cognitive architecture for interactive humanoid robots performing poetry on stage. However, there are fewer studies on public engagement with autonomous mobile robots. For instance, [10] investigated the nonverbal behaviors (proxemics) of audience members interacting with a mobile robot manipulator at a festival. [11] described the design of Fish-Bird, a kinetic artwork in the form of two wheelchairs that aim to investigate different forms of dialogue between two autonomous robots and their levels of engagement in human-robot interaction. [12] developed Ikit, an artwork comprised of three robot platforms that move autonomously towards people and make contact with them. [13] 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 with the aim to collect data from a multi-sensor kit and develop a new method to detect audience members' interactions with The Wheel and its miniature character.

At the same time, Sociology has a long history of interest in movement and visibility in public spaces [14, 15]. This includes the manner in which pedestrians glance at each other [16, 17], indicate objects of common interest [18], and navigate spaces occupied by vehicles and objects [19]. The sociologist Erving Goffman [20] conceived of public behavior as a performance in front of an audience and various forms of public 'spectacle' have been understood in terms of performance engagement [21]. He described 'civil inattention' as a key component of public behaviors, a normative strategy of 'polite' avoidance. Goffman identifies mitigation of inattention through a staged progression towards engagement and interaction through preparatory movements such as glances to and from a person or object of interest [18]. It is from the tension between public performance and polite avoidance that a social understanding of engagement with art-based autonomous devices is born. Drawing on Goffman, the applied approach of Embodied Conversation Analysis, understands how 'multimodal' communication elements, such as gaze and gesture, are resources for doing public interaction [22]. As robots and autonomous vehicles form part of everyday life, embodied resources have become a means of understanding interaction [2] and form a social understanding of human-device interaction [23]. This is seen in the prediction of human attention and engagement through the monitoring of behaviors such as gaze and gesture in the robot design literature [24]. It is akin to the research on the 'legibility' of robot motion [25].

3 Methods

The approach used here combines speed and trajectory produced through AlphaPose[26] with a social analytic approach called Embodied Conversation Analysis

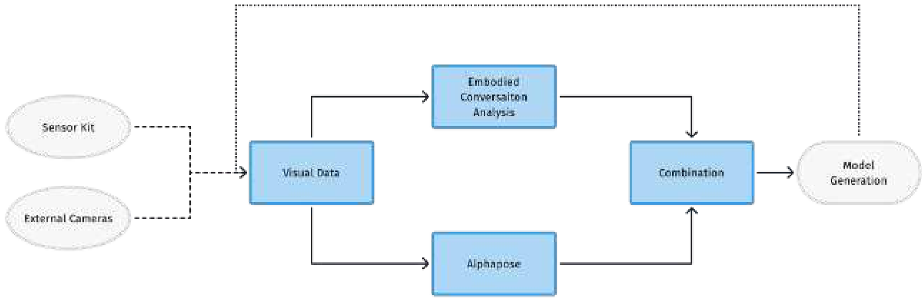


Fig. 2: Data processing pipeline.

(ECA) [27]. AlphaPose represents real-time onboard visual acuity while ECA provides a detailed description of the meaning of social behavior. The resulting work pipeline is shown in Fig. 2.

3.1 Data collection

Data collection for this study was conducted during the York Festival of Ideas, 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. 1). A data acquisition system, referred to as the sensor kit, which is installed on the structure, collected visual data in real-time, capturing the movement, behavior, and interactions of the audience as they followed and engaged with the installation throughout the event. 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 sensing range. The data were acquired using the ROS2 (Humble) stack running on Raspberry Pi OS (Bookworm 64-bit). Ethical approval was obtained from the University of York Ethics Committee.

3.2 Embodied Conversation Analysis

Embodied Conversation Analysis (ECA) is an approach in Sociology that details sequences of meaningful social interaction, typically in relation to talk [28]. For ECA, social interaction is premised upon the projection of upcoming elements such that - for example - a speaker anticipates the end of a co-speaker's turn, and prepares to respond at the appropriate moment [29]. Similar projection trajectories are identified in embodied activities, such as gesture and gaze production [30]. ECA follows a rigorous procedure of transcription and analysis. In the following transcripts, given that there is no recorded talk, the 'footwork' of the participants is used as a primary means to track the production of ongoing activity. Each step on either the left ("L") or right

("R") foot is separated into a 'preparatory' phase as the leg moves forward (indicated by periods) and a 'step' the period during which the foot is in contact with the floor (indicated by tildes). The combined movement is called a 'stride'. The same preparation and action 'phrasing' is used with eye movement (head pan, and aligned gaze) and gesture production (preparatory movement of the arm and then production of a pointing hand shape). Each character represents one tenth of a second. Corresponding images from the video are indicated with a star symbol and numbered. The transcription details the section of the video clip in the caption and a brief 'gloss' is included in double brackets justified right. Consecutive actions are separated by a semicolon and italicized descriptions relate to the italicized character in the transcript. The transcript is presented in a fixed width font so that the alignment of embodied elements can be seen.

3.3 AlphaPose

The position information of audience members were obtained through AlphaPose [26] by extracting keypoints data and using the midpoint between the shoulders as the reference point for determining the audience member's position. After pre-processing 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 and trajectory. The detailed method using AlphaPose for the velocity and trajectory generation is described in [31].

4 Analysis and Results

The following instances are analyzed from an Embodied Conversation Analytic approach and then detailed through the AlphaPose algorithm. The analysis follows a strategy of increasing behavioral complexity, starting with two people walking past and glancing at The Wheel, and then moving through increased levels of complexity in relation to behavior. The first instance is used to convey the marrying of the two methods, while the later more complex examples, are used to hint at potential ways to develop the analysis.

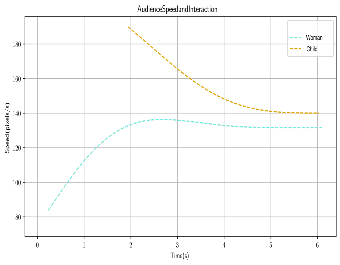
4.1 Instance 1: Gazing at an object

The onboard camera (pointed to the right) captured two people walking in parallel with The Wheel (Fig. 3b). Initially the woman can be seen in the frame (Fig. 3a, line 01), but then – at a slight distance – the second person (a child) appears (line 02). AlphaPose provides a reading of the speed and trajectory of the two people, shown in Fig. 3c and Fig. 3d, respectively. These show the relative movement of each person, as the second 'catches up' with the first, while maintaining a parallel trajectory. The transcript in Fig. 3a (lines 01 and 02) shows the woman's 'stride' ('preparation' plus 'step') to be on average larger and longer than the child's. On the fifth step (right foot) the child turns his gaze to the left (line 04), looking directly at the camera (on

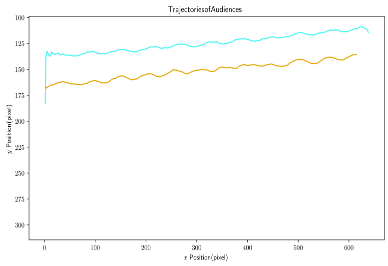
```
01 Wfeet:   L..~R..~L..~R...~L...~R...~~L...~R...~
02 Cfeet:               R~L.~R..~L..~R..~L.~R..
03 Whead:                                L-----      ((small pan left; look))
04 Chead:                               L------            ((pan left; look))

05 Wfeet:   L...~R...~~L..~
06 Cfeet:   ~L..~R..~L..~R~
07 Whead:     L.-----------                                  ((full pan left))
08 Chead:   ---------------
```

(a) Transcript 1 – Output Left 17-02 to 17-08 (onboard camera)

(b) Image



(c) Speed



(d) Trajectory

Fig. 3: Results of the ECA transcription and AlphaPose for Instance 1.

The Wheel). One tenth of a second later, on a left-foot step (line 03), the woman pans her gaze slightly to the left (possibly looking at person ahead of The Wheel, out of frame). During the latter part of a left foot preparation (line 06) she pans her gaze to The Wheel (line 07). The sequence can be summarized as two people walking past The Wheel. First one turns their gaze to look and then the other first pans towards something happening to her left and then fully turns her gaze to look at The Wheel. Both child and woman maintain this gaze alignment as they walk out of frame. Gaze realignment ("gazing-at") is an action that could preface more direct engagement, perhaps through a change of walking direction that brings them closer to The Wheel. Glancing at an object is a first stage action ("inattention" to attention) in a potential longer sequence of actions that culminates in engagement. It is precisely this type of indicative movement that a human uses to 'project' potential future actions. A similar sequential projection is seen in instance 2 (Fig. 4).

```
01 Mfeet:    R.~L...~~R...~~L...~~                        ((second L turned out))
02 Mhead:    ~~~~~~~~~~~~~~~L.....                    ((facing to left; pans left))
             *1            *2

03 Mfeet:    R...~~L....~~R..~~~L...
04 Mhead:    L.......    F....L..           ((down and up; to front; left & down))
                               *3

05 Mfeet:    ~~~R....~~L....~~~R.....~~
06 Mhead:    ~~~~~~~~~~~~~~~~~~~~~~~~~~~                        ((looking at Wheel))
             *4

07 Mfeet:    L...~~~R...~~~~                        ((Rf flat, no Left heel raise))
08 Mhead:    ~~~~~~~~~~~~~~~                                    ((looking at Wheel))

09 Mfeet:    (18.7)                                   ((stands still on both feet))
10 Mhead:    (18.7)                                             ((looking at Wheel))
```

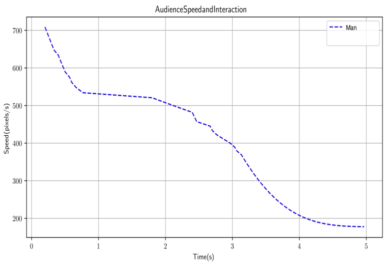
(a) Transcript 2 – VID00006 and Left Output 7:22 - 8:05 (onboard camera)

(b) Image 1

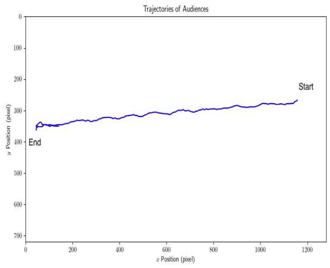
(c) Image 2

(d) Image 3

(e) Image 4



(f) Speed



(g) Trajectory

Fig. 4: Results of the ECA transcription and AlphaPose for Instance 2.

4.2 Instance 2: Walking towards, scene scanning, stride termination

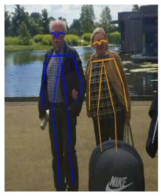
Here the person is already oriented towards The Wheel as he enters the frame (Fig. 4b) but the line of movement changes as each left foot stride turns the body to the left (Figs. 4c and Fig. 4d) and the person moves closer to The Wheel (Fig. 4e). This is seen in the trajectory graph (Fig. 4g) as he slows down (Fig. 4f). This "body re-orientation" (moving from a 'side-on' position to a 'facing' position) and speed reduction is a meaningful social action of 'walking-towards' The Wheel. Walking-towards is followed by 'scene-scanning' (looking down up, left, front) as shown in (Fig. 4a, line 04) as he looks at a person to the side of The Wheel and then turns his gaze to The Wheel itself (Fig. 4a, line 06). In Fig. 4a (line 07), the man comes to rest by transferring his weight into his right foot ('step') without beginning a preparation of the left foot ('no left heel

```
01 Wfeet:   ~~~~L.....~~~R....~~~                                ((Rf prep shorter))
02 Mfeet:   ~~~~L.....~~~R....~~~
03 Mgaze:   ((looking at Wheel))
04 Wgaze:   ......                                                  ((pan to Wheel))
05 Wbody:                 ........                          ((turns shoulders left))
            *1                      *2
06 Wfeet:   L.....~~R.~~~~~~
07 Mfeet:   L.......~~~R..                                              ((RF to LF))
08 Wgaze:   ........     ....                            ((pan to left; pan to W))
                        *3
09 Wfeet:   -------------------------                                ((looks at W))
10 Mfeet:   ------------L.....~~~R...                                ((looks at W))

11 Wfeet:   -------------------------                                ((looks at W))
12 Mfeet:   ~~~L.....~~~R.....~~~~L..

13 Wfeet:   -------------------~R.....  (0.5)R.((steps Lf, Rf to Lf, Rf step))
14 Wgaze           .....                                            ((pans to left))
15 Mfeet:   .....R.....|out of frame                              ((brings Rf to Lf))
```

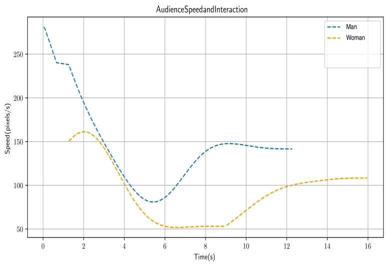
(a) Transcript 3 – VID00006 and Left Output 0:38 - 0:53

(b) Image 1

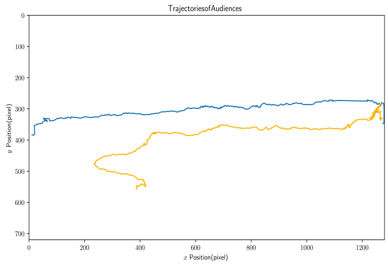
(c) Image 2



(d) Image 3



(e) Speed



(f) Trajectory

Fig. 5: Results of the ECA transcription and AlphaPose for Instance 3.

raise'). This action is a 'stride-termination'. Showing interest through gaze and body re-orientation, change of velocity and trajectory is indicative of potential engagement with an object-of-interest. Indeed, we see this engagement in Fig. 4a (lines 09 and 10) when the person stands still and looks at The Wheel (not shown in velocity graph). This behavior terminates the movement-towards The Wheel and hence mitigates the potential for path crossing and collision. To 'gazing-at' as an indicative movement, we can add 'walking-towards', 'scene-scanning', and 'stride termination' to a list of meaningful movements.

4.3 Instance 3: Footwork synchrony & engagement coordination

In the following instance, a man and a woman are walking in parallel with The Wheel. As they come into view (Fig. 5b) they are both stepping on their left then right foot (Fig. 5a, lines 01 and 02) in synchrony ('footwork synchrony' or 'walking together'). The man is already looking at The Wheel. A typical walking speed is 0.6 seconds for the preparation (indicated with periods) and step (indicated by tildes), yet both the man and woman's left foot strides are 0.9 seconds in length, meaning that they have slowed down together (see also speed graph in Fig. 5e). Overlapping the preparation of this stride, the woman pans her gaze to The Wheel (Fig. 5a, line 04), producing a shared attentional gaze with the man. She turns her shoulders and upper body towards The Wheel (line 05, Fig. 5c)('body re-orientation'). In Fig. 5a line 06, the woman produces a quicker stride preparation on her left foot, meaning she steps while the man is still moving his foot (line 07) ('footwork a-synchrony'). His step (line 07) coincides with her right foot preparation and the commencement of a step, meaning that she is transferring her weight into her right foot as he brings his feet together ("RF to LF")("stride-termination"). In overlap with this, the woman once again pans her gaze to The Wheel and the couple stop and look at it together (Fig. 5d) - indicated by a dashed line - for another second before moving off. The footwork, scene-scanning, and body re-orientation work to engineer stride termination and engagement coordination. The man then walks closer to The Wheel (Fig. 5a lines 10, 12, 15)('walking-towards'), while the woman first maintains her gaze and body alignment and then turns and walks away (to her left) (Fig. 5a lines 09, 11, 13)('walking-away'). As the couple are walking past The Wheel with synchronized strides ('footwork synchrony'), the man maintains gaze alignment ('gazing-at'), showing interest. Their footwork falls out of synchrony as the woman looks at The Wheel, then to away to the people around it, and then back to it ('scene-scanning'). They both stop and look at The Wheel. As he moves closer, she first stands looking and then turns away and moves in the opposite direction to him (Fig. 5f). This is one example of engagement coordination followed by differentiated interest and engagement and shows that while initial engagement might be coordinated, the nature and length of that engagement might change.

4.4 Instance 4: Gesture & engagement coordination

In this final instance, a man and child are walking at different pace-lengths (Fig. 6f) but similar speed, although the man's speed is initially slower relative to the camera (Fig. 6e). While they are both walking diagonally towards The Wheel, there are notable turns to the left on the left foot by the child in line 04 (Figs. 6a and 6d) and by the man in Fig. 6a (line 05), indicating a change in trajectory. In Fig. 6a line 03, the man produces a gesture preparation by raising his hand to his eyes (Fig. 6b), shielding them from the sun. This gesture acts to emphasize visual interest (an exaggerated 'gazing-at'- or what McNeill calls a 'metaphoric gesture') [32]. He then produces another gesture preparation to bring his left hand to a pointing ("deictic") gesture (Fig. 6c), angled towards the camera on The Wheel as shown in Fig. 6a line 06. A further gesture preparation has him bring his left hand back to his brow (Fig. 6d). We might ask

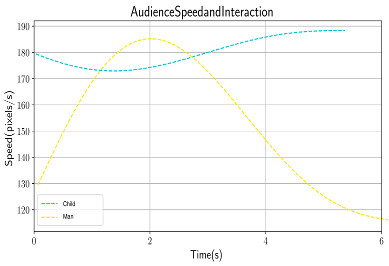
```
01 Cfeet:  L..~R..~~L..~~R..~~L.
02 Mfeet:  L..~~R...~~L...~~~R..
03 Mgest:   ....~~~~~~~~~~~~~....                                  ((above eyes))
               *1

04 Cfeet:  ..~~R..~~L...~~R...~~                              ((turning to left))
05 Mfeet:  .~~~L....~~~R...~~~L....                              ((turning left))
06 Mgest:  ~~~~~~~.....~~~~~~~~~~                          ((pointing; above eyes))
           *2        *3
```

(a) Transcript 4 – Instance 1: Left Output 0:10-0:17.

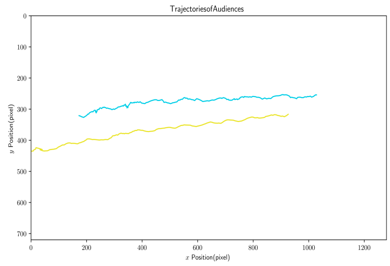
(b) Image 1

(c) Image 2

(d) Image 3



(e) Speed



(f) Trajectory

Fig. 6: Results of the ECA transcription and AlphaPose for Instance 4.

why the man produces this series of movements. After all, the child is already visually engaged with The Wheel and walking towards it. One answer is that this series of layered actions indicates a display of growing interest and hence a likely progression towards physical engagement with The Wheel. Importantly, this growing interest is communicative; it is produced for the child. It results in the reorientation of the child's body as he steps on his left foot in Fig. 6a line 04, angling his body further towards The Wheel (Fig. 6d), precisely following the end of the pointing gesture (Fig. 6a, line 06). To the list of meaningful actions we can add a sequence of gesture forms ('looking' and pointing then 'looking') that work to coordinate mutual attention and engagement ('engagement coordination').

4.5 Meaningful Social Actions

Table 1 provides a summary of the meaningful social actions observed in the instances' analysis as well as indicative computer vision techniques required to identify them

Meaningful Social Actions	Indicative techniques from computer vision
Gazing-at	Gaze tracking + speed estimation
Walking-towards	Speed + trajectory estimation
Scene-scanning	Gaze tracking + Head pose estimation
Stride termination	Leg pose + speed estimation
Body re-orientation	Whole body pose + trajectory estimation
Footwork (a)synchrony	Multi-persons leg pose + speed + trajectory estimation
Engagement coordination	Speed + trajectory estimation
Looking at	Head pose estimation
Pointing at	Hand gesture recognition + tracking

Table 1: Summary list of meaningful social actions from the observed social interactions with The Wheel and some indicative computer vision techniques needed to identify them.

autonomously. These actions have a sequential ordering, such that one typically precedes another. Generally, this follows a progression from gaze through gesture and body orientation to movement termination. For example, the action of stopping and looking, or what we have called 'stride termination', is then a finely tuned collective activity that is achieved through various activities that we have termed 'engagement coordination'. Moving through the instances in more detail, Instance 1 (Sec. 4.1) shows the 'gazing-at' action on the part of both woman and child. In that they are sequentially produced, one after the other, it shows a move from inattention to shared attention, and hence projects the potential for joint engagement. Gazing-at is also apparent in Instances 2 (Sec. 4.2) and 3 (Sec. 4.3). In the latter, the gazing-at of the man leads to a shared visual attentiveness that in turn results in stride termination. An alternative form of visual action is identified in Instance 2 with the man, who is already walking towards The Wheel, scanning the scene. While we cannot substantiate the point here, such actions are often oriented to recognition of existing shared interest on the part of others in the scene and working out what they are looking at. We see hints of this in the initial glance of the woman in Instance 1 and the scanning gaze of the woman in Instance 3. What comes to the fore in Instance 3 is the indication of upcoming stride termination through the move from footwork synchrony to a-synchrony. By shortening her stride and turning her shoulders (body-reorientation), the woman indicates and projects her intention to stop and look. Finally in Instance 4 (Sec. 4.4), engagement coordination is most apparent in the sequential production of an exaggerated looking pointing gestures. This informs engagement coordination.

5 Discussion

In each of the instances, the velocity and trajectory information produced through AlphaPose is combined with a descriptive analysis of the movements through the ECA approach. While the two forms of information provide insights into the relative movements of scene participants, the latter extends this insight through the identification of various socially meaningful 'actions' summarized in Table 1. This table currently contains a mix of informal and formal action descriptions which are not fully detailed here and are planned for future work. Translating a social understanding into a computer vision technique is a difficult process and one that highlights the complexities of human perception. While there may be some kind of 'mechanism' underpinning social perception and practice, revealing it is an ongoing process. Designing a computer version of human experience is a means to pursue such understandings and insights, this table is thus 'indicative' of such linkages, but it does not claim to be definitive.

The fine-grained action description provided by the ECA approach requires decades of expertise in human behaviour analysis and is therefore very time-consuming. The combined ECA approach with AlphaPose used here aims to reduce computation time and enable autonomous analysis and understanding of human behaviour. With existing approaches to supervised machine learning, video annotation involves identifying single actions which are then used to predict new instances [33]. ECA has the potential to improve processes of supervised learning by incorporating the incremental and procedural features of 'projectable' actions into training models. Future work aims to address that and also extends the analysis to additional examples of actions from various sources of data.

Acknowledgments

The authors would like to express their sincere gratitude to: (1) IOU Theatre (https://ioutheatre.org) for supporting this work with The Wheel and the data collection; (2) Rob Woolley who helped design the sensor kit.

References

[1] Fayyaz, M., González-González, E., Nogués, S.: Autonomous mobility: A potential opportunity to reclaim public spaces for people **14**(3), 1568. Publisher: MDPI. Accessed 2025-03-07

[2] Pelikan, H.R., Reeves, S., Cantarutti, M.N.: Encountering autonomous robots on public streets. In: Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction, pp. 561–571

[3] Camara, F., Bellotto, N., Cosar, S., Nathanael, D., Althoff, M., Wu, J., Ruenz, J., Dietrich, A., Fox, C.W.: Pedestrian models for autonomous driving part i: low-level models, from sensing to tracking **22**(10), 6131–6151. Publisher: IEEE. Accessed 2025-03-07

[4] Camara, F., Bellotto, N., Cosar, S., Weber, F., Nathanael, D., Althoff, M., Wu, J., Ruenz, J., Dietrich, A., Markkula, G.: Pedestrian models for autonomous driving part ii: high-level models of human behavior **22**(9), 5453–5472. Publisher: IEEE. Accessed 2025-03-07

[5] Hoffman, G., Weinberg, G.: Shimon: an interactive improvisational robotic marimba player. In: CHI '10 Extended Abstracts on Human Factors in Computing Systems, pp. 3097–3102. ACM. https://doi.org/10.1145/1753846.1753925

. https://dl.acm.org/doi/10.1145/1753846.1753925 Accessed 2025-03-07

[6] Sidnell, J.: The architecture of intersubjectivity revisited. In: Sidnell, J., Enfield, N.J., Kockelman, P. (eds.) The Cambridge Handbook of Linguistic Anthropology. Cambridge Handbooks in Language and Linguistics, pp. 364–399. Cambridge University Press. https://doi.org/10.1017/CBO9781139342872.018 . https://www.cambridge.org/core/books/cambridge-handbook-of-linguistic-anthropology/architecture-of-intersubjectivity-revisited/F0CDB2C89CE94A8DBD29322AE450E28D

[7] Abad, P., Franco, M., Castillón, R., Alonso, I., Cambra, A., Sierra, J., Riazuelo, L., Montano, L., Murillo, A.C.: Integrating an autonomous robot on a dance and new technologies festival. In: Ollero, A., Sanfeliu, A., Montano, L., Lau, N., Cardeira, C. (eds.) ROBOT 2017: Third Iberian Robotics Conference vol. 693, pp. 75–87. Springer. https://doi.org/10.1007/978-3-319-70833-1_7 . Series Title: Advances in Intelligent Systems and Computing. http://link.springer.com/10.1007/978-3-319-70833-1_7 Accessed 2025-03-07

[8] LaViers, A.: Make robot motions natural **565**(7740), 422–424. Publisher: Nature Publishing Group UK London. Accessed 2025-03-07

[9] Rodriguez, I., Astigarraga, A., Lazkano, E., Martínez-Otzeta, J.M., Mendialdua, I.: Robots on stage: A cognitive framework for socially interacting robots **25**, 17–25. Publisher: Elsevier. Accessed 2025-03-07

[10] 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. arXiv. https://doi.org/10.48550/arXiv.2105.13812 . http://arxiv.org/abs/2105.13812 Accessed 2025-03-07

[11] Velonaki, M., Rye, D.: Designing robots creatively. In: Herath, D., Kroos, C., Stelarc (eds.) Robots and Art, pp. 379–401. Springer. https://doi.org/10.1007/978-981-10-0321-9_19 . Series Title: Cognitive Science and Technology. http://link.springer.com/10.1007/978-981-10-0321-9_19 Accessed 2025-03-07

[12] Doepner, S., Jurman, U.: Robot partner—are friends electric? In: Herath, D., Kroos, C., Stelarc (eds.) Robots and Art, pp. 403–423. Springer. https://doi.org/10.1007/978-981-10-0321-9_20 . Series Title: Cognitive Science and Technology. http://link.springer.com/10.1007/978-981-10-0321-9_20 Accessed 2025-03-07

[13] Vorn, B.: I want to believe—empathy and catharsis in robotic art. In: Herath, D., Kroos, C., Stelarc (eds.) Robots and Art, pp. 365–377. Springer. https://doi.org/10.1007/978-981-10-0321-9_18 . Series Title: Cognitive Science and Technology. http://link.springer.com/10.1007/978-981-10-0321-9_18 Accessed 2025-03-07

[14] Simmel, G.: Sociology of the senses. visual interaction. In: Introduction to the Science of Sociology (Third Edition), 3rd edn., pp. 356–361. University of Chicago

Press

[15] Rosental, C.: Toward a sociology of public demonstrations **31**, 343–365 https://doi.org/10.1177/0735275113513454

[16] Sacks, H.: Lecture eleven: On exchanging glances **12**(3), 333–350. Publisher: Springer. Accessed 2024-09-25

[17] Macbeth, D.: Glances, trances, and their relevance for a visual sociology. In: Jalbert, P.L. (ed.) Media Studies: Ethnomethodological Approaches, pp. 135–170. University Press of America

[18] Goffman, E.: Behavior in Public Places: Notes on the Social Organization of Gatherings vol. 91194. Free Pr

[19] Conley, J.: A sociology of traffic: Driving, cycling, walking. Technologies of Mobility in the Americas (Intersections in Communications & Culture), pp. 219–236. Peter Lang

[20] Goffman, E.: The Presentation of Self in Everyday Life. Doubleday. Doubleday

[21] Benford, S., Crabtree, A., Flintham, M., Greenhalgh, C., Koleva, B., Adams, M., Tandavanitj, N., Farr, J.R., Giannachi, G., Lindt, I.: Creating the spectacle: Designing interactional trajectories through spectator interfaces **18**(3), 11

[22] Mondada, L.: Bodies in action: Multimodal analysis of walking and talking **4**(3), 357–403 https://doi.org/10.1075/ld.4.3.02mon

[23] Vinkhuyzen, E., Cefkin, M.: Developing socially acceptable autonomous vehicles **2016**(1), 522–534. Publisher: John Wiley & Sons, Ltd. Accessed 2025-04-11

[24] Dey, D., Walker, F., Martens, M., Terken, J.: Gaze patterns in pedestrian interaction with vehicles: Towards effective design of external human-machine interfaces for automated vehicles. In: Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 369–378. ACM. https://doi.org/10.1145/3342197.3344523 . https://dl.acm.org/doi/10.1145/3342197.3344523 Accessed 2025-04-11

[25] Dragan, A.D., Lee, K.C., Srinivasa, S.S.: Legibility and predictability of robot motion. In: 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 301–308 (2013). IEEE

[26] 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)

[27] Reed, D.J.: Turning heads and making conversation on twitch. Discourse, Context

& Media **60**, 100802 (2024)

[28] Hutchby, I., Wooffitt, R.: Conversation Analysis. Principles Practices and Applications. Polity Press. Polity Press

[29] Local, J., Kelly, J.: Projection and 'silences': Notes on phonetic and conversational structure **9**(2), 185–204

[30] Streeck, J., Jordan, J.S.: Projection and anticipation: The forward-looking nature of embodied communication. **46**(2), 93–102 https://doi.org/10.1080/01638530902728777 . Accessed 2022-04-17

[31] Wang, T., Camara, F., Woolley, R., Reed, D.: Identifying public engagement with autonomous art through human pose and speed detection. (under review)

[32] McNeill, D.: Hand and Mind: What Gestures Reveal About Thought. University of Chicago Press

[33] Zhang, J., Gould, S., Ben-Shabat, I.: Vidat—ANU CVML video annotation tool. https://github.com/anucvml/vidat