

Event-Aligned Social Robotic Co-Viewing for Sports: A Humanoid Peripheral Companion

Roy El-Helou and Matthew K.X.J. Pan
Ingenuity Labs Research Institute
Queen’s University
Kingston, Ontario, Canada

Abstract—This work investigates an event-centred framing of robotic companionship in which a physically embodied humanoid robot serves as a peripheral co-viewer during sports programming, reacting to unfolding game events. We present a compact whole-body motion library grounded in spectator affect theory, and a system architecture supporting Wizard-of-Oz, replay, and live operation modes for controlled and autonomous event-aligned reactions. Three pilot studies examine how an embodied co-viewer influences enjoyment, shared reality, and interpretation of affective alignment across individual and group viewing contexts. Exploratory pilot findings suggest that in group settings, the robot’s expressions increased pleasure and engagement without disrupting shared reality among human viewers, while effects were mixed in solitary viewing. Participants consistently perceived the robot as affectively aligned with the game, suggesting that contingent embodied reactions alone can support social attribution in shared leisure contexts. These pilots provide early observations that motivate future controlled evaluation that peripheral, non-assistive robotic co-viewers can shape collective media experiences and motivate future work on embodied companionship in everyday narrative settings.

I. INTRODUCTION

Social isolation is increasingly recognized as a serious societal problem with measurable impacts on well-being and social connectedness [1]. Beyond acute loneliness, isolation accumulates through ordinary routines that are now frequently experienced alone. Media consumption is a prominent example: emotionally salient content that was historically shared is increasingly encountered in solitude, reducing opportunities for collective interpretation and shared affect [2]. As a result, isolation is not merely episodic but embedded within everyday experience, motivating approaches that address how people co-experience moments rather than simply how they communicate.

Social robotics has been widely investigated as one such approach. Most existing systems mitigate isolation through user-centric interaction paradigms in which robots monitor individuals and provide support, guidance, or conversation, framing companionship as an outcome of sustained user-directed engagement [3]. This dominant model leaves a complementary question underexplored: can a robot contribute socially without assisting or monitoring the user, instead participating peripherally in a shared activity and reacting to the same unfolding narrative? Foundational theories of social presence and the media equation suggest that contingent embodied behaviour alone can be sufficient to elicit social attribution, even in the absence of dialogue [4], [5].

Attempts to recreate shared experience through digital mediation illustrate both the demand for co-experiencing and the limitations of disembodied channels. “Second-screen” behaviour exposes viewers to parallel reactions on social

platforms [6], yet mediated co-viewing consistently yields lower perceived co-presence and emotional engagement than physical co-location [7]. Social presence research attributes this deficit to the absence of embodied non-verbal cues [4], while studies of collective experiences show that shared affect is reinforced through visible bodily alignment and synchronized responses [8]. These findings point toward embodiment as a critical mechanism for restoring experiential togetherness.

Sports spectatorship provides a particularly revealing context to investigate this mechanism. Sporting events unfold through discrete, high-salience moments that evoke predictable affective trajectories and historically depend on collective reaction [9]. This structure enables isolation of a central question we wish to explore: whether embodied, event-aligned reactions alone—without dialogue, personalization,

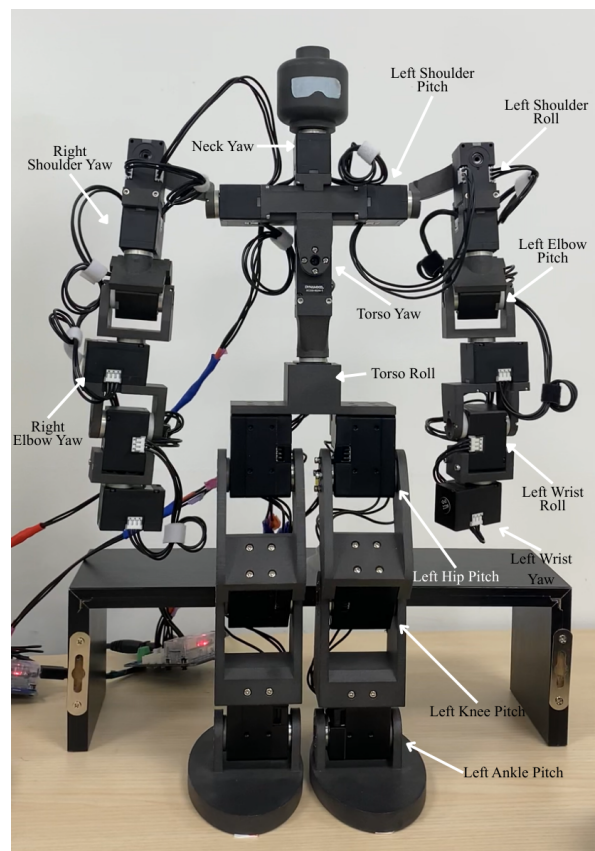


Fig. 1: Custom 23 DoF humanoid tabletop robot used for our social robotic sports co-viewing study. Labels indicate joint types; symmetric joints on the contralateral limb are omitted for clarity.

or assistance—can meaningfully shape the experience of watching.

In this work, we investigate an event-centred framing of robotic companionship in which a physically embodied humanoid robot acts as a peripheral co-viewer during sports programming. The robot (shown in Fig. 1) contributes solely through culturally legible spectator behaviours (e.g., celebration, protest, disappointment), expressed non-verbally in response to unfolding game events. We implement this framing through a compact whole-body motion library grounded in bodily expression and spectator affect [9]–[11], integrated within a system architecture supporting Wizard-of-Oz, replay, and live event-aligned operation.

We report three pilot studies spanning Wizard-of-Oz group co-viewing, individual replay, and group replay. We examine whether a peripheral robot co-viewer influences enjoyment and engagement, whether it disrupts or supports shared reality among human viewers, and how people interpret the robot’s affective alignment and perceived bias. Across pilots, participants consistently perceived the robot as affectively aligned with the game. In group settings, the robot increased pleasure and engagement without disrupting shared reality among co-viewers, whereas effects were mixed in solitary viewing. Together, these results provide initial evidence that contingent embodied reactions alone can support social attribution in shared leisure contexts and motivate future work on event-aligned robotic co-viewing beyond sports.

II. RELATED WORK

Social robotics has largely focused on reducing isolation through user-centric interaction paradigms, in which the robot monitors the person and responds with support, guidance, or assistance. In elder care, socially assistive robots are frequently designed and evaluated as interventions that reduce loneliness through structured engagement, positioning companionship as an outcome of sustained user-directed interaction [3]. Recent systems reinforce this orientation. As an example, ElliQ is a proactive social and health-coaching companion for older adults, providing daily conversational interaction and wellness prompts to address loneliness [12]. Companion doll robots, such as Hyodol, similarly emphasize emotional support and routine-based interaction and are evaluated as long-term home companions for mental-health and quality-of-life outcomes [13]. Across these systems, companionship is operationalized primarily through the robot’s supportive role toward the user, rather than through participation alongside the user in a shared narrative.

Alongside this work, a smaller body of research examines social connection through co-presence and shared context rather than assistance. Studies of virtual humans in mixed-reality environments show that agents which respond to events in the shared environment can increase perceived co-presence even when task interaction and dialogue are minimal [14]. Related findings from mediated social touch research demonstrate that affective alignment alone can support trust and social bonding without an explicit instrumental goal [15]. At the same time, experience-sampling studies comparing face-to-face and computer-mediated interaction consistently report lower positive affect and perceived social support in mediated settings [16]. Together, these findings suggest that shared affect and responsiveness matter, but that disembodied systems have limits in sustaining emotionally rich togetherness.

Within embodied HRI, there are emerging attempts to design robots as companions in shared media experiences, but the space remains fragmented and underspecified in relation to “event-centred” companionship—that is, participation through contingent responses to a shared unfolding narrative rather than user-directed interaction. Hoffman et al. introduced and experimentally evaluated the concept of *Robotic Experience Companionship (REC)* by examining how a small, autonomous desktop robot, synchronized with media content, influences people’s experience of listening to music and watching short videos [11]. Their studies show that simple, non-interactive bodily responses to media increased participants’ sense that the robot was co-experiencing the content with them, shaped enjoyment of the media, and led to more positive attributions toward the robot, even in the absence of dialogue or explicit interaction. More recent work on sports co-viewing, such as BleacherBot, moves in the opposite direction by centring conversational AI agents that explicitly interpret game context, manage emotional arousal, and align with user preferences through language-based interaction and personalization [17].

In contrast, the present work investigates a middle ground that remains underexplored in both lines of research: a physically embodied robot that serves as a peripheral co-viewer in live sports, engaging in nonverbal, event-aligned reactions without relying on dialogue, personalization, or explicit conversational engagement. Rather than shaping experience through semantic commentary or coaching, the robot’s contribution is constrained to culturally legible spectator behaviours, allowing us to examine how shared attention and embodied affective alignment alone support enjoyment and shared reality during a highly structured, emotionally dynamic narrative.

III. RESEARCH QUESTIONS AND CONTRIBUTIONS

Despite extensive work on socially assistive and conversational robots, companionship in HRI is typically framed as emerging from user-directed interaction in which the robot monitors, guides, or supports the individual. Far less explored is whether companionship can arise through peripheral participation in a shared activity, where an embodied agent contributes socially by reacting to an external unfolding narrative rather than the user themselves. To investigate this alternative framing, we study a physically embodied humanoid robot acting as a co-viewer during sports viewing and evaluate its influence on viewer experience through a set of pilot deployments.

We address the following research questions:

- **RQ1:** Does the presence of an embodied robot co-viewer influence enjoyment relative to viewing alone?
- **RQ2:** Does an event-aligned robotic co-viewing alter perceived shared reality among viewers?
- **RQ3:** How do viewers interpret the robot’s affective responses, including perceived alignment and team bias?

Through the exploration of these questions, we contribute the following:

- 1) An event-centred model of robotic companionship in which a physically embodied humanoid participates peripherally by reacting to a shared unfolding narrative rather than monitoring or assisting the user.
- 2) A robotic co-viewing platform that operationalizes event-aligned companionship through a compact interpretable whole-body motion library and multi-mode

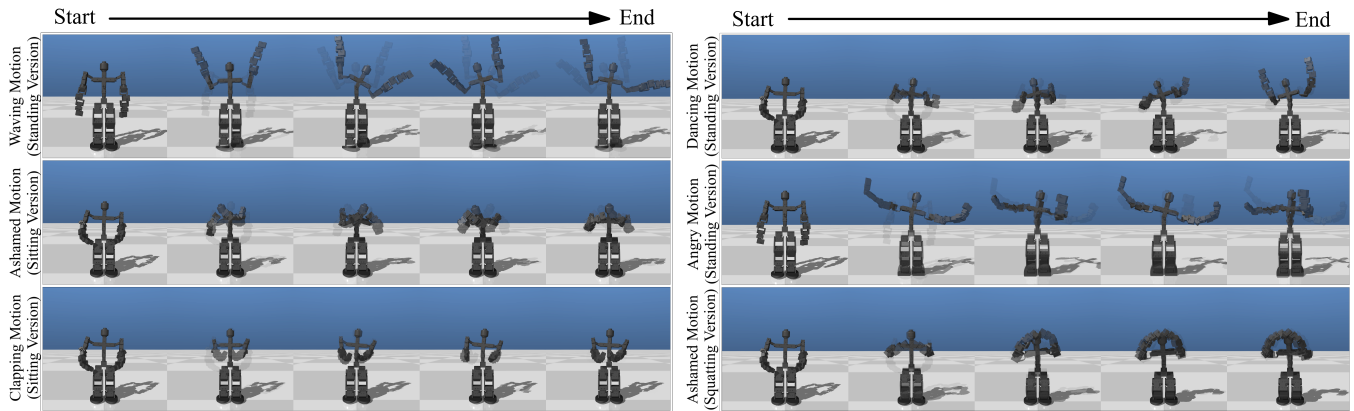


Fig. 2: Subset of whole-body reaction motions from the robot’s motion library, illustrating seated and standing variants across the five affective categories. Additional library motions include idle behaviours and seated/standing anchor postures used for posture transitions and baseline states.

execution framework (Wizard-of-Oz, replay, live), illustrating how embodied peripheral participation can be realized and experimentally controlled.

- 3) Exploratory observations and methodological insights on how peripheral embodied co-viewers influence enjoyment, shared reality, and perception of affective alignment through three pilot studies (spanning Wizard-of-Oz, individual replay, and group replay settings)

IV. METHODOLOGY

We propose a robotic co-viewing system designed to explore whether an embodied robot that reacts to unfolding sports events with expressive body movements can shape the viewing experience. To manage uncertainty while progressively increasing autonomy, the system was developed through three operational modes introduced sequentially: Wizard-of-Oz (WoZ), replay, and live. WoZ mode (Sec. IV-D.1) establishes stable motion execution and timing under direct human control; replay mode (Sec. IV-D.2) introduces autonomous behaviour selection while maintaining control over event timing through offline synchronization to recorded video; and live mode (Sec. IV-D.3) extends the same behaviour-selection and execution constraints to real-time operation using networked event acquisition and queued processing. Across all modes, the motion library and execution framework remain fixed, so differences arise primarily from how events are obtained, timed, and mapped to discrete behaviours. In this context, an event refers to a discrete entry in a sports play-by-play feed, typically consisting of a natural-language description, a score update, and game clock information.

A. Platform

The system is implemented on a custom 23-degree-of-freedom tabletop humanoid robot (shown in Fig. 1). The robot is actuated using two Dynamixel actuator models: the smaller joints are driven by Dynamixel XC330-M288-T servos, while the larger load-bearing joints (hips, knees, and torso) are powered by Dynamixel XM430-W210-T servos. All joints operate in position control mode. Low-level communication is handled in Python through a custom control interface built on the ROBOTIS Dynamixel SDK [18]. The interface initializes the actuator bus, addresses individual joint IDs, and transmits position commands directly to the motors.

B. Motion Library and Execution

As sporting events are conventionally associated with distinct emotional reactions expressed through recognizable body-language patterns, prior work has identified systematic relationships between such kinematic features and perceived robot expressivity [9], [10]. The motion library was therefore designed to translate these movement patterns into joint-space trajectories compatible with the current platform. Robot behaviour is constrained to a fixed library of 17 expressive motions and two anchor postures: seated and standing. Motions are authored offline using a keyframe-based process in which the robot is physically placed into desired configurations and joint positions are recorded. Full trajectories are then generated by linearly interpolating between keyframes. The resulting motion set spans a range of upper-body gestures, torso movements, and posture shifts:

- **Seated:** angry (downward arm thrust and torso roll), ashamed (forward hip flexion and raised arms to head), clap (bilateral arm convergence), dance (bent arms and torso roll oscillation), wave (bilateral arm elevation and torso yaw oscillation), lean-in (forward hip rotation) and squat-ashamed (lowered seated configuration).
- **Standing:** angry (thrusting leg movement), ashamed (slow oscillating legs), clap, dance and wave (all opposing oscillating leg motions).
- **Idle:** sit idle (slow arm stretch), stand idle 1 (slow arm stretch) and stand idle 2 (slow limited neck rotation).
- **Posture transitions:** sit-to-stand cheer (quick upright elevation of arms and body) and sit-to-stand protest (quick sideways elevation of arms and upright elevation of body).
- **Anchor:** seated (arms folded inwards) and standing (arms downwards).

Standing variants retain the same upper-body kinematic patterns as their seated counterparts, differing only in lower-body configuration to maintain upright balance. Refer to Fig 2 for samples of motions.

During execution, joint position targets are streamed at 200 Hz. Motions are serialized, so only one motion or transition runs at a time. Motions are not interrupted once initiated. This design avoids mid-trajectory discontinuities and preserves physical stability, as trajectories are authored as complete keyframe sequences without blending.

C. Event-to-Motion Mapping

A core challenge is mapping play-by-play events, which are provided as unstructured natural language and vary substantially in phrasing and context, to a small set of discrete robot behaviours. A rule-based mapping based on keywords is brittle under this variability and does not naturally account for context such as score state, recent event history, game momentum, or posture-dependent feasibility. We therefore employ GPT-4 as a semantic decision layer that converts each event description into a constrained action selection [19].

GPT-4 selects from predefined motion identifiers or a no-reaction option; it does not generate trajectories, change kinematic parameters, or issue joint-level commands. For each game, GPT-4 was instructed to select from a fixed vocabulary of predefined motion identifiers or a no-reaction option, with most routine play-by-play events (e.g., substitutions, timeouts, quarter transitions, low-salience possessions) expected to produce no reaction. For each subsequent event, the model is queried with the current play description along with recent event history, allowing selection decisions to reflect score state and unfolding game context. As a result, reactions were concentrated primarily around emotionally salient events such as scoring plays, turnovers, missed opportunities, and momentum shifts. To improve behavioural consistency and reduce inappropriate responses, replay schedules were manually reviewed after generation to remove erroneous mappings or undesirable posture transitions prior to experimental deployment. During execution, reactions were not queued: if a new event became due while another motion was still executing, the newer event was skipped to preserve motion continuity and physical coherence. In practice, skipped reactions were uncommon overall but occurred occasionally during dense gameplay sequences such as rapid scoring exchanges or late-game momentum shifts. This semantic mapping policy intentionally favours sparse reactions over continuous activity, examples of this are shown in Table I.

TABLE I: Example event-to-motion mappings generated under the constrained GPT-4 selection policy.

| Play-by-Play Event | Context | Selected Reaction |
|--|-------------------------|--------------------------------------|
| J. Randle (MIN) makes 25' 3pt jump shot, assisted by M. Conley | Q2, MIN trailing 31–26 | Seated Clap |
| O. Ighodaro (PHX) makes alley-oop dunk, assisted by C. Gillespie | Q2, MIN down 31–23 | Seated Anger |
| G. Allen (PHX) misses driving layup | Late Q2, MIN down 54–49 | Standing Wave + stand→sit transition |

To investigate RQ3 regarding perceived affective alignment and team bias, the system includes an explicit favoured-team parameter that conditions reaction polarity. Sports spectatorship is inherently partisan, and emotional expressions are typically interpreted relative to team allegiance. Without an explicit stance, reactions to scoring events would be ambiguous (e.g., celebration vs. protest), reducing interpretability of alignment. By parameterizing team support, the system enables controlled manipulation of perceived allegiance while keeping the motion library and execution constraints constant. This allows evaluation of whether observers infer a stable team preference from nonverbal reactions alone and whether such perceived bias influences enjoyment, shared reality, or interpretation of the robot’s role.

D. Operating Modes

1) *Wizard-of-Oz Mode*: In WoZ mode, reactions are triggered manually by an operator observing the broadcast video feed. The operator selects motions and transitions from the same library used in autonomous operation. WoZ is included to establish a baseline for reliable event-paced execution and posture management, free from uncertainty introduced by automated event interpretation or reaction selection.

2) *Replay Mode*: Replay mode uses GPT-4-based motion selection (as described in Sec. IV-C but controls timing through offline synchronization to recorded broadcast video). A completed game’s play-by-play event log is processed sequentially, and each event is assigned either a motion identifier or no reaction. Event timing is then aligned with the recorded video using a manual synchronization tool, in which the operator presses a key at each event onset while the video plays, producing a timestamped schedule. During playback, the system monitors elapsed video time and triggers scheduled reactions when their timestamps are reached. Execution follows the shared coherence policy: transitions are automatically applied when needed, motions play to completion, and overlapping due events are skipped.

3) *Live Mode*: Live mode uses the same GPT-4 selection and execution constraints described previously, but retrieves events in real time from the ESPN play-by-play feed [20]. Newly detected plays are inserted into a delayed processing queue, with a 45-second buffer applied to account for network variability, software overhead, and GPT inference time while allowing posture transitions to be staged before execution. Events are scheduled using game clock metadata (quarter and time remaining) rather than wall-clock arrival time, preserving relative spacing between plays even under fluctuating retrieval delays. The interface also provides operator controls for re-synchronization with the televised broadcast when drift occurs. When a play becomes due, GPT-4 selects either a predefined motion identifier or a no-reaction option under the constrained output policy. Execution follows the shared coherence constraints: motions are serialized, allowed to complete, and overlapping events are skipped rather than queued.

V. PILOT STUDIES

Three pilot studies were conducted to validate the experimental procedure, assess measurement instruments, and obtain preliminary insights into event-aligned robotic viewing across increasing levels of autonomy and social configuration. Pilot 1 employed the Wizard-of-Oz control mode to assess feasibility under ideal responsiveness conditions. Pilot 2 transitioned to replay mode to evaluate autonomous GPT-based reaction selection in an individual setting. Pilot 3 retained replay mode but reintroduced a group context to examine whether the robot’s influence differed in socially shared viewing.

Although a live mode was implemented as described in Section IV-D.3, it was not evaluated in these pilots. While operational, the live configuration introduces variability in event retrieval and broadcast synchronization that cannot be fully controlled. Because the objective of the pilots was to isolate social and experiential outcomes rather than evaluate system robustness, replay mode was selected to ensure consistent timing and identical event–reaction mappings across participants.

All pilots followed a two-condition within-subject structure consisting of a baseline viewing phase without the robot

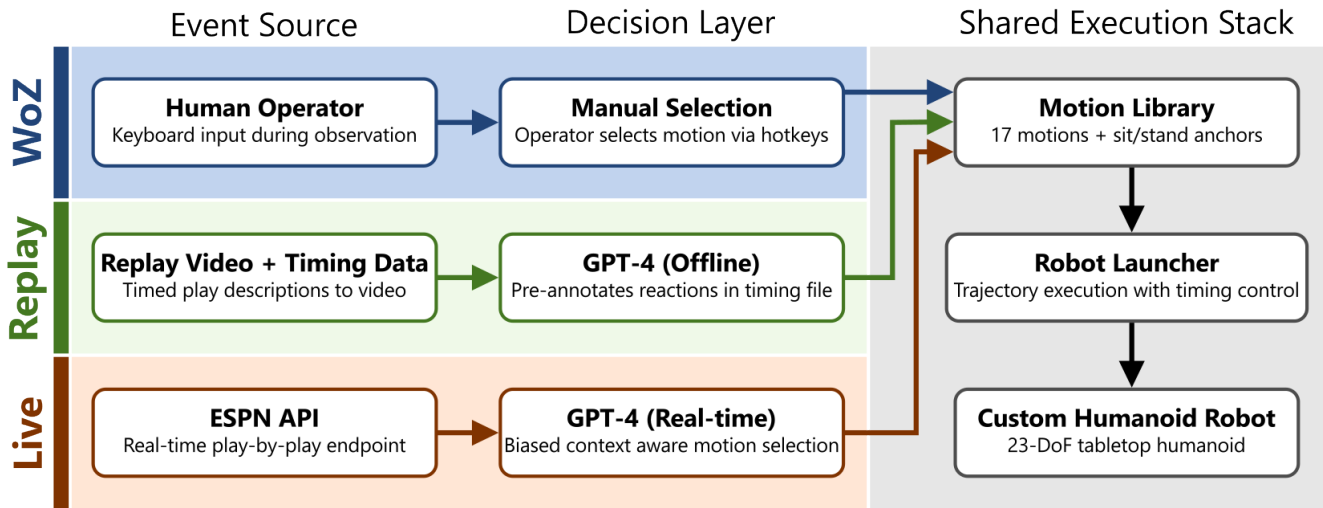


Fig. 3: System architecture for event-aligned robotic co-viewing. Three input modes (Wizard-of-Oz, Replay, Live) select reactions from the motion library, which are then executed on the custom humanoid platform.

and a second phase with the robot present. This structure operationalizes RQ1 and RQ2 by enabling within-participant comparison of enjoyment and shared reality in the absence and presence of the embodied co-viewer. Participants first completed the General Attributes Towards Robots Survey (GAToRS) [21] to characterize baseline attitudes toward robots. They then viewed the first half of a sports game without the robot and completed a post-condition questionnaire assessing enjoyment and, in multi-viewer studies, shared reality with other viewers. Enjoyment was measured using the ENJOY scale [22], addressing RQ1. Shared reality was measured using the Shared Reality toward a Target (SR-T) scale [23], addressing RQ2. In the robot-present condition, SR-T was additionally administered with the robot as the target to assess perceived affective alignment (RQ3).

The robot was introduced only in the second half of the game to frame it as an additive co-viewer rather than

an initially defining feature of the experience. This fixed ordering enhances ecological plausibility but introduces a progression confound related to game intensity; accordingly, results are interpreted conservatively, such that observed effects must exceed increases attributable solely to late-game dynamics. All measures were rated on 7-point Likert scales, with higher values indicating stronger agreement or experience.

A. Pilot Study 1: WoZ Group Setting

Pilot 1 evaluated the motion library and execution framework under human-controlled reaction selection in a one-to-many setting. This configuration established a feasibility baseline under ideal responsiveness and ensured that any perceptual effects could be attributed to embodied, event-aligned behaviour rather than to autonomy-level variability. Participants viewed a recorded Korea–Japan baseball

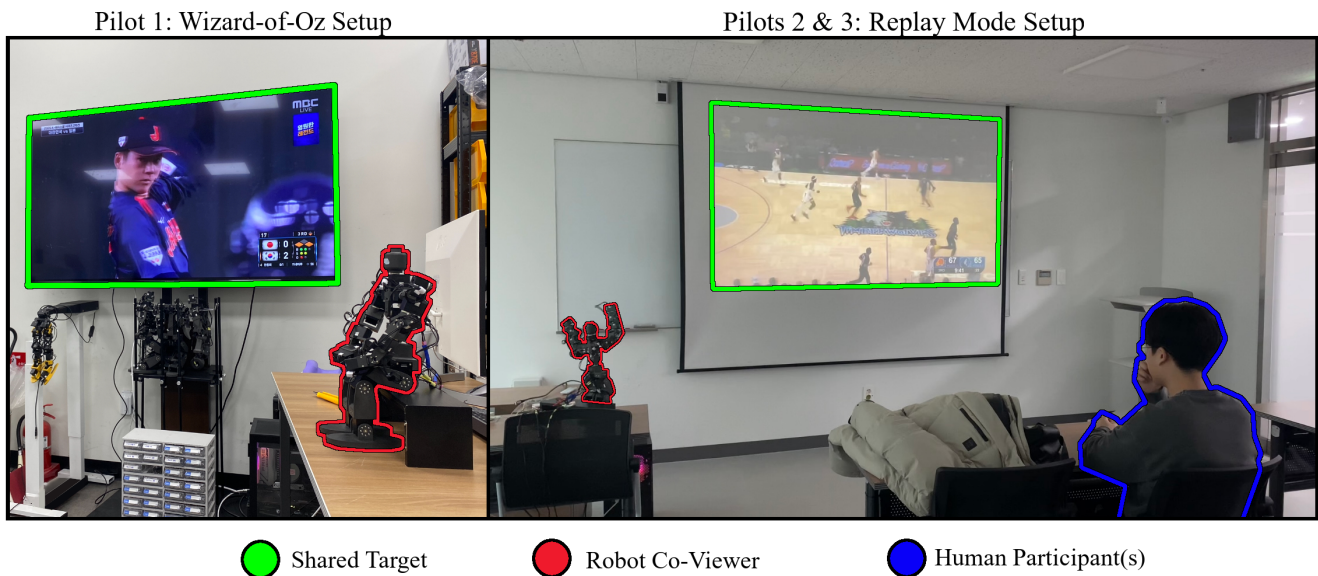


Fig. 4: Experimental setup for pilot studies 1 (left) and 2 & 3 (right).

game from November 16, 2025, as part of the Samurai Japan series. This matchup was selected because it carries strong regional and historical salience for viewers in Seoul, where the study was conducted. The intent was to assess whether the robot’s event-aligned embodied reactions could integrate into a highly affectively charged group viewing context, thereby serving as a test of the motion library under conditions likely to elicit pronounced spectator responses. The robot was positioned perpendicular to both the screen and the participants, as shown in Fig. 4 (left), allowing clear perceptual access to its reactions while maintaining the appearance of attending to the shared stimulus [11], [17]. In a multi-viewer setting, this orientation prioritized visibility of embodied responses across participants while framing the robot as a co-viewer rather than a performer. Open-ended questions were included in the post-condition questionnaire to capture qualitative impressions of the robot’s contribution to the viewing experience.

B. Pilot Study 2: Replay Mode (Individual)

Pilot 2 evaluated the replay configuration in an individual viewing context to examine how autonomous event-aligned reactions are interpreted in the absence of other human viewers. In contrast to Pilot 1, motion selection was no longer operator-controlled but followed the autonomous decision layer described in Section IV-D.2, while event timing remained fixed through replay synchronization to preserve consistent event–reaction mappings across the session. The participant viewed a recorded NBA game between the Phoenix Suns and Minnesota Timberwolves (December 8, 2025). Unlike the regionally salient Korea–Japan baseball game used in Pilot 1, this matchup was intentionally selected to reduce culturally specific emotional attachment and national identification effects, allowing the influence of the robot’s embodied reactions to be examined in a comparatively neutral viewing context. The robot was positioned facing the screen and slightly angled toward the participant (Fig. 4, right); this configuration emphasized the robot’s role as a co-viewer attending to the same broadcast rather than as an additional entertainment element directed toward the participant, while maintaining clear visibility of its reactions. In addition to the standard enjoyment and shared reality measures, the post-condition questionnaire included items probing perceived team alignment, recognition of bias, and perceived distraction, in order to assess whether autonomous reactions were perceptually salient, whether participants inferred a team preference from the robot’s behaviour, and whether such perceived alignment influenced interpretation of the game in a solitary viewing context.

C. Pilot Study 3: Replay Mode (Group)

Pilot 3 retained the replay configuration of pilot 2 but reintroduced a one-to-many viewing context to examine how the robot’s influence changes when embedded within a shared social environment. The same NBA game, synchronized timing file, and autonomous reaction selections were used as in pilot 2, ensuring that behavioural differences could not be attributed to changes in robot behaviour or autonomy level. The physical setup (Fig. 4 right), questionnaire content, and overall study design were identical to those of pilot 2, enabling comparison of measures between individual and group contexts under matching autonomous behaviour constraints.

VI. RESULTS

All pilots followed a two-condition within-subject design (baseline without robot, robot-present condition). For Pilots 1 and 3, baseline and robot-present scores were compared using paired-samples t-tests. Pilot 2 involved a single participant; therefore, only descriptive statistics are reported. Baseline attitudes toward robots were assessed using GAToRS prior to the viewing task (Table II). Enjoyment (ENJOY subscales) and shared reality (SR-T) outcomes are summarized in Tables III and IV. Pilot 1 (WoZ, group) included six participants (5 male, 1 female; $M = 26.0$ years, $SD = 2.45$). Pilot 2 (Replay, individual) involved one male participant aged 24 years. Pilot 3 (Replay, group) included five participants (3 male, 2 female; $M = 26.0$ years, $SD = 1.58$).

TABLE II: Baseline GAToRS scores by pilot (M (SD), 1–7).

| Pilot | P+ | P– | S+ | S– |
|----------------|-------------|-------------|------------|-------------|
| P1 ($n = 6$) | 4.67 (1.49) | 2.90 (1.66) | 5.93 (.73) | 4.83 (1.59) |
| P2 ($n = 1$) | 4.40 (1.74) | 5.00 (1.10) | 7.00 (.00) | 5.40 (1.74) |
| P3 ($n = 5$) | 4.96 (1.18) | 3.00 (1.33) | 5.92 (.89) | 4.72 (1.80) |

In pilot 1, higher pleasure and relatedness scores were observed in the robot-present condition, while engagement remained unchanged. Shared reality with other viewers showed a positive trend but did not reach statistical significance, as shown in Table III.

TABLE III: Pilot 1 (WoZ, group; $n = 6$). Baseline vs robot-present (M (SD)).

| Measure | Baseline | Robot | Effect Size | $t(5)$ |
|---------------|-------------|-------------|-------------|--------|
| Pleasure | 3.47 (1.30) | 4.23 (1.32) | -1.18 | -2.89* |
| Relatedness | 5.03 (.91) | 5.77 (.45) | -1.19 | -2.92* |
| Engagement | 4.03 (1.24) | 4.07 (1.38) | -.02 | -.05 |
| SR-T (Others) | 5.23 (.70) | 6.03 (.51) | -.98 | -2.39 |
| SR-T (Robot) | – | 5.40 (.63) | – | – |

* $p < .05$

For pilot 2 (Replay, individual), only descriptive statistics are reported due to the single-participant design. Pleasure decreased from baseline ($M = 5.20$) to the robot condition ($M = 4.80$). Relatedness increased slightly from 4.00 to 4.20, while engagement decreased from 3.00 to 2.60. Shared reality with the robot during the robot-present condition was 5.60. As no co-viewers were present, shared reality with other viewers was not meaningfully applicable in this study. Pilot 2 was retained not as a basis for statistical inference, but to verify that autonomous replay-based operation remained interpretable in the absence of human co-viewers before reintroducing a group context in Pilot 3.

TABLE IV: Pilot 3 (Replay, group; $n = 5$). Baseline vs robot-present (M (SD)).

| Measure | Baseline | Robot | Effect Size | $t(4)$ |
|---------------|------------|-------------|-------------|---------|
| Pleasure | 5.68 (.68) | 5.92 (.78) | -.46 | -1.04 |
| Relatedness | 5.84 (.52) | 5.96 (.65) | -.17 | -.37 |
| Engagement | 4.56 (.94) | 5.32 (1.02) | -2.91 | -6.52** |
| SR-T (Others) | 5.56 (.55) | 5.68 (.83) | -.11 | -.25 |
| SR-T (Robot) | – | 5.04 (1.37) | – | – |

** $p < .01$

In pilot 3, only engagement significantly increased in the robot-present condition, while pleasure, relatedness, and

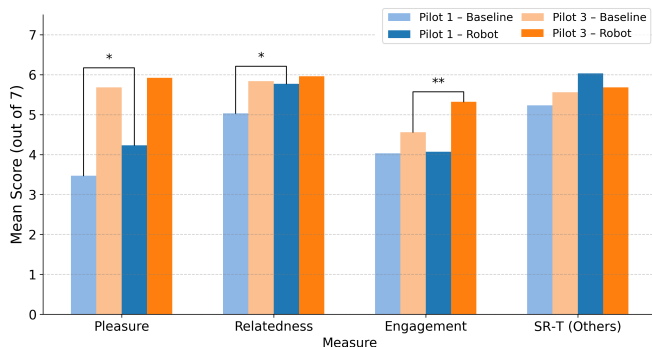


Fig. 5: Comparison of baseline and robot-present mean scores (out of 7) for pilot 1 and pilot 3. Thin brackets indicate statistically significant within-pilot differences: * $p < .05$, ** $p < .01$.

shared reality remained stable. Results from pilots 1 and 3 are visualized in Fig. 5. Interpretation of these results is limited by the small sample size of each pilot.

A. Team support and recognition of the robot’s team bias

In pilot 2, the lone participant reported supporting neither team and correctly identified the robot as supporting the Minnesota Timberwolves; they did not report that the robot influenced their team preference. In the group replay pilot, two said they supported neither team, two favoured the Minnesota Timberwolves, and one supported the Phoenix Suns. When asked which team the robot supported, four of five correctly answered Minnesota Timberwolves. When asked whether the robot changed which team they were supporting, only one participant answered “yes”, while four said “no”.

VII. DISCUSSION

This work initiates an initial exploration of an event-centred framing of robotic companionship in which a physically embodied humanoid robot participates peripherally in sports viewing by reacting to an external narrative rather than interacting directly with the user. Across three pilot studies, the results provide preliminary insights into RQ1 (enjoyment), RQ2 (shared reality), and RQ3 (interpretation of affective behaviour), while clarifying the contextual factors that shape when such contributions appear meaningful.

With respect to RQ1, the clearest positive effects on enjoyment were observed in the Wizard-of-Oz group setting (Pilot 1), where pleasure increased significantly when the robot was present. Engagement did not differ across conditions, suggesting that the robot did not intensify attentional focus on the game itself but instead shaped the affective quality of the viewing experience. This pattern is consistent with prior work on Robotic Experience Companionship, which shows that simple, non-interactive bodily responses synchronized to media content can enhance enjoyment without requiring dialogue or task engagement [11]. The present results extend this insight to a physically co-present humanoid and to sports spectatorship.

In the individual replay condition (Pilot 2), pleasure and engagement decreased slightly following the robot’s introduction. Although limited by the single-participant design, this pattern suggests that the robot’s event-aligned reactions did not add clear value in a solitary viewing context. Without other viewers present, the robot’s behaviour may have been

perceived as parallel to the viewing experience rather than as part of a shared activity, reducing its relevance to how the participant interpreted the game. In this setting, the robot’s reactions may have competed with the content for attention or highlighted their own artificiality, rather than supporting a sense of watching together. This contrasts with the group conditions, where the robot’s reactions were embedded within an already social environment and could be interpreted as contributing to a collective experience rather than standing alone.

Findings related to RQ2 suggest that the robot’s presence did not disrupt shared reality among human viewers. In the Wizard-of-Oz group pilot, a trend toward increased shared reality was observed. This could suggest that the robot’s reactions may have reinforced the collective interpretation of game events rather than replacing human-human alignment. In the group replay condition (Pilot 3), shared reality with other viewers remained stable, which may reflect that the robot did not disrupt existing social dynamics. This pattern is consistent with research on collective experiences and spectator emotions, which suggests that shared affect can be supported through visible, contingent reactions without requiring direct coordination or explicit interaction [8], [9].

Across all pilots, participants reported relatively high shared reality with the robot itself, directly addressing RQ3. Despite the robot’s constrained, non-verbal motion repertoire, viewers consistently interpreted its behaviour as affectively aligned with unfolding game events. This is consistent with foundational claims from social presence theory and the media equation, which argue that contingent bodily behaviour is sufficient for social attribution [4], [5]. At the same time, perceived alignment did not consistently translate into increased enjoyment or engagement, highlighting that recognizing co-experiencing and valuing it experientially are distinct outcomes.

Furthermore, participants consistently recognized which team the robot appeared to support, suggesting that its affective stance was both salient and interpretable despite the absence of verbal communication. However, most participants reported that this perceived bias did not alter their own team preferences. This suggests that while contingent embodied behaviour is sufficient to support social attribution and perceived alignment, it does not necessarily exert normative or persuasive influence within a peripheral framing. In this sense, the robot functioned less as an opinion-shaping agent and more as an affective amplifier of the unfolding narrative. This separation between interpretability and influence highlights a key design implication: peripheral robotic co-viewers may contribute socially by reinforcing shared attention and emotional framing without imposing social pressure or demanding user alignment.

Taken together, these findings suggest that event-centred robotic companionship can shape enjoyment and shared reality in group viewing contexts, while offering more limited or mixed benefits in solitary viewing.

VIII. LIMITATIONS AND FUTURE WORK

Several limitations constrain the conclusions that can be drawn from these pilot studies. First, sample sizes were small, particularly in the individual replay condition, limiting statistical power and generalizability. The results should therefore be interpreted as exploratory. Second, the fixed ordering of conditions, with baseline viewing preceding robot-present viewing, was chosen to preserve ecological validity

but introduces a temporal confound, as the progression of a game can naturally alter viewer affect and involvement over time. Because sports narratives often intensify toward the end of games, any increases observed in the robot-present condition cannot be cleanly separated from natural escalation in spectator affect.

Third, the pilots differed not only in autonomy and social configuration, but also in sport type and the cultural familiarity associated with the viewed content. Although all studies were conducted in Korea, the Wizard-of-Oz pilot featured a Korea-versus-Japan baseball game, a rivalry with strong regional and emotional salience for local viewers. In contrast, Pilots 2 and 3 used a replayed NBA game, which may carry different levels of familiarity, identification, and spectator norms for the same participant population. This may therefore have contributed to the stronger effects observed in the Wizard-of-Oz pilot, independently of control mode or group size. Future studies should systematically control for sport, league, and cultural context to disentangle these factors from the effects of robotic co-viewing. Furthermore, the semantic layer may struggle with multi-stage event logic, such as misinterpreting a foul as a purely negative occurrence when it actually results in a critical scoring opportunity (e.g., a free throw).

Future work should therefore focus on larger-scale studies directly comparing individual and group contexts, incorporating live viewing with robust synchronization, and examining longer-term exposure to assess habituation effects. Extending this event-centred companionship model beyond sports to other emotionally structured narratives would further clarify the scope and limits of peripheral robotic co-viewing.

IX. CONCLUSIONS

This work examined an alternative model of robotic companionship in which a physically embodied robot participates in sports viewing by reacting to unfolding events rather than interacting with the viewer. We implemented a compact whole-body motion library and a synchronized system supporting Wizard-of-Oz, replay, and live modes, and conducted three pilot studies across individual and group viewing contexts. Across pilots, participants interpreted the robot's behaviour as meaningfully aligned with game events, and in group settings, higher enjoyment and engagement scores were observed without disrupting shared reality among human viewers.

The main takeaway is that event-aligned embodied behaviour may function as a social cue. Participants sometimes interpreted the robot as contributing to the collective viewing atmosphere, though its impact is strongest when embedded within an already social viewing context rather than when substituting for human co-presence. These findings suggest that peripheral robotic co-viewers may be most effective as amplifiers of shared narratives rather than standalone companions. Future work should scale these results in larger controlled studies, disentangle the roles of embodiment and autonomy, and investigate long-term exposure and other narrative domains to clarify when and how such systems meaningfully contribute to everyday shared experiences.

X. ACKNOWLEDGEMENTS

The support of the Connected Minds Travel Award Program is gratefully acknowledged for making this project possible. Appreciation is also extended to Dr. Sungoon Choi

at Korea University for hosting and for providing valuable guidance and support throughout this work.

REFERENCES

- [1] "From loneliness to social connection: charting a path to healthier societies – Report of the WHO Commission on Social Connection," Jun. 2025.
- [2] V. D. Kannan and P. J. Veazie, "US trends in social isolation, social engagement, and companionship nationally and by age, sex, race/ethnicity, family income, and work hours, 2003-2020," *SSM - population health*, vol. 21, p. 101331, Mar. 2023.
- [3] J. Abdi, A. Al-Hindawi, T. Ng, and M. P. Vizcaychipi, "Scoping review on the use of socially assistive robot technology in elderly care," *BMJ Open*, vol. 8, no. 2, p. e018815, Feb. 2018.
- [4] F. Biocca, C. Harms, and J. K. Burgoon, "Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria," *Presence: Teleoperators and Virtual Environments*, vol. 12, no. 5, pp. 456–480, Oct. 2003.
- [5] B. Reeves and C. I. Nass, *The media equation: How people treat computers, television, and new media like real people and places*, ser. The media equation: How people treat computers, television, and new media like real people and places. New York, NY, US: Cambridge University Press, 1996, pages: xiv, 305.
- [6] F. Giglietto and D. Selva, "Second Screen and Participation: A Content Analysis on a Full Season Dataset of Tweets," *Journal of Communication*, vol. 64, no. 2, pp. 260–277, 2014.
- [7] J. Kim, K. Merrill Jr., and C. Collins, "Touchdown together: Social TV viewing and social presence in a physical co-viewing context," *The Social Science Journal*, vol. 61, no. 3, pp. 577–591, Jul. 2024.
- [8] G. Baranowski-Pinto, V. L. S. Profeta, M. Newson, H. Whitehouse, and D. Xygalatas, "Being in a crowd bonds people via physiological synchrony," *Scientific Reports*, vol. 12, no. 1, p. 613, Jan. 2022.
- [9] Y. Chang and Y. Inoue, "Spectator emotions in predicting psychological vigor: emotional meta experience and affect valuation perspectives," *Sport Management Review*, vol. 24, no. 4, pp. 594–619, Aug. 2021.
- [10] H. G. Wallbott, "Bodily expression of emotion," *European Journal of Social Psychology*, vol. 28, no. 6, pp. 879–896, 1998.
- [11] G. Hoffman, S. Bauman, and K. Vanunu, "Robotic experience companionship in music listening and video watching," *Personal and Ubiquitous Computing*, vol. 20, no. 1, pp. 51–63, Feb. 2016.
- [12] E. Broadbent, K. Loveys, G. Ilan, G. Chen, M. M. Chilukuri, S. G. Boardman, P. M. Doraiswamy, and D. Skuler, "ElliQ, an AI-Driven Social Robot to Alleviate Loneliness: Progress and Lessons Learned," *JAR life*, vol. 13, pp. 22–28, 2024.
- [13] O. E. Lee, H. Lee, A. Park, and N. G. Choi, "My Precious Friend: Human-Robot Interactions in Home Care for Socially Isolated Older Adults," *Clinical Gerontologist*, vol. 47, no. 1, pp. 161–170, Jan. 2024.
- [14] D. Pimentel and C. Vinkers, "Copresence With Virtual Humans in Mixed Reality: The Impact of Contextual Responsiveness on Social Perceptions," *Frontiers in Robotics and AI*, vol. 8, Apr. 2021.
- [15] J. B. F. van Erp and A. Toet, "Social Touch in Human-Computer Interaction," *Frontiers in Digital Humanities*, vol. 2, May 2015.
- [16] K. Kafetsios, D. Chatzakou, N. Tsigilis, and A. Vakali, "Experience of emotion in face to face and computer-mediated social interactions: An event sampling study," *Computers in Human Behavior*, vol. 76, pp. 287–293, Nov. 2017.
- [17] K. Kim, H. Song, J. Ryu, C. Oh, and B. Suh, "BleacherBot: AI Agent as a Sports Co-Viewing Partner," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, ser. CHI '25. New York, NY, USA: Association for Computing Machinery, Apr. 2025, pp. 1–31.
- [18] "ROBOTIS-GIT/DynamixelSDK," Feb. 2026, original-date: 2016-03-08T10:57:49Z. [Online]. Available: <https://github.com/ROBOTIS-GIT/DynamixelSDK>
- [19] OpenAI *et al.*, "GPT-4 Technical Report," Mar. 2024, arXiv:2303.08774 [cs].
- [20] "ESPN - Serving Sports Fans. Anytime. Anywhere."
- [21] M. Koverola, A. Kunnari, J. Sundvall, and M. Laakasuo, "General Attitudes Towards Robots Scale (GAToRS): A new instrument for social surveys," *International Journal of Social Robotics*, vol. 14, no. 7, pp. 1559–1581, 2022.
- [22] S. S. Davidson, J. R. Keebler, T. Zhang, B. Chaparro, J. Szalma, and C. M. Frederick, "The development and validation of a universal enjoyment measure: The enjoy scale," *Current Psychology*, vol. 42, no. 21, pp. 17733–17745, Jul. 2023.
- [23] M. Rossignac-Milon, B. Schmalbach, V. N. Keller, J. F. M. Cornwell, E. T. Higgins, and G. Echterhoff, "The role of target-specific shared reality in interpersonal interactions and protective health behaviours," *European Journal of Social Psychology*, vol. 54, no. 7, pp. 1431–1445, 2024.