Exploring Perception and Preference in Public Human-Agent Interaction: Virtual Human vs. Social Robot

Christian Felix Purps[1], Wladimir Hettmann[1], Thorsten Zylowski[1,2], Nathalia Sautchuk-Patrício[1], Daniel Hepperle[1,2], and Matthias Wölfel[1,2]

[1] Karlsruhe University of Applied Sciences, Karlsruhe, Germany
christian_felix.purps@h-ka.de
[2] University of Hohenheim, Stuttgart, Germany
matthias.woelfel@h-ka.de

Abstract. This paper delves into a comparison between virtual and physical agent embodiment. For our experiment we developed two agent embodiments, a virtual human and a mechanical social robot, that encourage passerby in public space to exercise squats through speech and non-verbal cues. We analyzed user behavior during the interaction with one of the distinct systems that differ in representation but share the same purpose and intent. The aim was to gain a better understanding how the two systems are perceived and interacted with in a public setting. We recorded 450 encounters in which a passerby listened fully to the agent's instructions. We used body tracking to analyze exercise engagement. At least one squat was performed in 145 encounters, which generally indicates fairly high system acceptance. Additional feedback came from 61 individuals (aged 13 to 74, 41 males, 20 females) through a questionnaire on perception of competence, autonomy, trust, and rapport. There was no significant difference found between the virtual human and the social robot concerning assessed factors. Responses to single questions indicate that interactions with the social robot were perceived as significantly more responsive, and gender differences in perceived interaction pressure emerged, with women reporting significantly higher values compared to men. Despite public space challenges, the agent systems prove reliable. Complexity-reducing technical and methodological simplifications and possible sampling biases are limitations. This work provides a glimpse into public interactions with virtually and physically embodied agents, and discusses opportunities and limitations for future development of such systems.

Keywords: human-agent interaction, virtual humans, social robotics, public space, physical exercise

1 Introduction

Recent advances in artificial intelligence (AI) have ushered a new era of human-computer interaction (HCI) in which virtual humans and social robots controlled

by intelligent agents could play an increasing role in further shaping our socio-technical landscape [1]. Voice-based agents (e.g. Alexa, Siri, Cortana etc.) have already become socially acceptable and ubiquitous [5]. In contrast, applications in which these software agents are embodied, e.g., by virtual characters, are still emerging or are, in the case of predominantly robotic embodiment, only a niche [10],[4]. However, these systems have the potential to significantly improve interaction quality by leveraging the multi-modality of communication channels, such as combining speech with non-verbal cues. This enables a more fluent and natural humanlike interaction, as embodied agents can convey meaning, emotions, and intentions more effectively through gestures, facial expressions, body language, and vocalizations and thus foster social acceptance [21].

Both, virtually embodied agents (e.g. virtual humans) and physically embodied agents (e.g. social robots) can take advantage of these features and serve similar application purposes. However, they differ in how users perceive these entities, even if they share a similar appearance, visual and behavioral fidelity, level of humanlikeness and other aspects (e.g. equivalent voice and speech communication). This may affect e.g. the generation of rapport and relation, trust, and perceived competence, which are all crucial facets of human communication [8]. For a better understanding of these distinctions, researchers have conducted studies and experiments in the field of HCI, comparing different embodiments for software agents. In addition to numerous applications in private usage, the potential utilization of these technologies in public spaces is also becoming increasingly conceivable [11]. However, there has been limited focus on conducting field tests involving comparisons between different types of embodied agents in public spaces. Thus, further study and research is needed to understand the implications and opportunities of using embodied agents in public settings, especially at a time when expectations for intelligent and accessible interactive systems are rising and, in many parts, remain unmet [3]. Conducting these types of experiments in public spaces can identify causal effects through randomization while studying people and groups in their natural settings, and brings a decidedly sociological perspective to the practice of experimentation by treating differences between people and places as a research opportunity rather than unwelcome threats to experimental control [2].

In our research, the primary objective is to compare two types of agent representations: one that is virtual and another that is physical. However, comparing these two types of embodiment poses challenges, especially when dealing with virtual humanlike characters and their physical counterparts (humanoid social robots). Constructing a high fidelity humanoid robot with the same appearance and movement quality as a virtual counterpart is difficult. Consequently, such robots often fall into the "uncanny valley", displaying unnatural traits, causing unease. Conversely, mechanical-looking social robots lack humanlike features (and thus avoid eeriness), but also make a direct comparison with realistic virtual humans problematic. However, while the proposed entities differ in appearance, their underlying behavior and intent may remain consistent. This justifies a comparison between them, despite the inevitable variation in the natural expression

of non-verbal cues depending on their respective embodiment. Rather than focusing on minor differences, our aim is to gain a general understanding of how people interact with these agents and how they perceive these very different entities in public space. We are particularly interested in exploring which type of embodiment, virtual or physical is preferred by the participants when interacting with these agents. Our investigation will deal with the participant's perception of the agent's competence, autonomy, and trustworthiness based on its embodiment. Furthermore, we will measure how well rapport is established and how long participants are willing to engage with these agents. By exploring these aspects, we aim to uncover insights into the attractiveness and effectiveness of both virtual and physical agent representations.

To conduct our study, we designed a simple interaction scenario to encourage participants to engage in physical activity with the agent in public space. The agent uses verbal and non-verbal cues and motivates participants to perform physical exercises (squats). Participants are greeted, instructed, and then asked to perform the exercises. The system tracks performance and provides verbal and non-verbal feedback. The agent cycles through different states depending on participants behavior and interactions aiming to provide a rewarding and motivating training experience.

2 Related Work

Social robotics has gained a valuable role in assisting, influencing and motivating human behavior in many HCI contexts [19]. Virtually embodied agents such as virtual humans may serve similar purpose and application while not requiring a physical representation. Similarities and differences in interacting with the distinct entities have been studied for different applications such as movie recommendation [16], socio-emotional interactions for children [6], and human decision-making in general [19].

Thellmann et al. compared virtual and physical agent embodiment (Nao robot and it's virtual representation) and it's effect on social interaction. Their investigations consider the relationship between physical and social presence. The results suggest that social presence of an artificial agent is important for interaction with people, and that the extent to which it is perceived as socially present might be unaffected by whether it is physically or virtually present [20].

Schneider and Kummert investigated the effect of an agent's embodiment type (humanoid social robot vs. virtual humans in three levels of humanlikeness) on motivation during the performance of versatile sport exercises. They figured out that participants tend to exercise significantly longer when interacting with a social robot than with a virtual embodied training partner. Additionally the participants found the robotic partner more likable than the virtual representation [18].

As likability, other factors, such as trust play a vital role in HCI research. One of the most commonly used definitions is that of Mayer et al. according to whom trust is *"The willingness of a party to be vulnerable to the actions of*

another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" [12]. Trust is established between two parties, and when designing reliable systems, attention is given to the qualities of both parties involved. On the computer side, considerations such as fairness, accuracy, and transparency come into play. As for the human aspect, traditional personal traits like introversion, extraversion, honesty, and affinity for technology become relevant. These human characteristics often remain outside the scope of technological systems. Van der Werff et al. present a model that links the motivation to trust with attributes derived from the Self-Determination Theory (SDT) [22]. By addressing these attributes through interface design, the human side of trust can be taken into account. SDT, originating from the work of Ryan and Deci, revolves around fulfilling psychological needs and encompasses competence, autonomy, and social relatedness as its key dimensions and is used, among other things, to explain motivation [17]. This motivation to trust can give an explanation why a person is willing to trust a system initially. As perceived competence, autonomy, and social relatedness increase, trust can be expected to increase as well.

3 Human-Agent Interaction System

We designed two distinct technical entities: A social robot, based on a one-arm robot with a smartphone attached, and a virtual human, displayed in life-size on a vertically arranged screen (see Fig. 4). To compare and evaluate both systems we created a trivial agent behavioral use case: a squat trainer application. Both systems are based on the same perceptive system (interlocutor localization and recognition of a squat physical exercise repetition) and rule based behavioral state machine and generate identical verbal output. Both forms of representation are capable of expressing a set of non-verbal social signals that are intended to be similar in meaning but differ in actual performance depending on whether the embodiment is virtual or physical.

3.1 Perception

To enable the agent interaction system to sense and interpret its environment, with a particular focus on observing and decoding the non-verbal signals of its interlocutors, optical sensors were used. In the context of a use case in public spaces, we were looking for a mobile and easily deployable perceptive system. Therefore, and since depth sensors have shown promising results in recognizing the shape of people, matching skeletons, and to recognize gestures, we decided to use Microsoft's Azure Kinect. The Kinect comes with a full body-tracking SDK[3] that allows extraction of joint information in real-time from the depth

[3] Body Tracking SDK for Azure Kinect enables segmentation of exposed instances and both observed and estimated 3D joints and landmarks for fully articulated, uniquely identified body tracking of skeletons. (http://www.azure.microsoft.com/en-us/services/kinect-dk)

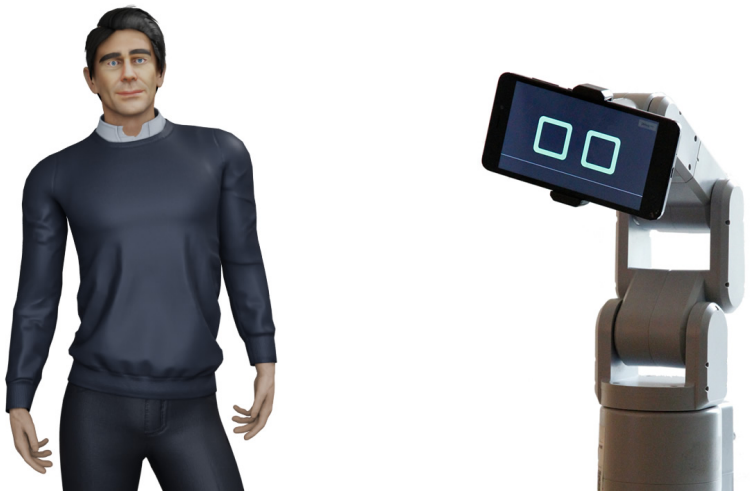

Fig. 1. Virtual Human and Social Robot. The image depicts the life size virtual human (left) while idling and the one-arm robot based social robot (right) during the "Seek Attention" animation.

image. To access data from an Azure Kinect, we use a self-developed middleware that seamlessly integrates the Kinetic Space[4] gesture-recognition module. Using Kinetic Space, we recorded squat exercise performances of different people to be used later for recognition. Kinetic Space can learn and recognize gestures from just few examples while the normalization of skeleton data ensures recognition accuracy by removing the influence of individual features, body orientation or localization. The middleware transmits social signals and spatial coordinates to control the agent's behavior for further processing.

3.2 Embodiment

For the embodiment of the two agents, we opted for fundamentally different representations. For the virtually embodied agent, we chose a virtual double of a real-life person, displayed life-size on a vertically arranged large screen. For the physically embodied agent, we created a very simple mechanical representation consisting of a robotic arm and a cell phone with abstract eyes and mouth.

Virtual Human The virtual embodied agent's representation in high fidelity (see Fig. 4) was created using Blender[5]. We chose largely neutral clothing for the agent instead of, for example, a sports outfit, so as not to create an additional

[4] Kinetic Space is an open-source tool that enables training, analysis, and recognition of individual gestures with a depth camera like Microsoft's Kinect family [23].
[5] https://www.blender.org/

bias in comparison with the social robot, where it is difficult to represent it in one way or another in an athletic manner. The avatar's body and rigging was extracted from a Mixamo[6] character and merged with a head, created based on a 3D scan and post-processing of a real person's head augmented with a fine-tuned facial rig for realistic facial expressions based on the proposed method of Purps et al. [15]. The avatar uses blend shape based lip synchronization for realistic mouth movements during speech. We used Mel Frequency Cepstral Coefficients (MFCC) to extract the phoneme profiles and assign the corresponding blend shape values. The virtual human is rendered in real-time using Unity[7] runtime application that connects with the agent behavioral system via localhost network connection. Vocal utterances and speech are played by the locally connected speaker.

Social Robot We created a very mechanical appearing social robot based on a robotic arm and a mobile phone. As the robotic arm base we used Elephant Robotics mechArm 270-Pi[8] and attached a mobile phone car holder to it. The movements of the robotic arm (rotation of it's six "joints") are controlled by a python application (that interfaces via C code to the serial port of the arm) and receives commands by the agent behavioral system (see Chapter 3.5) via TCP network connection. As "robotic face" serves a Xiaomi Mi 4 smartphone horizontally attached to the phone holder. The smartphone runs a 3D application created with Unity that displays two abstractly stylized eyes. These eyes can morph into different shapes depicting an abstraction of different facial expressions respectively emotions. If the robot performs vocal utterances or speaks to the interlocutor it plays a sine wave as a stylized mouth. The application receives commands via TCP Wi-Fi connection by the agent behavioral system, too. The mobile phone is connected via Bluetooth to a speaker to play vocal utterances and speech.

3.3 Non-verbal cues

While both representation, virtual human and social robot are congruent in their scripted behavior and speech, the performance of certain behavior significantly differs based on the possibilities provided by their embodiment. This does concern the appearance but mainly the possibilities of eliciting other non-verbal social signals to the conversation partner. The intention is for these distinct signals to be interpreted unequivocally by the agent's interlocutor, despite the marked differences between them (see Fig. 2). In states of idleness or focused attention on an interlocutor, the virtual human employs subtle idling behaviors,

[6] https://www.mixamo.com/
[7] https://www.unity.com/
[8] MechArm 270-Pi is a lightweight and compact 6-axis robotic arm manufactured by Elephant Robotics using Raspberry Pi as controller, with a payload of 250g, which is sufficient to lift an average mobile phone. (www.shop.elephantrobotics.com/en-de/collections/mecharm/products/mecharm)

while the social robot tilts its cell phone slightly back and forth. Additionally, the virtual human breaths and blinks frequently while the social robot simulates that through pulsing its abstract eyes. To encourage the potential participant to take action and join the interactive session for the exercise procedure, the virtual human greets with a friendly wave of the hand while the social robot performs a potentially similar associated gesture through arm movement and tilting/turning the mobile phone. To provide non-verbal feedback during the joint exercise execution, the virtual human simulates a squat, while the social robot moves its arm (and thus the attached mobile phone) up and down. Thus, the participant receives feedback if the squat was detected and executed correctly. In active conversation, the virtual human's lip movements simulate talking, while the social robot displays an audio sine wave on its cell phone. Smiling is depicted as a lip smile for the virtual human and a "happy" facial expression using its eyes and mouth for the social robot. In cheerful states, the virtual human shows a cheering/dancing animation, and the social robot engages in a side-to-side bouncing movement. In this way, the two representations lead to unique manifestations of identical behaviors.

3.4 Speech

To give the agents speech capabilities, we used the internal Windows speech synthesis API (SAPI[9]), which transcribes the assigned text content into an audio file using text-to-speech. In the embodied agent variant, we extended the speech synthesis with the use of the Unity Plugin uLipSync[10]. The playback of the audio file and the lip visemes for the pronounced vowels activate the required blend shapes in sync. Due to technical constraints for the social robot the text content to be used has been generated as an audio file with speech synthesis, so that the same voice, pronunciation, speed and pitch are used in both variants. The only difference is that in one case it is played back for each of the different states, and in the other it is transcribed and synchronized in real time. In addition, the pronunciation of the social robot is supported by the simulation of audio signal waves on the display. For each call and situation, the agent has a number of different phrases and utterances to choose from randomly.

3.5 Behavior

The core of our human-agent interaction system is the business logic for agent behavior. The business logic was implemented as Node.js application (social robot), respectively python application (virtual human) using the same finite state machine (FSM) architecture. The FSM is connected to the perception unit (see Chapter 3.1) via a local network and switches internal states based on detections. The proposed state machine consists of eight main states representing

[9] The Speech Application Programming Interface or SAPI is an API developed by Microsoft to allow the use of speech recognition and speech synthesis within Windows applications.

[10] https://github.com/hecomi/uLipSync

Fig. 2. Non-verbal cues of virtual human and social robot. It shows the social cues that describe the generic non-verbal behavior and how it is interpreted/performed by the two entities virtual human and social robot.

the general flow of interactions between agent and participant (see Fig. 3).

States and Transitions

- **Call to action:** The initial state where the system waits for a participant to be detected. As long as there is no participant found within the predefined interaction zone the agent performs a call to action animation (focusing passerby and waving for the virtual human and focusing a head tilting for the social robot). Once a participant is detected by the perceptive system, they are greeted, and the FSM transitions to the welcoming state.
- **Welcoming/General briefing:** In this state, the agent introduces itself and motivates the participant to start an exercise. After the short briefing, the FSM moves to the briefing for exercise state.
- **Briefing for exercise:** Here, the system demonstrates and explains the squat exercises. This means for the virtual human that it performs a demonstration squat while explaining the exercise verbally. The social robot uses its mobile phone to display a video of the virtual human who performs the squat while explaining the exercise verbally. Afterwards, the participant is asked to perform five squats as an exercise. A five-second countdown starts and the FSM transitions to the "Wait for performance" state.
- **Wait for performance:** In this state, the system waits for the participant to perform a squat while two timers, the Wait-Timer and Appreciation-Timer, are started. If the participant is inactive during the 5-second Wait-Timer, the FSM transitions to the intermediate motivation state. If the participant is inactive during the 15-second Appreciation-Timer, the FSM transitions to the appreciation state. Whenever a squat is detected, these timers are reset, the agent gives a visual animation feedback (squat performance) and utters the current squat count. After the participant has completed five squats the system transitions to the intermediate reward state.
- **Intermediate motivation:** When entering this state the agent holds a randomly (out of three) selected motivational speech.
- **Intermediate reward:** When entering this state the agent plays a rewarding animation (fixed) and holds a rewarding speech (randomly selected of three). After that the agent asks the participant to perform another set of five squats and then transitions to the wait for performance state.
- **Appreciation:** In this state the agent utters an appreciation text and thanks the participants for its performance.
- **Farewell:** The agent says goodbye to the participant.

4 Method

For our research we propose a method that aims to generally study passersby's perceptions, behaviors, and interactions with the virtual human and social robot in public space using the above stated human-agent interaction system. We aim

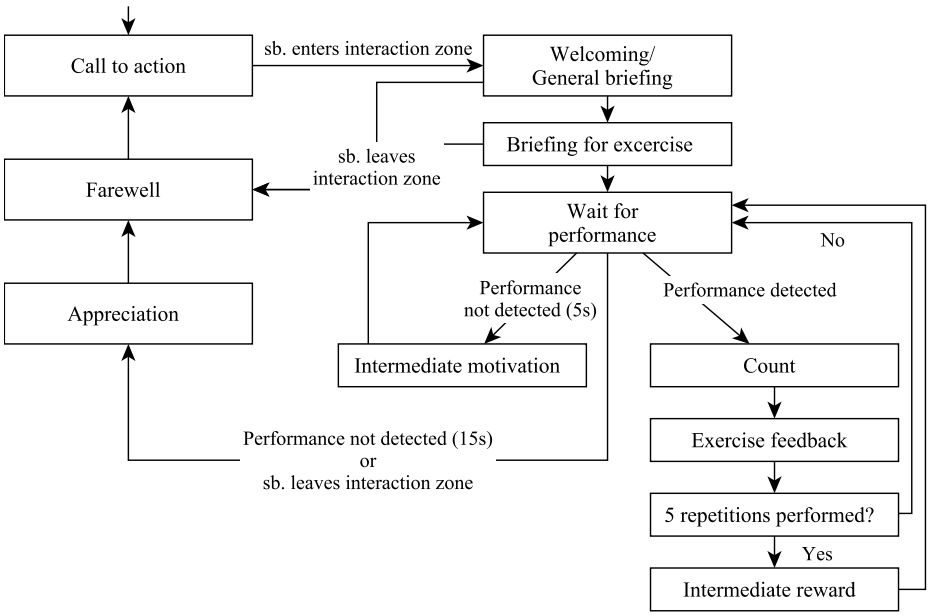

Fig. 3. System State Machine/Agent Behavioral Pattern. The agent initiates engagement by welcoming participants entering the interaction zone, guides squat exercises with feedback, offers motivation if squats aren't done, rewards every five repetitions, appreciates those who finish exercising and farewells those who leave.

to achieve this by combining quantitative performance metrics based on system tracking data with qualitative insights from questionnaires and behavioral observations.

4.1 Experiment Setup

The two installations have been placed side by side. The interaction zone was set at a distance of 1.3m from the embodied agent display and has a radius of 0.5m. The interaction zone for the social robot was centered around it with a radius of 2m. These dimensions were determined to reduce distraction or interference when using the installation simultaneously in a 5x5m room. Meanwhile, the social robot was placed on the far right, facing the entrance. Next to it, to the left of the social robot, approximately in the middle of the left side, facing the robot, was the embodied agent installation. To provide orientation and visual feedback, the two interaction zones were marked on the floor. A lateral boundary was also marked in the form of an open triangle leading from the agent to the sweet spot of the interaction zone. In addition, 0.5m diameter spaces were prepared to the right and centre of the entrance for participants to complete the questionnaires. These were also used for the final interviews at the end of the experiment.

4.2 Participants and Procedure

All participants in our study were passerby recruited ad hoc in public space at the "Effekte Festival" in Karlsruhe, Germany, a public event for universities and research institutions to present their scientific work. Passerby had the opportunity to engage with one of the showcased systems when approaching close enough to the social robot or virtual human. Upon their proximity, they were automatically registered as participants and the agent starts it's training procedure (see Fig. 3). Participants could interact with the systems for as long as they wished. The system reset when a participant left the interaction zone. In total, 450 times (214 social robot, 236 virtual human) passerby started an interaction with one of the distinct systems and remained in the tracking area long enough to listen fully to the exercise briefing. Noteworthy, our system did not identify individuals (e.g., through face identification), so the interaction count may not reflect the exact number of distinct individuals involved. From the 450 started interactions, 145 resulted in at least one squat execution that was detected by the system (78 social robot; 67 virtual human). After the interaction, participants were asked to fill out a post-experiment questionnaire. Overall, 66 participants completed the questionnaire. However, four participants did not fill out this correctly and for one participant the system logging was corrupted resulting in 61 cases for analysis (38 social robot, 23 virtual human). 41 participants identified themselves as male, 20 as female. The average age of the participants was 41 (13 youngest, 74 oldest).

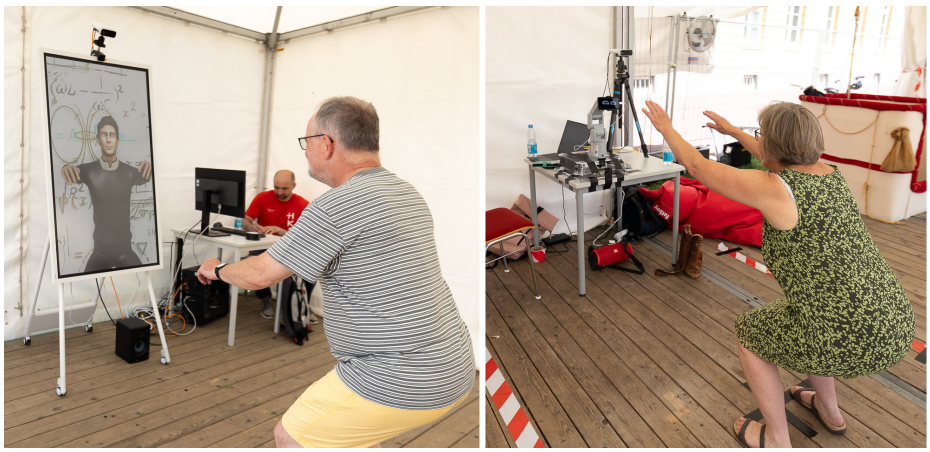

Fig. 4. Participants during Interaction. The participants perform a squat with the virtual human (left) and social robot (right).

4.3 Data Collection

For each participant, the system logged the interaction duration, the performed squat repetitions and the triggered number of rewards/motivations to evaluate performance numbers. Additionally, the participant's behavior was observed during participation for qualitative analysis. We created a questionnaire consisting of 17 questions (5 point Likert scale, 1=strongly disagree to 5=strongly agree) and assessed factors considering perception of the agent's **competence** (3 questions, "When using this agent I feel as a competent person", "If using this agent I feel effective", "I am convinced to be able to interact with this agent", $\alpha = 0.84$), **autonomy** (3 questions, "I can use my interactions freely when interaction with the agent", "I'm fully in control when interacting with the agent", "I can express my intentions when interacting with the agent", $\alpha = 0.74$), **rapport** (6 questions, "I perceive the agent as an independent person", " I feel attached to the agent", "I feel respected when using this agent", "It's fun to interact with the agent", "I will feel empathy with the agent", "I try to treat the agent as a human being", $\alpha = 0.79$) and **trust** (4 questions, "the agent is trustable", "I have a good feeling when relying on the agent", "I could trust information provided by the agent", "I trust this agent", $\alpha = 0.87$). Additionally, we asked "Do you feel under pressure while interaction with the agent?". The questionnaire is based on the Need Satisfaction Scale [7] and the Rapport–Expectation with a Robot Scale (RERS) [14].

5 Results

We calculated the arithmetic mean values for all data records and questionnaires. The respective results can be found in Tab. 1. A between-subjects one-way ANOVA was calculated to compare the perception of social robot and virtual human. There were no significant differences found between the virtual human and social robot for all factors assessed from the questionnaire nor system data (squat repetitions, number of rewards or motivations or interaction duration). However, the mean values for competence, autonomy, relationship, and trust are slightly higher (better) for the social robot. Moreover, we found significance with moderate effect size for single item ("I am convinced that I am able to interact with the assistant.", $F = 4.43$, $p = 0.04$, $\eta_p^2 = 0.07$). We also found a significant difference with large effect size between male and female gender in another item ("I feel pressured to behave in a certain way.", $F = 5.35$, $p = 0.007$, $\eta_p^2 = 0.15$). Eight participants performed squats in a way that was not recognized by our system. A between-subjects one-way ANOVA showed that rapport ($F = 3.36$, $p = 0.03$, $\eta_p^2 = 0.08$) and trust ($F = 5.67$, $p = 0.02$, $\eta_p^2 = 0.09$) were significantly lower rated in these cases.

6 Discussion

This research contribution investigated the public audience's preference for interaction partners, comparing virtual and physical embodiment. The study ex-

Table 1. Results of data evaluation (virtual human vs. social robot). The first group shows results (arithmetic mean, standard deviation, maximum values) based on our system logging. The second group shows the evaluated results of the questionnaire (arithmetic mean, standard deviation).

	Virtual human ($N = 23$)			Social robot ($N = 38$)		
	$\bar{x}$	σ	max.	$\bar{x}$	σ	max.
Interaction Duration (min.)	1:40	0:43	4:33	1:36	0:35	3:50
Squat Repetitions	5.48	5.34	20	6.76	5.70	30
Motivation tiggers	2.43	1.99	6	2.79	1.65	7
Reward triggers	1.04	1.07	4	1.13	1.26	6
Competence	3.16	1.19	-	3.54	0.91	-
Autonomy	2.17	0.79	-	2.54	1.07	-
Rapport	2.67	0.84	-	2.93	0.87	-
Trust	2.98	1.17	-	3.32	0.96	-

plored how perceived competence, autonomy, rapport and trust in the agent varied based on its embodiment, while also measuring the strength of generated rapport and engagement into the exercise task.

Although no significant distinctions emerged between the virtual human and the social robot across the assessed questionnaire factors or system data (squat repetitions, rewards, motivations, and interaction duration etc.), we measured slightly higher mean values in terms of competence, autonomy, rapport and trust for the social robot. This slightly better rating could possibly be attributed to the participants' increased confidence in their ability to effectively engage with the agent. This observation would be supported by significant differences in responses to the question "I trust myself to interact with the assistant," which has a moderate effect size. The fact that participants perceived responsiveness differently when interacting with the agent could be due to the robot's movements taking place in the dimensions of the physical world and thus being more discernible in nuances. This may play a role especially in public spaces, where uncontrolled conditions such as incident stray light, reflections, etc. make it harder to perceive small changes on an albeit large display. However, it is also possible that the strong degree of abstraction of the social robot is decisive here. At this point, further investigations (among others with comparisons a virtual embodiment of the same social robot) are interesting and necessary. We found a gender difference in the item “I feel pressured to behave in a certain way” with a large effect size indicating that women feel significantly more pressured during the interaction. This could be caused by the circumstance that women report higher subjective stress levels and arousal of emotional experience in HCI tasks [9]. This however, requires careful further investigation.

Our agent soft- and hardware and the perceptive system proved to be reliable and robust during the whole experiment duration even under the uncontrolled conditions in public space. However, some technological limitations and problems with the local setup were identified. The ambient noise in the public space occa-

sionally disrupted the exercise instructions, accentuated by the lack of acoustic isolation between the assistants, which ended up causing confusion among the participants when the interaction with the two assistants took place simultaneously. The lighting at the study site was not optimal for viewing the virtual human and changing over course of the day. Another technical problem caused the virtual human's cheering animation to not trigger regularly. The agents occasionally faced challenges in accurately detecting exercises, which were often influenced by the participants' squat execution, including factors like arm angles and squat depth. Additionally, instances of rapid squats occasionally led to detection hiccups, impacting the accuracy of counting. Furthermore, accurately identifying an individual's exercise among multiple people present within the detection area (camera field of view) occasionally posed a challenge for the perceptive system. Generally, instances where squats went unrecognized by the system (due to bad execution or sensor failure) or tracking was disturbed correlated with lower rapport and trust ratings, which is explainable by an increased frustration in these cases what could also be observed during the performance. Children preferred the social robot, displaying joyful behavior trying to figure out the boundaries and possibilities of the robot movements and body tracking. Unpleasantly, our system was neither trained nor calibrated to interact with persons below a certain body size. Thus, we cannot report reliable numbers considering actual interactions of children with the robot.

Overall, in approximately ⅓ (145 of 450) encounters in which passersby listened fully to the exercise briefing, the interaction was continued at least until the completion of the first squat. This shows a rather high acceptance rate for both systems, given the public space scenario [13]. It should be noted that the lack of face identification may cause the number of encounters to not reflect the exact number of distinct individuals involved. Participants interacted naturally with the agents, engaging in conversations, greetings, and waving. Initial movement detection problems frustrated some, causing them to leave early. Others, citing issues like heat or mobility, directly told the agents they wanted to stop. In interviews, participants consistently wanted agents to provide exercise feedback, especially for squatting techniques. Some were curious about the technology and expressed a desire for the agents to act as fitness instructors and homework assistants. It is important to note that the participants were not a random sample from the general population; rather, it is important to be aware that the individuals who participated most likely included many who had technical or scientific interests in one way or another due to the kind of event where they were recruited. There may also be bias in the sample because most participants were from Western cultures, particularly Germany. Interactions with agents were conducted in German, with observers translating for non-fluent participants. The questionnaire, solely in German, unintentionally acted as a participation criterion in this step. We also acknowledge biases in our study, particularly within qualitative observations. These biases, unintended and often subconscious, can inadvertently influence the observed phenomenon. Among the four observers,

three were men, one was a woman and all were Western-educated, which leads to interpretations influenced by this worldview.

7 Conclusion

In this paper we presented a simple interaction scenario in public spaces, where an agent embodied as a virtual human and a mechanic looking social robot encouraged participants to perform squats using verbal and non-verbal cues. The agent tracked their performance, and provided varied interactions to keep users engaged and motivated. In a study we examined how agent embodiment influenced perceived competence, autonomy, trust, rapport, and interaction engagement of participants. The study indicated a relatively high acceptance rate for both systems in a public space scenario. No significant differences emerged between the virtual human and the social robot across assessed questionnaire factors or system data. However, interactions with the social robot felt significantly more responsive to the participants. Notably, gender differences were observed in perceived pressure during interaction, with women reporting significantly higher levels. Our agent software, hardware, and perceptive system demonstrated reliability and robustness despite the challenges of an uncontrolled public environment. However, limitations related to technology, methodology, and sample bias were identified. Addressing the challenge of comparing vastly distinct embodiments (virtual and physical) demands innovative methodologies. Future research endeavors should encompass a repetition of the experiment under controlled conditions, effectively mitigating biases stemming from the turbulent and unregulated public space environment. If discrepancies arise, it would be worthwhile to examine the extent to which these can be attributed to either verbal or non-verbal feedback. Furthermore, there is a compelling need for studies that assess the participants' comprehension of the intended non-verbal cues emitted by the agent embodiments, especially if these are non-humanoid. Additionally, comparative investigations contrasting a virtual human with a virtual social robot hold promise for minimizing the influence of physical embodiment bias. Our path forward involves a commitment to ongoing research on the subject, with a focus on rectifying existing deficiencies. We also want to better understand how to generate easily interpretable artificial non-verbal signals for trivial physical embodiments such as the one presented, and ensure that our technology is suitable for future studies. An increasing use of embodied agents in public space is quite conceivable and part of the zeitgeist. Depending on the area of application and requirements, both virtual and physical agent embodiments (be they humanoid or mechanical) are potentially useful. Therefore, further efforts should be made to understand the mode of action of the different embodiments for agents and their respective communication capabilities, which have a crucial impact on natural interaction.

References

1. Aljaroodi, H.M., Adam, M.T., Chiong, R., Teubner, T., et al.: Avatars and embodied agents in experimental information systems research: A systematic review and conceptual framework. Australasian Journal of Information Systems **23** (2019)
2. Baldassarri, D., Abascal, M.: Field experiments across the social sciences. Annual review of sociology **43**, 41–73 (2017)
3. Cho, M., Lee, S.s., Lee, K.P.: Once a kind friend is now a thing: Understanding how conversational agents at home are forgotten. In: Proceedings of the 2019 on Designing Interactive Systems Conference. pp. 1557–1569 (2019)
4. Dereshev, D., Kirk, D., Matsumura, K., Maeda, T.: Long-term value of social robots through the eyes of expert users. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. pp. 1–12 (2019)
5. Hoy, M.B.: Alexa, siri, cortana, and more: an introduction to voice assistants. Medical reference services quarterly **37**(1), 81–88 (2018)
6. Jeong, S., Breazeal, C., Logan, D., Weinstock, P.: Huggable: the impact of embodiment on promoting socio-emotional interactions for young pediatric inpatients. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 1–13 (2018)
7. La Guardia, J.G., Ryan, R.M., Couchman, C.E., Deci, E.L.: Within-person variation in security of attachment: a self-determination theory perspective on attachment, need fulfillment, and well-being. Journal of personality and social psychology **79**(3), 367 (2000)
8. Li, J.: The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. International Journal of Human-Computer Studies **77**, 23–37 (2015)
9. Liapis, A., Katsanos, C., Sotiropoulos, D., Xenos, M., Karousos, N.: Stress recognition in human-computer interaction using physiological and self-reported data: a study of gender differences. In: Proceedings of the 19th Panhellenic Conference on Informatics. pp. 323–328 (2015)
10. Ling, E.C., Tussyadiah, I., Tuomi, A., Stienmetz, J., Ioannou, A.: Factors influencing users' adoption and use of conversational agents: A systematic review. Psychology & marketing **38**(7), 1031–1051 (2021)
11. Luria, M., Reig, S., Tan, X.Z., Steinfeld, A., Forlizzi, J., Zimmerman, J.: Re-embodiment and co-embodiment: Exploration of social presence for robots and conversational agents. In: Proceedings of the 2019 on Designing Interactive Systems Conference. pp. 633–644 (2019)
12. Mayer, R.C., Davis, J.H., Schoorman, F.D.: An integrative model of organizational trust. Academy of management review **20**(3), 709–734 (1995)
13. Narumi, T., Yabe, H., Yoshida, S., Tanikawa, T., Hirose, M.: Encouraging people to interact with interactive systems in public spaces by managing lines of participants. In: Human Interface and the Management of Information: Applications and Services: 18th International Conference, HCI International 2016 Toronto, Canada, July 17-22, 2016. Proceedings, Part II 18. pp. 290–299. Springer (2016)
14. Nomura, T., Kanda, T.: Rapport–expectation with a robot scale. International Journal of Social Robotics **8**, 21–30 (2016)
15. Purps, C.F., Janzer, S., Wölfel, M.: Reconstructing facial expressions of hmd users for avatars in vr. In: International Conference on ArtsIT, Interactivity and Game Creation. pp. 61–76. Springer (2021)

16. Rossi, S., Staffa, M., Tamburro, A.: Socially assistive robot for providing recommendations: Comparing a humanoid robot with a mobile application. International Journal of Social Robotics **10**, 265–278 (2018)
17. Ryan, R., Deci, E.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. The American psychologist **55**, 68–78 (2000)
18. Schneider, S., Kummert, F.: Comparing the effects of social robots and virtual agents on exercising motivation. In: Social Robotics: 10th International Conference, ICSR 2018, Qingdao, China, November 28-30, 2018, Proceedings 10. pp. 451–461. Springer (2018)
19. Shinozawa, K., Naya, F., Yamato, J., Kogure, K.: Differences in effect of robot and screen agent recommendations on human decision-making. International journal of human-computer studies **62**(2), 267–279 (2005)
20. Thellman, S., Silvervarg, A., Gulz, A., Ziemke, T.: Physical vs. virtual agent embodiment and effects on social interaction. In: Intelligent Virtual Agents: 16th International Conference, IVA 2016, Los Angeles, CA, USA, September 20–23, 2016, Proceedings 16. pp. 412–415. Springer (2016)
21. Turk, M.: Multimodal interaction: A review. Pattern recognition letters **36**, 189–195 (2014)
22. van der Werff, L., Legood, A., Buckley, F., Weibel, A., de Cremer, D.: Trust motivation: The self-regulatory processes underlying trust decisions. Organizational Psychology Review **9**(2-3), 99–123 (2019)
23. Wölfel, M.: Kinetic Space – 3D Gestenerkennung für Dich und Mich. Konturen **32** (2012)