2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) | 979-8-3315-1647-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISMAR62088.2024.00055

Perceived Empathy in Mixed Reality: Assessing the Impact of Empathic Agents' Awareness of User Physiological States

Zhuang Chang*
The University of Auckland, Auckland, New Zealand

Kangsoo Kim†
University of Calgary, Calgary, Canada

Kunal Gupta‡
The University of Auckland, Auckland, New Zealand

Jamila Abouelenin§
University of Pennsylvania, Philadelphia, PA, USA

Zirui Xiao¶
RingNet Pte. Ltd., Singapore

Boyang Gu‖
The University of Auckland, Auckland, New Zealand

Huidong Bai**
The University of Auckland, Auckland, New Zealand

Mark Billinghurst††
The University of Auckland, Auckland, New Zealand

Figure 1: The empathic agent conditions in the experiment: (1) **No Awareness Agent (NAA)**: The agent randomly praises users' shooting ignoring users' Skin Conductance Level (SCL) changes. (2) **Random Awareness Agent (RAA)**: The agent will randomly comment on users' SCL changes. (3) **Accurate Awareness Agent (AAA)**: The agent will provide accurate comments on the users' SCL changes whenever the arrow is displayed. The arrow in each picture lets users know the changes in their SCL. (4) The study setup - a participant wearing the Meta Quest 3 and Shimmer sensors.

Abstract

In human-agent interaction, establishing trust and a social bond with the agent is crucial to improving communication quality and performance in collaborative tasks. This paper investigates how a Mixed Reality Agent's (MiRA) ability to acknowledge a user's physiological state affects perceptions such as empathy, social connectedness, presence, and trust. In a within-subject study with 24 subjects, we varied the companion agent's awareness during a mixed-reality first-person shooting game. Three agents provided feedback based on the users' physiological states: (1) No Awareness Agent (NAA), which did not acknowledge the user's physiological state; (2) Random Awareness Agent (RAA), offering feedback with varying accuracy; and (3) Accurate Awareness Agent (AAA), which provided consistently accurate feedback. Subjects reported higher scores on perceived empathy, social connectedness, presence, and trust with AAA compared to RAA and NAA. Interestingly, despite exceeding NAA in perception scores, RAA was the least favored as a companion. The findings and implications for the design of MiRA interfaces are discussed, along with the limitations of the study and directions for future work.

Index Terms: Empathic computing, virtual agent, mixed reality, augmented reality, physiological state.

*e-mail: zcha621@aucklanduni.ac.nz
†e-mail: kangsoo.kim@ucalgary.ca
‡e-mail: kgup421@aucklanduni.ac.nz
§e-mail: jamilaaa@sas.upenn.edu
¶e-mail: zirui.xiao@ring.link
‖e-mail: bgu573@aucklanduni.ac.nz
**e-mail: huidong.bai@auckland.ac.nz
††e-mail: mark.billinghurst@auckland.ac.nz

1 Introduction

Recent advancements in Artificial Intelligence (AI) [69] and Mixed Reality (MR) [47] have catalyzed the development of new human-agent interactions. In these settings, Intelligent Virtual Agents (IVAs) [5] are embodied in MR environments as Mixed Reality Agents (MiRAs) [34]. Due to the seamless connection between the real and virtual worlds, these MiRAs engage dynamically with both virtual and physical environments [53, 52, 78, 25], marking a shift from traditional IVAs that operate on 2D screens. MiRAs, existing in the same 3D space with users, enhance perceptions of co-presence [61], social presence [40, 43], user experience [71] and reference [77] through their interaction with the physical world, which is often mediated by the Internet of Things (IoT) sensors [53].

A critical aspect of human-agent interaction is the formation of trust and social bonds, which are essential for effective collaboration in joint tasks [8]. Empathy in agents, or their ability to understand users' emotions, plays a significant role in personalizing interactions and enhancing user experience [31]. Although employing physiological sensors to gauge users' emotional and physical states is a promising approach to fostering empathy [24], the effectiveness of such strategies requires further exploration.

In this context, MiRAs must possess the empathic capacity to perceive and respond to users' affective states [56, 57]. This can be facilitated by physiological sensors, such as electromyograms (EMG) [15], electroencephalograms (EEG) [12], electrodermal activity (EDA)[73], electrocardiograms (ECG) [80], and photoplethysmograms (PPG) [38]. These sensors provide a deeper understanding of user states, potentially transforming MiRAs into Empathic Mixed Reality Agents (EMiRAs) [13, 11, 57]. However, how to effectively utilize physiological sensing to enhance perceived empathy in EMiRAs remains an open research question.

This paper investigates the impact of EMiRAs' empathic capabilities, informed by their awareness of users' physiological states, on users' social perceptions of these agents. We conducted a

979-8-3315-1647-5/24/$31.00 ©2024 IEEE
DOI 10.1109/ISMAR62088.2024.00055

within-subject study using the level of the agent's awareness as the independent variable. Participants played an MR shooting game, accompanied by an agent, under three conditions: (1) an agent *unaware* of the user's physiological states, (2) an agent that reported the user's physiological states *randomly*, sometimes inaccurately, and (3) an agent that monitored the user's physiological states *accurately* and reacted accordingly. Our findings suggest that the agent awareness of users' physiological states can enhance perceived empathy, and accurate monitoring further improves human-agent social interaction.

Our work contributes in three ways:

C1. Pioneering the integration of physiological sensors to foster empathy in MiRAs;

C2. Assessing how agents' recognition of users' physiological states influences perceived empathy;

C3. Offering new design implications for enhancing empathy in MiRAs.

2 Related Work

Our research extends earlier work on Empathic Agents, Mixed Reality Agents, and Physiological Biofeedback Systems. This section provides a summary of key related work in each of these areas.

2.1 Empathic Agents

Empathic agents are social agents that 1) show empathy towards users and/or 2) elicit user empathy towards them. For example, a virtual empathic therapist can make social conversation and advise users [66]. This empathic agent definition is based on the Perception-Action Mechanism of empathy [64], and a computational empathy model for empathic agents, comprising the empathic appraisal and response [68]. In human-agent social interaction, empathic agents can appraise the human user's emotional states and make appropriate empathic responses [21].

Empathic appraisal is a fundamental aspect of empathic agents and involves using sensors to recognize human emotions [30]. Physiological sensors, such as EEG, EDA, and PPG, can be used for automated emotion recognition [23]. For example, EEG, ECG, and facial expressions can be input to a conversational agent in cognitive behavior therapy to predict users' emotions [49]. Similarly, Gupta et al. [33] presented a real-time emotion prediction model based on EEG, EDA, and PPG signals and applied it to a context-aware empathic virtual agent. However, human emotion is complicated to model [39], and current technology cannot consistently and accurately evaluate human emotion [74]. The impact of an empathic agent's imperfect detection of users' physiological and emotional states on perceived empathy [79] remains unclear.

2.2 Mixed Reality Agents

MiRAs are agents embodied in MR that can make virtual-physical interactions with the MR environments [34]. For example, Dragone et al. [22] designed a MiRA by augmenting a virtual character into a physical robot with both virtual and physical body parts. The virtual part could gaze and point at a physical ball, while the physical part could move toward the ball, grab it, and take it back to the user. For a virtual agent embodied in MR [54], the virtual-physical interaction can be driven by IoT sensors and actuators [53]. For example, Kim et al. [41] presented a virtual human seamlessly integrated into the physical environment, which can perceive and respond to real-world fans. Similarly, Lee et al. [43] created a virtual human aware of the subtle movements in a shared real-virtual table.

However, these MiRAs primarily focus on the interaction between virtual agents and the physical environment. To expand the virtual agent's interaction capabilities from non-human physical environments to human users, physiological sensors [35] and tactile sensors [67] can be employed. For example, Prendinger and Ishizuka [63] introduced a virtual empathic companion to support users during virtual job interviews, where they used Galvanic Skin Conductance (GSR) and Heart Rate (HR) sensing to enable the agent to detect users' stress. While these studies demonstrate the potential of physiological sensing in enhancing virtual agents' empathic responses, they have primarily focused on 2D interfaces. The integration of such physiological sensing capabilities into virtual agents embodied in 3D immersive MR environments remains largely unexplored.

2.3 Physiological Biofeedback Systems

Biofeedback systems employ a range of sensors to monitor individuals' physiological and physical bodily functions, that they may not be typically conscious of [42]. Physiological biofeedback systems represent a major subset within the biofeedback systems, based on signals and parameters acquired from the neuromuscular, cardiovascular, respiratory, brain, skin, and other body systems [29]. Such biofeedback systems have been applied in medical diagnosis [6], rehabilitation [45], and even video games [51]. For example, Parnandi and Gutierrez-Osuna [59] presented an adaptive biofeedback game that aims to maintain the player's arousal by modifying game difficulty in response to the player's physiological state, as measured with EDA. Such adaptive games can engage players while not overwhelming them. Some games use physiological biofeedback through sensors like EDA, EMG, and PPG to offer opportunities for affective feedback [19].

Among those physiological sensors mentioned above, EDA or Skin Conductance Level (SCL) has been widely used to capture physiological arousal caused by stress and anxiety [62, 73, 72]. For example, Chiossi et al. [14] created a visual complexity adaptive system based on users' online changes in SCL compared with a dynamic baseline. In this work, we used the algorithm presented in [14] to capture users' online SCL states while playing an intensive MR shooting game, with a higher level of SCL indicating greater emotional arousal.

Compared to prior work on empathic agents, MiRAs, and physiological biofeedback systems, our work is novel in a number of ways. Inspired by the benefits of virtual humans' awareness of physical events in improving social presence [43, 61], we were interested in exploring users' social perception (see Sec. 4.6) of a MiRA that can sense and show awareness of users' physiological states. Unlike other systems that adjusted game difficulty level based on user stress, we wanted to employ a virtual human interface to deliver feedback on user physiological changes. So we created an empathic agent that could show awareness of users' physiological states at different accuracy levels and investigated how such awareness influences users' perceived empathy. This system is described in detail in the next section.

3 System Overview

Our hardware system comprises physiological sensors, one desktop computer, and one standalone MR headset (see Fig. 2). The software system consists of the physiological state monitoring module, the DroneRage game module, and an MR agent module. The physiological state monitoring module written in Python runs on the PC, while the other two modules created with Unity 3D run on an MR headset, i.e., Meta Quest 3. As shown in Fig. 1, users wearing physiological sensors and an MR headset interact with the virtual content powered by our developed software. Their physiological states will be monitored and displayed on the headset. Upon application launch, a 3D User Interface (UI) floated in front of users, showcasing crucial information through editable text boxes revealing the IP address of the physiological monitoring module and the participant's ID alongside toggles for experimental conditions. It also includes a start experiment button guiding users to the DroneRage game and MR agent module, where an MR human-like agent stands 30 degrees to the right and 2 meters away from the users, facing them directly (see Fig. 1 (1–3)). The agent acts as an empathic

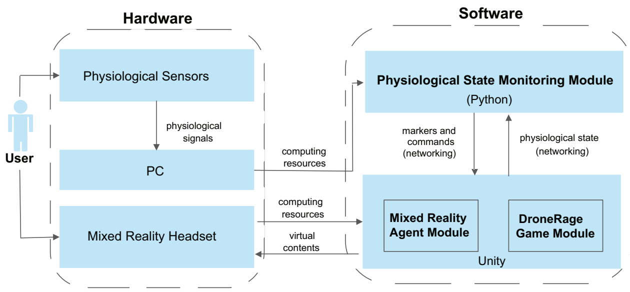


Figure 2: System overview in hardware and software.

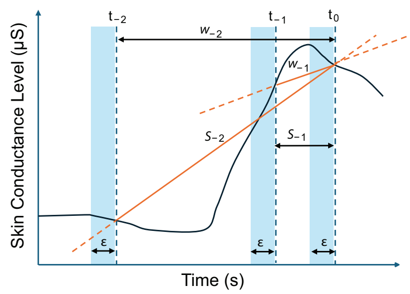


Figure 3: An overview of how SCL changes were calculated online using the algorithm proposed by Chiossi et al. [14].

game companion throughout the game, closely watching players' gameplay and providing feedback on their shooting behavior or physiological state. We will explain the three software modules in detail and their implementation in the following subsections.

3.1 Physiological State Monitoring

The physiological state monitoring module was designed to detect and calculate the player's Skin Conductance Level (SCL) changes online while the player was playing the DroneRage game (see Fig. 2). The calculated SCL state was sent to the DroneRage game module through the network (see Sec. 3.3 for further explanation). In addition to the online calculation of the player's physiological states, the physiological monitoring module can log raw physiological data to disk alongside markers received from the Unity side through the network, facilitating offline data analysis (see Sec. 4.6).

Inspired by the SCL-based adaptation strategy of [14], we adopted the method they used for online detecting SCL changes. We just visualized the detected SCL changes, which served as a ground truth to remind the player about the changes in their physiological state while playing the testing game.

$$\begin{cases} f(t_x) = (\sum_{i=t_x-\varepsilon}^{t_x} SCL(t_i))/\varepsilon \\ w_i = t_0 - t_i \\ s_i = (f(t_0) - f(t_i))/w_i \\ where\ i = -2,\ -1 \end{cases} \tag{1}$$

As shown in Fig. 3, their algorithm compares s_{-1} and s_{-2}, which are slopes of two lines identified by SCL points at t_0, t_{-2} and t_{-1}. The calculation of s_{-1} and s_{-2} are shown in Eq. (1), where $SCL(t_i)$ represents the SCL value at time point t_i and $f(t_x)$ denotes the mean value of SCL from time $t_i - \varepsilon$ to t_i. Based on Fig. 3 and Eq. (1), all SCL states are shown in Eq. (2), where θ is a threshold to activate the SCL changes.

$$SCLstate(s_{-1}, s_{-2}) = \begin{cases} increase & if\ s_{-2} < s_{-1} - \theta \\ decrease & if\ s_{-2} > s_{-1} + \theta \\ nochange & otherwise \end{cases} \tag{2}$$

Table 1: The DroneRage game parameters.

Wave Number	Total Number of Drones	Max Live Drones	Drone Spawn Time
1	1	1	4
2	5	2	2
3	9	3	1.5
4	13	4	1.4
5	17	5	1.3
6	21	6	1.25
7	25	7	1.21

Lastly, the SCL state was calculated in a preset time step (t_s) to avoid frequent updates in the MR Agent Module. We used a double-ended queue [55] to cache the ($w_{-2} + \varepsilon$) seconds data for calculating the SCL states. Once an SCL state is calculated, the physiological monitoring module will delete the earliest t_s seconds data in the queue and wait for another t_s seconds for new data to be enqueued.

3.2 The Mixed Reality Agent

In the MR agent module, we used a 3D virtual male character generated from the ReadyPlayerMe [3] . We downloaded an idle animation from Mixamo [2] and applied it to the virtual character to make its body move slightly. We used the Fimpossible Creatsons' eyes and look animator [17, 18] packages to create natural eye movement and looking behaviors. During the DroneRage game, the virtual agent would randomly look at the players or one of the enemies. The agent's verbal cues were enabled by audio clips pregenerated using the IBM Watson Text-to-Speech service [16] Each speech has at least three synonymous phrases of similar length and meaning to avoid stiffness.

The virtual agent's speech intention can be classified into greeting, self-introduction, praising users' shooting, and commenting on users' physiological states. Before the game starts, the virtual agent briefly greets users and makes a self-introduction. During the game session, by default, the virtual human randomly praises users for shooting when enemies take damage. Whenever the users' physiological states change, the virtual agent would either ignore or comment on the users' physiological state based on experimental conditions (see Fig. 1(1-3)). To avoid any direct intrusion into users' emotions, we crafted the speech structure of the virtual agent to first report objectively detected physiological changes and then infer feelings of stress. For example, one of the comments on users' physiological states was, "I've noticed your skin conductance level decreasing, indicating some decreased stress."

3.3 The DroneRage Game

The DroneRage game module is a shooting game in the MR environment, which was developed based on the open-source Discover project [65] As shown in Tab. 1, the game has seven rounds of enemy drones, with increasing totals and maximum live enemies. The game will upgrade to the next round after 5 seconds when all the enemies are eliminated. The game advances to the next round 5 seconds after all enemies are eliminated. Players use an MR headset, with the controller in their dominant hand as a virtual gun, shooting by pulling the trigger. A health bar on the gun keeps players aware of their health. We integrated the physiological state visualization feature into the DroneRage game. Whenever the DroneRage game module receives the player's physiological state from the physiological monitoring module (see Sec. 3.1), it activates a visualization event, displaying an arrow in the player's view for 5 seconds to signify an increase or decrease in the physiological state. The arrow, measuring 30 cm in height and 30 cm in width, was placed 1 meter away and 0.5 meters below the user's eye level (see Fig. 1). No arrow will be displayed if there is no change in the state. The physiological visualization event also notified the MR agent module to trigger different agent behaviors based on the experimental conditions shown in Sec. 4.1.

3.4 Implementation

The DroneRage game and MR agent modules were developed in Unity 2022.3.1f1 on a desktop computer powered by an Intel(R) Core i7-12700F CPU and NVIDIA GeForce RTX 3070 GPU, which was the same computer we used for running the physiological monitoring module. We used the Meta Quest 3 MR headset and the Shimmer3 GSR+ sensor, which includes both the EDA and PPG sensors. We used the ECL-Shimmer-C-API [70] to fetch real-time physiological signals and send the raw data to our monitoring module via Lab Streaming Layer (LSL) library [1] The detected physiological states were updated to the Unity side through a network message queue ZeroMQ [4].

4 User Study

This section describes our conducted user study aimed at addressing the following research question: ***How does an EMiRA's awareness of the users' physiological states impact users' social perception of such an agent?***

4.1 Experimental Design

We designed a within-subject user study to explore the impact of an agent's awareness of user SCL on user social perception of interaction with EMiRAs. The study involved three types of virtual agent: 1) No Awareness Agent (**NAA**), 2) Random Awareness Agent (**RAA**), and 3) Accurate Awareness Agent (**AAA**).

In the NAA condition, the agent would initially greet the player and introduce itself as a game companion capable of monitoring the player's physiological state. It would then invite the player to play the game. During the game session, the agent would praise the participant's shooting based on the logic outlined in Sec. 3.2, but would not respond to the user's SCL changes. In the RAA condition, when the user's SCL increased or decreased, the agent showed random awareness of such changes, with speech behavior randomly selected from options of increase, decrease, and no change, each with equal probability. For instance, when the user's SCL state was displayed as an upward arrow, the RAA might have a probability of $1/3$ of accurately verbalizing the user's SCL state, $1/3$ probability of wrongly reporting it, and $1/3$ probability of ignoring it. In the AAA condition, the agent was programmed to consistently provide accurate reports of the user's SCL states, as indicated by the arrow.

As detailed in Sec. 3.2, the virtual agent's dialogue regarding the player's SCL states involved objectively describing the SCL level and inferring the player's stress level. The NAA condition was established as a baseline, while the RAA condition simulated the limitations of current technology in detecting physiological and emotional states. Conversely, the AAA condition simulated the potential technological advancements for improved accuracy.

Based on previous research on the benefits of empathic agent systems [76, 63], we established our hypotheses as below:

- **H1** Showing awareness of the user's physiological states could enhance users' social perception of an EMiRA (as measured by perceived empathy, trust, etc.).
- **H2** Accurate awareness of the user's physiological states in the EMiRA would further improve users' social perception of such an agent.

For **H1**, we expected both the RAA and AAA conditions would receive higher ratings on the social perception measures presented in Sec. 4.6. For **H2**, we expected the AAA condition would exceed the RAA condition in users' social perception.

4.2 Participants

We recruited 24 participants (11 male, 12 female, and 1 non-binary) aged 19 to 39 ($M = 25.92$, $SD = 4.49$), from the university campus. We requested participants to share their experience with MR devices, virtual agents like Amazon Alexa, Microsoft Cortana, or Apple Siri, and smart wearable devices with physiological detection, like the Apple Watch. Their experiences with these technologies were assessed based on their frequency of use, categorized into three groups: never, occasionally (e.g., semesterly or annually), and frequently (e.g., daily, weekly, or monthly).

Regarding their experience with MR devices, 14 participants reported using them a few times, 7 had frequent usage, and only 3 reported never using such devices. In terms of using virtual agents, 16 participants reported occasional usage, 2 reported frequent usage, and 6 never used them. 11 participants reported occasional usage of smart devices capable of detecting physiological states, 2 frequent usage, and 6 never used them.

4.3 Experimental Setup

Inspired by [14], we ran rounds of preliminary tests to determine the values of parameters in Eq. (1) and Eq. (2) for calculating the SCL states in our experimental task to avoid frequent updates in SCL changes and virtual human speech. We set the w_{-2} at 30 seconds and the w_{-1} at 2 seconds to calculate the s_{-2} and s_{-1}. The ε was 1 second to calculate the SCL at a given time point. The sampling rate was set at 256 Hz. We cached 31-second ($w_{-2}+\varepsilon$) data for calculating the SCL states. Once the first SCL state was calculated, new SCL states were calculated and updated every 12 seconds. The θ in Eq. (2) was determined for each trial per participant by averaging the absolute differences between s_{-1} and s_{-2}, calculated during the one-minute baseline correction phase described in Sec. 4.5.

4.4 Experimental Task

Participants played DroneRage with the Quest 3, using their dominant hand to hold a controller, displayed as a gun in MR. The MR agent greeted and introduced itself, then invited participants to play. They pressed button "A" (right controller) or "X" (left controller) to start the baseline correction, during which their SCL slope was calculated to determine the θ value. To reduce EDA signal noise, participants were asked to stay still and relax. After baseline correction, the first wave of drone enemies spawned and flew in through the ceiling (Fig. 1). Participants shot these enemies by pulling the trigger and observed the virtual agent's behavior to see if it accurately perceived their physiological state. The game ended automatically 3 seconds after all waves were completed or if the player's in-game blood level reached zero.

4.5 Experimental Procedure

When participants entered the room, they were welcomed, seated at a computer, and given a slide explaining the experiment. After agreeing to participate, they signed a consent form and completed a demographic questionnaire. They were then seated in front of a wall and fitted with the Quest 3 headset. They entered their participant ID and experimental conditions on the UI (Sec. 3), while the experimenter started streaming EDA and PPG signals. Participants began the experiment by clicking the start game button, engaging in a task described in Sec. 4.4, with the MR agent's behavior manipulated based on the experimental condition.

Before the first condition, they completed a training session without the virtual agent but with the same task features, followed by a one-minute rest. Participants then experienced the experimental task in a Balanced Latin Square order, with the experimenter removing the headset after each trial for them to complete questionnaires (Sec. 4.6). Each participant completed three trials in about 90 minutes. Afterward, a semi-structured interview was conducted to understand their experience and perceptions of the agent.

4.6 Measurements

4.6.1 Subjective Measures

Perceived Empathy: Measuring perceived empathy in virtual agents lacked consensus on evaluation methods. Özge Nilay Yalçın [79] proposed to evaluate perceived empathy using system-level and feature-level evaluation methods. The system-level evaluations focus on the overall perception of empathy in social agents, which can be affected by the social agents' appearance, human likeness, and perceived intelligence [48]. These agent factors can be evaluated using the Godspeed questionnaire [7], originally developed to measure robots' anthropomorphism, animacy, likability, perceived intelligence, and perceived safety. In contrast to real robots that possess physical body parts, our EMiRAs are designed with only virtual human appearances, thereby minimizing safety risks. To assess the overall perceived empathy in our EMiRAs, we utilized the Godspeed questionnaire, omitting items that pertain to perceived safety.

Social Connectedness: We measured the social connectedness between users and the EMiRA using the "Inclusion of the Other in the Self" (IOS) questionnaire [28]. Social connectedness has been proven to be closely related to empathy [37].

Social Presence: We were interested in whether the virtual agent's awareness of human users' physiological states can also increase social presence. We measured the social presence with the questionnaire used by Leite et al. [44], which comprises six subscales: *co-presence*, *attentional allocation*, *perceived message understanding*, *perceived affective understanding*, *perceived affective interdependence*, and *perceived behavioral interdependence*.

Trust: Empathic capabilities in interactive agents are shown to lead to more trust [10]. In our study, the AAA condition accurately showed awareness of user physiological states, while the RAA was sometimes inaccurate. Therefore, we also measured users' trust in our EMiRAs as measured by the Trust in Automated Systems questionnaire [36].

Interview Feedback: For the qualitative measures, we used a semi-structured interview after the experiment, asking the participants the questions shown in Tab. 2. **Q1** and **Q2** were designed to query user preference, which can also reflect some insights from users to help understand the questionnaire results. The **Q3** was designed to understand users' concerns about privacy and trust in technology. This was motivated by the need to detect user physiological states in our study. **Q4** and **Q5** were designed to understand the user's expectations for creating MiRAs' empathic perception and response capabilities. The question **Q4** was motivated by the design of AAA's speech structure in our study shown in Sec. 3.2, where the AAA agent first reports users' objective physiological states detected by algorithms and sensor data, followed by an inference on their stress level based on the physiological states. Given that our experiment primarily centered around the agent's perception of users' physiological states, and the three agent conditions differed solely in speech behaviors, **Q5** was tailored to investigate further empathic responses within these agents.

4.6.2 Physiological Measures

We logged users' EDA and PPG signals with markers indicating baseline start/stop and game end. Offline analysis with Neurokit toolbox [46] checked if EMiRAs' awareness influenced these signals during the game.

After decomposing EDA into tonic and phasic components [9], we segmented the data into baseline and game intervals. We calculated the averaged SCL by subtracting baseline tonic SCL from game data and averaged phasic Nonspecific Electrodermal Responses (NS.EDRs) [27].

From PPG data, we derived Heart Rate Variability (HRV) features, focusing on RMSSD and SDNN [60, 33], commonly used in HRV analysis.

5 Results

5.1 Subjective Measures

5.1.1 Perceived Empathy

For the GodSpeed questionnaire, we only found significant differences in the *animacy* (see Fig. 4a) and *perceived intelligence* (see Fig. 4b) subscales. For *animacy* subscale, a Friedman test ($\chi^2(2) = 9.810$, $p = 0.007$) revealed an overall significant difference between the three agent conditions. However, the following Kendall's W test found only a small effect size ($W = 0.204$). The pairwise Wilcoxon signed-rank tests with Bonferroni correction showed the animacy rating in the AAA condition was significantly higher than NAA ($Z = -2.308$, $p = 0.021$). Similarly, the overall rating for *perceived intelligence* was significantly different between the three agent conditions($\chi^2(2) = 12.7$, $p = 0.002$), with a small effect size ($W = 0.265$) by the following Kendall's W test. Post-hoc tests further revealed the perceived intelligence in AAA condition was significantly higher than NAA ($Z = -3.428$, $p < 0.001$) and RAA ($Z = -2.697$, $p = 0.007$).

5.1.2 Social Connectedness

As shown in Fig. 4c, a Friedman test ($\chi^2(2) = 14.7$, $p < 0.001$) revealed significance between the three agent conditions on the IOS questionnaire responses. A moderate effect size ($W = 0.305$) was found by the following Kendall's W test. The Post-hoc pairwise Wilcoxon signed-rank tests with Bonferroni correction further revealed the social connectedness in the AAA condition was significantly higher than NAA ($Z = -3.090$, $p = 0.002$) and RAA ($Z = -1.969$, $p = 0.049$). The social connectedness in RAA was also significantly higher than NAA ($Z = -1.977$, $p = 0.048$).

5.1.3 Trust

Fig. 4d illustrates the statistical tests on the trust in agents as measured by the Trust in Automated Systems Questionnaire. A Friedman test showed an overall significant difference between the three agent conditions on the trust responses ($\chi^2(2) = 17.2$, $p < 0.001$), with a moderate effect size ($W = 0.359$) by the following Kendall's W test. Post-hoc tests further showed the overall trust in the AAA condition was significantly higher than NAA ($Z = -2.097$, $p = 0.036$) and RAA ($Z = -2.968$, $p = 0.003$).

5.1.4 Social Presence

Fig. 5 shows plots for the Social Presence questionnaire subscales with different significant main effects. For *co-presence* subscale, we found a significant main effect through a Friedman test ($\chi^2(2) = 17.6$, $p < 0.001$), and a moderate effect size ($W = 0.366$) through a following Kendall's W test. Post-hoc pairwise Wilcoxon signed-rank tests with Bonferroni correction showed the co-presence in the AAA condition was significantly higher than NAA($Z = -3.500$, $p < 0.001$) and RAA ($Z = -2.483$, $p = 0.013$).

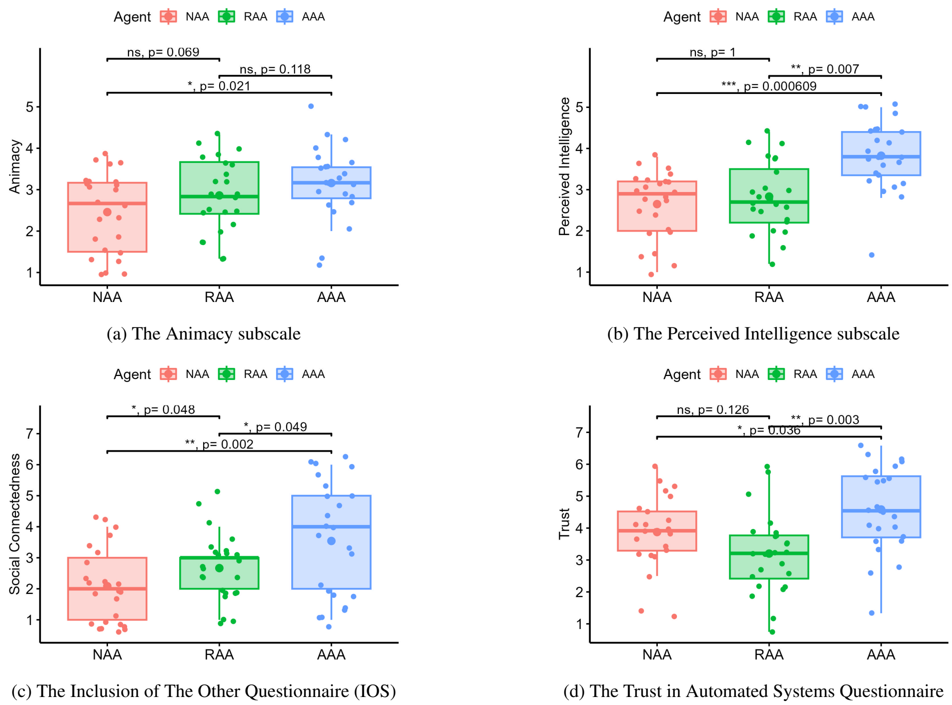
The *attentional allocation* subscale also showed a significant main effect ($\chi^2(2) = 17.6$, $p < 0.001$), with a moderate effect size ($W = 0.366$). Post-hoc tests further revealed the attentional allocation response in the AAA condition was significantly higher than NAA ($Z = -3.330$, $p < 0.001$).

Similarly, the *perceived message understanding* subscale also showed an overall significance between the three conditions ($\chi^2(2) = 16.8$, $p < 0.001$) with a moderate effect size ($W = 0.349$). Post-hoc tests showed the AAA condition was significantly higher than NAA ($Z = -2.968$, $p = 0.003$) and RAA ($Z = -2.878$, $p = 0.004$).

For the *perceived affective understanding*, a Friedman test ($\chi^2(2) = 18.3$, $p < 0.001$) showed an overall significant main effect with a moderate effect size ($W = 0.381$). Post-hoc tests revealed that the

Table 2: Interview questions and their motivations.

No.	Question	Motivation
Q1	Which agent do you like most, and why?	User preference
Q2	Which agent do you hate most, and why?	User preference
Q3	Are you glad to share your physiological data with a mixed reality agent to let it understand you better?	Privacy and trust in technology
Q4	Do you think the virtual human should accurately report your physiological state or further infer your emotional state?	Empathic perceptions
Q5	What behaviors do you expect from the virtual human in MR if it can understand your emotions?	Empathic responses



(a) The Animacy subscale

(b) The Perceived Intelligence subscale

(c) The Inclusion of The Other Questionnaire (IOS)

(d) The Trust in Automated Systems Questionnaire

Figure 4: Boxplots showing the subscales of Animacy and Perceived Intelligence from the GodSpeed Questionnaire, as well as measures of social connectedness from the IOS questionnaire and trust in agents from the Trust in Automated Systems Questionnaire.

AAA condition had significantly higher perceived affective understanding than the NAA condition (Z = -3.321, $p < 0.001$). Furthermore, the RAA condition also had significantly higher perceived affective understanding than NAA (Z = -2.878, p = 0.004). No significant difference was found between the AAA and NAA conditions.

Furthermore, we found the same trend in the *perceived affective interdependence* subscale with a significant main effect ($\chi^2(2)$ = 10.4, p = 0.006). However, only a small effect size (W = 0.217) was found by the following Kendall's W test. Similarly, post-hoc tests also showed the perceived affective interdependence in the AAA condition was significantly higher than that in the NAA condition (Z = -2.697, p = 0.007). Besides, the perceived affective interdependence in the RAA condition was also significantly higher than NAA (Z = -2.097, p = 0.036).

Lastly, for the *perceived behavioral interdependence* subscale, a Friedman test revealed significance over three conditions ($\chi^2(2)$ = 14.6, $p < 0.001$), with a moderate effect size (W = 0.304). Post-hoc tests further revealed the AAA condition had significantly higher perceived behavioral interdependence than NAA (Z = -3.090, p = 0.002) and RAA (Z = -2.086, p = 0.037).

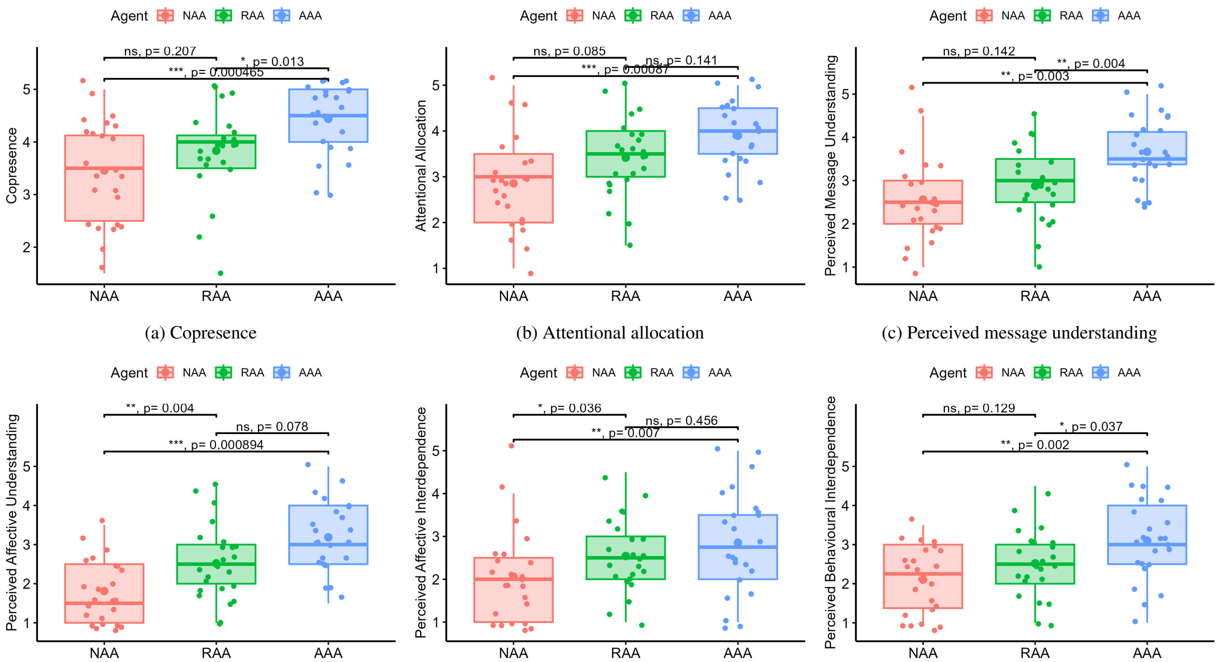
5.1.5 Interview Feedback

As shown in Tab. 2, **Q1** and **Q2** asked participants to rank the most and least liked conditions. A Chi-Square Goodness of Fit test yielded a significant difference against the most liked agent type ($\chi^2(2)$ = 15.844, $p < 0.001$). However, no significant difference was found on the most hated agent type (see Figure 6). Regarding **Q3**, after analyzing the interview recordings, we classified the participants' attitudes into "Definitely yes" (41.7 %), "Definitely no" (12.5 %), and "Depends" (45.8 %), while the responses to **Q4** were categorized into: "Just report accurate Physiological state"(41.7 %), "Just infer emotional state"(8.3 %), and "Report both"(50 %). For the **Q5**, participants' answers to this question were grouped into **(A)** general richer verbal and non-verbal behaviors and **(B)** personalized empathic behaviors. We will discuss these results in Sec. 6.3.

5.2 Physiological Measures

We analyzed EDA tonic and phasic components, and calculated HRV based on collected PPG data.

Regarding EDA, the Shapiro-Wilk test showed the EDA tonic feature SCL data violated normality, so we conducted the Friedman test with the SCL data. However, no significant main effect of the conditions was found ($\chi^2(2)$ = 0.25, p = 0.882). Similarly, the Shapiro-Wilk test also revealed the EDA phasic feature NS.EDR



(a) Copresence (b) Attentional allocation (c) Perceived message understanding

(d) Perceived affective understanding (e) Perceived affective interdependence (f) Perceived behavioural interdependence

Figure 5: Boxplots of subscales from the Social Presence Questionnaire.

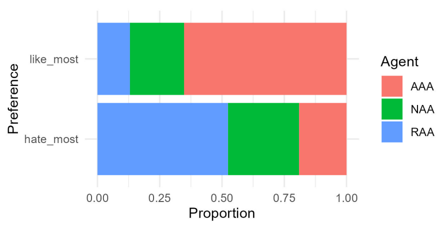


Figure 6: Results of the preference on most liked and hated agents.

broke normality. No significant difference was found among the three agent conditions with a Friedman test ($\chi^2(2) = 1$, $p = 0.607$).

For HRV, the Shapiro-Wilk test showed the HRV features of both RMSSD and SDNN data violated normality. No significant main effect of conditions on the RMSSD ($\chi^2(2) = 0.08$, $p < 0.959$) and SDNN data($\chi^2(2) = 0.674$, $p = 0.713$) through the Friedman test.

6 Discussion

The subsequent sections thoroughly examine the research outcomes and experimental observations and further discuss some observations in the interview.

6.1 Perceived Empathy

The results of perceived empathy-related measures partially support our **H1**. In RAA and AAA conditions, EMiRAs were able to show awareness of user physiological states with different levels of accuracy. Compared with the NAA, both RAA and AAA received higher ratings on social connectedness, perceived affective understanding, and perceived affective interdependence. This may be because users perceived that the speech from RAA and AAA, reflecting their physiological and stress states, implied an understanding and detection of their emotional states by these agents. Consistent with the findings of Desnoyers et al. [20] on the benefits of sharing physiological states among individuals for fostering a sense of connectedness, our study suggests that sharing physiological states between humans and virtual humans may similarly enhance social connectedness.

In addition to higher ratings in social connectedness and perceived affection subscales of social presence, the AAA condition also received significantly higher scores in animacy, perceived intelligence, trust, and all other aspects of social presence than the NAA condition. The Animacy subscale focuses on the extent of life in the agents, while the perceived intelligence subscale assesses the intelligence in the agents' human-like behaviors [7]. That is, the AAA's ability to precisely discern users' physiological and emotional states fostered a greater sense of human likeness, trust and social presence, which is important for building empathy between human and social agents [58, 26].

6.2 Human-Agent Social Interaction

The ratings of RAA and AAA on perceived intelligence, trust, social connectedness, and social presence subscales of copresence, perceived message understanding, and perceived behavioral interdependence partially supported the **H2**.

The AAA had higher perceived intelligence and trust ratings than the RAA. These were in line with the findings presented by [75], where an AI shopping assistant detecting user emotion with higher accuracy led to higher perceived intelligence and more willingness in consumers to follow the AI recommendations. Furthermore, Gupta et al. [32] also underscored the significance of agent accuracy in delivering assistance information as a crucial factor in fostering trust between users and agents. Our study also exemplifies how the accuracy of empathic agents in conveying users' physiological states can enhance perceived intelligence and user trust.

Moreover, the AAA also had higher social connectedness and social presence in copresence, perceived message understanding, and perceived behavioral interdependence. The social connectedness result aligned with the preference results. As commented by

P14, *"I like the agent that reflected my physiological state correctly and timely...."*(**Q1**). For the RAA, P18 commented, *"Once I found it was wrong, I didn't listen to him at all afterward."*(**Q2**). In addressing social presence, our study contributes to the literature by demonstrating how the accurate awareness of human physiological states by MiRAs impacts social presence. Previous findings regarding MiRAs' awareness of physical environment events, such as fan wind [41] and objects dropping [61], have been shown to enhance social presence. Combining our findings with previous research on MiRAs' virtual-physical interaction, we suggest that MiRAs' perception of both physical events and human physiological states contributes to fostering social presence.

No significant difference was found in the anthropomorphism and likeability of the GodSpeed questionnaire and the physiological measures. This could be because the three agent conditions had the same virtual human appearance. The only difference was the speech behavior, and the users' physiological states might mainly be influenced by the DroneRage shooting game, not the speech behavior. As reported by the P20, *"... I don't hate any of them...they looked pretty similar..."*(**Q2**). P08 said, *"...I was overwhelmed by the game and didn't pay too much attention to the agent..."*.

6.3 Privacy Concerns and Empathic Capabilities

Regarding the privacy and trust mentioned in **Q3**, the comments for "Definitely yes" were *"...It helps to let technology understand my emotion, why not..."* (P08) and *"...it's a bridge between the real world and the game world because emotions are something that games haven't encompassed"* (P16). The reasons for "Definitely no" might relate to participants' personalities and trust in technology. For example, P17 commented, *"I don't think I need a virtual human understanding my emotion."*. P23 commented, *"I don't believe AI technology can truly understand me... I don't want people to know my moods."*. The comments on "Depends" were mainly concerned about private data safety. For example, P05 said, *"Depends; I would be glad to share my physiological data with the virtual human only if the data privacy would be guaranteed."*. P02 mentioned, *"...Depends on whether my data will be protected."*.

Regarding **Q4**, reasons for the **I** type were mainly because participants treated the virtual human as a tool to monitor their physiological state. For example, P20 said, *"...I think the virtual human as an AI tool should be stronger than a real human in some way, like accurately detecting my physiological states."*. Some other participants had privacy concerns about letting the virtual human detect their emotional states. P10 said *"...I don't like the virtual human reporting my emotional state."*. However, for the type **II** responses, users thought they didn't know the true meaning of the physiological states so they preferred letting the machine tell them their emotional states directly. For instance, P08 said *"...just reporting physiological states makes no sense to me. The virtual human should directly tell me about my emotional state..."*. Lastly, for the **III** type responses, participants' reasons were a combination of type I and type II reasons. For example, P21 reported that *"...although I can interpret the physiological states myself, I would like the virtual human also infer my emotional states with some accuracy like over 80%."*. P11 reported, *"... I prefer the virtual human report, both my physiological state and the accuracy of detected emotional state..."*. Interestingly, we also found that some participants did not like the virtual human feeling, which was super human-like or accurate in detecting emotions. P14 commented, *"...if the virtual human can understand and respond to my emotion like a real human, I am afraid I will emotionally bond to this virtual human ..."*. P24 said, *"...I would feel it's horrible if the agent is too accurate in detecting my emotional state. I have no privacy in front of such a virtual human..."*.. This could be because of the uncanny valley effect [50] and privacy issues.

In **Q5**, an example for the group **(A)** responses is *"... the virtual human should have more variation of the tones, more body movements, and facial expressions..."* (P11). The P18's comments, *"...if I am in a sad mode, I don't expect the virtual human to change me from unhappy to happy. I would prefer it to leave me alone..."*, exemplifies the group **(B)** expectations on personalized empathic behaviors.

6.4 Implications on EMiRAs' Design

Based on our results, we identified design implications (DI) for future EMiRAs in stress-inducing contexts like the DroneRage game:

- **DI1.** Using physiological sensors to enable EMiRAs to perceive user physiological states helps build perceived empathy in EMiRAs (based on **H1**).
- **DI2.** Higher accuracy in detecting physiological states can further improve social perception of EMiRAs (based on **H2**).
- **DI3.** Users' physiological data safety and privacy should be handled properly (based on **Q3**).
- **DI4.** Prioritize guaranteeing EMiRAs' accuracy in detecting physiological states over emotional states (based on **Q4**).
- **DI5.** To enrich EMiRAs' biofeedback loop, incorporate non-verbal cues and personalized behaviors for more empathic responses (based on **Q5**).

7 Limitations

Our study has several limitations. First, the virtual human's communication lacks realism in both verbal and non-verbal aspects. Second, integration between the MR agent and game modules was suboptimal, resulting in occasional occlusion of drones by the virtual human's body. Third, we observed occasional mismatches between detected SCL changes and subjective user experiences. This discrepancy stemmed from the gradual nature of slow SCL changes and the algorithm's comparison of current data with SCL readings from 30 seconds prior. Finally, the simplicity of the game task and reliance on a single physiological measure limit the generalizability of our findings. Future research should address these limitations and explore more complex scenarios with multiple physiological indicators.

8 Conclusion and Future Work

In this work, we studied how EMiRAs' awareness of user physiological states affects perceived empathy. Our user study compared three virtual game companions with varying levels of physiological awareness. Agents demonstrating awareness (RAA and AAA) elicited higher perceived empathy, with AAA showing the most potential for enhancing human-agent interaction. User interviews revealed privacy concerns and considerations for implementing empathic capabilities in MiRAs. No significant differences were found in users' physiological data.

In future studies, we aim to investigate the impact of EMiRAs' intelligent capabilities, including awareness of both users' physiological states and physical events, such as door openings, phone calls, or temperature fluctuations, on perceived empathy.

Acknowledgments

This work was supported by the China Scholarship Council under Grant CSC NO. 202008320270. This research was also partly supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2022-03294, ALLRP 590555-23] and Alberta Innovates.

References

[1] Lab streaming layer library. https://github.com/sccn/liblsl?tab=readme-ov-file. doi: 10.5281/zenodo.7978343 4

[2] Mixamo. https://www.mixamo.com/. 3

[3] Readyplayerme. https://readyplayer.me/. 3

[4] Zeromq. https://zeromq.org/get-started/. 4

[5] R. Aylett and D. Ballin. *Intelligent virtual agents*. Springer, 2013. 1

[6] M. Z. Baig and M. Kavakli. A survey on psycho-physiological analysis & measurement methods in multimodal systems. *Multimodal Technologies and Interaction*, 3(2):37, 2019. 2

[7] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1:71–81, 2009. 5, 7

[8] T. Bickmore and J. Cassell. Relational agents: a model and implementation of building user trust. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p. 396–403, 2001. doi: 10.1145/365024.365304 1

[9] J. J. Braithwaite, D. G. Watson, R. Jones, and M. Rowe. A guide for analysing electrodermal activity (eda) & skin conductance responses (scrs) for psychological experiments. *Psychophysiology*, 49(1):1017–1034, 2013. 5

[10] S. Brave, C. Nass, and K. Hutchinson. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International journal of human-computer studies*, 62(2):161–178, 2005. 5

[11] Z. Chang. Using empathic mixed reality agents for remote collaboration. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 963–964. IEEE, 2023. 1

[12] Z. Chang, H. Bai, L. Zhang, K. Gupta, W. He, and M. Billinghurst. The impact of virtual agents' multimodal communication on brain activity and cognitive load in virtual reality. *Frontiers in Virtual Reality*, 3:179, 2022. 1

[13] Z. Chang, Y. S. Pai, J. Cao, K. Gupta, and M. Billinghurst. Emiras-empathic mixed reality agents. In *Proceedings of the 3rd Empathy-Centric Design Workshop: Scrutinizing Empathy Beyond the Individual*, pp. 1–7, 2024. 1

[14] F. Chiossi, Y. Turgut, R. Welsch, and S. Mayer. Adapting visual complexity based on electrodermal activity improves working memory performance in virtual reality. *Proc. ACM Hum.-Comput. Interact*, 7, 2023. 2, 3, 4

[15] M. Cifrek, V. Medved, S. Tonković, and S. Ostojić. Surface emg based muscle fatigue evaluation in biomechanics. *Clinical biomechanics*, 24(4):327–340, 2009. 1

[16] I. Cloud. Text to speech. https://cloud.ibm.com/catalog/services/text-to-speech. 3

[17] F. Creations. Eyes animtor. https://assetstore.unity.com/packages/3d/animations/eyes-animator-137246. 3

[18] F. Creations. Look animtor. https://assetstore.unity.com/packages/tools/animation/look-animator-122249. 3

[19] S. Cunningham, J. Henry, and J. Weinel. Augmenting virtual spaces: Affective feedback in computer games. *Technology, Design and the Arts-Opportunities and Challenges*, pp. 229–247, 2020. 2

[20] J. Desnoyers-Stewart, E. R. Stepanova, P. Pasquier, and B. E. Riecke. Jel: Connecting through breath in virtual reality. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–6, 2019. 7

[21] J. Dias and A. Paiva. Feeling and reasoning: A computational model for emotional characters. In *Portuguese conference on artificial intelligence*, pp. 127–140. Springer, 2005. 2

[22] M. Dragone, T. Holz, and G. M. O'Hare. Using mixed reality agents as social interfaces for robots. In *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 1161–1166. IEEE, 2007. 2

[23] A. Dzedzickis, A. Kaklauskas, and V. Bucinskas. Human emotion recognition: Review of sensors and methods. *Sensors*, 20(3):592, 2020. 2

[24] M. Egger, M. Ley, and S. Hanke. Emotion recognition from physiological signal analysis: A review. *Electronic Notes in Theoretical Computer Science*, 343:35–55, 2019. doi: 10.1016/j.entcs.2019.04.009 1

[25] C. Eghtebas, G. Klinker, S. Boll, and M. Koelle. Co-speculating on dark scenarios and unintended consequences of a ubiquitous (ly) augmented reality. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, pp. 2392–2407, 2023. 1

[26] A. Ferrari et al. Digital humanity. do users' gaming habits affect the perceived human-likeness of virtual agents in a simulated human interaction? 2021. 7

[27] S. for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures, W. Boucsein, D. C. Fowles, S. Grimnes, G. Ben-Shakhar, W. T. Roth, M. E. Dawson, and D. L. Filion. Publication recommendations for electrodermal measurements. *Psychophysiology*, 49(8):1017–1034, 2012. 5

[28] S. Gächter, C. Starmer, and F. Tufano. Measuring the closeness of relationships: a comprehensive evaluation of the 'inclusion of the other in the self'scale. *PloS one*, 10(6):e0129478, 2015. 5

[29] O. M. Giggins, U. M. Persson, and B. Caulfield. Biofeedback in rehabilitation. *Journal of neuroengineering and rehabilitation*, 10:1–11, 2013. 2

[30] V. P. Gonçalves, G. T. Giancristofaro, G. P. Filho, T. Johnson, V. Carvalho, G. Pessin, V. P. d. A. Neris, and J. Ueyama. Assessing users' emotion at interaction time: a multimodal approach with multiple sensors. *Soft Computing*, 21:5309–5323, 2017. 2

[31] J. Gratch, N. Wang, A. Okhmatovskaia, F. Lamothe, M. Morales, R. J. van der Werf, and L.-P. Morency. Can virtual humans be more engaging than real ones? In J. A. Jacko, ed., *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments*, pp. 286–297. Springer Berlin Heidelberg, 2007. 1

[32] K. Gupta, R. Hajika, Y. S. Pai, A. Duenser, M. Lochner, and M. Billinghurst. Measuring human trust in a virtual assistant using physiological sensing in virtual reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 756–765. IEEE, 2020. 7

[33] K. Gupta, Y. Zhang, T. S. Gunasekaran, N. Krishna, Y. S. Pai, and M. Billinghurst. Caevr: Biosignals-driven context-aware empathy in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 2, 5

[34] T. Holz, A. G. Campbell, G. M. O'Hare, J. W. Stafford, A. Martin, and M. Dragone. Mira—mixed reality agents. *International journal of human-computer studies*, 69(4):251–268, 2011. 1, 2

[35] M. Jacob Rodrigues, O. Postolache, and F. Cercas. Physiological and behavior monitoring systems for smart healthcare environments: A review. *Sensors*, 20(8):2186, 2020. 2

[36] J.-Y. Jian, A. M. Bisantz, and C. G. Drury. Foundations for an empirically determined scale of trust in automated systems. *International journal of cognitive ergonomics*, 4(1):53–71, 2000. 5

[37] J. A. Joireman, T. L. Needham, and A.-L. Cummings. Relationships between dimensions of attachment and empathy. *North American Journal of Psychology*, 4(1):63–80, 2002. 5

[38] P. Kalra and V. Sharma. Mental stress assessment using ppg signal a deep neural network approach. *IETE Journal of Research*, 69(2):879–885, 2023. 1

[39] D. Keltner, D. Sauter, J. Tracy, and A. Cowen. Emotional expression: Advances in basic emotion theory. *Journal of nonverbal behavior*, 43:133–160, 2019. 2

[40] K. Kim, C. M. de Melo, N. Norouzi, G. Bruder, and G. F. Welch. Reducing task load with an embodied intelligent virtual assistant for improved performance in collaborative decision making. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 529–538. IEEE, 2020. 1

[41] K. Kim, R. Schubert, J. Hochreiter, G. Bruder, and G. Welch. Blowing in the wind: Increasing social presence with a virtual human via environmental airflow interaction in mixed reality. *Computers & Graphics*, 83:23–32, 2019. 2, 8

[42] A. Kos and A. Umek. *Biomechanical biofeedback systems and applications*. Springer, 2018. 2

[43] M. Lee, K. Kim, S. Daher, A. Raij, R. Schubert, J. Bailenson, and G. Welch. The wobbly table: Increased social presence via subtle

incidental movement of a real-virtual table. In *2016 IEEE Virtual Reality (VR)*, pp. 11–17. IEEE, 2016. 1, 2

[44] I. Leite, C. Martinho, A. Pereira, and A. Paiva. As time goes by: Long-term evaluation of social presence in robotic companions. In *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 669–674. IEEE, 2009. 5

[45] A. Macintosh, N. Vignais, V. Vigneron, L. Fay, A. Musielak, E. Desailly, and E. Biddiss. The design and evaluation of electromyography and inertial biofeedback in hand motor therapy gaming. *Assistive Technology*, 34(2):213–221, 2022. 2

[46] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. H. A. Chen. NeuroKit2: A python toolbox for neurophysiological signal processing. *Behavior Research Methods*, 53(4):1689–1696, feb 2021. doi: 10.3758/s13428-020-01516-y 5

[47] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, vol. 2351, pp. 282–292. Spie, 1995. 1

[48] C. Misselhorn. Empathy with inanimate objects and the uncanny valley. *Minds and Machines*, 19:345–359, 2009. 5

[49] K. Miyamoto, H. Tanaka, K. Shidara, and S. Nakamura. Emotion prediction using multi-source biosignals during cognitive behavior therapy with conversational virtual agents. 2023. 2

[50] M. Mori, K. F. MacDorman, and N. Kageki. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2):98–100, 2012. 8

[51] D. Navarro, V. Sundstedt, and V. Garro. Biofeedback methods in entertainment video games: A review of physiological interaction techniques. *Proceedings of the ACM on Human-Computer Interaction*, 5(CHI PLAY):1–32, 2021. 2

[52] A. Nijholt. Towards social companions in augmented reality: Vision and challenges. In N. A. Streitz and S. Konomi, eds., *Distributed, Ambient and Pervasive Interactions. Smart Living, Learning, Well-being and Health, Art and Creativity*, pp. 304–319. Springer International Publishing, Cham, 2022. 1

[53] N. Norouzi, G. Bruder, B. Belna, S. Mutter, D. Turgut, and G. Welch. A systematic review of the convergence of augmented reality, intelligent virtual agents, and the internet of things. *Artificial intelligence in IoT*, pp. 1–24, 2019. 1, 2

[54] N. Norouzi, K. Kim, G. Bruder, A. Erickson, Z. Choudhary, Y. Li, and G. Welch. A systematic literature review of embodied augmented reality agents in head-mounted display environments. In *Proceedings of the International Conference on Artificial Reality and Telexistence & Eurographics Symposium on Virtual Environments*, 2020. 2

[55] C. Okasaki. *Purely functional data structures*. Cambridge University Press, 1999. 3

[56] A. Paiva. Empathy in social agents. *International Journal of Virtual Reality*, 10(1):1–4, 2011. 1

[57] A. Paiva, I. Leite, H. Boukricha, and I. Wachsmuth. Empathy in virtual agents and robots: A survey. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(3):1–40, 2017. 1

[58] S. Park and M. Whang. Empathy in human–robot interaction: Designing for social robots. *International journal of environmental research and public health*, 19(3):1889, 2022. 7

[59] A. Parnandi and R. Gutierrez-Osuna. A comparative study of game mechanics and control laws for an adaptive physiological game. *Journal on Multimodal User Interfaces*, 9:31–42, 2015. 2

[60] T. Pham, Z. J. Lau, S. H. A. Chen, and D. Makowski. Heart rate variability in psychology: A review of hrv indices and an analysis tutorial. *Sensors*, 21(12), 2021. doi: 10.3390/s21123998 5

[61] 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*, 8:25, 2021. 1, 2, 8

[62] R. G. Pizzie and D. J. Kraemer. The association between emotion regulation, physiological arousal, and performance in math anxiety. *Frontiers in psychology*, 12:639448, 2021. 2

[63] H. Prendinger and M. Ishizuka. The empathic companion: A character-based interface that addresses users' affective states. *Applied artificial intelligence*, 19(3-4):267–285, 2005. 2, 4

[64] S. D. Preston and F. B. De Waal. Empathy: Its ultimate and proximate bases. *Behavioral and brain sciences*, 25(1):1–20, 2002. 2

[65] M. Quest. Discover. `https://github.com/oculus-samples/Unity-Discover?tab=readme-ov-file`. 3

[66] H. Ranjbartabar, D. Richards, A. A. Bilgin, and C. Kutay. First impressions count! the role of the human's emotional state on rapport established with an empathic versus neutral virtual therapist. *IEEE transactions on affective computing*, 12(3):788–800, 2019. 2

[67] P. P. Ray. A review on tactile iot: Architecture, requirements, prospects, and future directions. *Transactions on Emerging Telecommunications Technologies*, 33(4):e4428, 2022. 2

[68] S. H. Rodrigues, S. F. Mascarenhas, J. Dias, and A. Paiva. "i can feel it too!": Emergent empathic reactions between synthetic characters. In *2009 3rd international conference on affective computing and intelligent interaction and workshops*, pp. 1–7. IEEE, 2009. 2

[69] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 3 ed., 2010. 1

[70] P. Sasikumar. Ecl-shimmer-c-api. `https://github.com/prasanthsasikumar/ECL-Shimmer-C-API`. 4

[71] S. Schmidt, O. J. A. Nunez, and F. Steinicke. Blended agents: Manipulation of physical objects within mixed reality environments and beyond. In *Symposium on Spatial User Interaction*, pp. 1–10, 2019. 1

[72] R. Sebastião. Classification of anxiety based on eda and hr. In *International Conference on IoT Technologies for HealthCare*, pp. 112–123. Springer, 2020. 2

[73] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert. Discriminating stress from cognitive load using a wearable eda device. *IEEE Transactions on information technology in biomedicine*, 14(2):410–417, 2009. 1, 2

[74] R. Sharma, H. Rajvaidya, P. Pareek, and A. Thakkar. A comparative study of machine learning techniques for emotion recognition. In *Emerging Research in Computing, Information, Communication and Applications: ERCICA 2018, Volume 2*, pp. 459–464. Springer, 2019. 2

[75] J. Van Der Tuyn. *Empathetic AI and Consumer Behavior*. PhD thesis, San Diego State University, 2020. 7

[76] A. Vrins, E. Pruss, J. Prinsen, C. Ceccato, and M. Alimardani. Are you paying attention? the effect of embodied interaction with an adaptive robot tutor on user engagement and learning performance. In *International Conference on Social Robotics*, pp. 135–145. Springer, 2022. 4

[77] I. Wang, J. Smith, and J. Ruiz. Exploring virtual agents for augmented reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019. 1

[78] F. Weidner, G. Boettcher, S. A. Arboleda, C. Diao, L. Sinani, C. Kunert, C. Gerhardt, W. Broll, and A. Raake. A systematic review on the visualization of avatars and agents in ar & vr displayed using head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 2023. 1

[79] Ö. N. Yalçın. Evaluating empathy in artificial agents. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 1–7. IEEE, 2019. 2, 5

[80] P. Zontone, A. Affanni, R. Bernardini, A. Piras, and R. Rinaldo. Stress detection through electrodermal activity (eda) and electrocardiogram (ecg) analysis in car drivers. In *2019 27th European Signal Processing Conference (EUSIPCO)*, pp. 1–5. IEEE, 2019. 1