Do relative positions and proxemics affect the engagement in a Human-Robot collaborative scenario?

Abstract

This paper investigates the effects of relative position and proxemics in the engagement process involved in Human-Robot collaboration. We evaluate the differences between two experimental placement conditions (frontal vs. lateral) for an autonomous robot in a collaborative task with a user across two different types of robot behaviours (helpful vs. neutral). The study evaluated placement and behaviour types around a touch table with 85 participants by measuring gaze, smiling behaviour, distance from the task, and finally electrodermal activity. Results suggest an overall user preference and higher engagement rates with the helpful robot in the frontal position. We discuss how behaviours and position of the robot relative to a user may affect user engagement and collaboration, in particular when the robot aims to provide help via socio-emotional bonding.

INTRODUCTION

Recently, the use of domestic robots has increased rapidly and can be found in various settings such as home companions[1], artificial tutors[2], caring[3] and more. In these types of domestic situations involving close interaction between humans and robots, the relative position of a robot is likely to have implications on user engagement and bonding because gaze (Feldman, 2007) and proximity (Bowlby, 1970) are known to be key factors in human-human socio-emotional bonding. This is true especially for cases that involve learning (Wall, 1993) or collaboration (Zaga, Truong, Lohse & Evers, 2014). For example, the position and engagement levels of a teacher in school environments impact the learning ability and motivation of the students, as well as their engagement (Cruickshank, Jenkins & Metcalf, 2009). Additionally, students' learning capabilities are highly dependent on their affective state, and as such, focusing on strategies that promote tutor-student engagement is of paramount significance (Robison, Mcquiggan and Lester, 2009). Similarly, identifying the affective state of a user and selecting an appropriate strategy is a task that could also be performed by an artificial robotic tutor

[1] Acceptable robotiCs COMPanions for AgeiNg Years, http://accompanyproject.eu/
[2] Embodied Perceptive turors for empathy-based learning, http://www.emote-project.eu/
[3] Managing Active and healthy aging with use of caRing servIce robots, http://www.mario-project.eu/

(Castellano, Paiva, Kappas, Aylett, Hastie, Barendregt & Bull, 2013). The use of robots instead of virtual agents has shown that physical embodiments are generally preferable for users as they are more enjoyable, improve performance, and enhance social presence (Kennedy, Baxter & Belpaeme, 2014). In terms of collaboration, humans and robots can coordinate their actions in order to accomplish shared goals such as in tasks that involve social mediators in remote communications (Papadopoulos, Dautenhahn & Ho, 2013), tele-operation platforms (Horiguch, Sawaragi & Akashi, 2000), collaborative control systems (Fong, Thorpe & Baur, 2002) and more.

In this paper we present a study that aims to evaluate the effects of robot position and behaviours in a collaborative scenario. In our setup, users are required to work together with a robot on a simple yet easily scalable and highly controlled memory-type task on a computer touch screen. The robot is able to provide help to the users by executing a number of behaviours depending on the experimental condition (helpful vs. neutral). The helpful behaviours of the robot were carefully designed from a pedagogical and psychological perspective in order to enhance the bond and engagement between the user and the robot, as well as to provide practical aid. In this study, users interacted together with the robot on a task, while we utilised a number of sensors to capture users' expressions and affective states, physiological data, relative distances and gaze behaviour. In order to evaluate the effect of robot behaviours, position and proxemics to the user, 85 participants took part in a study that exposed them to two different positions and behaviours of robot. In the frontal position, we placed the robot directly opposite the user at the wide end of a rectangular touch table, while in the lateral position the robot was placed on the short side of the table to the right of the user (see Figure 3, page 7).

The present paper thus focuses on answering the following research questions:

R1) Which robot position is more preferable and engaging to the users in a Human-Robot collaborative scenario?

R2) How does the robot position affect task engagement and social engagement in a collaborative scenario?

R3) Can different robot behaviours affect HRI in a collaborative scenario?

BACKGROUND RESEARCH

There are multiple research areas this study draws from, such as task and social engagement in Human-Robot interaction Corrigan, Peters, Küster & Castellano, 2016), proxemics in classrooms and HRI and affect recognition. This section presents related research from the perspectives of the aforementioned areas.

Task and Social Engagement

Humans tend to interact with each other extremely efficiently (Küster & Kappas, 2014) and they typically do so in an automatic manner that pays little conscious attention to the micro-level behavioural mechanics of the interaction (Bargh, 1988). However, in order to enable robots to attain a similarly effortless proficiency in HRI, we have to understand how engagement works and what types of engagement can exist in the collaborative scenario that this paper focuses on. As proposed by Sidner et al., social engagement is "the process by which two (or more) participants establish, maintain and end their perceived connection during interactions which they jointly undertake" (Sidner, Kidd, Lee & Lesh, 2004) while task engagement is considered to be present when users experience a sense of flow in their interactions, including a certain level of attention, enjoyment and concentration with the learning task (Corrigan, Basedow, Küster, Kappas, Peters & Castellano, 2015). Flow occurs when humans are highly involved with the activity and it offers an enjoyable experience. In a collaborative task, both social and task engagement play important roles. Yet, the collaboration itself has the potential to affect the overall user engagement, e.g., when it requires to shift the focus of attention between the task and the robot (Csikszentmihalyi, 1990). In turn, the level of engagement experienced by a user can influence how much effort and attention their prepared to give to the collaboration, and ultimately, their decision to stop or continue.

Proxemics and positions in a framework of socio-emotional learning

Relative proxemics and positions are likely to play an important role in human-robot collaborative learning because they impact how easy it is for the robot to engage the learner via socio-emotional bonding and micro-level social behaviours. Our framework for asserting an importance of socio-emotional bonding in engagement and learning has a wide basis in human attachment and developmental theory (Bowlby, 1970). In bridging fundamental attachment theory and work on its biological underpinnings, recent studies (Feldman, 2007) have taken this approach a step further to study affiliative social bonds in the context of micro-level social behaviours, such as gaze, affective responses, and touch – suggesting that these types of micro-level behaviours indeed play a role binding individuals together. Thus, an empathic robot should be able to facilitate learner engagement via socio-emotional bonding (Castellano et al. 2013), which might ultimately lead to the type of advantages found in school for trusting teacher-student relationships (Hattie, 2009). However, in order to better understand these types of affective dynamics in HRI, fundamental contextual factors such as proxemics and relative positioning of the robot require further attention.

Proxemics in HRI is a growing field that dictates the settings of the interaction with the robot. For example, Kim and Multu (Kim & Mutlu, 2014) performed two studies to evaluate the effect of power and proxemics distances between the robot and the users. In the first study with the robot acting as supervisor, they discovered that participants reported a more positive user experience with the robot being close to

them rather than distant while they obtained the opposite results when the robot was subordinate. In the second study, when the robot was cooperating with the users, participants had more positive experiences when the robot was distant than close while exactly the opposite happened when the robot was competing against the users. The findings above suggest that proxemics may indeed play a vital role –and that a tutoring robot may require a carefully calibrated distance to the user, at least with regards to subjective user experience.

Affect recognition

In order for an intelligent system to detect affective cues, hardware sensors capable of perceiving socio-emotional cues are required. Electrodermal activity (EDA) has been among the most widely used response systems in psychophysiological research, and it has historically been closely linked with psychological concepts of arousal, attention, and emotion in particular (Dawson, Schell & Filion, 2007). It is based on changes in the electrical conductivity of the skin that result from the opening and closing of sweat glands (Dawson, Schell & Filion, 2007), that in turn is related to arousal (Boucsein, Fowles, Grimnes, Ben-Shakhar, Roth & Filion, 2012). An interesting example for ambulatory measurement is a wrist sensor used by researchers in order to record and analyse the EDA of participants inside and outside the laboratory, including long-term measurement (Picard, Fedor, & Ayzenberg, 2016), as well as a variety of other tasks, such as the identification of frustration during a math exercise (Prendinger, Mayer, Mori, & Ishizuka, 2003), or the detection of stress levels (Picard & Healey, 2000). In combination with a measurement of eyebrow position and heart rate, EDA has further been used to assess a number of more differentiated emotional states elicited during a computer game (Conati, 2002). The literature above signifies the importance of affect recognition in scenarios where an autonomous system (i.e., robot) is able to predict user's emotions and trigger an appropriate action.

In view of the extant research on proxemics in HRI, there still appears to be a need for further research on collaborative HRI scenarios involving help on emotionally meaningful tasks. In particular, we argue that in order to elicit meaningful socio-emotional bonds and engagement, it is essential to understand more about how the potential aid provided by a robot may impact the user depending on the proxemics of relevant micro-level behaviours. We further suggest that this may be most effectively investigated via comparatively simple yet emotionally significant paradigms. With this paper we intend to take a step in this direction by analysing and evaluating the physiological data of the users interacting with a collaborative robot on a task that was specifically designed and framed to facilitate an affective link between user and robot.

SYSTEM DESCRIPTION

In order to close the affective loop between robot and user, the system can attempt to bond with the user very directly if the task is properly framed to elicit care and empathy from the user. We developed a simple memory card game that requires collaboration between the user and the robot in order to complete it and can last for up to ten minutes, although it can be completed sooner depending on user performance (Figure 1). The robot begins the interaction by explaining that his battery is damaged and needs help to build a new one, thereafter, the human and the robot work together in an attempt to find all the components. This framing of the task was designed to elicit empathy from the user, and it is reinforced by additional utterances made by the robot during the game. The participant is required to help the robot find the components it needs to build a new battery. The task is displayed on a flat touch screen in a horizontal position to allow the users to interact effortlessly both with the robot and the task.

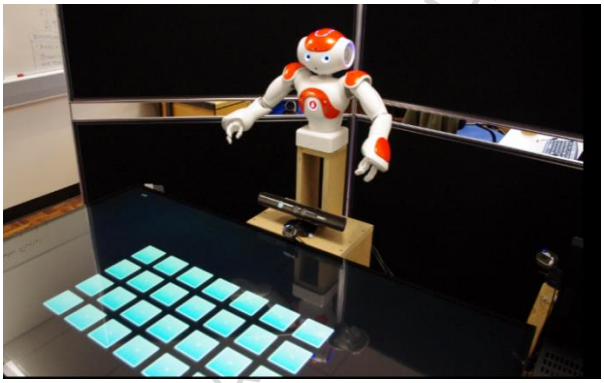


Figure 1: Memory task and the robot in Frontal position

For the role of the robot, we decided to use a torso version of the NAO[4] robot from Aldeberan (Figure 2). This robot has 14 degrees of freedom, consisting of two 5-degrees-of-freedom arms with two hands and a head with 2-degrees-of-freedom for pan and tilt. The head of the robot is equipped with multi-colour LED lights, two speakers in its ears for producing synthesised sound through text-to-speech, and two cameras that can capture videos up to 30fps. These characteristics provided us with an embodiment capable of exhibiting simple expressive behaviours towards the user.

[4] http://www.aldebaran.com/en

Figure 2: Robot Nao Torso

METHODOLOGY

In this study, we evaluated the effect of robot position and behaviours to the users in a collaborative empathy-eliciting scenario. In order to collect and analyse the data, we designed two different setups that utilise a number of sensors capable of recording the interactions and expressive states of the users.

Participants

Eighty-five adult participants (53 males – 32 females) were recruited at the University of Birmingham. In order to evaluate the experimental conditions, we divided the participants randomly into 2 conditions, 42 participants with the robot in a lateral position and 43 in a frontal position (see **Figure 3** for robot positions). Additionally, we balanced the exposure of 2 different robot behaviours, namely, helpful and neutral to the participants. Then gender distribution was equally balanced across the conditions. Due to technical constrains and limitations of the sensors we dropped the data of 5 participants as the recorded data failed to capture continuously. The participants were either students or employees at the University aged between 18 and 57 (M= 24.4, SD= 6.08). All participants provided informed consent prior to participating in the study.

Scenario and robot behaviours

Participants collaborated with the NAO robot in a memory-type game designed to help the robot. We defined two counterbalanced conditions that entailed different robotic behaviours in order to assess the potential impact of helpful bonding-oriented behaviours, as well as to balance any possible positive or negative short-term emotional effects resulting from the initial novelty of being introduced to a robot. We categorised these behaviours based on the type of assistance offered by the robot:

<u>Helpful, Bonding-oriented, and Instructive:</u>

In this group, users experienced a robot that was designed to be helpful, instructive, and socially connective. The robot expressions were designed to be encouraging and personable towards the participant in a way that was designed to establish a certain

level of socio-emotional bonding with the user. For example, in the introduction the robot described the reason they needed to work together in order to build their battery while at the same time it was looking directly at the participant and addressing them by their name. Furthermore, it aimed to emphasize a shared emotional experience and sense of "togetherness" in collaborating on the task (sample utterances: "I'm really worried that we won't recover the component in time"; "I'm really glad we are doing this as a team". This mode emphasised the collaboration and praised the participant for performing well in the task.

Neutral and Partially Instructive:

In this mode, the robot was designed to be neutral and to provide less help. The robot did not address the participants by their name and altered the utterances for collaboration by removing words socially-oriented utterances and individual words such as "our" or "we" during the interaction. Additionally, we disabled the automatic head tracking behaviours of the robot to reduce the sense of a direct social connection between the robot and the user.

Experimental setup

We configured our setup to optimally compare the two different robot positions and behaviours and capture the physical reactions from the users while interacting with our system (Figure 3). We sketched both conditions and positioned the sensors and cameras around the robot in order to prevent any possible obstruction in the camera view and maximise the viewing angle as much as possible. The cameras were placed at the table's trim level facing towards the user's face to maximise the viewing angle of the face when the user faces the robot or the screen.

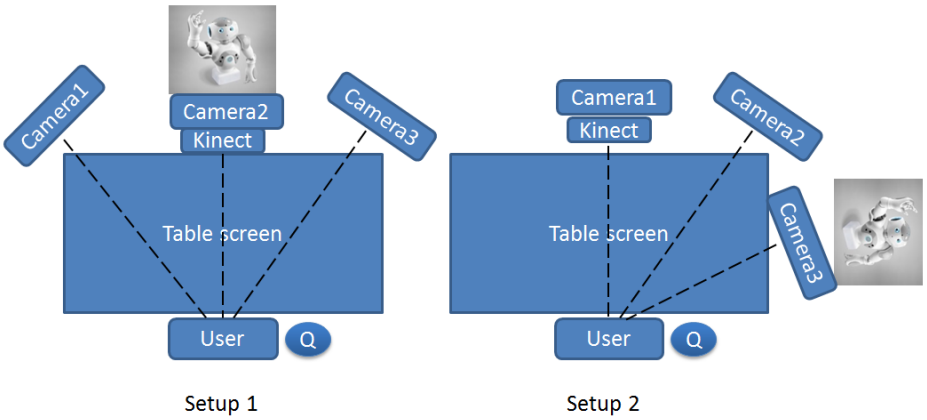


Figure 3: System setups

The robot was placed either in the frontal or lateral position as shown in Figure 3 on top of camera 2 or 3. The Kinect sensor was fixed on the long side of the table in order to fully capture the skeleton data and user's distance from the table.

Procedure

Before each session, participants filled out the short questionnaire and demographic form while a wrist sensor[5] able to pick electro-dermal activity data was attached to their wrist in order to warm up and provide accurate readings during their subsequent interaction with the system. After the questionnaire, each participant approached the touch screen and was given instructions about the task. Participants performed a simple touch/tap task before the main task in order to familiarize themselves with the touch table. During each session, three web-cameras facing the participant from different viewing angles started recording automatically in a synchronised manner with all other sensors. Participants were informed they had the freedom to quit the study without any repercussions (all users opted to continue through to the end). At the end of each session, participants were given a small reward (£5 Amazon voucher) as compensation.

Data collection

In order to capture and evaluate data from the study, we used a number of sensors capable of perceiving the environment and the users. We utilised three web cameras in different locations around the touch table, a Microsoft Kinect[6] sensor, and an Affectiva Q Sensor (Figure 3). The three web cameras pointed towards the participant and saved synchronized compressed video files. The Q Sensor was transmitting information via Bluetooth with an 8 Hertz sampling rate. Finally, the Kinect sensor extracted facial characteristics in real-time such as head direction information and facial action units along with the depth information.

The Kinect module uses the Microsoft Kinect sensor to extract head gaze information, depth and Kinect facial Animation Units (AUs; (Ahlberg, 2001). For this study, we decided to use the sensor for detecting the gaze of the user as results from laboratory tests provided sufficient performance for gaze estimation (D'Mello, & Graesser, 2010). The sensor can recognize the participant's head orientation in the X and Y axis thereby estimating their gaze/attention. The gaze estimation is limited to the head direction as the eye gaze is not accounted for as head orientation, and because it is one of the most reliable features to determine the focus of user's attention (Ba & Odobez, 2006; Sidner, Kidd, Lee & Lesh, 2004)). Finally, the sensor also provided information

[5] Affectiva Q Sensor, http://qsensor-support.affectiva.com/
[6] Microsoft Kinect Sensor, http://www.microsoft.com/en-us/kinectforwindows/

regarding the depth of view to predict the average distance of the user from the sensor.

The Q Sensor module utilises an electronic wrist sensor for capturing users' electro-dermal activity during their interactions with the system. This sensor has been used to record electro-dermal activity in HRI in previous research (Leite, Henriques, Martinho & Paiva, 2013) and it can be configured to transmit data in real-time to a computer for further processing. In addition, the sensor transmits a skin temperature measurement along with the hand's acceleration readings. Electro-dermal activity (EDA) can be used to estimate user's emotional arousal while interacting with the task or the robot (Chanel, Rebetez, Bétrancourt & Pun, 2008; Dawson Schell and Filion, 2007).

Data analysis

The study collected a large corpus of data, including videos and sensor logs. This section describes the steps and methodology taken to analyse and evaluate the recorded information.

Cameras and video processing

We used three web cameras around the touch table in order to capture the user from different angles and select the best performing camera that captures the users' facial characteristics. The video recordings were synchronised with the log files by the server to have the same starting and ending points. The camera module saved the recordings from each task individually for offline analysis.

For the offline video analysis, we used the OKAO SDK[7] in order to extract facial characteristics such as expression and smile estimation, eye gaze information and blink estimation. For this study, we only used the smile values because they can be measured most reliably in this system, and because they have previously been linked to user engagement (Malta, Miyajima and Takeda, 2008). However, we caution that the smile is of course *not* a pure readout of enjoyment since it can, depending on the context and circumstances, alternatively relate to social motives (Fridlund, 1994), communicative functions such as politeness, as well as unrelated mechanisms such as speech (Tassinary, Cacioppo and Vanman, 2007). We developed a batch tool extractor that utilises the OKAO libraries and extracts the facial characteristic values into CSV files for further statistical analysis. The OKAO library can recognise the smile value of the user along with the confidence levels for each timeframe. In order to get accurate smile data, we extracted the values from the videos and sorted them based on smile intensity in 3 different groups: 11-30% for mild smiles, 31-50% for moderate smiles

[7] OMRON OKAO SDK, http://www.omron.com/ecb/products/mobile/

and finally 51-100% for high smiles. Values below 10% were disregarded as this group contained false positives. The aggregated smile detection is the output of the smile value multiplied by the confidence value. The confidence value (1-1000) is generated from the OKAO software and indicates how accurately the smile has been captured. Additionally, we performed a simple test with the OKAO libraries in order to confirm the performance of the cameras by assessing the confidence levels of the SDK. In the frontal condition of the robot, camera 2 has shown higher average confidence level than camera 1 or 3 in capturing user's face (1= 47, 2= 375, 3= 50). In the lateral condition, camera 1 had higher degree of confidence in the interaction than camera 2 and 3 (1= 407, 2= 333, 3= 10). For that reason, we selected Cameras 1 and 2 for the Smile estimation analysis (see Figure 3, page 7).

Electro-dermal Activity analysis

The Q Sensor device transmits signals via a Bluetooth module to a local computer at a predefined rate of 8Hz. Since the interactions with the system lasted between 5 to 10 minutes, we decided to transform the samples by binning averages for each condition into 5 slots (20%, 40%, 60%, 80% and 100% of interaction times). The averaged data for each participant were standardized using *Z*-scores and later analysed in SPSS[8] to compare the frontal with the lateral position and helpful with neutral behaviours of the robot (Ben-Shakhar, 1985).

Gaze estimation procedure

In order to evaluate and distinguish gaze between the task and the robot, we developed an offline gaze detector tool using a simple Neural Network (NN). To successfully detect user's gaze, it is necessary to train a NN with various postures and facial expressions for both experimental conditions (frontal and lateral). We developed a gaze classifier using the Kinect sensor as input and a computer keyboard to define the gaze as an output for the NN. We recruited two male participants from our department and instructed them to look at various locations throughout the classification process and at the same time to vary their standing posture in terms of height to represent a variety of user's height. The classifier uses three inputs from the Kinect sensor to detect gaze: head pan, tilt and position in the vertical axis (Height). The coding was done using a keyboard by an external observer. The log files were later fed into a simple NN using hyperbolic tangent activation function (TANH) with three inputs (pan, tilt and height), three active and nine hidden layers and finally one output. For training, we used the Rprop (resilient backpropagation) (Riedmiller & Braun, 1993) supervised algorithm as it produced the most accurate results compared to other learning methods. The generated network was utilised with the Kinect data from the main study in order to estimate users' gaze for each interaction. The NN tool labels the output according to the three inputs into three categories: gaze at screen, gaze at

[8] IBM SPSS Software, http://www-01.ibm.com/software/uk/analytics/spss/

robot, and gaze elsewhere. This approach has some inherent limitations as the gaze is calculated using head orientation information while ignoring eye gaze. However, our lab tests revealed that head orientation is adequate for gaze approximation in our setup as the mapping consists of only 3 different locations that require head movement.

Hypotheses

Based on the literature review and previous studies we derived the following hypotheses:

H1: Users that interact with the robot in the frontal position will find it more preferable and enjoyable than in the lateral position.

H2: Users will be more engaged with the robot when it is located in the frontal position.

H3: The helpful and instructive behaviours of the robot will have a positive and engaging impact in the Human-Robot interaction.

RESULTS

In this section, we present our findings structured by type of data delivered from the sensors more specifically, the electrodermal activity and temperature data from the Q Sensor, gaze behaviour and distance from Kinect sensor, and finally smile estimation from the OKAO libraries. We analysed these data in order to examine and evaluate how robot position could affect the Human-Robot interaction, and overall the user engagement with the robot. Additionally, we present the effect of different robot behaviours in the interaction (neutral vs. helpful behaviours).

Electro-dermal activity

To simplify and reduce the number of samples, we divided the study time into 5 bins and averaged the EDA values as described in the Data analysis section. Figures 4 and 5 show the results of the EDA activity and skin temperature of the user respectively.

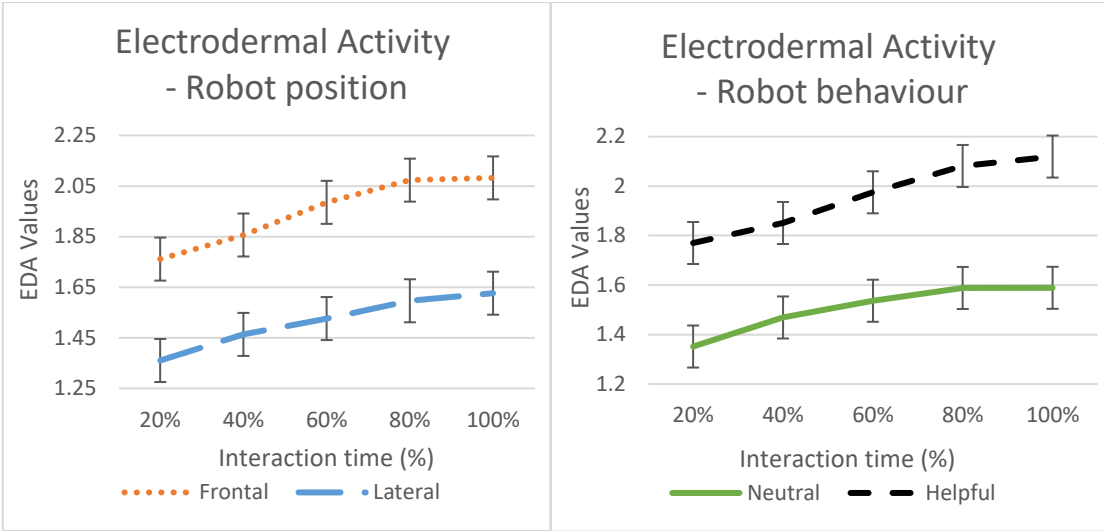


Figure 4: Electrodermal activity values taken from the Q Sensor for both robot positions and behaviours. The X axis indicate the elapsed time from the beginning of the interaction using percentages (% of time). The Y axis shows the average electro-dermal activity values. The error bars represent standard error of the mean (*SEM*).

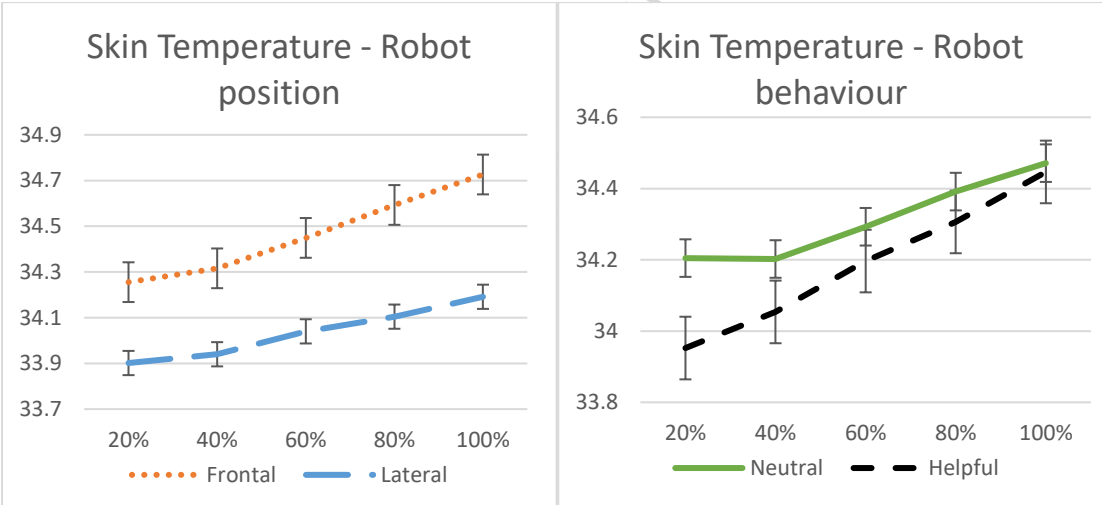


Figure 5: Skin temperature average values taken from the Q Sensor. The X axis shows the time while the Y axis the temperature in Celsius units.

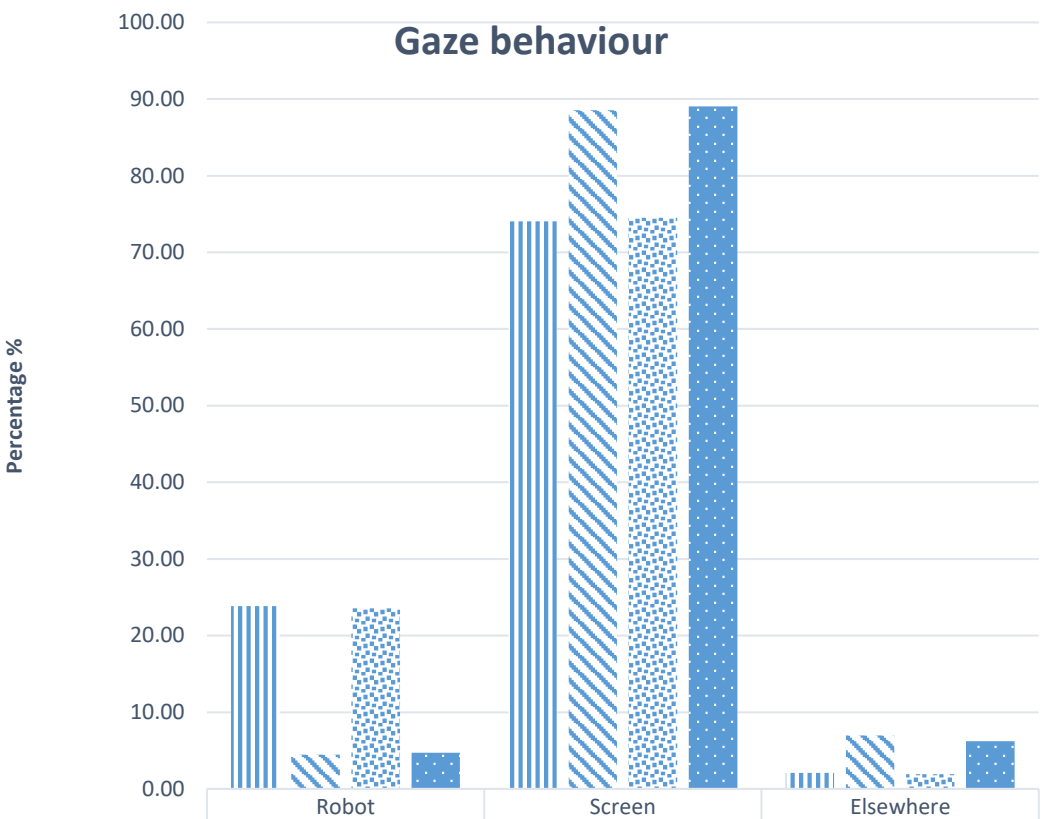
The values in the graphs represent the means of all participants for each time slot. Repeated measures 2-Way mixed factorial ANOVAs with position and behaviour as between subjects' factors, and percentage of time as the repeated measure were carried out for EDA and Skin temperature measures in order to investigate the effect of both factors over time. Significant differences were observed for the robot position as measured by skin temperature, $F(4,73) = 3.245$, $p = .0165$, $\eta^2 = .581$, and EDA, $F(4,73) = 2.975$, $p < .0246$, $\eta^2 = .465$. These findings clearly supported our hypothesis that a frontal position of the robot was physiologically more activating. In comparison, the robot behaviours appear to have had less of an effect on physiological arousal but were still significant for EDA, $F(4,73) = 2.492$, $p = .0503$, $\eta^2 = .170$, though not

significant for temperature, $F(4,73) = 1.611$, $p = .1804$, $\eta^2 = .267$. However, out of the two measures, EDA is arguably the psychologically more important one, with a very long and well established history in the assessment of emotional arousal (Dawson, Schell & Filion, 2007). The Interaction effect between the two factors (robot position and behaviour) was not significant (EDA: $F(4,73) = .427$ $p = .516$ $\eta^2 = 2.046$; Temp.: $F = 2.779$ $p = .100$ $\eta^2 = 20.049$). Thus, while there was a statistical trend in the temperature data, we found no statistical evidence for a significant amplification effect between robot placement and robot behaviour. Rather, both factors appeared to have the potential to make a separate positive contribution to these psychophysiological engagement parameters.

The EDA and temperature data suggest that users were more activated and engaged by the interaction with the robot in the frontal position. In addition, they were more activated by the helpful robot compared to the neutral robot. We suggest that when the robot was in the frontal position, users found it easier to pay attention to the instructions and directions whilst having a direct contact with the robot as it occupied the centre of their vision and attention. The different robot behaviours significantly affected EDA, suggesting that the helpful robot indeed increased the arousal state of the user during the interaction. With respect to the dynamics of these changes, we further observed that the differences generally already emerged at a very early stage of the interaction and tended to persist throughout the interaction – with the exception of the behaviour-based temperature differences that appeared to dissipate over time.

Gaze behaviour

The gaze estimation provides useful information regarding users' attention and we simplified that model by assigning different labels that could be used to distinguish engagement and attention with the task and the robot. Figure 6 represents the overall time (%) the users spent gazing toward the robot, screen or somewhere else for both experimental conditions and robot behaviours. The percentages were calculated using the output of the NN tool and show the difference in terms of gazing between the frontal and the lateral positions and different behaviours of the robot. It is evident that users spent more time gazing at the robot when it was placed in front of them. A 2-Way ANOVA statistical analysis was performed in order to confirm our observations for significant differences between the two robot positions and behaviours. (Position: $F(3,74) = 65.346$, $p < .0001$, Behaviours: $F(3,74) = .798$, $p = .499$). The Interaction effect between the two factors (Robot position and behaviour) was not found to be significant ($F = .033$ $p = .883$ $\eta^2 = 5751$).



	Robot	Screen	Elsewhere
Frontal - Helpful	23.86	74.04	2.10
Lateral - Helpful	4.48	88.55	6.98
Frontal - Neutral	23.57	74.52	1.90
Lateral - Neutral	4.73	89.03	6.24

Figure 6: Gaze estimation: The average gaze time for all users between the two experimental modes presented into three gaze groups

Similar to the EDA and skin temperature readings, the gaze extraction confirms our hypothesis that users should be in the frontal position when the aim is a collaborative scenario with the robot. In our scenario, the user's objective was to collaborate with the robot on a specific task that took place on the touch table. For that reason, users had to divide their attention between focusing on the task while at the same they had to follow the instructions delivered and feedback as well as look at it in order to follow its nonverbal behaviour. This type of divided attention during problem solving can be challenging, and users generally prefer to reduce their effort when they have to frequently switch the target of their gaze between two objects. In consequence, they are likely to prefer the frontal position of the robot (Papadopoulos, Dautenhahn & Ho, 2012). With the robot in a frontal position, users pay more attention to the instructions, which may additionally increase their social bonding (Belpaeme, 2012). In the present data, the different robot behaviours did not significantly affect the gazing behaviour of the users. Importantly, with view to future research in

collaborative robotics, the rather dramatic difference in percentage of gaze directed at the robot should be noted. We assume that, for the frontal condition, the rough distribution of one quarter of the time spent paying attention directly at the robot (75% at the screen) may indeed have been a good balance for managing attention. However, we argue that in HRI, less than 5% of time spent paying direct attention to the robot would most likely not be sufficient for most purposes in social robotics. Therefore, a lateral placement of the robot should generally be avoided.

Distance from the sensor

The depth values from the Kinect sensor represent the distance of the user from the sensor located at the end of the screen (see Figure 3). We assumed that users might tend to get into closer proximity with the sensor, and thus to the robot, when the robot is located in the frontal position. However, the statistical analysis failed to find a significant difference on this criterion. Standing closer and leaning towards the screen indicates that users were more engaged with the task during their interaction with the system. Participants were instructed to stay in the middle of the screen as it was necessary to reach all the buttons on both sides of the screen.

The graphs below show the distance values between the Kinect sensor and the user while interacting with the system. The horizontal axis represents the study time divided into 5 equal slots. Overall, users appeared to stand slightly closer to the task when the robot is present in front of them (frontal position). However, a repeated 2-way factorial ANOVA analysis did not support any of our hypotheses as statistically significant, with p =.130 for differences in robot position, and p = .427 for robot behaviours (Position: $F(4,78)$= 1.839, p = .130, Behaviour: $F(4,78)$ = .974, p = .427. The Interaction effect between the two factors was not found to be significant (F = 1.285 p = .260 η^2 = 62855).

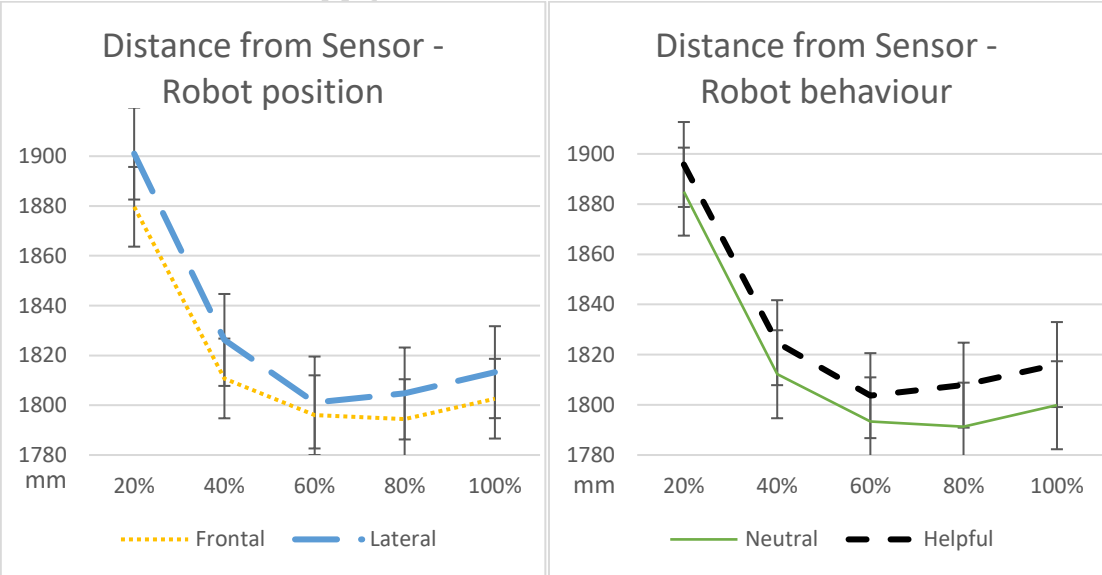


Figure 7: Distance (mm) from the Sensor divided into 5 times slots. The error bars represent standard error of the mean (*SEM*).

Despite the lack of statistical significance in the present analysis, it is still possible that more pronounced differences might be observed during more specific moments of the interaction, such as when direct eye contact can be achieved with minimal effort, or after a longer period of time in the interaction, when the degree of social bonding has increased. In order to further investigate these findings, a somewhat longer-term study would be required with a sufficient number of participants, and in which these subtle situational differences are systematically integrated in the design.

Smile estimation

The smile estimation via OKAO is expected to be related to enjoyment, albeit with some limitations (Mauss & Robinson, 2009; Hollenstein, T., & Lanteigne, 2014). We extracted these values for both conditions to evaluate the most enjoyable and engaging position and behaviour type of the robot. For this analysis, we categorised the smile intensity in 3 different groups: 11-30% for mild smiles, 31-50% for moderate smiles and finally 51-100% for high smiles. Results suggest a number of subtle differences in smiling behaviour when smile intensity is taken into account. Figure 8 shows that users generally smiled mildly during the game and had a positive experience with the robot. A repeated measures 2-way factorial ANOVA with position and behaviour as the between-subjects factors, and smile-intensity (3 levels) as the repeated measure showed a significantly higher overall smiling intensity in the frontal vs. lateral position of the robot, $F(2,71) = 3.883$, $p = .025$. No differences were found for robot behaviour ($p = .246$).

We further analysed the different smiling patterns by smile intensity to identify what types of smiles were most affected by the manipulation. Users who experienced the frontal position of the robot expressed a significantly higher number of mild smiles (11 - 30% intensity) than those in the lateral position during the interaction with the system, $F(3, 70) = 7.873$, $p = .006$. We suggest that the robot may have had a greater effect on the users in the frontal vs. lateral position, because users were able to visually perceive it better, prompting them to show at least a mild or polite type of smiling. No significant differences found in the moderate and high intensity smiles in terms of robot position (Moderate $p = .169$, High $p = .424$). Similarly, the different robot behaviours did not significantly affect the individual smiling groups in the study (Mild $p = .096$, Moderate $p = .727$, High $p = .757$).

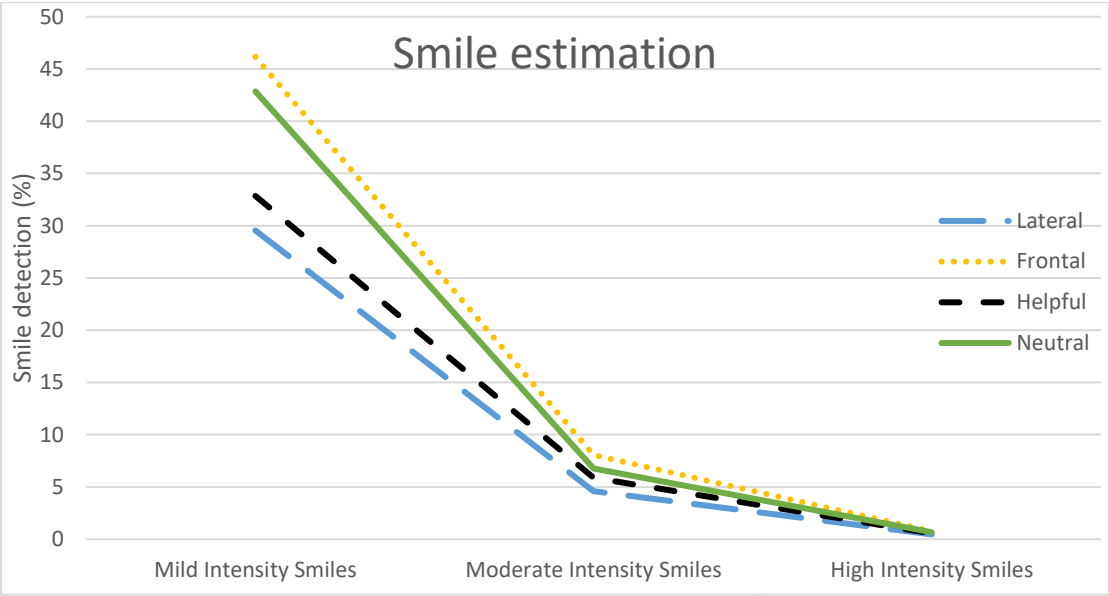


Figure 8: Smile percentages binned into 3 categories for both robot position and behaviours. The horizontal axis represents the groups for smile categories: Mild (11-30% intensity), Moderate (31-50% intensity) and High (51-100% intensity) while the vertical axis shows the smile degree detection in percentage (%).

Overall, based on the significant results taken from the skin conductance sensor (Q Sensor), smiling, and the gaze activity, users appeared to show higher attention levels and greater engagement with the robot for the frontal position with helpful robot behaviours compared to the results of the lateral position of the robot. In particular, the increased smiling, albeit at relatively mild intensity, suggests that participants found it easier to connect socially with the robot.

Discussion and Conclusions

In this paper we present a study that evaluated the effects of robot position and behaviour in a collaborative scenario. We designed and developed a simple interaction that required a human user to collaborate with a robot during a task, whilst also utilising a number of sensors to capture facial expressions, physiological data, gaze direction and relative distances from the robot during the interaction. The systematic manipulation of frontal vs. lateral placement of the robot allowed us to examine the effects of proxemics on user engagement and enjoyment. Furthermore, we introduced two different robot behaviours to the users to examine possible differences in HRI throughout the study. In summary, placing a robot capable of expressing helpful behaviours opposite the user may benefit the interaction and the engagement process as a whole mainly because users spend more time facing towards

the robot, thus increasing the interaction and collaboration. This may reduce the subjective cost of effort and divided attentional resources and thus may make it easier for users to attach emotionally to the robot. Additionally, the electrodermal activity and skin temperature readings were higher in the frontal position. These findings along with the gazing pattern and smile estimation confirm our hypotheses (H1 and H2) as the overall preference and engagement related indicators were higher in the frontal position of the helpful robot (H3). However, since the study involved a relatively short and simple task that lasted only 5-10 minutes, it is possible that more pronounced engagement and attentional effects would have emerged over a longer time span as a function of improved socio-emotional bonding. Thus while the present study presents some clear arguments for the effectiveness of our manipulation, a more long-term study would be required to fully explore the effects of an autonomous robot on a Human-Robot collaborative task.

The results of this study suggest a particular importance of gazing behaviour for collaborative work between humans and robots in general. The mutual gaze behaviour is a powerful indicator of engagement therefore, detecting such behaviour could greatly enhance and target the robot behaviours towards the user (D'Mello, Chipman & Graesser, 2007). Currently, our system analyses the gazing behaviour using offline video data but a similar approach could be used for real time automatic detection of gaze using the trained neural networks. Additionally, exploratory offline video analysis of our present data suggest that individual users may show substantial inter-individual differences in their overall pattern of gazing behaviour. For example, for some users, there appeared to be a dissociation between eye gaze and head gaze, such as that although they were facing the robot, their actual eye gaze appeared to be directed at the screen. Therefore, an automatic detection algorithm based on a combination of head orientation and eye gaze information would substantially improve the reliability and performance of the system. We suggest that a successful implementation of such algorithms may have a substantial positive impact on human-robot collaboration, for example in the workplace.

The results of this study signify the importance of robot behaviours and position relative to the user in collaborative scenarios. Participants in our study generally reacted positively to the system that they used to collaboratively interact with the robot. However, our findings are limited in that they only present data on a relatively short period of interaction with the robot, and thus, as in many HRI experiments, some of our findings may have been influenced by the novelty of first exposure to this type of technology. Ultimately, we hope the findings from this study will inform other researchers who aim to develop scenarios with autonomous robots as a collaborative tutor, whether in a collaborative game context, as addressed in this paper, or in other contexts such as robotic tutors or other contexts and application areas where robots can collaboratively interact with people. We recommend that, in all of these areas,

careful attention should be paid not only to the relative placement of the robot as such, but to the overall proxemics and contextual factors that might further contribute to the establishment of a sense of easy connection and socio-emotional bonding in HRI.

ACKNOWLEDGMENT

This work was partially supported by the European Commission (EC) and was funded by the EU FP7 ICT-317923 project EMOTE. The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

REFERENCES

Ahlberg, J. (2001). CANDIDE-3 -- an updated parameterized face (Report No. LiTH-ISY-R-2326). Dept. of Electrical Engineering, Linköping University, Sweden.

Ba, S. O., & Odobez, J. M. (2006). Head pose tracking and focus of attention recognition algorithms in meeting rooms. In: Stiefelhagen, R., Garofolo, J.S. (eds.) CLEAR 2006. LNCS, vol. 4122, pp. 345–357. Springer, Heidelberg.

Bargh, J. A. (1988). Automatic information processing: Implications for communication and affect. In *L. Donohew, H. Sypher, & E. T. Higgins (Eds.), Communication, social cognition and affect* (pp.9–32). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc

Belpaeme, T., Baxter, P., Read, R., Wood, R., Cuayáhuitl, H., Kiefer, B., ... Humbert., R., (2012). Multimodal Child-Robot Interaction: Building Social Bonds. *Journal of Human-Robot Interaction*, *1*, 33-55.

Ben-Shakhar, G. (1985). Standardization within individuals: A simple method to neutralize individual differences in skin conductance. *Psychophysiology*, *22*, 292-299.

Boucsein, W., Fowles, D.C., Grimnes, S., Ben-Shakhar, G., Roth, W.T., & Filion, D.L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*, 1017-1034.

Bowlby, J. (1970). Disruption of affectional bonds and its effects on behavior. *Journal of Contemporary Psychotherapy*, *2*, 75-86.

Castellano, G., Paiva, A., Kappas, A., Aylett, R., Hastie, H., Barendregt, W., Nabais, F., Bull, S. (2013). Towards empathic virtual and robotic tutors. In: *Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013*. LNCS, vol. 7926, pp. 733–736. Springer, Heidelberg

Chanel, G., Rebetez, C., Bétrancourt, M., & Pun, T. (2008). Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. In *Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era* (pp. 13-17). ACM.

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence*, *16*, 555-575.

Corrigan, L.J., Basedow, C., Küster, D., Kappas, A., Peters, C., Castellano, G. (2015). Perception matters! Engagement in task orientated social robotics*, in Robot and Human Interactive Communication (RO-MAN*), 2015 24th IEEE International Symposium, KOBE, Aug. 31 2015-Sept. 4, pp.375-380.

Corrigan, L.J., Peters, C., Küster, D., Castellano, G. (2016). Engagement Perception and Generation for Social Robots and Virtual Agents, *in* Esposito, *A.,* Jain, L.C., (Eds.), *Toward Robotic Socially Believable Behaving Systems – Modelling Emotions* - Intelligent Systems Reference Library, Vol. 105 - In Print

Cruickshank, D. R., Jenkins, D. B., & Metcalf, K. K. (2009). *The act of teaching*. (5th ed.), Boston: McGraw-Hill Higher Education.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York: Harper Perennial, (5th ed.), Boston: McGraw-Hill Higher Education.

D'Mello, S., Chipman, P., & Graesser, A. C. (2007). Posture as a predictor of learner's affective engagement. *In Proceedings of the 29th annual cognitive science society* (Vol. 1, pp. 905-910). Cognitive Science Society, Austin, TX.

D'Mello, S. K., & Graesser, A. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction, 20*, 147-187.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 159–181). New York: Cambridge University Press.

Feldman, R. (2007). Parent–infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions. *Journal of Child psychology and Psychiatry*, *48*, 329-354.

Fong, T., Thorpe, C., & Baur, C. (2002). Robot as partner: Vehicle teleoperation with collaborative control. *In Proceedings from the 2002 NRL Workshop on MultiRobot Systems, Washington, D. C*

Fridlund, A. J. (1994). Human facial expression: An evolutionary view. San Diego, CA: Academic Press.

Hattie, J. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. New York, Routledge.

Hollenstein, T., & Lanteigne, D. (2014). Models and methods of emotional concordance. *Biological psychology*, *98*, 1-5.

Horiguch, Y., Sawaragi, T., & Akashi, G. (2000). Naturalistic human-robot collaboration based upon mixed-initiative interactions in teleoperating environment. *In Systems, Man, and Cybernetics, 2000 IEEE International Conference on* (Vol. 2, pp. 876-881). IEEE.

Kennedy, J., Baxter, P., & Belpaeme, T. (2014). Comparing robot embodiments in a guided discovery learning interaction with children. *International Journal of Social Robotics*, *7*, 293-308.

Kim, Y., & Mutlu, B. (2014). How social distance shapes human–robot interaction. *International Journal of Human-Computer Studies*, *72*, 783-795.

Küster, D., & Kappas, A. (2014). What could a body tell a social robot that it does not know? In A. Holzinger, S. H. Fairclough, D. Majoe, & H. P. da Silva (Eds.), In *Proceedings of the International Conference on Physiological Computing Systems* (pp. 358-367). SciTePress Digital Library. doi:10.5220/0004892503580367

Leite, I., Henriques, R., Martinho, C., & Paiva, A. (2013). Sensors in the wild: exploring electrodermal activity in child-robot interaction. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction* (pp. 41-48). IEEE Press.

Malta, L., Miyajima, C. and Takeda, K. (2008) Multimodal estimation of a driver's affective state. *In Workshop on Affective Interaction in Natural Environments* (AFFINE), ACM International Conference on Multimodal Interfaces (ICMI'08), Chania, Crete, Greece.

Mauss, I. B., & Robinson, M. D. (2009). *Measures of emotion: A review*. Cognition and emotion, 23, 209-237.

Papadopoulos, F., Dautenhahn, K., & Ho, W. C. (2012). Exploring the use of robots as social mediators in a remote human-human collaborative communication experiment. *Paladyn*, *3*, 1–10. doi:10.2478/s13230-012-0018-z

Papadopoulos, F., Dautenhahn, K., & Ho, W. C. (2013). AIBOStory–Autonomous Robots supporting Interactive, Collaborative Story-telling. *Paladyn, Journal of Behavioral Robotics, 4*, 10-22. Chicago.

Picard, W., & Healey, J. A. (2000). Wearable and automotive systems for affect recognition from physiology*, MIT, Tech. Rep.*

Picard, R. W., Fedor, S., & Ayzenberg, Y. (2016). Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emotion Review*, *8*, 62-75.

Prendinger, H., Mayer, S., Mori, J., & Ishizuka, M. (2003). Persona effect revisited: Using bio-signals to measure and reflect the impact of character-based interfaces. In *Proceedings of the 4th International Working Conference on Intelligent Virtual Agents*, (IVA-03), pages 283–291, Kloster Irsee, Germany

Riedmiller, M., & Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The RPROP algorithm. *Proceedings of the 1993 IEEE International Conference on Neural Networks (ICNN 93*), vol. 1, San Francisco, pp. 586–591.

Robison, J. L., Mcquiggan, S. W. and Lester, J. C. (2009). Modeling Task-Based vs. Affect-based Feedback Behavior in Pedagogical Agents: An Inductive Approach, In *Proceedings of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling*, Amsterdam, The Netherlands, The Netherlands, pp. 25–32.

Sidner, C. L., Kidd, C. D., Lee, C., & Lesh, N. (2004). Where to look: a study of human-robot engagement. In *Proceedings of the 9th international conference on Intelligent user interfaces* (pp. 78-84). ACM.

Tassinary, L. G., Cacioppo, J. T. and Vanman, E. J. (2007). The skeletomotor system: Surface electromyography. In J. T. Cacioppo, L. G. Tassinary and G. G. Berntson (Ed.), *Handbook of Psychophysiology* 3rd ed. (pp. 267-299). New York: Cambridge University Press.

Wall, A. (1993). How Teacher Location in the Classroom Can Improve Students' Behavior, *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 66 (5), 299-301.

Zaga, C., Truong, K. P., Lohse, M., & Evers, V. (2014). Exploring child-robot engagement in a collaborative task. In: *Proceedings of the Child-Robot Interaction Workshop: Social Bonding, Learning and Ethics*, 17 Jun 2014, Aarhus, Denmark. Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa (INESC-ID).