”Oh! I am so sorry!”: Understanding User Physiological Variation while Spoiling a Game Task

Roxana Agrigoroaie, Arturo Cruz Maya, Adriana Tapus

”Oh! I am so sorry!”: Understanding User Physiological Variation while Spoiling a Game Task

Roxana Agrigoroaie, Arturo Cruz-Maya, and Adriana Tapus

Abstract— **This paper investigates how individuals react in a situation when an experimenter (human or robot) either tells them to stop in the middle of playing the Jenga game, or accidentally bumps into a table and makes the tower fall down. The mood of the participants and different physiological parameters (i.e., galvanic skin response (GSR) and facial temperature variation) are extracted and analysed based on the condition, experimenter, and psychological questionnaires (i.e., TEQ, TEIQ, RST-PQ). This study was a between participants study with 23 participants. Our results show that multiple GSR parameters (e.g., latency, amplitude, number of peaks) differ significantly based on the condition and the experimenter the participants interacted with. The temperature variation in three regions of interest (i.e., forehead, left, and right periorbital regions) are good indicators of how ready an individual is to react in an unforeseen situation.**

I. INTRODUCTION

In their everyday lives humans interact with each other. They might notice the others around them, or they may choose to ignore them. The way that an individual reacts to another individuals actions defines the concept of empathy [1]. Empathy is an important construct especially in social interactions. The reactions can be of mostly two types: cognitive or emotional [1].

Can the concept of empathy be used in the interaction between a human and a robot? One possible answer to this question was given by the authors of [2]. They state that eventhough a machine cannot feel empathy, it could display a behavior that is empathetic. In order to accomplish this, the authors of [2] state that ”a robotic system should be capable of recognizing the user’s emotional state, communicating with people, displaying emotion”.

A literature search shows that there are multiple studies that investigate the role of empathy while an individual interacts with a robot [3], [4]. In [3] the authors have used a friendship questionnaire to assess the relationship between the empathy level displayed by a robot and the perceived level of friendship towards a robot. The authors of [4], have shown that a robot that displays an empathetic behavior facilitates the interaction with an individual.

In this paper, we are focusing on the understanding of the emotional state of an individual when it is interacting with a robot around a task. A literature research shows that there are multiple physiological parameters that can be used to determine an individual’s emotional state (e.g., heart rate, facial temperature variation [5], galvanic skin response (GSR) [6], blinking). Our research focuses on two of these parameters: GSR, and the facial temperature variation.

The GSR has been used for many years in the research of psychophysiology [7]. It was shown that it can be successfully used to determine the arousal level [8], the cognitive load [9], the emotional state of an individual [6], to differentiate between stress and cognitive load [10]. In the current study, the GSR is used to determine the arousal level of the participants.

Research has shown that the emotional state of an individual has an effect on different physiological parameters [11], [5]. In [5], the authors have associated the emotional state of an individual with the variation of the facial temperature in different regions of interest. The six most important regions of interest for determining the emotional state of an individual are: the nose, cheeks, periorbital region, chin, maxillary area, and the forehead [12].

In the current study, it is investigated how individuals react in an unforeseen situation. More specifically, while doing a certain task (i.e., playing the Jenga game), the participants are abruptly interrupted either by knocking over their tower, or by telling them that they have to stop. Of interest for this research, is how different physiological parameters vary (i.e., GSR, facial temperature variation) based on who the participants are interacting with (either a robot or a human), if their tower is knocked over or not, their empathy level, their emotional intelligence, and their personality.

For the personality, out of the multiple personality theories found in the literature (e.g., Eysenck Personality Theory [13], Big 5 model [14], Reinforcement Sensitivity Theory (RST) [15]), RST has been chosen. This model was chosen, as it relates the personality traits to physiological parameters. Moreover, the model is based on a system that is responsible for how individuals react in unforeseen situations (i.e., the Fight Flight Freeze System).

Regarding the emotional intelligence, the model proposed by Petrides [16] was chosen, as the model considers the emotional intelligence as a personality trait. The empathy is measured by using the Toronto Empathy Questionnaire [17], which was shown to be reliable, and it considers empathy as an emotional process.

The paper is structured as follows: Section II presents the experimental design. Section III presents the methodology of how the data was extracted and analysed. The results are summarized in Section IV. A short discussion of the results is provided in Section V. While the conclusions and a perspective on future works are part of Section VI.

*This work was funded and conducted in the context of the EU H2020 ENRICHME projet, Grant Agreement No:643691

The authors are at the Autonomous Systems and Robotics Laboratory, U2IS, ENSTA-ParisTech, Université Paris-Saclay, Palaiseau, France
firstname.lastname@ensta-paristech.fr

II. EXPERIMENTAL DESIGN

A. Robotic platform and sensors

For this experiment the Pepper robot, developed by SoftBank Robotics (former Aldebaran), was used. Data from two sensors was recorded: an Optris PI640 thermal camera and a Grove[1] GSR sensor. The thermal camera was placed in front of the participants, while the GSR sensor was placed on the ring and middle finger of the left hand of each participant.

B. Questionnaires

All participants had to fill the following questionnaires:

Toronto Empathy Questionnaire (TEQ) [17] is an instrument to measure the empathy level. By empathy, it is meant the consequences of accurately perceiving how another individual is feeling. The more empathic an individual is, the better it understands what the person it interacts with is feeling, and the more appropriately it can react.

Trait Emotional Intelligence Questionnaire (TEIQ) [16] measures the emotional intelligence of an individual and it considers it as a personality trait. The model is based on 4 traits: emotionality, sociability, well-being, and self-control. For example, an individual with high self-control is capable of better controlling its impulses than an individual with low self-control.

Reinforcement Sensitivity Theory-Personality Questionnaire (RST-PQ) [18] measures the personality traits as defined by RST [15]. The theory proposes that the personality traits are based on three neurobehavioral systems: the Fight Flight Freeze System (FFFS), the Behavior Activation System, and the Behavior Inhibition System. The FFFS is responsible with how individuals react towards aversive stimuli.

Positive And Negative Affect Schedule (PANAS) [19] was developed to measure both positive and negative affect. It consists of 20 items, each measured on a scale from 1 ("Not at all") to 5 ("Very much"). High positive affect is characterized by high energy, full concentration, while negative affect is characterized by unpleasurable engagements.

Post-questionnaire a custom designed questionnaire was given to the participants in order to assess how they perceived the experimenter (e.g., friendly, motivating, empathetic), the task (i.e., stressful, difficult) and if the breathing exercise helped them relax or not. Each question was measured on a scale from 1 ("Strongly disagree") to 5 ("Strongly agree").

For all traits of the TEQ, TEIQ, and RST-PQ, the scores are considered either low or high, by using a threshold equal to the median of the participants that took part in this study.

C. Participants

A total of 23 participants (4 female and 19 male, mean age of 27.48, SD = 5.76) agreed to take part in this experiment. Most of the participants have a technical background (21 out of 23), one has social and cognitive sciences background, and one has non technical background. Before the experiment, each participant's knowledge of robotics was measured on a scale from 1 ("Not at all") to 5 ("Very much") (see Table I).

[1]http://wiki.seeed.cc/Grove-GSR_Sensor

TABLE I: Participants distribution based on questionnaires

RST-PQ

Category	FFFS	BIS	BAS RI	BAS RR	BAS GDP	BAS I
low	12	13	12	12	12	13
high	11	10	10	11	11	10

TEIQ

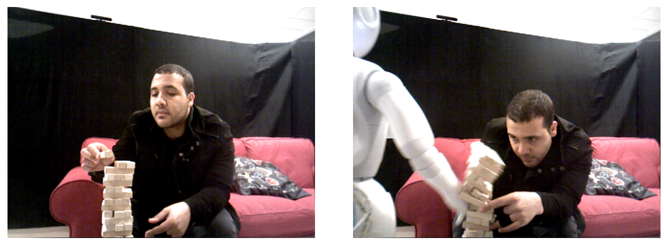
Category	Well being	Sociability	Emotionality	Self Control
low	15	13	14	13
high	8	10	9	10

TEQ

Category	Empathy
low	13
high	10

Robotics knowledge

Not at all	A little	Somewhat	Much	Very much
1	4	6	5	7



(a) Participant placing a Jenga piece

(b) Experimenter knocks over the tower

Fig. 1: Jenga game

Table I summarizes the results of the participants to the questionnaires presented in Section II-B. Based on these results, the following questionnaire results were considered for further analysis: all traits of RST, empathy, sociability and self control (from TEIQ).

D. Scenario

For this experiment, the participants had to play the Jenga game (see Figure 1a). Given a tower of 54 pieces of wooden blocks (3 blocks per layer), the purpose of the game is for the user to extract a block from any of the layers of the tower and place it on the top of the tower. The game end when the tower falls. The participants sat at a table, where the tower was already built. They were instructed to play the game and to do their best to build the tallest tower. Each participant interacted with either a robot or a human experimenter. During the game, the experimenter would periodically give encouragements and cheer the participant.

Once seated at the table, and before the experiment started, there was a 5 minute relaxation period, in which each participant had to follow a breathing exercise. Once the relaxation was done, the experimenter explained to the participants what they had to do. After the instructions, each participant filled the PANAS questionnaire and the interaction with

TABLE II: Participants distribution based on the conditions

Condition	Low empathy	High empathy
C1	5	2
C2	3	3
C3	2	2
C4	3	3

the experimenter started. First, there was a short dialogue between the experimenter and the participant. The purpose of this, was so that the participant familiarizes itself with the experimenter. Both the human and the robot experimenters had the same behavior, which is explained in Section II-D.1. After approximately two minutes of playing the game, the experimenter approaches the participant. In some of the cases it will bump into the table and the tower falls (see Figure 1b), while in some other cases it just approaches the participant and it informs him/her that the game is over. When the experiment was finished, each participant had to fill a new PANAS questionnaire.

In our experiment four conditions were developed. More specifically:

Condition C1: Interaction with Robot experimenter with a strong bump that makes the tower fall

Condition C2: Interaction with Robot experimenter with a bump not strong enough to make the tower fall

Condition C3: Interaction with Human experimenter with a strong bump that makes the tower fall

Condition C4: Interaction with Human experimenter with a bump not strong enough to make the tower fall

The participants were assigned a certain condition based on their empathy levels. In the end, the distribution shown in Table II was obtained.

Next, the behavior of the robot is going to be presented.

1) Robot behavior: For the dialog between the robot and the participant the Text To Speech (TTS) and the Automatic Speech Recognition (ASR), provided by the NaoQI Framework were used. The TTS was used to generate the speech of the robot, while the ASR was utilized to recognize the instruction of the participants to "start" the game. The body gesture of the robot was designed using Choregraphe.

The robot started the interaction having a pre-programmed speech with waiting times between phrases. The waiting times used are: "Hello", wait (1 sec), "now you can proceed to complete the questionnaire on the table", wait (3 secs), "When you finish it please let me know saying **start**", wait (until participant said "start"), "How are you?", wait (6 secs), "I'm fine, my name is Pepper, and I am here to stand by you during the game", wait (6 secs), "What is your name?", wait (8 secs), "It is very nice to meet you, what do you think about our game Jenga?", wait (7 secs), "Ok. This game is funny, and even more when somebody is motivating you. I hope you do very well. You can start now", wait (7 secs).

Every twenty seconds, the robot moved autonomously backwards and forwards (it started by first moving backwards) covering a distance of 50 cm. Then, after ten seconds

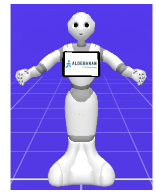


Fig. 2: Robot posture when cheering up the participants and knocking down the tower

TABLE III: Angles of arm joints of the robot when knocking down the tower

	Elbow Yaw	Hand	Shoulder Pitch	Shoulder Roll	Wrist Yaw
Right	70.5°	0.61°	36.8°	-40.5°	43.2°
Left	-70.5°	0.61°	36.8°	40.5°	-43.2°

it said a phrase to cheer up the participant while opening the arms (as shown in Fig. 2). The angles of the arm joints of the robot used for the postures are shown in Table III. Some examples of the encouraging phrases said by the robot are: "It seems you are a good player", "That was a great movement", "You are doing it very well".

Once 110 seconds past since the participant said "start", if in condition C1 (making the tower fall), the robot moved forward with open arms while saying "look, you are awesome". The robot continued moving until bumping into the table with its base and knocking down the tower with its arms (see Figure 1b). Then, the robot said "Oh! I am so sorry!", wait (5 secs), "Ok. I think you cannot keep playing this game, I'm sorry", wait (10 secs), "Please fill the questionnaire on the table". In condition C2, the robot just informed the participant that the time was over, and asked him/her to fill the PANAS questionnaire that was on the table.

For conditions C3 and C4, with the human experimenter, the experimenter behaved in the same way as the robot. It followed the same steps and questions with the dialog, and it encouraged the participant every 20 seconds. When the 110 seconds were elapsed, the experimenter either bump into the table, or not, depending on the condition.

E. Hypotheses

Based on the information presented so far the following hypotheses were developed:

H1. Individuals with a low empathy level will have longer recovery times for the event based analysis than individuals with high empathy level.

H2. Participants in conditions C1 and C3 (where the tower fell) will show a greater physiological and mood change response, than participants in the conditions C2 and C4 (where the tower does not fall) (as measured by GSR event based analysis and PANAS questionnaire).

III. Data extraction and analysis

In this section, the methods used to extract and analyze the data are described.

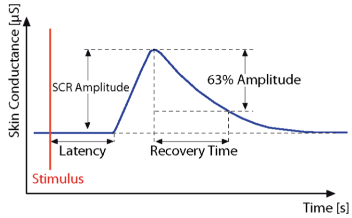


Fig. 3: Ideal GSR signal with the computed features [10]

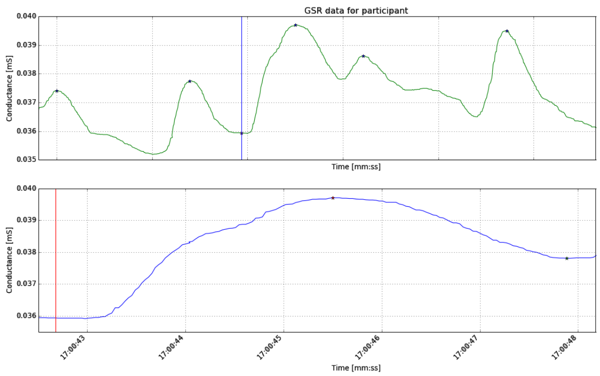


Fig. 4: Typical data for event based analysis. The vertical lines represent the bump events.

A. GSR

There are two types of analysis that can be performed on the GSR data: event based analysis and analysis on the entire interaction. For the entire interaction analysis two parameters were extracted: accumulative GSR (AccGSR) [9], and the total number of peaks. For the event based analysis there are three parameters of interest [10]: latency time, amplitude, and recovery time (see Figure 3). For this study, the event considered is the moment when the experimenter (either robot or human) bumped into the table (for conditions C1 and C3) and when the experimenter informed the participants that the game is over (for conditions C2 and C4). The latency time was computed as the total time (in seconds) it took for the signal to increase with at least 5% compared to the level at the bump time. The amplitude represents the difference between the maximum value and the level at the bump time. The recovery time (in seconds) is computed as the time difference between the time when the signal reaches a level of 63% of the amplitude and the time when the maximum value is reached.

As the output of the GSR sensor that was used is the resistence of the skin, first the data had to be converted to conductance. Next, the algorithm presented in [9] was applied to extract the AccGSR for the relaxation period, and the entire interaction. First, each participant's signal had to be normalized, by dividing the signal during one interaction by the mean value of all interactions of the participant (Eq (1) of [9]). The AccGSR was extracted from the normalized signal (Eq (2) in [9]). For the peaks, only the peaks that were at least 2% of the total range of values were extracted.

A typical GSR signal is shown in Fig. 4. The upper part of the figure presents the variation of the GSR signal and the detected peaks, while in the lower part of the figure is shown the GSR signal corresponding to the bump event. In both figures, the vertical line, represents the time at which the bump occurred.

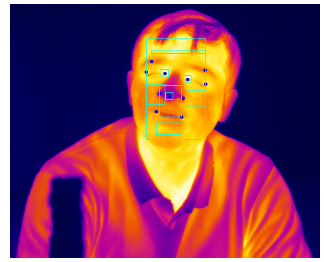


Fig. 5: Example of ROIs

B. Facial Temperature variation

The temperature variation across different regions of interest (ROI) provides good insight into the current internal state of an individual. Therefore, the thermal data was used to extract these temperature variations.

No open source face detector has been found for thermal images. As a result, a facial detector for thermal images was trained [20]. To detect the ROIs, a feature point detector was trained for 11 feature points [20]. The 11 feature points of interest (i.e., the middle of the eyebrows, the inner and outer corners of the eyes, the corners and the tip of the nose, and the corners of the mouth) were selected as these are sufficient to define nine ROIs on the face (see Fig. 5): the entire face, the forehead, the left and right periorbital regions, the nose, the left and right cheek, the chin, and the perinasal region.

The ROIs were defined based on the distance between the inner corners of the eyes (eyes_dist) as follows (see Fig. 5):

1) the forehead region: width equal to the distance between the middle of the eyebrows; and the height equal to the distance between the eyes and the nose.
2) the left, and right periorbital regions: both regions were defined as square regions around the inner corners of the eyes with the side equal to 1/3 of eyes_dist.
3) the nose: a square region around the tip of the nose, with the side equal to 1/3 of eyes_dist
4) the perinasal region: width equal to the distance between the corners of the mouth; and a height equal to the distance between one corner of the nose and the mouth at which a distance of 1/3 of eyes_dist was added in order to include the nostrils too.
5) the chin region: width equal to the distance between the corners of the mouth, and the height equal to eyes_dist
6) the left, and right cheek regions: rectangular regions with the width equal to the length between the corners of the eyes, and the height equal to the distance between the midpoint of the corner of the eyes and nose, and the corner of the mouth and nose

Once the ROIs were defined, the mean temperature could be extracted together with the timestamp at which it occurred. A Butterworth low pass filter was applied on the

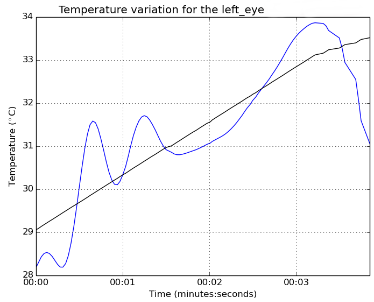


Fig. 6: Temperatature variation over time

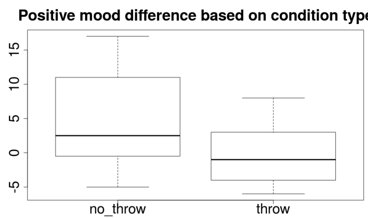


Fig. 7: Positive mood difference based on the condition type (throw/no throw)

data ($F_s = 30, order = 6, cutoff_f = 2.1$). Next, a least-square regression was applied in order to fit a linear model on the data. Figure 6 shows an example of filtered data for the left periorbital region for bump event. The result of the linear regression was overlapped on the temperature data (the model was fitted with $r^2 = 0.817$ and $p < 2.12\mathrm{e}{-28}$). In this case, the temperature increases with $0.0398°$C/s.

IV. Results

Next, the main results that were obtained are presented.

A. Panas questionnaire

As previously mentioned, the participants completed a PANAS [19] questionnaire before, and after the experiment. Positive and negative mood differences were computed between the two questionnaires. Both parameters, positive difference and negative difference, showed a normal distribution ($p = 0.06; p = 0.22$). Therefore, an ANOVA analysis could be applied on the data. Statistical analysis yielded the following significant results.

The condition type parameter approaches significance ($F(1, 21) = 4.09, p = 0.056$), with participants in the no fall conditions (C2 and C4) showing a greater positive mood difference than participants in the fall conditions (C1 and C3) (see Figure 7). As a result, **Hypothesis H2 is partially validated**.

As shown in section II-B, each of the following statements had to be rated by the participants on a scale from 1 ("Strongly disagree") to 5 ("Strongly agree"): (a) "The experimenter was polite", (b) "The experimenter was friendly", (c) "The experimenter was helpful", (d) "The experimenter was motivating".

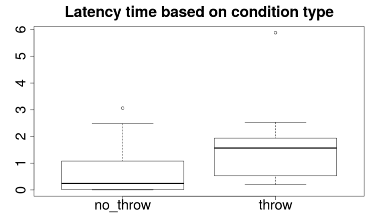


Fig. 8: Latency based on the condition type (throw/no throw)

Considering statement (a), participants that gave a rating of 5 showed a significantly lowered negative mood after the experiment compared to the participants that gave a rating of 4 ($p = 0.0043$). The same result was found for statement (c) ($p = 0.012$).

Regarding statement (b), participants that rated the experimenter with a 5 showed a significantly lower score for the negative mood after the experiment than the participants that rated the experimenter with a 4 ($p = 0.029$), or with a 3 ($p = 0.032$).

For statement (d), participants that gave a rating of 5 felt significantly better after the experiment than the participants that neither agreed nor disagreed with the statement ($p = 0.0009$). A significant interaction between how the participants rated statement (d) and condition was found ($F(5, 11) = 5.075, p = 0.01$). Furthermore, the same significant interaction was found between how the participants rated statement (d) and the experimenter ($F(3, 15) = 4.106, p = 0.0259$).

Regarding the self control scores of the participants, a significant result was found when using the negative difference as dependent variable ($F(1, 21) = 6.323, p = 0.02$). The participants with high self control scores showed a lower negative mood score than participants with low self control. On the other hand, participants with high sociability scores showed a significantly higher positive mood difference than participants with low sociability scores ($F(1, 21) = 9.436, p = 0.0057$). The same relationship was found between positive mood difference and RST-FFFS ($F(1, 21) = 9.514, p = 0.0056$).

B. GSR

The GSR analysis was performed on the latency time, amplitude, and recovery time, for the event based analysis, and the AccGSR and the number of peaks for the relaxation period, and the entire interaction.

A Shapiro-Wilk normality test showed that the latency was not normally distributed ($p = 0.0002$). Therefore, a Kruskal-Wallis test was applied on the latency variable by using the different variables as factors. A result that approaches significance (see Figure 8) was found for the condition type (throw / no throw) ($\chi^2 = 3.649, p = 0.056$). Participants in the throw conditions showed higher latency times than participants in the no throw conditions.

As the amplitude parameter shows a normal distribution, an Anova analysis was performed. The significant results

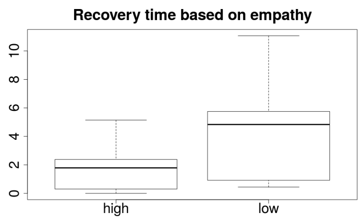


Fig. 9: Recovery time based on empathy level

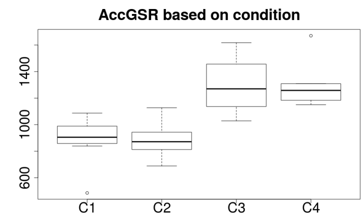


Fig. 10: Accumulative GSR based on condition

were found for the condition ($F(3,19) = 4.44, p = 0.015$) and the condition type. Significant differences were found between C1 and C2 ($p = 0.019$), and a difference that approaches significance for C2 - C3 ($p = 0.058$). In conditions C1 and C3, the participants showed a greater amplitude than the participants in C2. For the condition type, a significant difference was found between the throw and no throw conditions ($F(1,21) = 897, p = 0.0068$), with participants in the throw condition showing a significantly higher amplitude than participants in the no throw conditions.

The recovery time parameter did not show a normal distribution ($p = 0.032$), therefore a Kruskal-Wallis analysis was performed. A significant result was found for the condition (C1-C2($p = 0.049$), C1-C3($p = 0.03$), C1-C4($p = 0.014$)). Participants in condition C1 had a significantly higher recovery time than the participants in the other three conditions. Significant results were found for the condition type ($\chi^2 = 7.674, p = 0.0056$), with the participants in the throw conditions showing a significantly higher recovery time than the participants in the no throw conditions.

Considering all three event based GSR parameters (i.e., latency time, amplitude, recovery time), participants in the throw conditions showed an increased latency time, higher amplitude, and higher recovery time, than participants in the no throw conditions. Taking into account the result of the PANAS questionnaire, it can be stated that **hypothesys H2 can be validated**.

Another significant result was found for the empathy type of the participants ($\chi^2 = 3.94, p = 0.047$), with participants that showed low empathy levels having a significantly higher recovery time than the participants with high empathy scores (see Figure 9). Therefore, **hypotheses H1 can be validated**.

Regarding the AccGSR during the entire interaction, a Shapiro test showed that the data is normally distributed ($p = 0.72$), therefore an Anova analysis of variance was performed. Significant results were found for the condition ($F(3,19) = 9.05, p = 0.0006$), and the experimenter ($F(1,21) = 29.99, p = 1.96e-05$). The significant differences between conditions (see also Figure 10) were found for C1-C3 ($p = 0.01$), C1-C4 ($p = 0.0047$), C2-C3 (0.01), and C2-C4 ($p = 0.006$). Participants in conditions C3 and C4 showed significantly higher AccGSR than participants in conditions C1 and C2. Participants that interacted with a human experimenter (C3, C4) showed a significantly higher AccGSR than the participants that interacted with the robot experimenter (C1, C2).

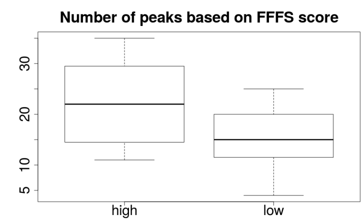


Fig. 11: Number of peaks based on RST-FFFS score

The peaks extracted from the entire interaction also displayed a normal distribution ($p = 0.48$). The Anova analysis showed the same significant results as with the AccGSR: condition ($F(3,19) = 3.123, p = 0.05$), and experimenter ($F(1,21) = 5.28, p = 0.03$). The participants interacting with the human experimenter had a significantly higher number of peaks than participants interacting with the robot experimenter. Another significant result was found for RST-FFFS ($F(1,21) = 4.67, p = 0.042$); participants with high scores displayed a significantly greater number of peaks than participants with low scores (see Figure 11).

No significant results were found for the GSR parameters extracted during the relaxation period.

C. *Facial Temperature variation*

For the temperature variation, the rate of change of the temperature variation was analysed. The temperatures were extracted a few seconds before the event and up to 10 seconds after the event.

When considering the RST-FFFS as factor, the following three regions of interest showed significant results: the forehead ($\chi^2 = 4.9, p = 0.02$), the left periorbital region ($F(1,21) = 4.42, p = 0.047$), and the right periorbital region ($\chi^2 = 7.0, p = 0.008$). For all three regions, individuals with high RST-FFFS scores had a significantly higher slope than participants with low scores. Furthermore, participants with high scores showed a positive slope, while participants with low scores showed a negative slope.

For the self control factor, significant results were found for the right periorbital region ($\chi^2 = 7.78, p = 0.005$) , for the nose region ($F(1,21) = 4.388, p = 0.048$), the perinasal region ($\chi^2 = 4.44, p = 0.03$), and a result that approaches significance for the left periorbital region ($F(1,21) = 4.25, p = 0.0517$). In all four regions, individuals with high self control scores had significantly higher slopes than participants with low scores.

V. DISCUSSION

In this study the main emphasis was put on the GSR physiological parameter, as it was proved in the literature to be a good indicator of an individual's arousal level [8], or of its emotional state [6]. For the GSR data significant results were found both for the event based analysis and for the entire interaction. Significant differences were found between conditions, condition types (either throwing or not throwing the tower). As the latency, amplitude and recovery were extracted for a specific event, it should not be very surprising that no results were found for any of the psychological questionnaires.

When looking at the entire interaction, significant results were found for FFFS (for the number of peaks). This is an indicator that individuals with high FFFS scores are more aroused than individuals with low scores. Moreover, participants that interacted with the human experimenter were more aroused than the individuals interacting with the robot (both for the number of peaks, and for the AccGSR).

As shown in Section II-B, the FFFS is responsible with how individuals react in "get me out of here" situations. As a result, it is expected of individuals with high scores to be better prepared to flee from a situation than an individual with low scores. Therefore, as shown in [12], this situation should be characterized by a greater electrodermal activity. When looking at the temperature variation, significant results were found between how the temperature varies in the forehead, and the periorbital regions and the FFFS. According to [12], individuals with high scores should have a lower increase in temperature than individuals with low scores. Considering that the temperature was extracted a little before and after the event, our results can be explained by the fact that individuals with high scores were ready to fight or flee (therefore they showed an increase in temperature), while individuals with low scores froze (they showed a decrease in temperature).

The main limitations of our work consists in the relatively small number of participants (i.e., 23 participants) that took part in our study. Furthermore, more investigation is needed in order to confirm the universality of our results.

VI. CONCLUSION

In this paper, a between participants study was presented that was carried out with 23 participants, with the purpose of finding out if the empathy level of an individual would influence how it would react in a situation where an experimenter accidentally knocks over a Jenga tower. Four conditions have been developed, two with a robot experimenter and two with a human experimenter, in which the experimenter (either robot or human) either knocked over the tower of the participant or not. Different psychological questionnaires were used to measure the participants empathy level, emotional intelligence, and personality (i.e., TEQ, TEIQ, RST-PQ). Also, different physiological parameters (GSR, temperature variation) were extracted to better understand the reactions of the participants. **The results confirmed all our hypotheses**.

Most of our results are based on the GSR parameters. We found evidence that the event based analysis parameters (i.e., latency time, amplitude, recovery time) are dependent on the condition, and if the experimenter makes the tower fall or not. Furthermore, the AccGSR and the peaks are significantly different depending on who the participants are interacting with. Moreover, how the temperature varies in three regions of interest across the face (i.e., forehead, left, and right periorbital regions) is a good indicators of how ready an individual is to react in an unforeseen situation.

Some of our future work include the extraction of other physiological parameters from the recorded data (e.g., heart rate, respiration rate) and to analyse the variation of these parameters based on the condition and the experimenter. Furthermore, the temperature should be extracted from the entire interaction in order to show if there is a connection between how the temperature varies and the FFFS.

REFERENCES

[1] M. H. Davis, "Measuring individual differences in empathy: Evidence for a multidimensional approach." *Journal of personality and social psychology*, 1983.

[2] A. Tapus and et.al., "Socially assistive robotics [grand challenges of robotics]," *IEEE Robotics & Automation Magazine*, 2007.

[3] A. Pereira and et.al., "Using empathy to improve human-robot relationships," in *International Conference on Human-Robot Personal Relationship*. Springer, 2010.

[4] A. Tapus and M. J. Mataric, "Socially assistive robots: The link between personality, empathy, physiological signals, and task performance." in *AAAI spring symposium: emotion, personality, and social behavior*, 2008, pp. 133–140.

[5] S. Ioannou, V. Gallese, and A. Merla, "Thermal infrared imaging in psychophysiology: Potentialities and limits," *Phychophysiology*, 2014.

[6] C. L. Lisetti and F. Nasoz, "Using noninvasive wearable computers to recognize human emotions from physiological signals," *EURASIP Journal on Advances in Signal Processing*, 2004.

[7] J. T. Cacioppo, L. G. Tassinary, and G. Berntson, *Handbook of psychophysiology*. Cambridge University Press, 2007.

[8] V. D. Nebylitsyn and J. A. Gray, *Biological bases of individual behavior*. Academic Press, 1972.

[9] N. Nourbakhsh and et.al., "Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks," in *Proceedings of the 24th Australian Computer-Human Interaction Conference*, 2012.

[10] C. Setz and et.al., "Discriminating stress from cognitive load using a wearable eda device," *IEEE Transactions on information technology in biomedicine*, 2010.

[11] R. W. Picard and et.al., "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE transactions on pattern analysis and machine intelligence*, 2001.

[12] S. Ioannou, P. Morris, H. Mercer, M. Baker, V. Gallese, and V. Reddy, "Proximity and gaze influences facial temperature: a thermal infrared imaging study," *Frontiers in psychology*, 2014.

[13] H. J. Eysenck, "Dimensions of personality: 16, 5 or 3? - criteria for a taxonomic paradigm," *Personality and Individual Differences*, 1991.

[14] L. Goldberg, "An alternative 'description of personality': the big-five factor structure," in *Journal of Personality and Social Psychology*, 1990.

[15] J. Gray and N. McNaughton, *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system*, 2000.

[16] K. V. Petrides, "Psychometric properties of the trait emotional intelligence questionnaire," in *Assessing emotional intelligence*, 2009.

[17] R. N. Spreng and et.al., "The toronto empathy questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures," *Journal of personality assessment*, 2009.

[18] P. Corr and A. Cooper, "The reinforcement sensitivity theory of personality questionnaire: Development and validation." *Psychological Assessment*, 2016.

[19] D. Watson and L. Clark, "The positive and negative affect schedule: Panas questionnaire," *Journal of personality and social psychology*.

[20] R. Agrigoroaie and A. Tapus, "Physiological parameters variation based on the sensory stimuli used by a robot in a news reading task," in *Accepted In Proceedings of RO-MAN*, 2018.