Cross Validation of Emotional Features with a non-Anthropomorphic Platform

ABSTRACT

Robots should be able to express emotional states to interact with people as social agents. Emotions are a peculiar characterization of humans and usually conveyed through face expression or body language, however there are cases where robots cannot reproduce anthropomorphic shape: they have to perform tasks which require a specific structure: this is the case of home cleaners, package carriers, and many others platforms. Therefore, emotional states need to be represented by exploiting other features, such as movements and shape changes. The work presented in this paper studies emotion expression in non-anthropomorphic platform and how it is perceived by humans. The work uses results obtained from a previous experiment, which was done to assess precise values for linear velocity, angular velocity, oscillation angle, direction and orientation that could be used to express happiness, angry, fear and sadness. Results show that people can recognize some emotional expressions better than others and that additional features should be added to differentiate some emotions. Additionally, it was done a small scene to verify if people prefer emotions with or without emotion.

Keywords

Human-Robot Interaction; Case Study; Experiment Cross-validation; Emotion Projection

1. INTRODUCTION

Emotions and mental states are not just expressed through facial expression but also by body through postures and other features [1]. Many psychological studies have been focused on understanding the role of human face in emotion projection [2], [3] and mental states. This trend has been followed by the robotics community, where anthropomorphic faces (e.g., [4], [5]) and bodies (e.g., [6], [7], [8]) have been used to convey emotions. However in many situations presence of anthropomorphic elements would be out of place and not justified by the main robot's functionalities. Most of the current and future robotics platforms on the market will not require anthropomorphic faces [5] or limbs [9], for instance in floor cleaning robots, anthropomorphic characteristics could even be detrimental to robot's task accomplishment. This generates the necessity to study other mechanisms that could help to project emotions, which could give people an idea about the robots' state, and engage the user in long term relations.

It has been noticed that the amount of works studying non-anthropomorphic features in robotics (e.g., [10, 11, 12, 13]) remains still small in comparison to those exploiting anthropomorphic features. Moreover, these works do not give specific range of values for the characteristics used to expressed emotions implemented. For example Suk and collaborators [14] in their study give specific values for acceleration and curvature, but their connection to specific emotions is not explicited. Rather, they establish a relationship between their values and pleasure/arousal dimensions. This could help to select specific features to convey certain emotions, but it would be even better to know the precise range of values to be assigned to each characteristics to avoid misinterpretation with wrong emotions. These values could be used in social robotics to show emotions and as consequence increase their acceptance.

This paper is organized as follows. Section 1 introduce some previous works done in conveying emotion. Then the experiment is briefly described in section 3. Section 4 presents the platform used in the case study and the Emotional Enrichment System. Finally, sections 5 and 6 presented case study's design and results.

2. RELATED WORK

A direct consequence of the abundance of works in face elicitation in humans is the amount of works done in Human-Robot Interaction (HRI) focused also on faces. One of the most well-known expressive robots is Kismet [5], a robotic face able to interact with people and to show emotions. The face had enough degrees of freedom to portray the basic emotions suggested by Ekman [2] (*happiness*, *surprise*, *anger*, *disgust*, *fear*, and *sadness*) plus *interest*. Despite the complex system behind Kismet, the emotion's projection evaluation was done using videos with a very limited number of participants. In this experiment, each participant was asked to look at seven videos, one for each implemented emotion. After a video was projected, each participant selected a emotion from a list, which contained seven implemented emotions. The results showed high percentage of recognition (over 57%).

Saerbeck and Christoph [10] analyzed the relationship be-

HRI '17 Vienna, Austria

DOI: 10.1145/1235

tween robots' motion characteristics and perceived affect. They first did a literature review to select characteristics that could be used to show affection. From their review, they decided to use acceleration and curvature on robot's trajectory. Using these two characteristics, they did an experiment with eighteen participants. They used two different platforms (iCat and Roomba) to verify if the embodiment had any impact on affection determination. Each participant was exposed to both embodiments and all possible variable combinations. To assess participants' perception, they used PANAS [15] and Self-Assessment Manikin (SAM) [16]. Their results show that there was no significant difference between the two embodiments used and that participants were able to assign different emotions to the movement patterns shown to them. More importantly, they found out that acceleration is correlated with the perceived arousal but not with valence.

Cañamero and collaborators [6, 17] studied the perception of key body poses designed to show emotions using a NAO [18]. They suggested that techniques used to convey emotions in virtual characters could not be used in robots, due to the fact that virtual characters have no physical constraints, while robots are constrained by their physical capabilities [19, 6]. They proposed a set of poses that could be used to express emotions with a NAO, paying a particular attention to head's contribution. Their results showed a recognition rate of 88%, 85%, 92%, 88%, 73%, and 73% for anger, sadness, fear, pride, happiness and excitement respectively. The major finding from their experiments is that moving the head up improved the identification of pride, happiness, and excitement. While moving the head down improved the identification of anger and sadness.

Daryl [4] is an anthropomorphic robot, without facial expression, nor limbs, but with mobility capabilities. This platform was used to test whether it is possible to project emotions without using cues used in a human-like platform (e.g., tail, ears and head). This robot has head, ears, ability to generate colors in a RGB-LED set positioned in its chest, and a speaker system. The emotions implemented were: happiness, sadness, fear, curiosity, disappointment, and embarrassment. During the experiments done by the researchers, special attention was put to the final distance between the robot and the subject. The subjects were exposed to a sequence of movements. For each sequence the subject was asked to rate the intensity of each emotion enlisted in a five-point Likert scale questionnaire. The enlisted emotions were the six implemented plus anger, disgust and surprise. Their results showed that participants gave a higher intensity to the desired emotion, also when similar ones emotions were presented, such as sadness and disappointment.

Using the platform WABIAN-2R, Destephe and collaborators [20] studied the attribution of emotion to a robot's gait. To obtain robot's movements, the researchers asked two professional actors to walk in a room conveying anger, happiness, sadness, and fear with different intensities (low, regular, normal, high, and exaggerated). All the actors' walking were recorded using a Cortex motion capture system. These data were later reported to robot's embodiment. Each subject was exposed to a series of videos. After each video, the subject was asked to select one emotion from a list and state the intensity of the selected emotion. The videos showed were not made with the real robot, but with a virtual model of the robot. Their results showed that sadness was recognized 73.81% (average) of the times with an intensity of 21.43%, happiness was recognized 66.67% (average) of the times with an intensity of 30.95%, anger was recognized 61.9% of the times with an intensity of 26.19%, and fear was recognized 83.33% (average) of the times with an intensity of 28.58%. These results show that people could reasonably well recognize emotions from the robot's gait.

Brown and Howard focused their interest on head and arm movements using a DARwIn-OP platform [21]. Their hypothesis was that using some principles, that they determined as important, it is possible to express happiness and sadness in a way that people could recognize. These principles establish that to show happiness is necessary to move robot's head up, arms up, and it should be done fast. While to express sadness the head goes down, arms down, and it should be done slow. To test their hypothesis they conducted an experiment with thirteen participants. Each participant was exposed to fifteen sequences of poses and he/she had to answer for each sequence a five-point Likert scale (very-sad to very happy, with neutral in the middle). They obtained a 95%, 59%, and 94% of accuracy for happiness, neutral and sadness respectively.

Sharma and collaborators [12] used a quadrotor to study how different Laban's effort [22] parameters could impact on the perception of affection. A professional Laban certified actor was asked to generate 16 different paths, for each one changing one of the four Laban's parameters (space, weight, time, and flow). Each generated path was recorded using the Vicon motion-tracking system. Then, they did an experiment were people were asked to assess each path using the Self-assessment Manikin (SAM) [16]. In order to analyse the results, they mapped them on the circumplex model of emotion, and evaluated the contribution of each Laban's parameter in the 2-D circumplex model (arousal and valence). The results show that it is possible to increase both valence and arousal by using a more indirect space, or by performing the motion more quickly, and to decrease valence or arousal by a more direct use of space, or a more sustained motion. Although they suggest the use of Laban's description as a tool to specify affection movements in Human-Robot Interaction, how to use these parameters in the actual implementation remains an open question, since Laban defined them qualitatively, with reference to human people, and they leave very open questions about the most appropriate numerical values.

Suk and e.t. studied the human emotional interpretation for speed, smoothness, granularity and volume of a non-human or animal like object [14]. Two experiments were designed to determine the relationship among these features and the emotional interpretation. To assess participants' emotional response they used SAM. The first experiment was focused on speed and smoothness movement features, selecting five different values for each one of these two features. Each participant was exposed to twenty five movements. After a participant observed a movement, the participant marked the two SAM graphic figures (pleasure and arousal). The results from this first experiment show that the arousal increases as speed increases, but that there is not any clear tendency for smoothness. In the second experiment they evaluated the other two features (granularity and volume) using the same procedure followed in the first experiment. Their results show that granularity is positively correlated with pleasure and arousal. On the other volume is negatively correlated with pleasure and positively correlated with arousal. As overall result, they found a major contribution of speed on arousal and minor contribution of granularity and volume.

To improve the coordination between humans and robots Novikova and Watss [13] proposed the use of emotion. They did studies using a own built platform called E4 to verify whether people could detect emotions from a non-human like platform. The E4 platform was constructed using a Lego Mindstorms NTX and it was based on a Phobot robot's design. In all their experiments they used six emotional expressions (scared, surprised, excited, angry, happy, sad) and neutral. The emotion list used in their experiments was balanced using the 2-D circumplex model (arousal and valence), selecting three from each quadrant for a total of twelve emotions, plus the options of "other" and "don't know". They obtained a recognition rate of 52%, 42%, 41%, 36% and 15% for surprise, fear, sadness, happiness, and anger respectively.

3. THE EXPERIMENT

An experiment was done to assess precise values for linear velocity, angular velocity, oscillation angle, direction and orientation that could be used to express happiness, angry, fear and sadness, which correspond to four of six basic emotions suggested by Ekman [23]. These features were selected after study emotion projection in robotics, humans (i.e. previous work and theatre) and previous case studies. The robotic platform used in the experiment is holonomic, which are characterized by the possibility to move in any direction without necessity to have a specific orientation, i.e., they are free to move taking any desired orientation. The experiment was performed at the university campus during the months of June and July of 2015. A total of 49 volunteers were involved: 12 female and 37 male. The average age of the participants was 25.28 with standard deviation of 2.8, with a minimum age of 20 and maximum of 32.

A Google form was used to collect participants' answers. The form included four emotions plus two mental states (i.e. excitement and tenderness) and option "other". These mental states were added to see if people could confuse the desired emotion with them. Each participant was exposed to twenty over one hundred and ninety five possible treatments. The whole process, including a brief explanation and assessment lasted from 10 to 15 minutes. This was decided because each subject was a volunteer and would not perceive any monetary remuneration, so the time dedicated to the experiment had to be kept limited. The twenty treatments were selected picking a number without replacement from 1 to 195. For each presented treatment, participants were asked to give an intensity perceived for each option in the form.

After collecting all the data, it was created a table for each treatment. Each table contained the following information mean, standard deviation, and median. ANOVA test was not possible to be used over the data because the assumption of normality is not achieved in the collected data. This was check using the Shapiro-Wilk Test. Additional to these table, a contingency table for each emotion was generated in each treatment as it is depicted in Table 1, where the intensity for the other emotions is calculated as the mean of them, including the option of "other". For all tables, including the contingency, were calculated the Krippendorff's alpha agreement [24] (α), which is a reliability coefficient to measure the agreement among different participants. Unlike other coefficients (Kappa), α is a generalization of several known reliability indices, and it applies to:

Table 1: Contingency table formula used for each emotion and treatment. Where k is the kth treatment, j is jth emotion for the kth treatment, n is the total number of participants, and $Value$ is the intensity given by a participant.

Participant	Desire Emotion	Other Emotions
1	$Value^1_{k,j}$	$\frac{\sum_{i=1}^{i<=7\wedge i\neq j}(Value^1_{k,i})}{\sum_{i=1}^{i<=6\wedge i\neq j}(1)}$
2	$Value^2_{k,j}$	$\frac{\sum_{i=1}^{i<=7\wedge i\neq j}(Value^2_{k,i})}{\sum_{i=1}^{i<=6\wedge i\neq j}(1)}$
...	...	...
n	$Value^n_{k,j}$	$\frac{\sum_{i=1}^{i<=7\wedge i\neq j}(Value^2_{k,i})}{\sum_{i=1}^{i<=6\wedge i\neq j}(1)}$

- Any number of observers.
- Any number of categories.
- Any type of data.
- Incomplete or missing data.
- Large and small sample sizes.

This calculation was done using the R package *irr*. To improve table interpretation, it was decided to just record the emotions' alpha values that had a mean greater than zero. Therefore, tables with a top ten ranking have been set up. The raking considered: (i) the mean of the respective emotions, (ii) the alpha agreement for the respective emotion, and (iii) the alpha agreement for the treatment. The decision to give more importance to emotion's alpha rather than intensity average was taken basing on the consideration that most participants agreed on their observation. From the results was possible to notice:

- Fear was the only emotion that had six over ten movements obtaining both general and specific alpha agreement over 0.41, which is the lower bound for moderate agreement [25]. It seems that people perceive as fear when the robot is looking at them and moving far from them fast.
- It seems that people attribute sadness to slow velocities with slow angular velocity and small oscillation angle. Regarding the other two features, there is not a concrete pattern that could lead to make a generalization.
- Happiness is attributed to different values of the independent variables. It seems that happiness is mainly attributed to fast angular velocities and big oscillations angles. However specific agreement among the other features is not present.
- Anger seems to be attribute to fast velocities, both angular and linear, small angle of oscillation and the robot facing the person when it is approaching them.

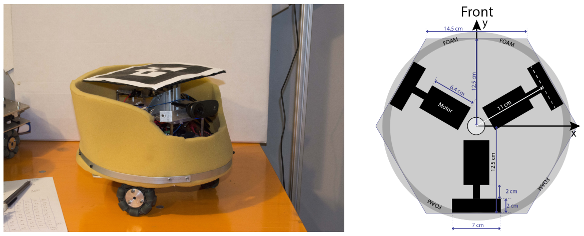

Figure 1: Platform used in the case study (left), and holonomics's blue prints (right). The arrows represent robot's frame of reference.

4. SYSTEM

The same system used in the experiment was used in the case study presented in this paper and they are explained in the following subsection.

4.1 Robotic Platform

A non-anthropomorphic robotic platform was created to do not have any bio-inspired appearance. The holonomic platform was built using Odroid U3, Arduino Due, and 3 metal gear motors with 64 CPR encoders and omniwheels. The platform could be observed in Figure 1. The Arduino Due is in charge to be the interface between the hardware (e.g. motors) and Odroid, which host all the Emotion Enrichment System.

4.2 Emotion Enrichment System

The main idea of Emotional Enrichment System is to blend an emotion and a action to generate the effect of "emotional action". However this is not far as previous works have done so far. Therefore, the system has been envisioned to also enable:

- Interoperability among different platform.
- Introduction of new parameters and emotions.
- Interface with diverse action decision system.

In other words, the system could be thought as a black box that receives a desire action, action's parameters and emotion to blend them together without paying particular attention on who or how the decision to execute a particular action was made. This is achieved through the use of messages to describe the action that should be executed, as well its parameters, emotion and emotion's intensity. These information are divided into two messages: one for the action and its parameters, and other for the emotion and its intensity. This is done to do not affect the action's execution.

Every time the system receives a new action, it verifies if the action exists and parameters corresponds to the desire action. If these two conditions are met, then the system checks if the action is compound or simple. If it is a simple, the system proceeds to add all the required actions (emotional actions) and change their parameters to convey the desired emotion. On the other hand if the action is compound, the action is first decomposed in simple actions (mandatory actions). Each one of these actions is then process to add their emotional actions. However, this addition is done taking in consideration to not add any action that was previously added. Once all the set of simple actions and their parameters are determined, the system proceeds to verify the drivers' existence to execute these actions in the desired platform. If the system finds out that one or more mandatory actions could not be executed, it triggers a message to inform that the desire action could not be executed.

4.2.1 *Simple and Compound Actions*

To achieve actions' enrichment with emotions, two different types of actions are used: simple and compound actions. Simple actions are actions that could be considered as "atomic"; these actions can be enriched with emotion changing their specific parameters. For example, consider two simple actions: speak and move body. The human action of speaking is related to changes in the vocal cords to produce different sounds, while move body is directly related to the body movement. Therefore, for the speaking action the following parameters are expected: text, pitch and tone. On the other hand, the action move body would have as parameters the desired destination, velocity, and trajectory constraints. Since each action is characterized by different parameters, parameters' modifications to obtain an emotional action will be different for each of them. On the other hand, compound action is an action created from other compound or simple actions. For example, suppose that it is required that the robot has to *recite* some dialogue and *walk* to a different positions, as doing a trajectory. A possible implementation could consist on speak simple action in parallel to two consecutive action walk, which is also a compound action that is composed by three parallel actions: balance left arm, balance right arm, and move body.

Although compound actions could be used to generate diversity of actions from simple actions, the main drawback of this approach is that all possible combination of actions should code in the system to be used. Using the example previously described, if it is decided to have three positions instead of two, then it would mean that a new compound action should be implemented to have an action in which could be specified three positions. To overcome this limitation, it was created a language to specify actions that are not implemented in the system. A computational representation of these language is used to execute the actions described in this language.

4.2.2 *Emotional Execution Tree*

The *Emotional Enrichment System* is grounded on the use of *Emotional Execution Tree* (EXT), which is a computational representation of desired actions that must be executed. EXT is a connected acyclic graph with vertices and edges. The root and non-leaf nodes could be *parallel* or *sequence* type. The parallel node could be one out of four different sub-types: action and emotion synchronous, action synchronous and emotion asynchronous, action asynchronous and emotion synchronous, or action and emotion asynchronous. Sequential nodes could just be one of two sub-types: emotion synchronous or asynchronous. Action synchronous means that each time that a parallel node receives a finish notification (success or failure), it will send a finish message to all the nodes that derived from it. On the other hand, emotion synchronous means that each time that a node (sequence or parallel) receives an emotion synchronization message, it will propagate the message to all the actions in the branches. This distinction creates the possibility to synchronize emotional changes without affecting the normal execution of the desired action. Finally, the leaf nodes could only be simple action nodes. All the nodes can assume two levels: principal or secondary. If a node is principal, it will notify its predecessor about the messages that it has received, while the secondary cannot propagate any message to its predecessor.

4.2.3 *Implementation*

Figure 2 presents the design used to achieve the desire goals. As it could be observed, the system is divided in two parts. The first part in charge to enrich a given action with a desired emotion, and the second in charge to hide the real execution of the action.

The action enrichment is achieved through the following phases:

1. *Generation of emotional execution tree:* this phase starts every time that a new action message is received. The process begins by parsing the format, verifying that the actions described on it exist in the system, and that the parameters correspond to the ones expected by each action described in the message. This parameters' verification is done on the implemented description for each simple action, which describes the parameters that are mandatory and those that are optional. When the verification is done, and all the action exists and the parameters correspond, an EXT is created.

2. *Emotion addition:* uses the EXT created in the previous phase. In this phase new simple actions are added to the EXT and the simple action's parameters are modified following emotions' descriptors, which are loaded from files. This process is broken down in two steps. First, all actions that are required to convey the desired emotion, and that are not yet present in the branch are added. Second, the emotional parameters are modulated based on the emotion's intensity and character traits.

3. *Execution:* this is the last phase and it is done after the EXT is "coloured" with emotional characteristics (actions additions and emotional parameters). The decision to have two different communication channels, one for action parameters and another for the action emotional parameters, was taken to enable the possibility to update the emotional parameters without interfering with the current execution. In this phase is maintain a reference to the mandatory and emotional action, thus when a new emotion is received the system stops the previous emotional action and start executing the new ones without affecting the general execution of the actions.

5. CASE STUDY

The case study was done at Researcher's Night 2015 with two main objectives (i) cross-validate the findings obtained from the experiment, and (ii) use the Emotional Enrichment System to verify whether the participants would prefer scenes when the robot expresses emotions or rather moves without any emotion expression. The Emotional Enrichment System was used in both cases. To reduce scene variability was used two web-cams and eight Alvar tags to informed a Kalman filter to improve robots localization in the stage. The detection of the AR tags was done through the use of the ROS package ar_track_alvar [26]. The distribution of the web-cams and the tags are depicted in Figure 3.

5.1 Emotion Description

The parameters selected for each of four emotions (*Anger*, *Happiness*, *Sadness* and *Fear*) are shown in Table 2. The main considerations to select the two implementations for each emotion were: (i) the linear velocity should be greater than 0. In other words the robot should show some linear displacement. And (ii) it should be in the top 10 list of the emotion obtained in the experiment.

Table 2: Parameters' values selected from the experiment.

Emotion	Direction (rad)	Orientation (rad)	Linear Velocity (mm/s)	Angular Velocity (rad/s)	Angle (rad)
Happiness 1	0	0	500	3	0.349
Happiness 2	0	0	900	3	0.174
Anger 1	π	0	500	3	0.087
Anger 2	0	0	900	1	0.087
Fear 1	π	π	900	2	0.174
Fear 2	π	π	500	2	0.087
Sadness 1	π	0	200	1	0.349
Sadness 1	0	π	200	1	0.349

5.2 Scene

Stage discretization was used to give zones of movements instead of absolute positions. This idea was brought from human theatrical actors, who prepared their movements based on zones on the stage [27]. This allows them to adapt their position based on other actors. Therefore, the stage was discretized in 9x9 matrix as is shown in Figure 4. Robot's movements are given in terms of the matrix positions to the Emotional Enrichment System. This allows the adaptation to different stage dimensions because robot's final position is calculated by the Emotional Enrichment System during execution, which takes under consideration the stage dimensions given. For instance, during the scene's preparation in the laboratory the stage was 3 meters per 3 meters, but in the final presentation the stage was 2.5 meters per 2.5 meters.

The scene's description is the following: the robot starts in the middle of the stage to move to the upstage right (See [28]), close to the right wing. Then, the robot moves to upstage right center and rotates by $\pi/2$ left (See [29]). Next the robot moves to the right center to then go to the center. When it arrives there, it turns full back and move backwards to downstage center with a full front orientation. There, it turns full back to move to center. Finally the robot turns to profile right and it does a step back; then it goes to the upstage center and then upstage right. The sequence of movements programmed to the robot are depicted in Figure 5.

The relation between emotion and movement is as follow: movements one, two, three, four and five are expressed without any emotion. Movements six, seven, eight, nine and ten show fear. Movement eleven depicts happiness, and the remaining movements depict sadness. The two scenes are

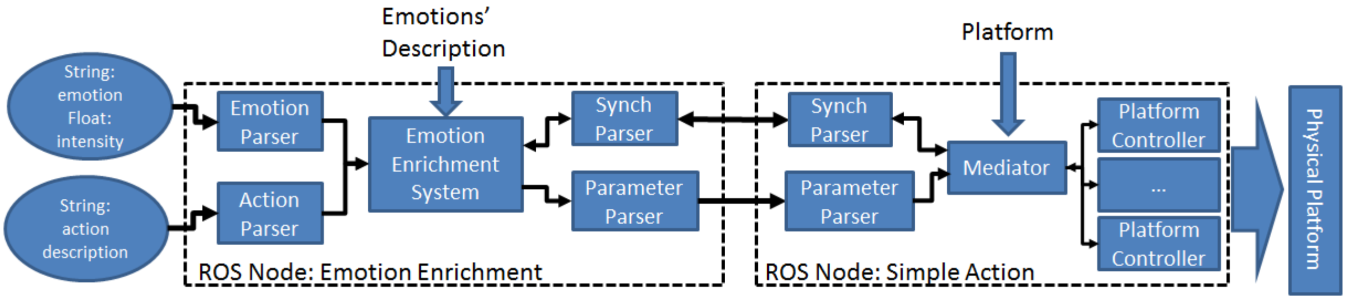

Figure 2: General system design. Each simple action corresponds to one ROS node, and there is just one node for the emotion enrichment system. The ovals represent the ROS topic parameters, rectangles represent *black boxes*, and texts outside containers represent input files that contain the system parametrization.

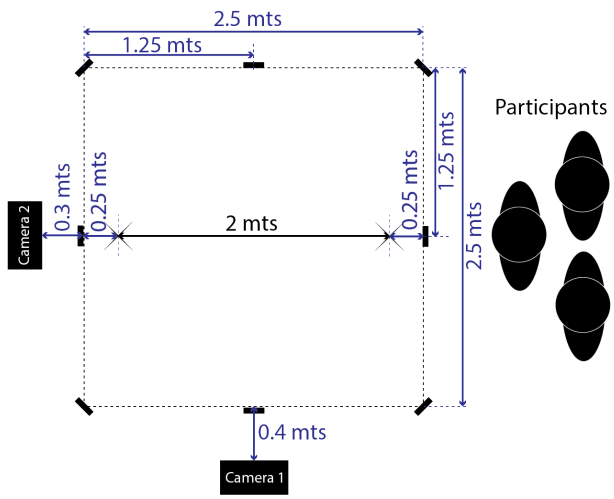

Figure 3: Environment setup for the case study.

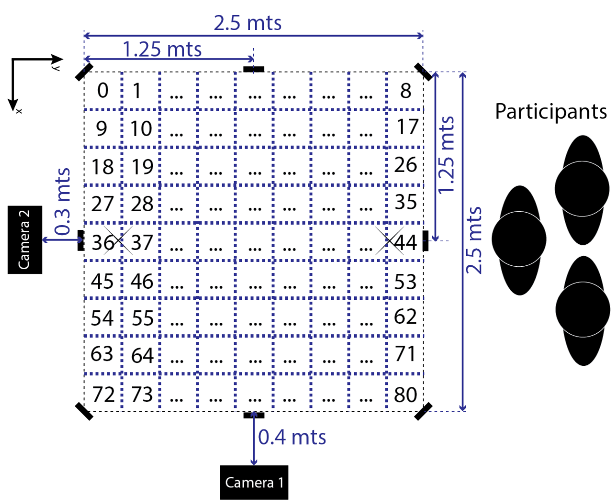

Figure 4: Stage discretization used for the small scene. The blue squares correspond to the each zone, while the numbers correspond to the ID given to each zone.

executed by the Emotional Enrichment System using the same "script" and the emotion selection is done manually via graphical interface (Figure 6).

5.3 Study

This case study was done during Researchers' Night, 2015. During a period of two days, people were asked to participate to this study. Each subject was exposed to two rounds, in each one the robot was performing a different emotion. And they were also exposed twice to a small scene, one with emotion and other without emotions. The emotions showed in each trial and the order of the scenes (with or without emotion) were generated randomly beforehand. The total number of volunteers was 256: 128 males, 126 females, and 2 that chose not to specify their gender. The average age was 27.29 years, with standard deviation of 16.58, minimum age was 4 and maximum 76.

6. RESULTS

Table 3 summarizes results obtained during the case study.

Table 3: Summary of the answers obtained in the case study.

Presented/Reported	Happiness	Anger	Fear	Sadness	Excitement	Tenderness	Other	Total
Happiness 1	8	16	7	4	16	4	7	62
Happiness 2	11	11	6	2	19	3	1	53
Anger 1	7	5	6	2	21	7	1	49
Anger 2	14	29	13	2	13	3	2	76
Fear 1	6	2	28	1	9	6	0	52
Fear 2	7	3	37	2	20	4	1	74
Sadness 1	3	5	17	14	5	16	5	65
Sadness 2	5	5	15	28	6	15	7	81

It could be observed that two implementations of *Happiness* were confused with *Anger* and *Excitement*. In a similar way, the first implementation of *Anger* was mostly confused

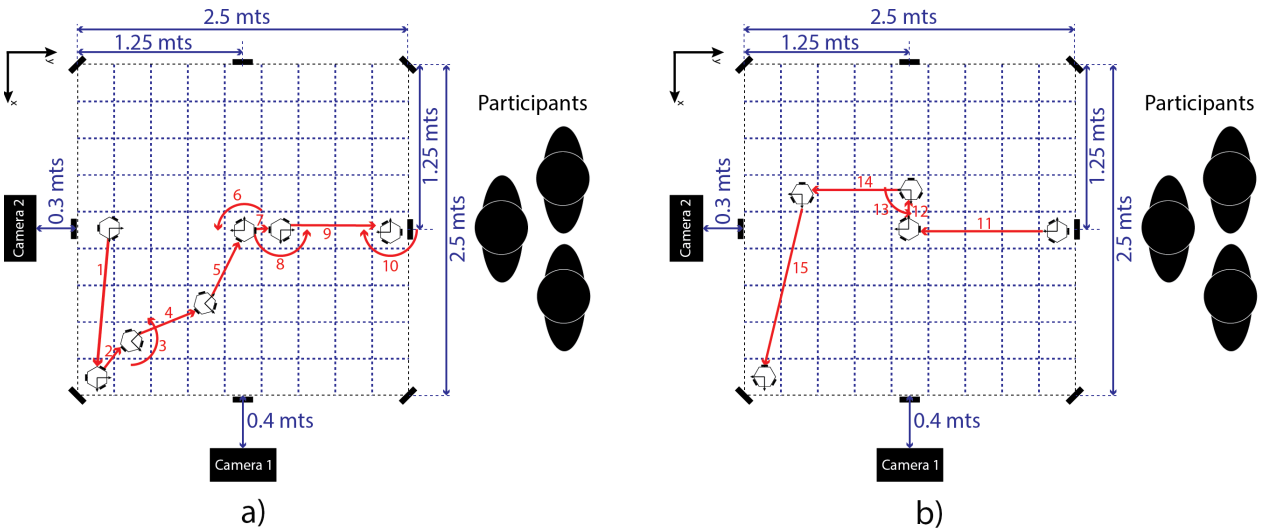

Figure 5: Sequence of movements done by the robot. The red arrows show the trajectory done by the robot, while the numbers show the order among the movements. a) The first ten movements b) The last five movements

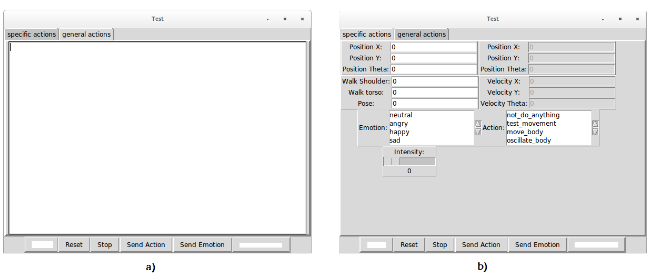

Figure 6: Graphical interface used to communicate with the Emotional Enrichment System. a) It is the interface used to send the actions sequence. b) It is the interface used to send a emotion and its intensity.

with *Excitement*, which was voted twenty one over forty nine subjects. The second implementation of *Anger* showed an improvement of perception from 10% to 38%. This implementation was perceived also as *Happiness*, *Fear* and *Excitement*. Both implementations of *Fear* had a high level of recognition 54 % and 50 % and mostly confused with *Excitement*, which was voted nine times for the first implementation and twenty times for the second implementation. Finally, the two implementation of *Sadness* was confused with *Fear* and *Tenderness*.

To verify these misinterpretations among implemented emotions, a Fisher's exact test was applied for ten different combinations. Additionally, a Holm-Bonferroni correction was applied for multiple comparisons to get a better p-value estimation. The results are shown in Table 4. As this analysis suggest, two implementation of *Anger* were perceived as two different emotions. Also shows that two implementation of Happiness were perceived to be similar to the second implementation of Anger.

A contingency matrix was created for each emotion to analysi each one. For each table the positive predictive

Table 4: Pair comparison among all the implemented emotions using Fisher's exact test for both questionnaires with $\alpha = 0.05$ for the case study. The * indicates that the p-value was adjusted using the Holm-Bonferroni correction for multiple comparisons.

Pair Compared	p-value	p-value*
Happiness 1 vs Happiness 2	0.38	1.0
Anger 1 vs Anger 2	7.3e-4	4.4e-3
Anger 2 vs Happiness 1	0.137	0.69
Anger 2 vs Happiness 2	0.157	0.69
Fear 1 vs Fear 2	0.74	1.0
Sadness 1 vs Sadness 2	0.665	1.0
Fear 1 vs Sadness 1	8.35e-5	5.8e-4
Fear 1 vs Sadness 2	5e-7	4e-6
Fear 2 vs Sadness 1	2e-7	1.8e-6
Fear 2 vs Sadness 2	1e-7	1e-6

value, accuracy and a Pearson's χ^2 were computed. The results are shown in table 5. They show that there is significant evidence to conclude that second implementation of *Anger*, both of *Fear* and *Sadness* have an impact in the perception of the emotion and they are considered as different implementation respect other implementations. While both implementation of *Happiness* and first of *Anger* are considered as similar to the other implementation.

Table 5: Accuracy, precision and results of Pearson's χ^2 for each contingency matrix with $\alpha = 0.05$ for the case study.

Presented Emotion	Positive Predicted Value	Accuracy	$\chi^2(1)$	p-value
Happiness 1	0.13	0.79	0.11	0.74
Happiness 2	0.21	0.81	3.7	0.054
Anger 1	0.1	0.8	3.8e-29	1
Anger 2	0.38	0.81	34.4	4.47e-9
Fear 1	0.54	0.8	36.2	1.8-e9
Fear 2	0.5	0.78	35.8	5.3e-10
Sadness 1	0.22	0.85	27.4	1.63e-7
Sadness 2	0.35	0.85	72.9	2.2e-16

For each of the contingency tables the classification accuracy and the no-information rate (NIR), i.e. the accuracy that had been obtained by random selection, are reported in table 6. The results reveal that only implementation with enough statistical evidence is *Sadness*' second implementation. Nevertheless, it is important no notice that the results were obtained using the lower part of the robot without any change in shape. Another important factor to highlight is the impact words enlisted in the questionnaire have on the perception rate. As it was expected, mental states *Excitement* and *Tenderness* were confused with emotions with similar arousal level. In this precise case the emotions Anger, Happiness and Fear were confused with Excitement, and Sadness was confused with Tenderness. Despite the bias generated by the two mental states listed in the questionnaire, the recognition rate of five out of eight implementations was over 35%, being the two implementation of *Fear* the implementations with the higher recognition rates (54% for the first and 50% for the second).

Table 6: Classification accuracy of the presented emotions by the single panels, computed as mentioned in the text, with corresponding 95% confidence interval, no-information rate, and p-value that accuracy is greater than the NIR.

Presented Emotion	Classification Accuracy	95% CI	No-Information Rate	P-Value [Acc > NIR]
Happiness 1	0.79	(0.75,0.82)	0.89	1.0
Happiness 2	0.81	(0.77,0.84)	0.88	1.0
Anger 1	0.8	(0.76,0.83)	0.88	1.0
Anger 2	0.89	(0.76,0.84)	0.83	0.95
Fear 1	0.79	(0.75,0.83)	0.88	1
Fear 2	0.78	(0.73,0.81)	0.83	0.99
Sadness 1	0.85	(0.81,0.88)	0.84	0.47
Sadness 2	0.85	(0.81,0.88)	0.81	0.035

The results obtained form the small scene are presented in Table 7. A chi-squared test with one degree of freedom with an alpha of 0.5 was done to verify if there was enough statistical evidence to accept our hypotheses: (i) people prefer scenes with emotions and (ii) gender has no impact on the preference. The results of the tests show that there is enough statistical evidence to accept our first hypothesis and reject the second one, with p-values of $1.42E-6$ and 0.85, respectively.

Additionally, the Emotional Enrichment System was used in the two parts of the case used. Although, there was not done any measure of any variable of the system, two things could be said about the system. First it enables the possibility to adapt same script to different stage measures with any impact in the script. Second, that it does not block the execution of an action when an emotion is changed.

Table 7: Answers obtained for the small scene.

Gender	With Emotion	Without Emotion	Total
Male	84	43	127
Female	81	45	126

7. CONCLUSIONS AND FURTHER WORK

The case study presented in this paper was done to cross validate the findings in the experiment and verify whether the participants would prefer scenes when the robot expresses emotions or rather moves without any emotion expression. For each one of four emotions (i.e. Anger, Happiness, Sadness and Fear) studied in the experiment were selected a two set of parameters. These parameters were described in files and used as input to the Enrichment Emotional System, which was in charge to combine actions with emotions. The results show that both implementations of happiness were confused with anger and excitement, while one implementation of anger was just confused with excitement. Both implementations of sadness were confused with tenderness and fear. Both implementations of fear had a recognition rate over 50%. Scene's results show that people prefer scenes with emotional movements and there is not any difference in gender.

Additionally to the results already mentioned, there are words that could bias participants' perception. For instance happiness and anger were considered as excitement. This misinterpretation should not be a surprise given the fact that there is not a unique definition of emotion [30, 31], and each person would interpret a situation differently, so they will give a different label to the presented movement. Moreover a misinterpretation of Happiness and Anger could suggest that additional features (e.g. trajectory or shape) should be added to increase differentiation between these two emotions. For example, Venture and collaborators [32] had found out that in human bodies the recognition rate of anger and fear are increased when torso and head are downwards. On the other hand, they found that happiness perception is increased when the torso and head are move upwards. This example could bring some insight to possible body changes that could occur in non-human like bodies, but it should tested in this kind of platforms to confirm if the same impact is reached.

8. REFERENCES

[1] de Gelder, B.: Why bodies? twelve reasons for including bodily expressions in affective neuroscience. Philosophical Transactions of the Royal Society of London B: Biological Sciences **364**(1535) (2009) 3475–3484

[2] Ekman, P.: Emotions Revealed : Recognizing Faces and Feelings to Improve Communication and Emotional Life. Owl Books (March 2004)

[3] Kleinsmith, A., Bianchi-Berthouze, N.: Affective body expression perception and recognition: A survey. Affective Computing, IEEE Transactions on **4**(1) (2013) 15–33

[4] Embgen, S., Luber, M., Becker-Asano, C., Ragni, M., Evers, V., Arras, K.O.: Robot-specific social cues in emotional body language. In: Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'12), USA, IEEE Computer Society (September 2012) 1019–1025

[5] Breazeal, C.: Designing Sociable Robots. MIT Press, Cambridge, MA, USA (2002)

[6] Beck, A., Hiolle, A., Mazel, A., Cañamero, L.: Interpretation of emotional body language displayed by robots. In: 3rd ACM Workshop on Affective Interaction in Natural Environments. (2010) 37–42

[7] Häring, M., Bee, N., André, E.: Creation and evaluation of emotion expression with body movement, sound and eye color for humanoid robots. In: IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man). (2011)

[8] Destephe, M., Maruyama, T., Zecca, M., Hashimoto, K., Takanishi, A.: Improving the human-robot interaction through emotive movements: a special case: walking. In: HRI. (2013) 115–116

[9] Li, J., Chignell, M.H.: Communication of emotion in social robots through simple head and arm movements. I. J. Social Robotics **3**(2) (2011) 125–142

[10] Saerbeck, M., Bartneck, C.: Perception of affect elicited by robot motion. In: Proceedings of the 5th ACM/IEEE International Conference on Human-robot Interaction. HRI '10, Piscataway, NJ, USA, IEEE Press (2010) 53–60

[11] Lakatos, G., Gacsi, M., Konok, V., Bruder, I., Berczky, B., Korondi, P., Miklosi, A.: Emotion attribution to a non-humanoid robot in different social situations. PLos ONE (December 2014)

[12] Sharma, M., Hildebrandt, D., Newman, G., Young, J.E., Eskicioglu, R.: Communicating affect via flight path: Exploring use of the laban effort system for designing affective locomotion paths. In: Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction. HRI '13, Piscataway, NJ, USA, IEEE Press (2013) 293–300

[13] Novikova, J., Watts, L.: Towards artificial emotions to assist social coordination in hri. International Journal of Social Robotics **7**(1) (2015) 77–88

[14] NAM, T.J., LEE, J.H., PARK, S., SUK, H.J.: Understanding the relation between emotion and physical movements. International Journal of Affective Engineering **13**(3) (2014) 217–226

[15] Watson, D., Clark, L.A., Tellegen, A.: Development and validation of brief measures of positive and negative affect: the panas scales. Journal of Personality and Social Psychology **54** (1988) 1063–1070

[16] Lang, P.J., Bradley, M.M., Cuthbert, B.N.: International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8, The Center for Research in Psychophysiology, University of Florida, Gainesville, FL (2008)

[17] Beck, A., namero, L.C., Bard, K.: Towards an affect space for robots to display emotional body language. In: IREE RoMan Conference. (2010)

[18] Robotics, A.: Nao. http://www.aldebaran-robotics.com/en/

[19] Saerbeck, M., van Breemen, A.J.N.: Design guidelines and tools for creating believable motion for personal robots. In: RO-MAN. (2007) 386–391

[20] Destephe, M., Zecca, M., Hashimoto, K., Takanishi, A.: Perception of emotion and emotional intensity in humanoid robots gait. In: Robotics and Biomimetics (ROBIO), 2013 IEEE International Conference on. (2013) 1276–1281

[21] Brown, L., Howard, A.M.: Gestural behavioral implementation on a humanoid robotic platform for effective social interaction. In: Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on Robot and Human Interactive Communication. (Aug 2014) 471–476

[22] Laban, R., Ullmann, L.: Modern educational dance. 2d ed., rev. by lisa ullmann. edn. Praeger New York (1968)

[23] Ekman, P.: Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (Revised and Updated Edition). 2 rev sub edn. W. W. Norton & Company (September 2001)

[24] Krippendorff, K.: Computing Krippendorff's Alpha Reliability. Technical report, University of Pennsylvania, Annenberg School for Communication (June 2007)

[25] : Understanding Interobserver Agreement: The Kappa Statistic. Family Medicine **37**(5) (May 2005) 360–363

[26] Niekum, S.: ar_track_alvar. http://wiki.ros.org/ar_track_alvar

[27] Wilson, E., Goldfarb, A.: Theatre: The Lively Art. McGraw-Hill Education (2009)

[28] Kids, M.T.: Stage directions for actors

[29] Drama, A.: Learning about the stage

[30] Plutchik, R.: The Nature of Emotions. American Scientist **89**(4) (2001) 344+

[31] Cacioppo, J., Tassinary, L., Berntson, G.: Handbook of Psychophysiology. University Press (2000)

[32] Venture, G., Kadone, H., Zhang, T., GrÃÍzes, J., Berthoz, A., Hicheur, H.: Recognizing emotions conveyed by human gait. International Journal of Social Robotics **6**(4) (2014) 621–632