Cross Validation of Emotional Features with a non-Anthropomorphic Platform

Julian M. Angel-Fernandez
Automation and Control Institute, Vienna
University of Technology
1932 Wallamaloo Lane
Vienna, Austria
jangelfe@tuwien.ac.at

Andrea Bonarini
Dipartimento di Elettronica, Informazione e
Bioingegneria, Politecnico di Milano
P.O. Box 1212
Milan, Italy
andrea.bonarini@polimi.it

ABSTRACT

Keywords

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

1. INTRODUCTION

2. RELATED WORK

3. THE EXPERIMENT

An experiment was designed 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 basic emotions suggested by Ekman [**?**]. These features were selected after the study on emotion projection in robotics, humans 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 the necessity to have a specific orientation, i.e., they are free to move taking any desired orientation. The experiment was performed at Politecnico di Milano, campus Leonardo 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.

For each table were calculated the mean, standard deviation, and median. It was not possible to use ANOVA test over the data because the assumption of normality is not achieved in the collected data. This was check using the Shapiro-Wilk Test. Additionally, 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 [**?**] (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 the 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 [**?**]. It seems that people perceive as fear when the robot is looking at them and moving far from them fast.

HRI '17 Vienna, Austria

DOI: 10.1145/1235

- 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.

4. SYSTEM

4.1 Robotic Platform

4.2 Emotion Enrichment System

5. CASE STUDY

The case study was done at Researcher's Night 2014 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.

5.1 Design and Setup

This case study uses the results obtained in the experiment and was designed to have two parts (i) emotion expression through changes in linear velocity, angular velocity, oscillation angle, orientation and direction. (ii) Presentation of a small scene to verify whether the participants would prefer scenes when the robot shows emotions or not. the Emotional Enrichment System was used in both cases (with and without emotions). Two web-cams and eight Alvar tags were added to use Kalman filter to improve robots localization in the stage. The detection of the AR tags is done through the use of the ROS package ar_track_alvar [**?**]. The distribution of the web-cams and the tags are depicted in Figure 1.

5.1.1 *Emotion Description*

The parameters selected for each of the 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.

5.1.2 *Scene*

As it was already mentioned, actors should adapt to different circumstances during the performance. Following this approach, the stage was discretized in 9x9 matrix as is shown in Figure 2. Thus, the movements of the robot are given in terms of the matrix positions. This allows the adaptation to different stage dimensions because the robot's final position is calculated by the Emotional Enrichment System during execution taken under consideration the stage dimensions.

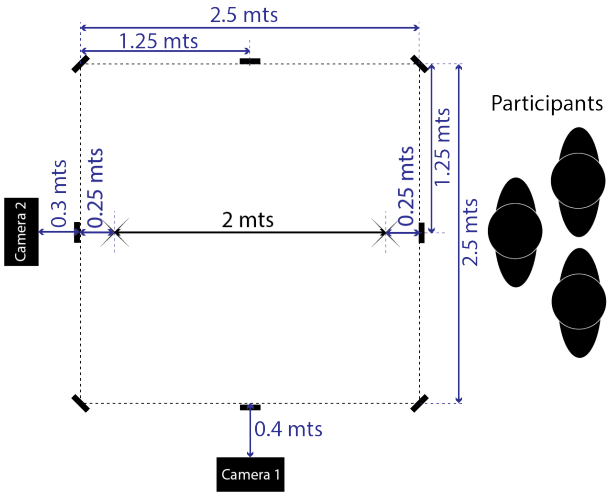

Figure 1: Environment setup for the case study.

Table 2: Parameters' values selected from the experiment.

Emotion	Direction (rad)	Orientation (rad)	Linear Velocity (mm/s)	Angular Velocity (rad/s)	Angle (rad)	Treatment ID
Happiness	0	0	500	3	0.349	29
	0	0	900	3	0.174	38
Anger	π	0	500	3	0.087	144
	0	0	900	1	0.087	31
Fear	π	π	900	2	0.174	113
	π	π	500	2	0.087	102
Sadness	π	0	200	1	0.349	130
	0	π	200	1	0.349	52

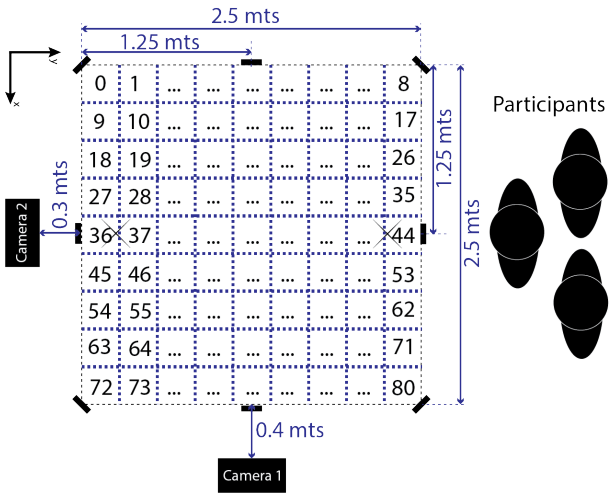

Figure 2: Stage discretization used for the small scene. The blue squares correspond to the each zone, while the numbers correspond to the ID given to each zone.

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 (Figure **??**), close to the right wing. Then, the robot moves to upstage right center and rotates by $\pi/2$ left (Figure **??**). 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 3.

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 executed by the Emotional Enrichment System using the same "script" and the emotion selection is done manually via graphical interface (Figure 4).The "script" were written in JSON using the language described in the Section **??**.

5.2 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 the results obtained during the case study.

It could be observed that the two implementations of *Happiness* were confused with *Anger* and *Excitement*. In a similar way, the first implementation of *Anger* was mostly confused with *Excitement*, which was voted twenty one over forty nine subjects that were exposed to the first implementation of *Anger*. 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 the 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, the two implementation of *Anger* were perceived as two different emotions. Also shows that the two implementation of Happiness were perceived to be similar to the second implementation of Anger.

An analysis was done for each emotion, therefore it was created a contingency matrix such as was done in the previous studies. For each of these tables, the positive predictive 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 the rest of implementations. While both implementation of *Happiness* and first of *Anger* are considered as similar to the other implementation.

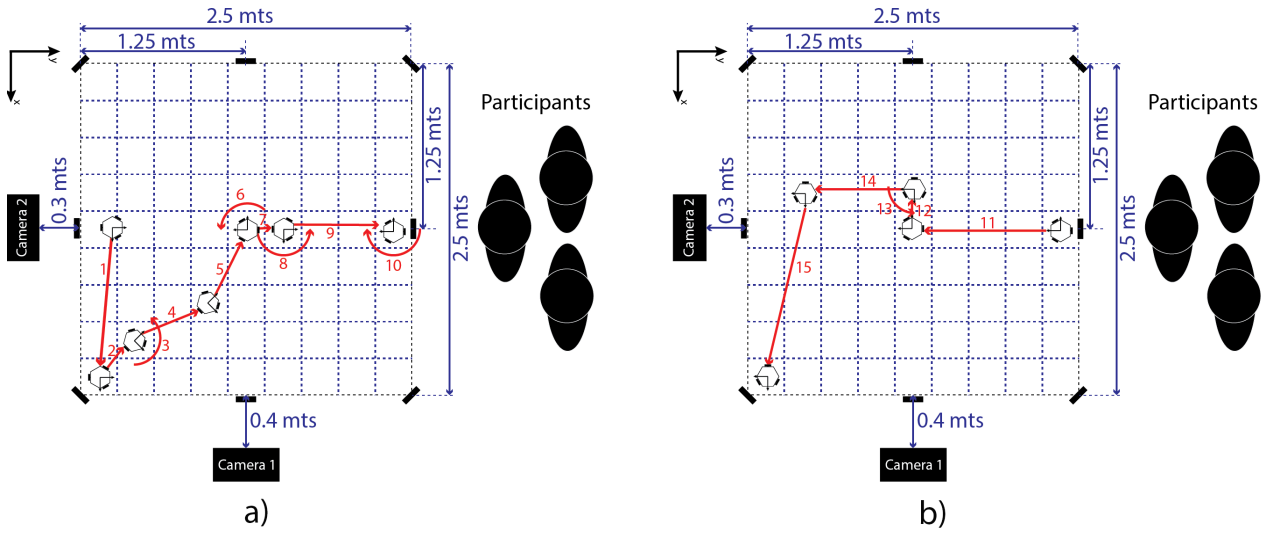

Figure 3: Sequence of movements done by the robot. The red arrows show the trajectory done by the robot, while the numbers show the order among the movements. a) The first ten movements b) The last five movements

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

Presented/Reported	Features					Emotions							Total
	Direction (rad)	Orientation (rad)	Linear Velocity (mm/s)	Angular Velocity (rad/s)	Angle (rad)	Happiness	Anger	Fear	Sadness	Excitement	Tenderness	Other	
Happiness	0	0	500	3	0.349	8	16	7	4	16	4	7	62
	0	0	900	3	0.174	11	11	6	2	19	3	1	53
Anger	π	0	500	3	0.087	7	5	6	2	21	7	1	49
	0	0	900	1	0.087	14	29	13	2	13	3	2	76
Fear	π	π	900	2	0.174	6	2	28	1	9	6	0	52
	π	π	500	2	0.087	7	3	37	2	20	4	1	74
Sadness	π	0	200	1	0.349	3	5	17	14	5	16	5	65
	0	π	200	1	0.349	5	5	15	28	6	15	7	81

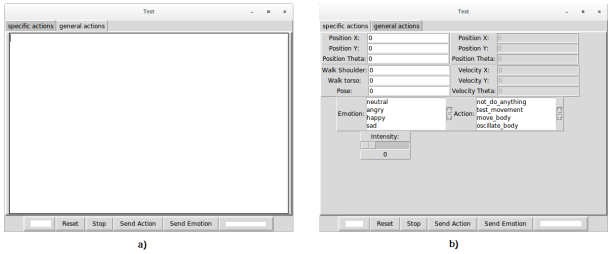

Figure 4: Graphical interface used to communicate with the Emotional Enrichment System. a) It is the interface used to send the actions sequence. b) It is the interface used to send a emotion and its intensity.

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

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 **??**. The results reveal that the only implementation with enough statistical evidence is the second implementation of *Sadness*. 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 in the experiment, *Excitement* and *Tenderness* were confused with other emotions with similar arousal level. In this precise case the emotions Anger and Happiness were confused with Excitement, and Sadness and Fear emotions were confused with Tenderness. Despite the bias generated by the two mental states enlisted 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).

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.

7. CONCLUSIONS AND FURTHER WORK

The case study presented in this chapter was done to cross validate the findings in the experiment. For each one of the four emotions studied in the experiment were selected two set of parameters. 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%. Additionally to the cross validation, it was done a small scene to check if people have preference to scenes where emotional movements are presented or not. The results show that people prefer scenes with emotional movements.

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

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	**Features**					**Classification Accuracy**	**95% CI**	**No-Information Rate**	**P-Value [Acc > NIR]**
	Direction (rad)	**Orientation** (rad)	**Linear Velocity** (mm/s)	**Angular Velocity** (rad/s)	**Angle** (rad)				
Happiness	0	0	500	3	0.349	0.79	(0.75,0.82)	0.89	1.0
	0	0	900	3	0.174	0.81	(0.77,0.84)	0.88	1.0
Anger	π	0	500	3	0.087	0.8	(0.76,0.83)	0.88	1.0
	0	0	900	1	0.087	0.89	(0.76,0.84)	0.83	0.95
Fear	π	π	900	2	0.174	0.79	(0.75,0.83)	0.88	1
	π	π	500	2	0.087	0.78	(0.73,0.81)	0.83	0.99
Sadness	π	0	200	1	0.349	0.85	(0.81,0.88)	0.84	0.47
	0	π	200	1	0.349	0.85	(0.81,0.88)	0.81	0.035

Table 7: Answers obtained for the small scene.

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