

Evaluating the influence of the physical embodiment of robot on the recognition accuracy of the emotional tone

Infantdani Abel Rayan
College of Computing
Georgia Institute of Technology
Atlanta, United States
irayan@gatech.edu

Abstract—The ability to communicate emotions is essential for a natural interaction between robots and humans because it is *not possible not to communicate*. Hence, if a robot utters something, it is likely that it would have conveyed some emotion through it. Hence, my research was focused on finding out which of the seven emotions -Anger, Disgust, Fear, Happiness, Sadness, Surprise and Neutral is more readily perceptible by people and to find the influence of embodiment of a physical robot on recognition accuracy of the emotional tone.

Keywords- *emotion recognition, influence of physical embodiment*

I. INTRODUCTION

Robots are increasingly used for a variety of services for people. It can be used for doing repetitive tasks with great consistency and efficiency. It gets interesting when robots speak, especially pet-like robots. Pets have long been associated with humans, there is a huge set of emotional interaction possible with pets [1]. Toys are often modeled to look pet-like and children love to interact with such “toybots”. Moreover, there are studies that show autistic children are interested by the movements made by the robots, and enjoy interacting with these devices [2]. Research relevant to the emotional issues [3] shows that emotional expression in speech is a crucial issue in social interaction, and lack of it is one major problem for people with ASD. Using robots in learning/training is a type of strategy to avoid a contrived learning situation such as sitting in front of a desktop computer.

Pleo robot by Innvo Labs Corporation was used for the study. The motivation behind this study was to find whether a commercial toybot like Pleo accurately conveys the intended emotion by uttering emotional sentences. If it can, then which of the emotions contained in sentences are perceived readily and what is the effect of physical embodiment on the recognition accuracy. It will be useful to further explore how a Pleo like pet robot can be used for interacting with children with Autism. Autism is often associated with problems in expressing emotion. If robots like Pleo with voice embodiment could accurately convey “an emotion”, then it can be a very small step towards realizing a robot that can hear the monotone voice of Autistic individuals and express it with animated

emotions or such that anyone can understand them. Moreover, among the high functioning autistic individuals there is a trend to rely on strategies other than facial expressions in the recognition of emotions. [6] [7]. Hence, one significant advantage of using Pleo for emotional recognition study is that it has no facial expression. Thus, this study can help further investigation into how robots can convey effective emotions to autistic children without relying on facial expressions.

There are six basic emotions as classified across cultures by Ekman[4] [5] . These emotions are Anger, Disgust, Fear, Happiness, Sadness and Surprise. However, for the purpose of the study a seventh choice of “neutral” was included, this will help mitigate the effects of forced choice and also one's tone can be monotonic without any emotions.



The first question is which of the seven emotions is easily perceptible when Pleo utters it. To answer this question, a recorded sentence in all seven emotions was embedded to Pleo and participants were recruited to classify it. The second question is what is the influence of physical embodiment of Pleo on these emotions. To answer this, a virtual simulated Pleo which can perform the same actions with same control code was used. Thus, the real physical Pleo and the virtual Pleo were programmed to do the same task and users were asked to classify the emotions in both the cases. Furthermore, the users were asked subjective questions regarding their experience with Virtual Pleo and Physical Pleo.

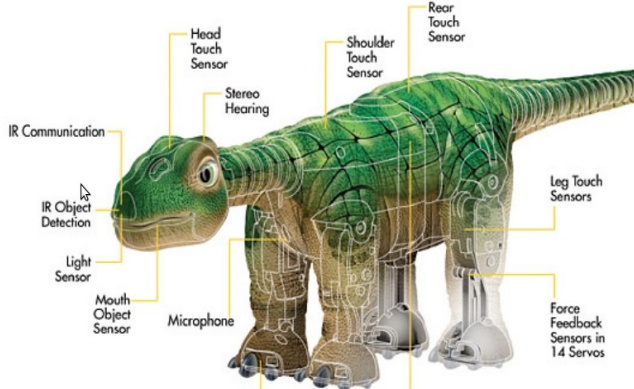
II. PLATFORM AND RESOURCES

A. Voice

The researchers at the School of Psychology, Sonification Lab of Georgia Institute of Technology helped to generate the voice for embedding on Pleo. Four people were chosen randomly from the lab and were asked to utter three different sentences in seven different emotions (six basic emotions and neutral) in a recording studio with sound-proof walls. They did two to three trials or until they were convinced about the emotions themselves. Out of these voices, one sentence uttered by a North American male aged 31 was found suitable in all the seven emotions and was the choice for embedding on to Pleo. Noise cancellation filter cool-edit pro 2.0 version was used to further remove noise, if any.

B. Pleo

Pleo is a commercial robot manufactured by Innvo Labs Corporation. It has a variety of programmable motors and sensors.



(image source: Pleo Developer manual)

Pleo runs on “Life OS”. It is the complex software platform that enables Pleo’s mechanical, electronic, sensory, and AI systems to interact, to allow Pleo to react to people and his environment. Life OS is a collection of a number of different resources such as sounds, motions, commands, scripts, and properties. For this study it was necessary to program only the motions and sound sensory system.

The build tools use the built-in Python audioop module to process incoming WAV files into raw PCM or ADPCM formatted USF files (UGOBE Sound File [USF] format.). The Sound System can currently play 11k, 8-bit mono sound data.

Motions are resources that contain data on how and when to move joints in Pleo. Internally, Life OS animates using a 3ds Max model whose animation data we export using a custom MaxScript. This results in a CSV file, with a special header with some meta-information about the motion – such as frame rate, duration, and the joints that are used. This CSV file is processed into a UGOBE Motion File with a special Python script.

C. Pleo PC Simulation

The Pleo MySkit software runs on the PC. It has a “P3D performance player” which is intended for reviewing performance skits on the PC before uploading them to PLEO and has controls for navigating the performance -- similar to the Windows Media player. Motion & sound are synchronized together. It uses OpenGL for drawing a 3D PLEO preview. A fast machine is not required. However, for simulated playback a fast machine (800 Mhz+ CPU) with a 3D graphics card is recommended.

Thus the same control code used to run the Pleo can be played by the Pleo “P3D performance player” with same speed and accuracy.

D. Pleo Programming

Pleo development Kit makes it possible to creatively interact with Pleo on the programming level, to modify his behaviors, tweak an animation, design a new neural network learning algorithm and also add sound effects. MySkit Performance Editor for PLEO helps to create custom PLEO dances & performances. There are many features, effects, and safeguards to make editing easy. MySkit directly supports initializing SD flash cards using a card reader. YAPT can be used to define or modify the behavior logic of the personality. YAPT does not create or modify resources like sounds or motions, but exploits the resources from an existing personality. The generated voice must be integrated with suitable action so that the Pleo appears to be embedding the voice. Mouth movements alone is not the right option. At the same time we are not “animating” the pleo since that would require one to design proper “animations” according to proper principles such that it does not contribute to the noise instead of aiding the emotion recognition and there was lack of time for such a pursuit. Hence a neutral path for actions is adopted. Pleo walks, closes eyes and does a tail wag, he utters the sentence and then finally does another tail wag. He does this for all seven emotions for the sentence.

III. EXPERIMENT DESIGN

A. Hypothesis

1) Accuracy of Recognizing emotions

H0: The accuracy of recognizing any of the seven emotions does not exceed 50%

H1: The accuracy of recognizing at least one of the seven emotions is significantly greater.

The rationale behind choosing the 50% as hypothetical mean was to find those only those emotions among the seven that exhibit statistically significant accuracy recognition percentage than mere chance.

2) Influence of Embodiment

H0: There are no differences in recognizing emotions because of Physical Embodiment

H1: There are significant differences in recognizing the emotions because of Physical Embodiment.

The rationale was to do an statistical analysis and compare the recognition accuracy of the emotions (see section for experimental analysis)

B. Experiment Design

- Independent Variables:
 1. Seven Emotions
 2. Two types of Embodiment (Virtual and Physical)
- Dependent Variable:

Success rate of delivery of intended emotions - Accuracy of recognizing emotions.
- Within Subjects design for 14 participants
- Half of them encountered the Physical Pleo first and then the Virtual Pleo while half of them, the other way.

C. Randomization using the Latin Square

The presentation of Anger (1) Disgust(2) Fear(3) Happiness (4) Neutral(5) Sadness(6) and Surprise (7) was randomized using the Latin square, which is a randomized block

Seq1: 1 3 5 7 4 6 2

Seq2: 2 4 6 1 7 3 5

Seq3: 3 5 7 2 1 4 6

Seq4: 4 6 1 5 3 2 7

Seq5: 5 1 3 6 2 7 4

Seq6: 6 7 2 4 5 1 3

Seq7: 7 2 4 3 6 5 1

This sequence was further presented with Physical embodied Pleo (Phy) and virtual Pleo (Virt) in following order to the 14 participants (P1 to P14)

P1: Phy_Seq1 and Virt_Seq2

P2: Virt_Seq1 and Phy_Seq2

P3: Phy_Seq3 and Virt_Seq4

P4: Virt_Seq3 and Phy_Seq4

.

.

P14: Virt_Seq6 and Phy_Seq7

The rationale was to avoid doing the emotions in same order and same embodiment sequence by any two participants.

D. Subjective Questions

A number of subjective questions were asked as a follow-up interview and the responses were noted. Some of the questions asked were:

Q1: Is the physical presence of Pleo engaging and watchful than virtual agent ?

(participants were explained the meaning of words “engaging and watchful”)

Q2: Is a pet like dino robot having human voice awkward ?

Q3: Will you prefer Physical Pleo to virtual for education ?

(education for preteens through skits and role play interaction)

Q4: Can the Pleo feel his emotion ?

IV. EXPERIMENT

Participants were recruited by sending emails to specific lists in Georgia Tech. Some of them received credits in their respective courses for participating in this user study.

A. Characteristics

The following is population characteristics:

a) Age: Ranged from 20 to 26

b) Average is 21.5

c) Gender: 9 males and 5 females

d) Undergraduates: 10 Graduates: 4

e) 1 had “lot” of experience with Robots,

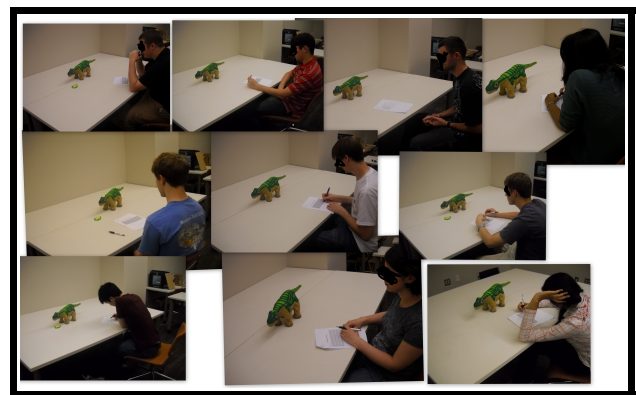
4 had “none”, rest had “little”

f) Everyone had “lot” of experience with Computers

B. Experimental Setup

The experiment was carried out at the Robotic Intelligence Lab at Georgia Tech. The lab was set up and was quiet without any other parallel activities. The responses were collected on a sheet of paper in both the cases.

1) Physical Pleo



Pleo was placed at a suitable distance, such that the participant can view it without any disturbance.

2) Virtual Pleo



The virtual pleo was presented in a laptop, with properly adjusted sound levels and suitable distance.

In both the cases, the participants were seated in front of Pleo and were asked to observe Pleo as it performs its walk and utters the sentence. The participants were asked to mark the answer on the given sheet of choices. They were told not to bias their current choice based on their previous choice. That is, they were told to treat each emotion independently and it was not necessary for them to have heard all seven emotions.

Then they were asked to answer the follow-up subjective questions.

C. Analysis

- Raw Percentages for recognition accuracy - This will give an overall idea regarding how emotions are perceived in both the cases – virtual and physical
- Paired Sample t-test between Physical and Virtual cases for the accuracy of emotion recognition. This will provide us with answers for the influence of embodiment.
- One sample t-test with 50% as Hypothetical Mean. This will provide us with those set of emotions that are readily perceptible and those which are not in both the cases.
- Subjective analysis. This will help us understand, what are the other reasons for preferring Physical Pleo over Virtual or vice-versa.

V. RESULTS AND DISCUSSION

1) Raw Percentages for recognition accuracy:

The Figure 1 shows the distribution of raw recognition percentages. One can see that for the emotions which are easily recognizable (Anger, Fear, Sadness) the virtual case performs better than the physical case. However, for the

emotions with lower percentage of recognition accuracy (Disgust, Happiness, Surprise) the virtual case performs worse than the physical case.

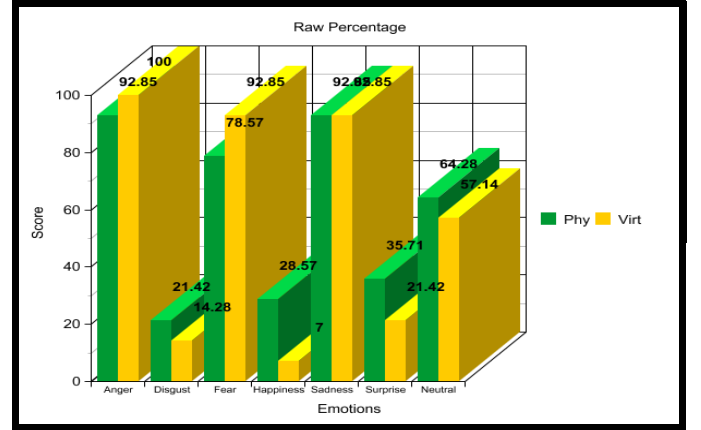


Figure 1: Percentage Accuracy of Emotion Recognition

2) Paired Sample t-test between Physical and Virtual cases

A paired-sample t-test was performed between the recognition accuracies of Physical and Virtual Pleos. Table I shows the results.

TABLE I. PHYSICAL PLEO VS VIRTUAL PLEO PAIRED SAMPLE T-TEST

	Physical	Virtual
Mean	4.07	3.86
SD	1.49	0.95
SEM	0.4	0.25
N	13	13

Intermediate values used in calculations : $t = 0.5247$ $df = 13$
standard error of difference = 0.408

P value and statistical significance: The two-tailed P value equals 0.6086. By conventional criteria (<0.05), this difference is considered to be not statistically significant.

Confidence interval: The mean of Physical minus Virtual equals 0.21 and 95% confidence interval of this difference is from -0.67 to 1.10

Thus the difference between Virtual and Physical Pleo cases for emotion recognition accuracy is not statistically significant.

3) One sample t-test with 50% as Hypothetical Mean

TABLE II. PHYSICAL PLEO EMOTION CLASSIFICATION

Emotion	Emotions with Positive Difference from mean of 50%		
	P-Value	Difference	t-value
Anger	<.0001	42.85	6
Sadness	<.0001	42.85	6
Fear	0.06	21.42	1.71
Neutral	0.15	14.28	1.08
	Emotions with Negative Difference from mean of 50%		
	P-Value	Difference	t-value
Disgust	0.01	-28.57	-2.51
Happiness	0.06	-21.42	-1.71
Surprise	0.15	-14.28	-1.07

TABLE III. VIRTUAL PLEO EMOTION CLASSIFICATION

Emotion	Emotions with Positive Difference from mean of 50%		
	P-Value	Difference	t-value
Anger	<.0001	50	infinity
Sadness	<.0001	42.85	6
Fear	<.0001	42.85	6
Neutral	0.3	7.14	0.52
	Emotions with Negative Difference from mean of 50%		
	P-Value	Difference	t-value
Disgust	<0.001	-35.71	-3.67
Happiness	<0.001	-42.85	-6
Surprise	0.01	-28.57	-2.51

Thus one can see that in the case of Physical Pleo, among the emotions with positive difference from mean of 50%, only Anger and Sadness are statistically significant, whereas in case of Virtual Pleo Anger, Sadness and Fear are statistically significant.

Among the emotions with negative difference from the mean of 50%, only disgust is statistically significant for the Physical Pleo, whereas Disgust, Happiness and Surprise are statistically significant for the Virtual Pleo. Thus with respect to emotion classification, the Virtual Pleo either does very good for some emotions or does very bad for some emotions. The Physical Pleo degrades much gracefully.

4) Subjective analysis

A number of questions were asked to the participants and some interesting answers were noted.

Q1: Is the physical presence of Pleo engaging and watchful than virtual agent ?

Interesting Answers:

“The noise of the motors helped me pay more attention to the physical”

“Physical Pleo was more watchful, since I felt that we can reach out and touch Pleo”

“immediately drew my attention, it was like seeing a toddler walk, a dog play, as opposed to a screen”

“People care about moving-physical objects than virtual”

Some participants, just gave the answer as “agree” for physical Pleo being more engaging, but they were not able to come up with reasons for justification. On prompting, their answer was “I don't know, but it was more engaging”

TABLE IV. PHYSICAL PRESENCE OF PLEO WAS ENGAGING AND WATCHFUL THAN VIRTUAL AGENT

Strongly Agree	6
Agree	6
Middle	1
Disagree	0
Strongly Disagree	1

There were significant number of participants who “agreed” (including strongly). Thus Physical Pleo was more engaging and was liked to be watched. (12 out of 14)

Q2: Is a pet like dino robot having human voice awkward ?

Interesting Answers:

“Exposed to Cartoons, always wanted animals to talk”

“Wanted to be cartoon-like, it is a baby dino”

“Virtual Pleo, with human voice is fine – seen in games and cartoon”

“It is not awkward, because it is the first voice I heard attached to Pleo”

TABLE V. A PET LIKE DINO ROBOT HAVING HUMAN VOICE IS AWKWARD

	Physical	Virtual
Strongly Agree	1	0
Agree	6	2
Middle	0	5
Disagree	6	5
Strongly Disagree	1	2

The two-tailed P value equals 1.00. By conventional criteria, this difference is considered to be not statistically significant. Hence there is no significant difference in preference of voice.

Q3: Will you prefer Physical Pleo to virtual for education ?
(education for preteens through skits and role play interaction)
Interesting Answers:

“It is too distracting to Play with a toy and listen to it, as a teacher”

“Virtual feels like mere recording, the physical pleo gives the feeling that someone is talking to you”

“It is more fun and cool, and it will attract children to spend more time with it

TABLE VI. I WILL PREFER PHYSICAL PLEO TO VIRTUAL FOR EDUCATION

Strongly Agree	8
Agree	5
Middle	0
Disagree	1
Strongly Disagree	0

There were significant number of participants who “agreed” (including strongly) over preference of Physical Pleo for education (13 out of 14)

Q4: Can the Pleo feel his emotion ?

Interesting Answers:

“It is in laptop, everything a computer does look programmed, but this is a physical thing”

“Physical Pleo seemed like it embedded the voice, where as in virtual it looked disconnected”

“On Computer I know it is just a piece of code running”

“Physical Pleo may have some sensors”

TABLE VII. PLEO CAN FEEL HIS EMOTION

	Physical	Virtual
Strongly Agree	0	0
Agree	4	2
Middle	6	3
Disagree	2	6
Strongly Disagree	2	3

The two-tailed P value equals 1.00. By conventional criteria, this difference between physical and virtual cases is considered to be not statistically significant.

VI. CONCLUSION

Although, from the study there was no statistical significant improvement of recognition accuracy of emotional tone. The subjective analysis showed users enjoyed to interact more with the Physical Pleo. This cannot be solely attributed to the novelty effect of the robots. The follow-up questions show that people are more attracted to “moving things” rather than something confined to a monitor, they like to see the real movements of the robot (the sound of limb movements) and is

more captivating. Of course, there are huge number of differences between the the physical and the Virtual Pleo and this study by no means is exhaustive enough to find what difference between the Virtual and Physical Pleos will contribute to its more engaging presence. Some participants, just gave the answer as “agree” for Physical Pleo being more engaging, but there were not able to come up with reasons for justification.

Analysis for emotion recognition shows us that for the Physical Pleo, the accuracy of recognizing anger and sadness exceeds 50% and is statistically significant. For the Virtual Pleo, the accuracy of recognizing anger, fear and sadness exceeds 50% and is statistically significant. For the Physical Pleo, the accuracy of recognizing disgust does *not* exceed 50% and is statistically significant. For the Virtual Pleo, the accuracy of recognizing disgust, happiness and surprise does *not* Exceed 50% and is statistically significant.

From paired samples t-test, it fails to reject that there are no differences in recognizing emotions because of Physical Embodiment.

This pilot study just sheds light on possible approaches for embedding emotional voices on toybots like Pleo and how it is perceived by people. In future, this work with additional implementation with respect to voice and interaction can be extended to analyze how it is perceived by children with ASD. This is particularly exciting since it is known that children with ASD have difficulty in understanding facial expressions [6] [7], but Pleo conveys emotions without any expression.

Thus, the future direction would be to identify a population with ASD for which interaction with Pleo can be useful. Determine the accuracy of recognizing emotions maybe aid it with music [8]

Use subjective measures to find out, which kind of voice is preferred (older kid voice, male or female, cartoon). Find number of trials to recognize and time of recognizing emotions. Identify strategies like - auditory cues, rhythmic clapping, to aid emotion recognition. Identify music - to facilitate transitions, to facilitate listening as previous studies shows it can very useful intervention for children with autism [8]. One could even come with a game design for intervention. This study was very exiting and useful.

REFERENCES

- [1] DOGS: A Startling New Understanding... by Raymond and Lorna Coppinger, 2002, ISBN: 978-0684855301 Dog Language by Roger Abrantes, 3rd Ed. 2001, ISBN: 978-0966048407
- [2] Michaud, F., & Theberge-Turmel, C. (2002). Mobile robotic toys and autism. In Dautenhahn, K. eds. Socially Intelligent Agents – Creating Relationships with Computers and Robots, Springer, 125-132.
- [3] El Kaliouby, R., Picard, R. W., & Baron-Cohen, S. (2006). Affective computing and autism: Progress in convergence. In W. S. Bain-bridge and M. C. Roco (Eds.), Annuals of the New York Academy of Sciences, 1093, 228-248.
- [4] Ekman, P. An argument for basic emotions. Cognition and Emotion, 6, 169-200.
- [5] Ekman, P. Basic Emotions. In T. Dalgleish and T. Power (Eds.) The Handbook of Cognition and Emotion Pp. 45-60. Sussex, U.K.: John Wiley & Sons, Ltd.

- [6] Christiane Bormann-Kischkel, Markus Vilsmeier, Beate Baude, "The Development of Emotional Concepts in Autism", *Journal of Child Psychology and Psychiatry* Vol 36 No 7 Pg: 1243-1259 (1995)
- [7] Peter Hobson "The Autistic Child's Appraisal of Expressions Of Emotion: A Further Study" *Journal of Child Psychology and Psychiatry* Vol: 27 No: 5 Pg: 671-680 (1986)
- [8] Goodman & Williams, 2007 - Interventions for increasing the academic engagement of students with autism spectrum disorders in inclusive classrooms. *TEACHING Exceptional Children*, 39, 53-61.)
- [9] *My Doggie Says...; Messages from Jamie* by Fred Haney ISBN 0-9785515-0-8
On Talking Terms With Dogs: Calming Signals by Turid Rugaas ISBN 0-9674796-0-6.
- [10] Hailpern, J. (2007). Encouraging speech and vocalization in children with autistic spectrum disorder, *SIGACCESS NEWSLETTER*, 89, Sept.
- [11] Bartneck, C., & Reichenback, J. (2005). Subtle emotional expressions of synthetic characters, *Int. J. Human-Computer Studies*, 62, 179-192.
- [12] Beer, J. M., Fisk, A. D., & Rogers, W. A. (2009). Emotion recognition of virtual agents facial expressions: The effects of age and emotion intensity. In *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting (HFES 2009)*