

Using simulated body language and colours to express emotions with the Nao robot

Wouter van der Waal
S4120922
Bachelor Thesis Artificial Intelligence
Radboud University Nijmegen

Supervisor:

Khiet Truong

Department of Artificial Intelligence
Donders Institute for Brain, Cognition, and Behaviour
Radboud University, Nijmegen

Abstract

During this study I used the Nao-robot to show emotions, which were then classified by participants. There were two different models used to show the emotions. The first model used human body language, the second model used differently coloured LEDs to show emotions. This experiment shows that the body language is easier to recognize by human subjects than the lights, but using the lights in addition to the body language generally reaches the highest recognition rate. Therefore LEDs can be a valid addition when showing emotions with a robot.

In addition, I tested if the location of the LEDs influenced the recognition rate. I found that lights located in the eyes of the Nao-robot were significantly easier to recognize than lights located at the chest. So the location is important to consider when using lights to show emotions.

Contents

- Abstract 2
- 1 Introduction..... 4
 - 1.1 Natural Modalities..... 4
 - 1.2 Artificial Modalities 4
- 2 Previous Research 5
 - 2.1 Poses..... 5
 - 2.2 Colours..... 5
 - 2.3 Comparison 6
- 3 Research Questions and Hypotheses 6
- 4 Research Method 7
 - 4.1 Material and Software..... 7
 - 4.2 Conditions..... 8
 - 4.3 Task..... 8
- 5 Experiment 8
 - 5.1 Participants..... 8
 - 5.2 Procedure 9
- 6 Results 9
 - 6.1 Research Question 1..... 9
 - 6.2 Research Question 2..... 11
- 7 Discussion 13
 - 7.1 Research question 1 13
 - 7.2 Research question 2 14
- 8 Conclusion 14
- 9 Future Work 14
- 10 References..... 15
- 11 Appendix..... 15

1 Introduction

For a robot to be trusted and used outside of an industrial environment, it must exhibit social behaviour (Brooks, 1997). One way of making a robot able to express emotions is by imitating humans. Humans can show emotions in multiple ways. Facial expressions, body language and paralanguage can all be used to express emotions (Saldien, Goris, Vanderborght, Vanderfaeillie, & Lefeber (2010).

It has been shown that computer models using multiple modalities to classify human emotions perform better than models using only one modality. This is the case for Facial Expressions in combination with Speech (Busso et al., 2004) and Facial Expressions and Body Language (Gunes & Piccardi, 2007). Therefore, a robot expressing emotions can also use multiple modalities.

1.1 Natural Modalities

One of the most effective ways to let a robot express emotions is by imitating certain aspects of a human's facial expressions, such as the height of the eyebrows or eyelids (Kobayashi & Hara, 1993). However, to use this type of simulation you do need a specialized face robot, which makes this way largely unattractive for most purposes.

Beck, Stevens, Bard, & Cañamero, (2012) have shown that robots can effectively show emotions by copying an actor's body language. This has the advantage of being easier to use, because the body language can be displayed effectively by using, for instance, the Nao robot from Aldebaran instead of a specialized robot.

1.2 Artificial Modalities

A problem with biologically inspired designs is that they often suffer from the uncanny valley principle (Fong, Nurbakhsh, & Dautenhahn, 2003). The more something looks like a human, the more humans will like and trust it, but if it looks very close to human, people will like it less than objects that aren't human at all (Figure 1). One way to solve this problem is to not try to reach perfect human likeness, but to create functional, artificial oriented ways to express emotions, such as the use of colours to express emotions (Plutchik, 2001).

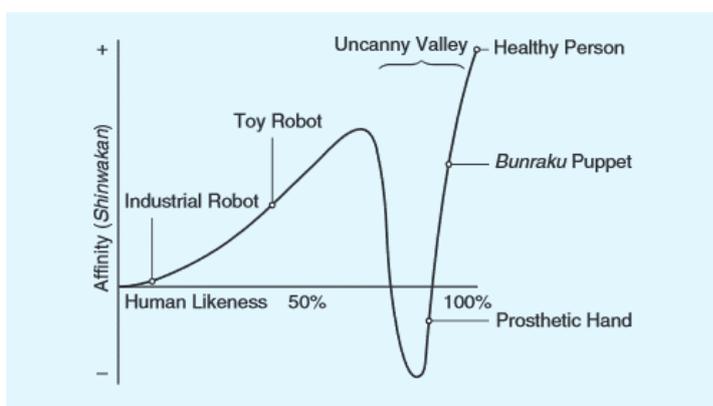


Figure 1: This graph visualises the uncanny valley principle. (Mori, MacDorman, & Kageki, 2012)

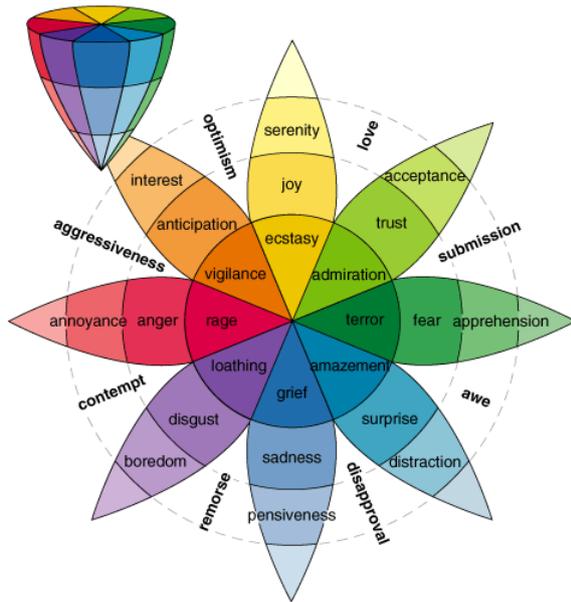


Figure 2: Plutchik's wheel of emotions. Plutchik (2001)

Plutchik (2001) proposed a way to model emotions through colours (Figure 2). In his wheel of emotions Plutchik suggested eight primary emotions. Each had two emotions added, a weaker and a stronger variant. For instance Anger, with the weaker variant Annoyance and the stronger variant Rage. Emotions could be mixed just like colours.

It should be noted that the colours of the model are not based on research, but chosen by Plutchik (2001). This model does not give a reason why the colours work, but research shows that they do work.

2 Previous Research

2.1 Poses

As mentioned before, body poses are an effective way to show emotions for artificial systems. During this study I will make use of emotional body language, specifically the poses described by Beck et al. (2012).

Beck et al. (2012) obtained the poses by recording a professional actor who was instructed by a professional director to perform ten emotions using only static body language. Based on these motion captured performances, six key poses were constructed for the Nao; Anger, Sadness, Fear, Pride, Happiness and Excitement. The poses were slightly altered to improve stability for the Nao. Participants, who classified these emotional poses, recognized all emotions far better than chance. The emotions that were hardest to recognize were Happiness and Excitement, which reached a recognition rate of 73%. The same poses displayed by the actor reached 92% on average. The emotions are easier to read from humans than from an artificial platform, but the artificial platform still performs a lot better than chance level.

Beck et al. (2012) made their model used in their study freely available online. This model is in the form of a toolkit for the Nao robot. In this toolkit, the emotions can be shown using sliders. A slider contains 2 emotions. Each slider can be positioned in the middle for a neutral state, one of the extremes for the corresponding emotion or anywhere in-between for a weaker version of the emotion. By using multiple sliders, it is possible to mix the emotions. For instance, 70% sadness, 30% pride.

2.2 Colours

Terada, Yamauchi, and Ito, (2012) used LEDs (light-emitting diodes) to implement the aforementioned colour model by Plutchik (2001). The robot they used changed the colour of its body to simulate emotions. Base emotions were simulated by colour and the intensity of the emotion was simulated by changing the waveform. The faster the LED pulsed the stronger the emotion.

During this study, participants categorised the emotions of the robot. The exact correct recognition rates were not reported, though Terada et al. (2012) did mention that four of the eight emotions were uniquely recognised. Specifically Happiness, Trust, Sadness and Fear were uniquely recognizable. Anger, Anticipation and Surprise were not. While this implies a recognition rate lower than simulating faces or body language, this model does have the major advantage of being very cheap to implement.

2.3 Comparison

These two models do have a small difficulty in comparing them. The main problem is that they use different emotions. Only Anger, Happiness, Sadness and Fear exist in both models, so the comparison is limited to those four emotions.

Table 1 shows The comparison between the models. Both perform well on Fear and Sadness. The pose model performs well on Anger, but not on Happiness (Beck et al., 2012). The coloured LED model performs opposite on these emotions; bad on Anger, but good on Happiness.

Emotion	Anger	Fear	Happiness	Sadness
Poses (Beck et al., 2012)	+	+	-	+
LEDs (Terada et al., 2012)	-	+	+	+

Table 1: Comparison between the pose model by Beck et al. (2012) and the coloured LED model by Terada et al. (2012). A '+' sign indicates an easily recognizable emotion a '-' sign a hard recognizable emotion.

3 Research Questions and Hypotheses

Both modalities have different strong and weak points. The first goal of this project was to investigate if these modalities complement each other;

Research question 1:

Can a humanoid robot, such as the Nao, equipped with a multi-model system of both body language and lights, express emotions better than a robot, equipped with only one of these modalities?

Hypothesis 1:

Because multiple modalities give more information, I would assume that the combination of modalities is recognised better. This effect should be more clear at Anger and Happiness, since the models show conflicted results on those emotions.

When humans show emotions through facial language, some areas of the face are used more than other areas to show different emotions (Fong, Nourbaksh, & Dautenhahn, 2003). Therefore the second research question is;

Research question 2:

Is the recognition of emotions, expressed by a robot, such as the Nao, influenced by the position (e.g. the eyes or chest) of the relevant lights on the robot?

Hypothesis 2:

The eye lights are located at a place where humans show a lot of emotion, the face. This location could therefore be more natural and better recognisable than lights located at the chest.

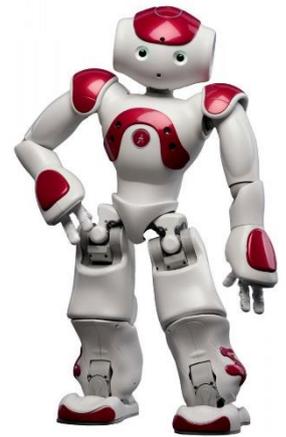


Figure 3: The Nao robot. Source Aldebaran

4 Research Method

4.1 Material and Software

4.1.1 Body Language

The Nao-robot is designed to show human-like movements. For the actual poses, I made use of the toolkit by Beck et al. (2012) These static poses were used in the same experiment. For this experiment, I only used single 100% emotions, so there were no combinations of multiple emotions. All poses used can be found in Figure 4.

The emotions were designed to be static, because of that the transitions between the poses are not stable. Therefore, the recording of the movies was only done using the static poses, without the transition from neutral to the pose.



Figure 4: The Nao robot in four different stances:

Top-Left: Anger, Top-Right: Happiness, Bottom-Left: Sadness, Bottom-Right: Fear

4.1.2 Coloured LEDs

The Nao has multiple RGB LEDs, located on the “eyes”, “chest” and “feet”. I made use of the LEDs on the eyes and chest. The colours and period of the LEDs were the ones reported by Terada et al. (2012). I used the most extreme forms of the emotions. So Rage instead of Anger or Annoyance. The exact setting for each colour can be found in Table 7 in the appendix.

4.1.3 Movies

For each condition I recorded a movie using an HD webcam. All recorded movies were converted to animated gif-pictures. The duration of each gif was just long enough to show the blinking of the lights. By using gifs instead of movies I can determine the response times of the participants more easily. This way I can start the timer at the start of the animation and participants can take as long as they need to decide on an emotion.

4.2 Conditions

For the first research question, there are two modalities, pose and lights. When using lights, both eye lights and chest lights were used (Condition 1,2 and 3). For the second research question, there are two modalities, eye lights and chest lights (Condition 3,4 and 5). See Table 2 for all conditions.

Condition	Pose	Eye Lights	Chest Lights
1	Yes	Yes	Yes
2	Yes	No	No
3	No	Yes	Yes
4	No	Yes	No
5	No	No	Yes

Table 2: All condtions used for each emotion

Because I combined two different models, I only used the four emotions both models contain: Anger, Fear, Happiness and Sadness. All movies were recorded at two different locations. Each location had a different angle from where the emotion was recorded. In total there were 4 emotions, 2 locations/angles and 5 conditions for a total of 40 movies.

4.3 Task

During the experiment, participants were shown one gif per page on which they had to choose one primary emotion from the list. They also had the option to indicate a secondary emotion. These emotions were all emotions from the pose model by Beck et al. (2012) and the colour model by Plutchik (2001) combined; Anger, Disgust, Excitement, Fear, Happiness, Neutral, Pride, Sadness, Surprise, Trust or Other:.... The other emotion could be filled in by the participant. They were instructed to imagine they were having a conversation with the Nao. The response times for the full page, containing both the primary and the secondary classification was recorded.

5 Experiment

5.1 Participants

25 participants were recruited, mostly by email (9 females and 16 males) ranging in age from 20 to 57 (M=34.4, SD = 12.63). 1 participant had a Georgian nationality, all others a Dutch nationality.

5.2 Procedure

Participants performed the task at home, or at a public place, at a time of their own choosing. Before the actual experiment, the participants were shown two neutral gifs to test if their browser and computer could show those without problems. They were also shown the list of emotions they could choose from. Directly after they were shown all 40 gifs in succession on separate pages.

6 Results

6.1 Research Question 1

The first research question was about the comparison between poses and lights. The primary emotions participants indicated was used to determine the correct recognition rate. Response times were also used in this analysis.

6.1.1 Accuracy Analysis

I used a one-way ANOVA with repeated measures, with modality (qualitative: body language, LEDs, both) as the within-subject factor and correct recognition rate (quantitative) as the dependent variable.

6.1.2 Accuracy

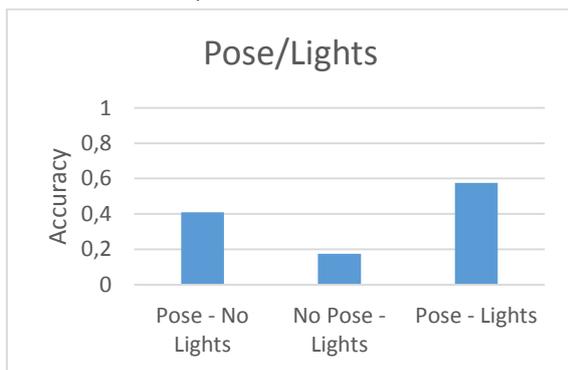


Figure 5: The difference in accuracy between poses, lights and both

A repeated measures ANOVA with a Greenhouse-Geisser correction determined that mean recognition rate differed statistically significantly between the conditions ($F(1.804, 43.297) = 63.733$, $p < 0.0005$). This effect is strong ($\eta^2 = 0.726$). Post hoc tests revealed that the difference between all conditions was statistically significant in all cases ($p < 0.0005$ for all cases). Therefore we can conclude that emotions simulated by poses have a higher recognition rate than emotions simulated by lights, but using both poses and lights reaches the highest recognition rate.

6.1.3 Recognition Rates per Emotion

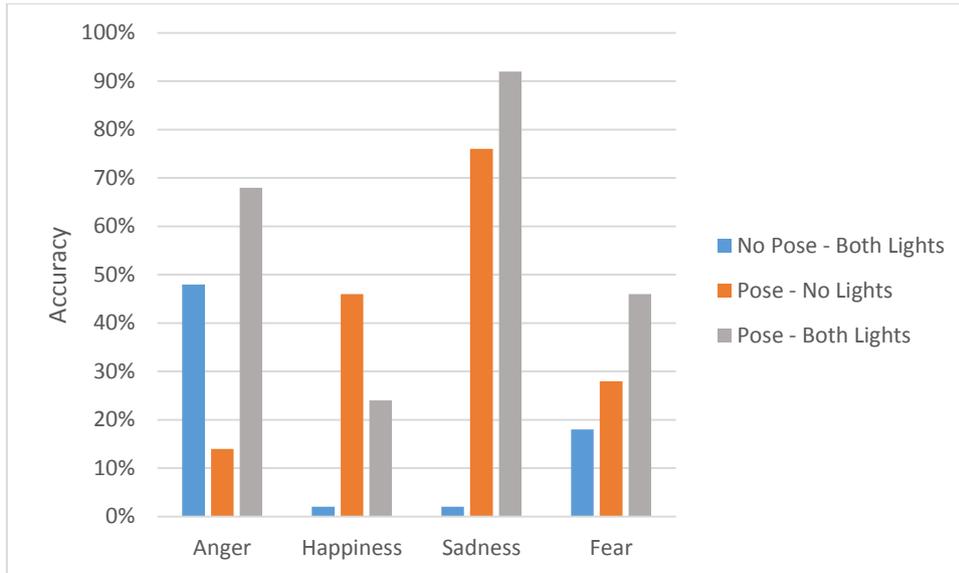


Figure 6: The recognition rate per emotion for RQ1. Exact values can be found in Table 5 in the appendix.

As shown in Figure 6, only when using lights, two emotions (Happiness and Sadness) are recognised below chance level (10%). In all other cases the recognition rate is above chance level. In addition using pose outperforms using lights, except for anger. The combination outperforms pose, except for Happiness, which is the only emotion that is harder to recognize when adding lights to the poses.

6.1.4 Interpretations

Modality	Emotion	Most Common Primary Interpretation	Most Common Secondary Interpretation
Lights	Anger	Anger 48%	None 74%
	Happiness	Surprise 36%	None 76%
	Sadness	Neutral 26%	None 78%
	Fear	Neutral 28%	None 74%
Pose	Anger	Trust 22%	None 84%
	Happiness	Happiness 46%	None 76%
	Sadness	Sadness 76%	None 94%
	Fear	Fear 28%	None 60%
Both	Anger	Anger 68%	None 88%
	Happiness	Happiness 24%	None 50%
	Sadness	Sadness 92%	None 90%
	Fear	Fear 46%	None 60%

Table 3: The primary and secondary interpretations for each emotion for lights, pose and both.

As shown in Table 3, when using lights, both Sadness and Fear are perceived as Neutral and Happiness as Surprise. Using a pose, Anger is primarily recognised as Trust. In all other cases the correct emotion was the primary interpretation. Secondary interpretations were primarily not used.

6.1.5 Reaction Time Analysis

I used a one-way ANOVA with repeated measures, with modality (qualitative: body language, LEDs, both) as the within-subject factor and correct recognition rate (quantitative) as the dependent variable.

6.1.6 Reaction Times

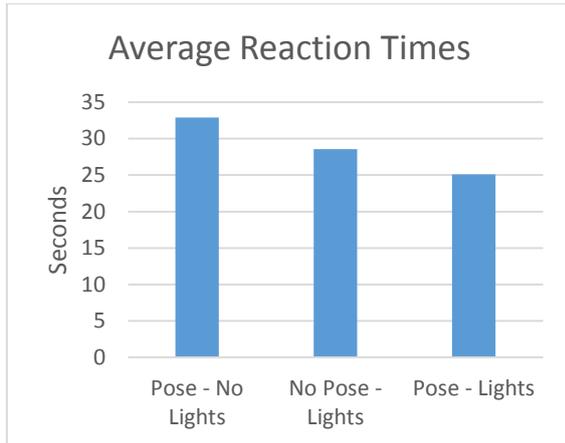


Figure 7: The average reaction time for pose, lights and the combination.

A repeated measures ANOVA with a Greenhouse-Geisser correction determined that mean reaction time did not differ statistically significantly between the conditions ($F(1.650, 39.608) = 1.824, p > 0.1$). Therefore we can conclude that there is no real difference in reaction time between emotions shown by poses, lights and the combination of them.

6.2 Research Question 2

The second research question was about the comparison of different locations of the LEDs. The primary emotions participants indicated was used to determine the correct recognition rate. Response times were also used in this analysis.

6.2.1 Accuracy Analysis

I used a one-way ANOVA with repeated measures, with modality (qualitative: eyes, chest, both) as the within-subject factor and correct recognition rate (quantitative) as the dependent variable.

6.2.2 Accuracy

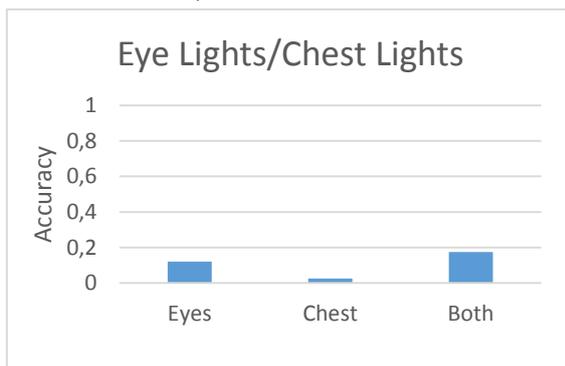


Figure 8: The average accuracy for eye lights, chest lights and both lights.

A repeated measures ANOVA with a Greenhouse-Geisser correction determined that mean recognition rate differed statistically significantly between the conditions ($F(1.673, 40.143) = 18.760, p = 0.0005$). This effect is medium ($\eta^2 = 0.439$). Post hoc tests revealed that the difference between all conditions was statistically significant in all cases ($p = 0.018$ for eyes vs both, $p < 0.0005$ for all other cases). Therefore we can conclude that emotions simulated by lights on the eyes have a higher recognition rate than lights on the chest, but using both eye lights and chest lights reaches the highest recognition rate.

6.2.3 Recognition Rates per Emotion

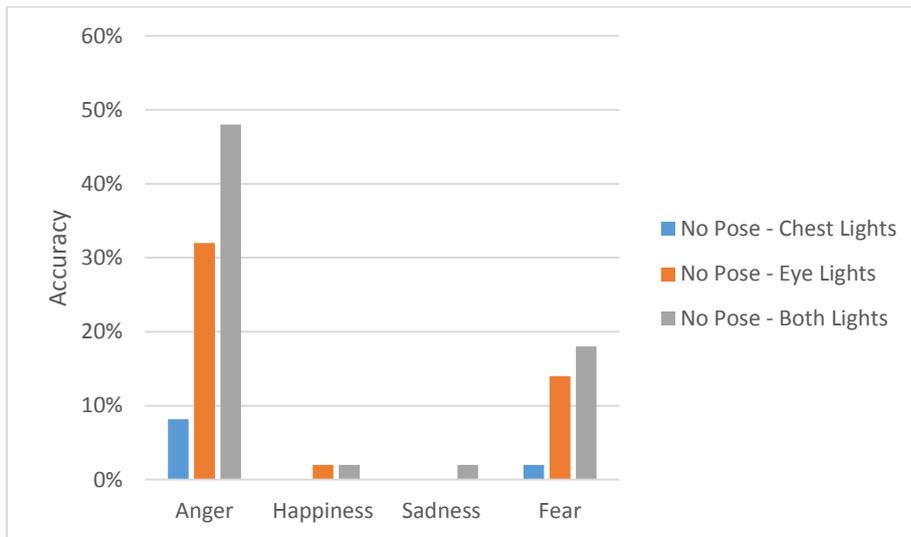


Figure 9: The recognition rate per emotion for RQ2. Exact values can be found in Table 6 in the appendix.

As shown in Figure 9, only for the emotions Anger and Fear the recognition rate is higher than chance level (10%), but only when using at least the eye lights. Using both lights seems better recognisable.

6.2.4 Interpretations

Modality	Emotion	Most Common Primary Interpretation		Most Common Secondary Interpretation	
		Primary	Percentage	Secondary	Percentage
Chest	Anger	Neutral	48%	None	84%
	Happiness	Neutral	62%	None	82%
	Sadness	Neutral	64%	None	84%
	Fear	Neutral	56%	None	80%
Eyes	Anger	Anger	32%	None	82%
	Happiness	Neutral	34%	None	80%
	Sadness	Surprise	30%	None	78%
	Fear	Surprise	30%	None	84%
Both	Anger	Anger	48%	None	74%
	Happiness	Surprise	36%	None	76%
	Sadness	Neutral	26%	None	78%
	Fear	Neutral	28%	None	74%

Table 4: The primary and secondary interpretations for each emotion for chest lights, eye lights and both.

As show in Table 4, only Anger is primarily recognised, if at least the eye lights are used. In all other cases, the emotion is incorrectly perceived as either Neutral or Surprise. Secondary emotions were primarily not used.

6.2.5 Reaction Time Analysis

I used a one-way ANOVA with repeated measures, with modality (qualitative: eyes, chest, both) as the within-subject factor and reaction time (quantitative) as the dependent variable.

6.2.6 Reaction Times

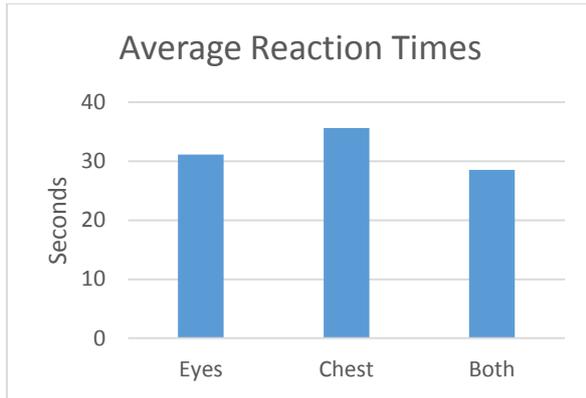


Figure 10: The average reaction time for pose, lights and the combination.

A repeated measures ANOVA with a Greenhouse-Geisser correction determined that mean reaction time did not differ statistically significantly between the conditions ($F(1.635, 39.241) = 0.705, p > 0.1$). Therefore we can conclude that there is no real difference in reaction time between emotions shown by lights on the eyes, on the chest or on both.

7 Discussion

7.1 Research question 1

As expected, using both poses and lights is best recognisable, with the exception of Happiness. In that case, the primary emotion Happiness decreases. At the same time, Surprise increases. A few participants noticed that when using eye lights, the Nao seemed surprised, because the eyes seemed larger. Though this has only a major influence for Happiness. Beck et al. (2012) also found that the Happiness pose was mostly confused with Surprise.

While recognition rate did increase, the difference in reaction times was not statistically significant. So even if the emotions are better recognised, they are not faster recognised.

I did not replicate the findings of the studies by Beck et al. (2012) and by Terada et al. (2012). Anger was hard to recognise using poses, and easy when using lights, and Happiness vice-versa. The aforementioned studies found the exact opposite. Poses might be harder or easier to recognise because the movies shown were from fixed angles, while in the previous study by Beck et al. (2012), the Nao could be examined from all angles.

The only difference for the study by Terada et al. (2012), is that during this research relatively small LEDs were used as opposed to a larger orb that changes its whole colour, which can influence the recognition rates for emotions.

7.2 Research question 2

As expected, the eye lights were easier to recognise than the chest lights, and the combination was best recognisable. However, while the recognition rate does increase, not every emotion is recognisable above chance level, except for Anger and in a lesser amount Fear. On further inspection, chest lights are usually recognised as a neutral state, and eye lights as a surprised state.

As was the case with emotions, even when recognition rate did increase, the difference in reaction times was not statistically significant. So even if the emotion are better recognised, they are not faster recognised.

Participants reported that a chest lights sometimes looks like a power light, so they did not recognise an emotion in it. As mentioned before, the eyes seemed to grow when using those lights, resulting in a surprised emotion.

8 Conclusion

Using lights to show emotions can work quite well, but usually only when used in combination with poses. Usually the combination of these two modalities results in participants more accurately recognizing emotions. While the recognition rate does increase, the response times of the participants does not change. The LEDs do have the advantage of being both cheap and easy to implement, so LEDs can be used quite often to improve the recognisability of emotions.

This study shows that lights located on the chest are generally not perceived as an emotion, as opposed to lights located at the eyes. So, when using lights to show emotions, the location of the lights it is important.

9 Future Work

This research was performed using movies from two angles. Previous research placed the Nao directly in front of participants. This can influence the participants' perception of the poses. To improve this, I would suggest a hands-on experiment.

This study was limited to four emotions. While it does show that lights and poses can be used to express these emotions, there is no information about more emotions. Studying more emotions would show if these modalities work for more emotions.

All body language in this experiment was designed to be static. This is a bit unnatural and impractical. Therefore I would suggest designing new body language that is less static, or optimally body language that is combined with another task, for instance walking with emotional body language.

This research showed that lights can be a useful addition to poses when showing emotions, but whether or not this is useful for more modalities than body language has not been researched. Comparing lights to more modalities, such as facial emotions, voiced emotions or artificial modalities, would give more information about how the lights can be used in more practical applications.

It is very well possible that the colours can be learned over a short amount of time, if people see lights in combination with a natural modality. So whether or not there is some kind of learning effect would be a possibility for future research.

10 References

- Beck, A., Stevens, B., Bard, K. A., & Cañamero, L. (2007). Emotional body language displayed by artificial agents. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(1), 2.
- Brooks, R. A. (1997). From earwigs to humans. *Robotics and autonomous systems*, 20(2), 291-304.
- Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Leem S., Neumann, U., & Narayanan, S. (2004). Analysis of emotion recognition using facial expressions, speech and multimodal information. *Proceedings of the 6th international conference on Multimodal interfaces*, 205-211.
- Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*, 42(3), 143-166.
- Gunes, H., & Piccardi, M. (2007). Bi-modal emotion recognition from expressive face and body gestures. *Journal of Network and Computer Applications*, 30(4), 1334-1345.
- Kobayashi, H., & Hara, F. (1993). Study on face robot for active human interface-mechanisms of face robot and expression of 6 basic facial expressions. *Robot and Human Communication, 1993. Proceedings., 2nd IEEE International Workshop*, 276-281.
- Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *Robotics & Automation Magazine, IEEE*, 19(2), 98-100.
- Plutchik, R. (2001). The nature of emotions. *American Scientist*, 89(4), 344-350.
- Saldien, J., Goris, K., Vanderborght, B., Vanderfaeillie, J., & Lefeber, D. (2010). Expressing emotions with the social robot probot. *International Journal of Social Robotics*, 2(4), 377-389.
- Terada, K., Yamauchi, A., & Ito, A. (2012). Artificial emotion expression for a robot by dynamic color change. *RO-MAN, 2012 IEEE*, 314-321.

11 Appendix

Modalities	Anger	Happiness	Sadness	Fear
No Pose - Lights	48%	2%	2%	18%
Pose - No Lights	14%	46%	76%	28%
Pose - Lights	68%	24%	92%	46%

Table 5: The recognition rate per emotion between poses, lights and both. Chance level is 10%

Modalities	Anger	Happiness	Sadness	Fear
Chest Lights	8%	0%	0%	2%
Eye Lights	32%	2%	0%	14%
Both Lights	48%	2%	2%	18%

Table 6: The recognition rate per emotion between chest lights, eye lights and both. Chance level is 10%

Emotion	Hue	Colour Name	Period (ms)
Anger (Rage)	1	Red	312
Happiness (Ecstasy)	42	Yellow-Red	748
Fear (Terror)	280	Purple	1220
Sadness (Grief)	242	Blue-Purple	3124

Table 7: The hue colour and period settings for the LEDs for each used emotion. The settings originate from Terada et al. (2012)