The Effects of Eyebrows On Perceived Emotional Expression

A Case Study Using Pluggable Eyebrows on NAO Robots

HEATHER KEMP, University of Iowa, USA

One of the leading robots in the field of robot-human interaction is the NAO robot. While the NAO offers many useful features like facial recognition, gestures, text-to-speech audio, and changing eye colors, it notably tends to lack the ability to express emotions without utilizing these features. This work builds off of a previously proposed notion of pluggable eyebrows for the NAO robot to see if, in a variety of experiments, static eyebrows would assist in expressing emotions more clearly or would even cause a change in emotion perception during a performance and while remaining still.

Additional Key Words and Phrases: Facial expressions, emotions, NAO robot, eyebrows, humanoid robots

1 INTRODUCTION

Eyebrows or facial muscle movement tend to be key indicators of what emotion is being expressed by a person or an object. When it comes to the NAO robots, however, neither of these features are available, making it increasingly challenging to convey a wide variety of emotions with only movement and speech. Due to the inability to change any facial muscles on the robot, it was decided that attempting to add eyebrows would be the best way to begin expressing a wider breadth of emotions. The studies described in this paper are a preliminary attempt at analyzing if the inclusion of pluggable eyebrows on the NAO robot will assist with its ability to convey different emotional expressions. It should be noted that due to the preliminary nature of this study that only static eyebrows were utilized.

These studies were performed at a STEM night at Lucas Elementary School in Iowa City. As elementary school students and their parents stopped by, their were asked to either participate in the static robot study (Section 2) or the performing robot study (Section 3). Students and their parents in attendance were given one sheet based on a random time during the night. To observe their right to decline, participants were allowed to turn in a blank sheet if they did not wish to participate in the study. This study was thus a single-blind experiment.

2 EXPERIMENTS WITH STATIC ROBOTS

The first set of experiments are interested in the emotions that the robot can convey without movement or sound. In this paper questionnaire, which was filled out by 49 people in varying age groups, the participants were asked the following questions:

- How do you think that Taylor (Blue Robot) is feeling?
- How do you think that Alex (Orange Robot) is feeling?
- How do you think that Erin (Blue Robot) is feeling?
- How do you think that Jackie (Orange Robot) is feeling?

With the following answers as checkbox answers:

- Anger
- Disgust
- Fear

Author's address: Heather Kemp, University of Iowa, Iowa City, IA, 52242, USA, 2heatherk@gmail.com.

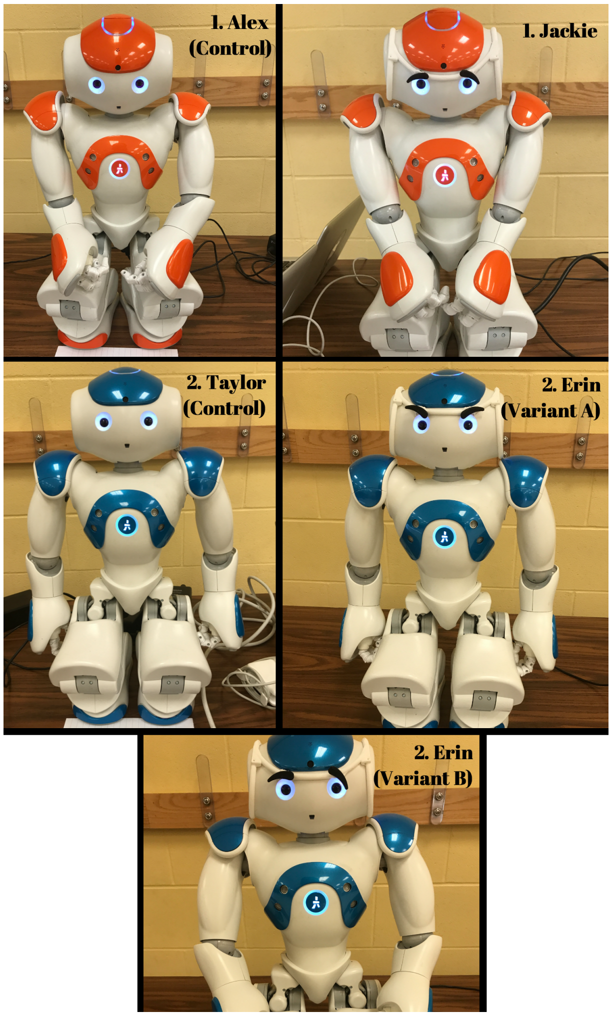

Fig. 1. Robot pair 1 (Alex and Jackie) are attempting to express sadness. Robot pair 2 (Taylor and Erin) are attempting to express anger in variant A and confusion in variant B.

- Happiness
- Sadness
- Surprise
- Confusion
- Neutral/No Emotion
- I Don't Know

In this questionnaire, participants had to identify the robot emotion(s) without any suggested bias from the surveyors. "I don't know" was provided as an option to ensure that participants were not answering at random.

Both robots in a sample pair were in the same position and were of the same color, with the only difference being the inclusion of eyebrows (Figure 1). The robot without eyebrows was considered to be the control robot in this study, against its corresponding pair, which was attempting to convey

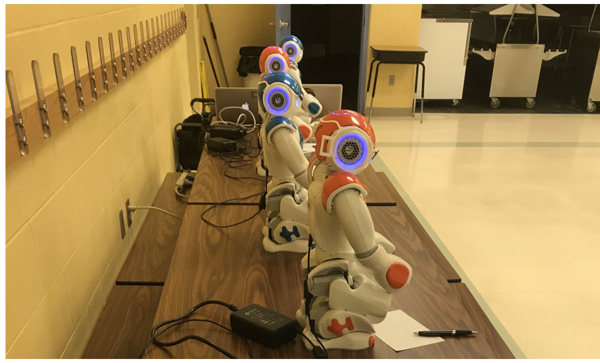

Fig. 2. The robots placed on the table (in order from closest to furthest): Jackie, Erin, Alex, Taylor

Robot Name	Resulting Data Set
Jackie	0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0
Taylor [Angry]	0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
Taylor [Confused]	0,0,0,0,0,0,0,0,0,0,0,0
Alex	0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
Erin [Angry]	0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0
Erin [Confused]	1,0,0,0,0,0,0,0,0,0,0

Table 1. Converted data for expressing *any* emotion

the same emotion. All robots were placed stationary on a table in a random order with name tags to prevent confusion between the different robots.

2.1 Experiment 1 - Expressing Any Emotion

In this experiment, we compare the results of the NAO robot expressing any emotion with or without the eyebrows in its color pair.

We convert the survey data as follows: If any of the emotions provided were selected, a value of 0 is assigned to that survey. In the case of blank entries, no response was recorded. In the case of multiple selected values, if at least one emotion was selected, without an "I Don't Know", the value of 0 was assigned; otherwise, a 1. Only "I Don't Know" answers were given a 1.

Thus, our resulting data is as seen in Table 1, which, when the appropriate formulas are applied to each of these data sets, produces the figures seen in Table 2.

In these tests, we will be comparing against a hypothesized mean of 0, or that is to say, our null hypothesis is that there is no difference between a robot with eyebrows and a robot without eyebrows. Simply put, if we fail to reject the null hypothesis, then there was no difference in that case, while if we do reject the null hypothesis, the result is significant enough to indicate that the eyebrows did change something.

In these t tests, we compute the t-value as follows in Algorithm 1.

Robot Name	Sample Size	Mean $\bar{x}$	Sample Standard Deviation s
Jackie	48	0.020833333333	0.144337567
Taylor [Angry]	36	0.027777777778	0.166666666667
Taylor [Confused]	12	0	0
Alex	49	0.0408163265	0.1999149479
Erin [Angry]	36	0	0
Erin [Confused]	12	0.08333333333	0.28867513

Table 2. Sample size, mean, and standard deviation for the data on expressing *any* emotion

$$t = \frac{sampleMean - hypothesizedMean}{sampleMeanStandardError} \quad (1)$$

2.1.1 Jackie Vs Alex.

Results. This is a comparison of means test, for which we'll be using a t-test. To compare the two "sad" robots, Jackie and Alex or the robot with eyebrows and the robot without eyebrows respectively, we compute the value of the t-test statistic as follows:

$$H_o : \bar{x}_a - \bar{x}_j = 0$$

$$\bar{x}_a = 0.0408163265 \qquad \bar{x}_j = 0.020833333333$$
$$s_a = 0.1999149479 \qquad s_j = 0.144337567$$
$$n_a = 49 \qquad n_j = 48$$

The resulting t-statistic, -0.564, with 95 degrees of freedom, when cross referenced with a T-table, produce the P-value = 0.5741.

Using a .05 significance level, we fail to reject H_o (since 0.5741 > .05), suggesting that the difference witnessed between these two samples is not significant in this case. That is, in this case, the eyebrows in a sad manner did not alter the detecting of *any* emotion.

2.1.2 Erin Vs Taylor [Variant A: Angry.

Results. This is a comparison of means test, for which we'll be using a t-test. To compare the two "angry" robots, Erin and Taylor, or the robot with eyebrows and the robot without eyebrows respectively, we compute the value of the t-test statistic as follows:

$$H_o : \bar{x}_e - \bar{x}_t = 0$$

$$\bar{x}_e = 0 \qquad \bar{x}_t = 0.027777777778$$
$$s_e = 0 \qquad s_t = 0.166666666667$$
$$n_e = 36 \qquad n_t = 36$$

The resulting t-statistic, -1.000, with 70 degrees of freedom, when cross referenced with a T-table, produce the P-value = 0.3208.

Using a .05 significance level, we fail to reject H_o (since 0.3208 > .05), suggesting that the difference witnessed between these two samples is not significant in this case. That is, in this case, the eyebrows in an angry manner did not alter the detecting of *any* emotion.

2.1.3 Erin Vs Taylor [Variant B: Confused].

Robot Pair	Resulting Data Set
Alex Vs Jackie	0,0,0,0,0,0,0,0,0,0,0,1,1,1,0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0
Erin Vs Taylor [Variant A: Angry]	0,0,0,0,0,0,1,0,0,0,0,0,1,1,0,0,0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
Erin Vs Taylor [Variant B: Confused]	0,0,1,0,0,0,0,0,0,0,0,0

Table 3. Converted data for change in emotional perception

Robot Pair	Sample Size	Mean $\bar{x}$	Sample Standard Deviation s
Alex Vs Jackie	48	0.104166666667	0.3087092782
Erin Vs Taylor [Variant A: Angry]	36	0.08333333333	0.2803059
Erin Vs Taylor [Variant B: Confused]	12	0.08333333333	0.28867513

Table 4. Sample size, mean, and standard deviation for converted data on change in emotional perception

Results. This is a comparison of means test, for which we'll be using a t-test. To compare the two "confused" robots, Erin and Taylor, or the robot with eyebrows and the robot without eyebrows respectively, we compute the value of the t-test statistic as follows:

$$H_o : \bar{x}_e - \bar{x}_t = 0$$

$$\bar{x}_e = 0.08333333333 \qquad \bar{x}_t = 0$$
$$s_e = 0.28867513 \qquad s_t = 0$$
$$n_e = 12 \qquad n_t = 12$$

The resulting t-statistic, -1.000, with 22 degrees of freedom, when cross referenced with a T-table, produce the P-value = 0.3282.

Using a .05 significance level, we fail to reject H_o (since 0.3282 > .05), suggesting that the difference witnessed between these two samples is not significant in this case. That is, in this case, the eyebrows in a confused manner did not alter the detecting of *any* emotion.

2.2 Experiment 2 - Change in Emotional Perception

In this experiment, we compare the results of the NAO robot expressing the same emotion with or without the eyebrows in its color pair.

We convert the survey data as follows: If the emotions identified match between the two robots in a color pair, then a value of 1 is assigned to that survey. In the case of blank entries, the entire survey response was ignored. In the case of multiple responses, a 1 was assigned if there was at least one emotion commonly identified, with a 0 being given otherwise.

Thus, our resulting data is as seen in Table 3, which, when the appropriate formulas are applied to each of these data sets, produces the figures seen in Table 4.

In these tests, we will be comparing against a hypothesized mean of 0, or that is to say, our null hypothesis is that there is no difference between a robot with eyebrows and a robot without eyebrows. Simply put, if we fail to reject the null hypothesis, then there was no difference in that case, while if we do reject the null hypothesis, the result is significant enough to indicate that the eyebrows did change something.

In these t tests, we compute the t-value as seen in Algorithm 1.

2.2.1 *Alex Vs Jackie.*

Results. This is a comparison of means test, for which we'll be using a t-test. To compare the change in emotion perception between the two "sad" robots, Jackie and Alex, or the robot with eyebrows and the robot without eyebrows respectively, we compute the value of the t-test statistic as follows:

$$H_o : \bar{x}_{ja} = 0$$
$$\bar{x}_{ja} = 0.104166666667$$
$$s_{ja} = 0.3087092782$$
$$n_{ja} = 48$$

The resulting t-statistic, -2.338, with 47 degrees of freedom, when cross referenced with a T-table, produce the P-value = 0.0237.

Using a .05 significance level, we reject H_o (since 0.0237 < .05), suggesting that the difference witnessed between these two samples is significant in this case. That is, in this case, the eyebrows in a sad manner altered the probability of detecting a different emotion.

2.2.2 Erin Vs Taylor [Variant A: Angry] .

Results. This is a comparison of means test, for which we'll be using a t-test. To compare the change in emotion perception between the two "angry" robots, Erin and Taylor, or the robot with eyebrows and the robot without eyebrows respectively, we compute the value of the t-test statistic as follows:

$$H_o : \bar{x}_{et} = 0$$
$$\bar{x}_{et} = 0.08333333333$$
$$s_{et} = 0.2803059$$
$$n_{et} = 36$$

The resulting t-statistic, 1.784, with 36 degrees of freedom, when cross referenced with a T-table, produce the P-value = 0.0831.

Using a .05 significance level, we fail to reject H_o (since 0.0831 > .05), suggesting that the difference witnessed between these two samples is not significant in this case. That is, in this case, the eyebrows in an angry manner did not alter the probability of detecting a different emotion.

2.2.3 Erin Vs Taylor [Variant B: Confused].

Results. This is a comparison of means test, for which we'll be using a t-test. To compare the change in emotion perception between the two "confused" robots, Erin and Taylor, or the robot with eyebrows and the robot without eyebrows respectively, we compute the value of the t-test statistic as follows:

$$H_o : \bar{x}_{et} = 0$$
$$\bar{x}_{et} = 0.08333333333$$
$$s_{et} = 0.28867513$$
$$n_{et} = 12$$

The resulting t-statistic, 1.000, with 11 degrees of freedom, when cross referenced with a T-table, produce the P-value = 0.3388.

Using a .05 significance level, we fail to reject H_o (since 0.3388 > .05), suggesting that the difference witnessed between these two samples is not significant in this case. That is, in this case, the eyebrows in a confused manner did not alter the probability of detecting a different emotion.

2.3 Experiment 3 - Detecting the Right Emotion

In this experiment, we compare the results of the NAO robot expressing the correct emotion with or without the eyebrows in its color pair.

Robot Name	Resulting Data Set
Jackie	1,1,0,0,1,1,0,1,0,1,1,1,0,1,0,0,0,1,1,0,0,0, 0,1,1,0,1,0,0,0,1,0,1,1,0,1,1,1,1,0,1,0,1,0,0,0,0,1
Taylor [Angry]	0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
Taylor [Confused]	0,0,1,0,0,0,0,1,0,0,0,0
Alex	0,0,0,1,0,0,0,0,0,1,0,0,1,1,1,0,1,1,0,0,0,0,1,1, 0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,1,1,0,0
Erin [Angry]	1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0, 0,0,1,0,0,1,1,1,1,1,1,1,1,1,1,0,1,1,1
Erin [Confused]	0,0,1,1,1,0,1,0,1,1,0,0

Table 5. Converted data for detecting the right emotion

Robot Name	Sample Size	Mean $\bar{x}$	Sample Standard Deviation s
Jackie	48	0.5	0.50529115
Taylor [Angry]	37	0.027027027	0.164398987
Taylor [Confused]	12	0.1666666666667	0.3892495
Alex	49	0.26530612245	0.4460712856
Erin [Angry]	36	0.7777777778	0.421637021356
Erin [Confused]	12	0.5	0.5222329678671

Table 6. Sample size, mean, and standard deviation for the converted data on detecting the right emotion

We convert the survey data as follows: We mark the surveys with the emotion we wish to convey as a 1, and all other emotions as a 0. Responses with multiple emotions marked were counted as 1 if the correct emotion was among the chosen emotions, with a single 0 being given otherwise. Neutral/No Emotion and I Don't Know were marked as 0's. Blank responses were not included.

Thus, our resulting data is as seen in Table 5, which, when the appropriate formulas are applied to each of these data sets, produces the figures seen in Table 6.

In these tests, we will be comparing against a hypothesized mean of 0, or that is to say, our null hypothesis is that there is no difference between a robot with eyebrows and a robot without eyebrows. Simply put, if we fail to reject the null hypothesis, then there was no difference in that case, while if we do reject the null hypothesis, the result is significant enough to indicate that the eyebrows did change something.

In these t tests, we compute the t-value as follows in Algorithm 2.

$$t = \frac{\bar{x_1} - \bar{x_2}}{\bar{x_1} - \bar{x_2}} \tag{2}$$

With the standard error being calculated as follows in Algorithms 3 and 4.

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \tag{3}$$

$$se(\bar{x_1} - \bar{x_2}) = s * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{4}$$

2.3.1 *Jackie Vs Alex.*

Results. This is a comparison of means test, for which we'll be using a t-test. To compare the two "sad" robots, Jackie and Alex, or the robot with eyebrows and the robot without eyebrows respectively, we compute the value of the t-test statistic as follows:

$$H_o : \bar{x}_a - \bar{x}_j = 0$$

$\bar{x}_a = 0.26530612245$ $\qquad$ $\bar{x}_j = 0.5$

$s_a = 0.4460712856$ $\qquad$ $s_j = 0.50529115$

$n_a = 49$ $\qquad$ $n_j = 48$

The resulting t-statistic, -2.486, with 95 degrees of freedom, when cross referenced with a T-table, produce the P-value = 0.0147.

Using a .05 significance level, we reject H_o (since $0.0147 < .05$), suggesting that the difference witnessed between these two samples is significant in this case. That is, in this case, the eyebrows in a sad manner altered the ability to detect the correct emotion.

2.3.2 *Erin Vs Taylor [Variant A: Angry.*

Results. This is a comparison of means test, for which we'll be using a t-test. To compare the two "angry" robots, Erin and Taylor, or the robot with eyebrows and the robot without eyebrows respectively, we compute the value of the t-test statistic as follows:

$$H_o : \bar{x}_e - \bar{x}_t = 0$$

$\bar{x}_e = 0.7777777778$ $\qquad$ $\bar{x}_t = 0.027027027$

$s_e = 0.421637021356$ $\qquad$ $s_t = 0.164398987$

$n_e = 36$ $\qquad$ $n_t = 36$

The resulting t-statistic, -10.074, with 71 degrees of freedom, when cross referenced with a T-table, produce the P-value < 0.0001.

Using a .05 significance level, we reject H_o (since $p < 0.0001 < .05$), suggesting that the difference witnessed between these two samples is highly significant in this case. That is, in this case, the eyebrows in an angry manner altered the ability to detect the correct emotion.

2.3.3 *Erin Vs Taylor [Variant B: Confused].*

Results. This is a comparison of means test, for which we'll be using a t-test. To compare the two "confused" robots, Erin and Taylor, or the robot with eyebrows and the robot without eyebrows respectively, we compute the value of the t-test statistic as follows:

$$H_o : \bar{x}_e - \bar{x}_t = 0$$

$\bar{x}_e = 0.5$ $\qquad$ $\bar{x}_t = 0.1666666666667$

$s_e = 0.5222329678671$ $\qquad$ $s_t = 0.3892495$

$n_e = 12$ $\qquad$ $n_t = 12$

The resulting t-statistic, -1.773, with 22 degrees of freedom, when cross referenced with a T-table, produce the P-value = 0.0901.

Using a .05 significance level, we fail to reject H_o (since $0.0901 > .05$), suggesting that the difference witnessed between these two samples is not significant in this case. That is, in this case, the eyebrows in a confused manner did not alter the ability to detect the correct emotion.

2.4 Experiment 4 - Recognition Improvement

In this experiment, we compare the results of the change in correctness for the NAO robot expressing the correct emotion with or without the eyebrows in its color pair.

$$\begin{array}{cc} Eyebrows.Correct & Eyebrows.Incorrect \\ \end{array}$$
$$\begin{pmatrix} 3 & 10 \\ 21 & 14 \end{pmatrix} \begin{array}{l} NoEyebrows.Correct \\ NoEyebrows.Incorrect \end{array}$$

Fig. 3. Alex Vs Jackie

$$\begin{array}{cc} Eyebrows.Correct & Eyebrows.Incorrect \\ \end{array}$$
$$\begin{pmatrix} 1 & 0 \\ 27 & 8 \end{pmatrix} \begin{array}{l} NoEyebrows.Correct \\ NoEyebrows.Incorrect \end{array}$$

Fig. 4. Taylor Vs Erin [Angry]

$$\begin{array}{cc} Eyebrows.Correct & Eyebrows.Incorrect \\ \end{array}$$
$$\begin{pmatrix} 1 & 1 \\ 5 & 5 \end{pmatrix} \begin{array}{l} NoEyebrows.Correct \\ NoEyebrows.Incorrect \end{array}$$

Fig. 5. Taylor Vs Erin [Confused]

We convert the survey data as follows: we create a matrix broken up into NoEyebrows.Correct, NoEyebrows.Incorrect, Eyebrows.Correct, and Eyebrows.Incorrect, with values being tallied into the corresponding square based on the correctness as defined in Section 2 Experiment 3.

Thus, our resulting data is as seen in Figures 3, 4, and 5.

In these tests, our null hypothesis is that the tables are symmetric, or the probability of cell [i, j] is equal to the probability of cell [j, i]. If we reject this hypothesis, we are saying the tables are in-fact asymmetric, or that there was a significant change from without eyebrows to with eyebrows.

In these tests, we use a nominal symmetry test as defined by Salvatore S. Mangiafico.

2.4.1 Jackie Vs Alex.

Results. H_o : The table in Figure 3 is symmetrical

When applying a nominal symmetry test for the results from comparing the two "sad" robots, Jackie and Alex, or the robot with eyebrows and the robot without eyebrows respectively, we compute the p-value as 0.0708.

Using a .05 significance level, we fail to reject H_o (since 0.0708 > .05), suggesting that the difference witnessed between these two samples is not significant in this case. That is, in this case, the eyebrows in a sad manner did not cause a significant change in perception.

2.4.2 Erin Vs Taylor [Variant A: Angry].

Results. H_o : The table in Figure 4 is symmetrical

When applying a nominal symmetry test for the results from comparing the two "angry" robots, Erin and Taylor, or the robot with eyebrows and the robot without eyebrows respectively, we compute the p-value as 0.0000000149.

Using a .05 significance level, we reject H_o (since 0.0000000149 < .05), suggesting that the difference witnessed between these two samples is significant in this case. That is, in this case, the eyebrows in an angry manner caused a significant change in perception.

2.4.3 *Erin Vs Taylor [Variant B: Confused].*

Results. H_o : The table in Figure 5 is symmetrical

When applying a nominal symmetry test for the results from comparing the two "confused" robots, Erin and Taylor, or the robot with eyebrows and the robot without eyebrows respectively, we compute the p-value as 0.219.

Using a .05 significance level, we fail to reject H_o (since 0.219 > .05), suggesting that the difference witnessed between these two samples is not significant in this case. That is, in this case, the eyebrows in a confused manner did not cause a significant change in perception.

3 EXPERIMENTS WITH PERFORMING ROBOTS

The second set of experiments are interested in the emotions that the robot can convey during a performance. In this paper questionnaire, which was filled out by 39 people in varying age groups, the participants were asked the following questions:

- In the skit just performed, what is the emotion expressed by the orange robot?
- In the skit just performed, what is the emotion expressed by the blue robot?

With the following answers as checkbox answers:

- Anger
- Disgust
- Fear
- Happiness
- Sadness
- Surprise
- Confusion
- Neutral/No Emotion
- I Don't Know

In this questionnaire, participants had to identify the robot emotion(s) without any suggested bias from the surveyors. "I don't know" was provided as an option to ensure that participants were not answering at random.

Similarly to the first set of experiments, both robots in a sample pair were given the same Choreograph program and were of the same color, with the only difference being the inclusion of eyebrows (Figure 1). The robots without eyebrows were considered to be the control robots in this study, against its corresponding pair, which was attempting to convey the same emotion in the performance. To prevent confusion between the different robots, robots of opposite colors were using during the skits.

It should be noted that a fourth performance with eyebrows were given, but during said performance, the eyebrows were reversed, or that is, the eyebrows were attempting to convey the emotion that the other robot performer was attempting to convey.

3.1 Experiment 1 - Detecting the Right Emotion

In this experiment, we compare the results of the NAO robot expressing the same emotion with or without the eyebrows in its color pair during a performance.

Robot	Resulting Data Set
Jackie [Sad Eyebrows, Angry Performance]	0,0,1,1,1
Jackie [Angry Eyebrows, Angry Performance]	1,1,0,1,1,1,1,0,1,1,1,1,1,1
Erin [Angry Eyebrows, Sad Performance]	0,0,0,0,1
Erin [Sad Eyebrows, Sad Performance]	1,0,0,1,1,1,1,0,1,0,1,0,0,0
Taylor [Sad]	0,0,0,0,1,0,0,0,1,0,0,0,0,0,1,1,0,1,0,0
Alex [Angry]	0,0,0,0,0,0,0,0,1,1,1,0,0,0,1,0,0,0,0,0

Table 7. Converted data for detecting the correct emotion during a performance

Robot	Sample Size	Mean $\bar{x}$	Sample Standard Deviation s
Jackie [Sad Eyebrows, Angry Performance]	5	0.6	0.54772256
Jackie [Angry Eyebrows, Angry Performance]	14	0.85714286	0.3631365
Erin [Angry Eyebrows, Sad Performance]	5	0.2	0.4472136
Erin [Sad Eyebrows, Sad Eyebrows]	14	0.5	0.5188745
Taylor [Sad]	20	0.25	0.444262
Alex [Angry]	20	0.2	0.41039134

Table 8. Sample size, mean, and standard deviation of the converted data for detecting the correct emotion during a performance

We convert the survey data in the same manner as we convert the data in Section 2.3, which gives us the data seen in Table 7, which, when the appropriate formulas are applied to each of these data sets, produces the figures seen in Table 8.

In these tests, we will be comparing against a hypothesized mean of 0, or that is to say, our null hypothesis is that there is no difference between a robot with eyebrows and a robot without eyebrows. Simply put, if we fail to reject the null hypothesis, then there was no difference in that case, while if we do reject the null hypothesis, the result is significant enough to indicate that the eyebrows did change something.

In these t tests, we compute the t-value as seen in Algorithm 2, with the standard error being calculated as seen in Algorithms 3 and 4.

3.1.1 *Erin [Angry Eyebrows] Vs Taylor.*

Results. This is a comparison of means test, for which we'll be using a t-test. To compare the two "sad" robots, Erin and Taylor, or the robot with angry eyebrows and the robot without eyebrows respectively, we compute the value of the t-test statistic as follows:

$$H_o : \bar{x}_e - \bar{x}_t = 0$$

$$\bar{x}_e = 0.2 \qquad \bar{x}_t = 0.25$$
$$s_e = 0.4472136 \qquad s_t = 0.444262$$
$$n_e = 5 \qquad n_t = 20$$

The resulting t-statistic,0.225, with 23 degrees of freedom, when cross referenced with a T-table, produce the P-value = 0.8241.

Using a .05 significance level, we fail to reject H_o (since 0.8241 > .05), suggesting that the difference witnessed between these two samples is not significant in this case. That is, in this case, the eyebrows in an angry manner on a sad robot did not alter the ability to detect the correct emotion.

3.1.2 Erin [Sad Eyebrows] Vs Taylor.

Results. This is a comparison of means test, for which we'll be using a t-test. To compare the two "sad" robots, Erin and Taylor, or the robot with sad eyebrows and the robot without eyebrows respectively, we compute the value of the t-test statistic as follows:

$$H_o : \bar{x}_e - \bar{x}_t = 0$$

$\bar{x}_e = 0.5$	$\bar{x}_t = 0.25$
$s_e = 0.5188745$	$s_t = 0.444262$
$n_e = 14$	$n_t = 20$

The resulting t-statistic, -1.507, with 32 degrees of freedom, when cross referenced with a T-table, produce the P-value = 0.1416.

Using a .05 significance level, we fail to reject H_o (since 0.1416 > .05), suggesting that the difference witnessed between these two samples is not significant in this case. That is, in this case, the eyebrows in a sad manner on a sad robot did not alter the ability to detect the correct emotion.

3.1.3 Jackie [Sad Eyebrows] Vs Alex.

Results. This is a comparison of means test, for which we'll be using a t-test. To compare the two "angry" robots, Jackie and Alex, or the robot with sad eyebrows and the robot without eyebrows respectively, we compute the value of the t-test statistic as follows:

$$H_o : \bar{x}_j - \bar{x}_a = 0$$

$\bar{x}_j = 0.6$	$\bar{x}_a = 0.2$
$s_j = 0.54772256$	$s_a = 0.41039134$
$n_j = 5$	$n_a = 20$

The resulting t-statistic, -1.829, with 23 degrees of freedom, when cross referenced with a T-table, produce the P-value = 0.0804.

Using a .05 significance level, we fail to reject H_o (since 0.0804 > .05), suggesting that the difference witnessed between these two samples is not significant in this case. That is, in this case, the eyebrows in a sad manner on an angry robot did not alter the ability to detect the correct emotion.

3.1.4 Jackie [Angry Eyebrows] Vs Alex.

Results. This is a comparison of means test, for which we'll be using a t-test. To compare the two "angry" robots, Jackie and Alex, or the robot with sad eyebrows and the robot without eyebrows respectively, we compute the value of the t-test statistic as follows:

$$H_o : \bar{x}_j - \bar{x}_a = 0$$

$\bar{x}_j = 0.85714286$	$\bar{x}_a = 0.2$
$s_j = 0.3631365$	$s_a = 0.41039134$
$n_j = 14$	$n_a = 20$

The resulting t-statistic, -4.812, with 32 degrees of freedom, when cross referenced with a T-table, produce the P-value < 0.0001.

Using a .05 significance level, we reject H_o (since p < 0.0001 < .05), suggesting that the difference witnessed between these two samples is significant in this case. That is, in this case, the eyebrows in an angry manner on an angry robot altered the ability to detect the correct emotion.

4 CONCLUSIONS

Looking at the results of this study raised many interesting questions in regards to what can impact a perceived emotional expression.

Specifically, the comparison between sadness and anger results during the experiment with static robots (Section 2.4). While both results were significant in getting the correct answer, only anger was significant in converting answers from either wrong to right or right to wrong. It's possible that due to the posture of the robots (the sad robot was looking down while the angry robot was looking straight forward), the full impact of the eyebrows was washed out. This could be corroborated with the results of Section 3.1, where eyebrows conveying the opposite emotion did not hinder the ability to detect the correct emotion. The gestures and position of the robots may play a heavier role in perceiving emotions than was anticipated.

This results of Section 3.1, the performing robots, also brought another interesting observation to our attention. While the eyebrows tended to be significant while stationary, the results for the eyebrows during a performance tended to be less significant. Posture and gestures were already discussed above, but a third impacting factor may be the fellow audience members that exist during a performance. During the skit, many of the adults observing tended to let out a surprised laughter at pointed comments from the "angry" robot, and laughed at the song that played ('Love Hurts') for the "sad" robot. During the take with the "angry" eyebrows on the "sad" robot, it was interesting to note that happiness was unanimous, happiness also remained present in many of the different takes. It should be noted that adults were not at every performance, and there were more adults at some performances than others. This, of course, opens up the door to a new research question, which would be how strong of an impact are fellow audience members in your perception of a piece, along with other context clues like voice, posture, and music, as discussed above.

While these results lead to some interesting implications, it was not without its faults. Many of our samples had a small size due to the nature of the event. If one were to test the power of the sample size, its highly probable that most of these samples require more data to indicate any higher significance. Due to the preliminary nature of these experiments, only static eyebrows were used, while many emotions are heightened by the movement of eyebrows as opposed to their initial position. This flaw also lead to eyebrows moving without our intention, which is what caused the branching off into separate takes like confusion for the stationary robots and the reversed emotions for the performing robots. However, even with these flaws, many results turned out to be significant, and thus, further studies are warranted.

5 ACKNOWLEDGEMENTS

We would like to thank Albert De Beir et al from the paper "Enhancing Emotional Facial Expressiveness on NAO: A Case Study Using Pluggable Eyebrows" for the base designs of the eyebrows and headset. Furthermore, we would like to thank Carl Clement from Emotion Robotics for additional debugging support during the creation of this headset and eyebrows.

In regards to the study itself, we would like to thank Lucas Elementary School, along with the students and parents in attendance, for their cooperation in this study, along with Dr. Shelia Barron for assistance with design of the expermient.