Exploring Behaviours for Social Robots

Joana Sá[1] and João Silva Sequeira[1]

Instituto Superior Técnico, University of Lisbon, 1049-001 Lisbon, Portugal,
{joana.sa, joao.silva.sequeira}@tecnico.ulisboa.pt

Abstract. The paper details a set of experiments to assess the perception of basic behaviors for social robots. These include (i) the blinking of the eyes, (ii) the shape of the mouth, (iii) the way the whole body moves, (iv) the motion of the arms, and (v) a collection of non-verbal sounds. The goals are (i) to validate behaviors that can be useful for social robots with simple anthropomorphic features, and (ii) to explore the perception of the audiences selected to interact with the robot, namely how long do they last and the perceived emotions, i.e., the extent to which small behaviours are perceived by people observing the robot. Online questionnaires containing media displaying a Monarch MBot robot undertaking several actions are used to assess people's perception. Inequality indicators (Gini Index, *pq*-means, and Hoyes Index) are used as analysis indicators, in addition to classical statistical tests.

Keywords: Social Robot, Social Behaviours, Gini Index, Likert Questionnaires, Hoyes Index, Inequality indexes

1 Introduction

The research on the field has pointed to numerous aspects, namely acceptance of the robots by people. The uncanny valley paradigm, though not universally accepted, already established a link between the appearance of a robot and the emotional response of people observing it.

The paper presents a collection of experiments aim at understanding how people perceive small physical differences between behaviors thought relevant for social purposes and how important they are for acceptance of social robots. The robot considered is a Mbot, developed within the European project FP7 Monarch, to interact with children. in a hospital environment[1] (a omnidirectional platform with 1-dof arms and neck and simple led-based facial features). The good acceptance of this robot in previous experiments and its basic anthropomorphic features make it interesting for this kind of study. Furthermore, a major part of the populations involved are not new to the robot and hence novelty effects are minimized.

The experiments in the paper aim at demonstrating the importance of small or short-duration behavioural features to the perception of people observing the robot in social contexts.

[1] In the text the robot is named *Casper*, after the name given by the children at the hospital.

The paper is organized as follows. Section 2 presents an overview of the state of the art. Section 3 to 7 detail the Likert questionnaires[2] used to assess people's perception and respective analysis. Section 8 discusses the results obtained and points to future developments.

2 Brief overview of Behaviours for Social Robots

Social robots can be utilitarian (or service robots) or affective social robots (or socially assistive robots), [2]. They are often endowed with anthropomorphic features to facilitate integration and acceptance in human environments, e.g., having a robot moving its arms during motion facilitates the recognition by humans as walking [4].

The effect of gait speed in the upper body kinematics and intersegmental coordination between upper and lower body in healthy individuals was studied in [15], including shoulder, elbow, thoracic, and pelvic movements for 20 healthy subjects walking at six speeds ranging from extremely slow to very fast. As speed increases, there's a significant increase in range of motion (RoM); the extremely slow walking speed has a peak-to-peak difference of roughly $10°$, whereas the very fast walking speed has a difference of about $40°$.

The behaviour of robotic eyes has been recognized as relevant and useful in the field of social robotics [1], e.g., to relay social cues, for example during a conversation, or to convey the psychological state of the person (or robot) exhibiting the behavior [13]. Although complex eye behavior includes gaze, saccades, and blinking, the Mbot has static LEDs for eyes and hence it is only capable of mimicking blinks. It is generally acknowledged that blinking patterns vary greatly depending on several factors including the person itself, the environment, and the task at hand [14]. The work in [8] reported that both spontaneous eyeblink rate (SEBR) and inter-eyeblink interval (IEBI) change substantially depending on the task the subject is doing. Facial expressions are an important medium for communicating emotions between people [12]. The MBot's facial features are limited. Aside from the robot's eye color, the only configurable feature of the robot's face is its mouth LED matrix.

Mouth movements and shapes for the Mbot were shown in [5] and reported that having a mouth the robot is perceived as more life-like and less sad. Additionally, users prefer the human-like mouth, which was rated as friendliest, as opposed to the wavelike mouth. Using a virtual environment, [9] studied users' recognition of eight different emotions displayed by the MBot.

In general, the literature agrees that basic emotions "should be discrete, have a fixed set of neural and bodily expressed components, and a fixed feeling or motivational component that has been selected for through longstanding interactions with ecologically valid stimuli" [17].

[2] Five-point scales are used in all questionnaires.

3 Questionnaires basics

The participants in the online surveys were not required to interact directly with the robot. Media contents displaying the adequate features were included on the questionnaires.

The questionnaires were shared, in person, where people were approached and asked to answer a quick set of questions to evaluate a given behavior being performed by a social robot. Access was granted through QR codes and also distributed through social media, direct messaging, and group chats.

The university environment resulted in most of the answers being from people in the 18 to 24 years old age group.

Evidence that participants' engagement in a questionnaire dropping significantly as the median completion time of the questionnaire increased has been reported in [3]. Therefore, the duration of all questionnaires was kept under five minutes.

Demographic or classification questions were left to the end of the questionnaire to reduce the effect of attention fading [3].

3.1 An argument against standard HRI questionnaires

The trait-specific questions do not follow any HRI questionnaire standards such as the Godspeed or the RoSAS. These were not used since they tend to be extensive and poorly adapted to audiences that quickly shift attention if the duration exceeds a short period.

The semantic complexity of the Godspeed and the RoSAS is also a relevant aspect. Some of the questionnaires presented had children audiences. [18] justify children's (namely girls from the ages of five to eight years old) inability to recognize shame and contempt with the fact that they are unable to conceive the complexity of these emotions and understand the verbal labels given to them. The Godspeed and RoSAS use terms such as "quiescent", "compassionate", and "organic" which may be difficult for children to understand. Finally, as referred in [19], it is still common within the field of HRI to use custom surveys to assess users' subjective perceptions of a robot.

4 Survey 1: Arm movements for a social robot

This questionnaire was designed to assess users' perceptions of the walking arm movement trait. Two variations of arm movement are proposed. A slower one, with a smaller RoM, which will be referred to as "small arm movement". A faster one, with a larger RoM which will be referred to as "large arm movement".

This questionnaire had 202 participants, with a balanced gender representation (51/49%) and ages 18-34 being representing 64% of the participants.

For the trait-specific questions, the users were presented with three videos (one without arm movements and two with the different arm movements) showing the robot navigating between two goals in an indoors lab. The questionnaire contained a single question: "How naturalistic was the robot's arm movement?". This question was shown in the questionnaire a total of 3 times, once for each movement type. This feature was created to anthropomorphize the MBot's "walking" style by modulating a human-like arm movement that rises in amplitude and frequency proportionate to walking speed [15].

Hypothesis under test are the following. (H_1) The users will consider the robot performing either "small arm movement" or "large arm movement" to be more naturalistic than when there is no arm movement; (H_2) The "large arm movement" was designed to be more noticeable since the "small arm movement" was considered imperceptible; the users will consider the "large arm movement" more natural than the "small arm movement".

Table 1 shows the corresponding descriptive statistics. There is a clear user preference for robotic arm movement, as opposed to no arm movement and also that there is a slight preference for the large over the small arm movement.

Table 1. Descriptive statistics for "no arm movement" (No), "small arm movement" (Small) and "large arm movement" (Large).

Statistic	No	Small	Large
Mean	2.324	3.58	3.88
Mean rank	1.35	2.18	2.46
Mode	2	4	4
Median	4	2	4
Standard deviation	1.188	0.934	1.022

Although there is some debate over the use of parametric methods to analyze Likert scale data (see [16]) the interpretation of this data as ordinal, implying the use of nonparametric methods, is acceptable [7]. Therefore, nonparametric statistical methods were also used to interpret the Likert scales data.

A Friedman test was used to prove that there is a relevant difference in participants' opinions about the three arm movements (the samples are dependent since they were drawn from the same group of people).

The null hypothesis is that there is no statistically relevant difference between the three samples, i.e. there is no difference in perceived robotic arm movement naturalness between the three arm movement types. A common level of risk, $\alpha = 0.05$, was selected [7]. Friedman test was significant ($F_{r(2)} = 158.029, p < 0.001$), meaning that the null hypothesis can be rejected.

The Wilcoxon signed ranks and the Sign tests were used for the three comparisons (no movement versus small movement, small movement versus big movement, and no movement versus big movement). However, it is also relevant to quantify the magnitude of the difference between groups, which can be achieved by calculating the effect size, ES (see [7]). ES, ranges from 0 to 1 and grows with the difference between groups, can be classified as small, medium and large when values are approximately 0.1, 0.3 and 0.5, respectively [7], [6].

Table 2 shows the results. The null hypotheses was that there was no statistically relevant difference between the two compared samples, i.e. there is no difference in perceived robotic arm movement naturalness between the two arm movement types. The risk level is set to $\alpha = 0.05/3 \approx 0.0167$, by applying the Bonferroni procedure. All three comparisons (for both tests) are statistically significant with p-values below the established α, which means that the null hypotheses can be rejected.

Table 2. Wilcoxon signed ranks and Sign tests. The z-scores were computed based on negative ranks for both tests, thus negative values indicate that the first rank in the calculation is higher than the second one.

	Wilcoxon signed ranks			Sign test		
	Small - No	Large - Small	Large - No	Small - No	Large - Small	Large - No
z-score	-9.497	-3.337	-9.914	-8.869	-3.705	-10.628
p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
ES	0.668	0.235	0.698	0.624	0.261	0.748

The results indicate that participants clearly noticed a difference between the three movements. This is confirmed by the z-scores and ES values. There are large ES values, i.e. bigger than 0.5, with corresponding highly negative z-scores for comparisons “Small-No” and “Large-No”. Since these z-scores were computed based on negative ranks, this means that the ratings given to small and large arm movements were significantly higher than ratings given to no arm movement.

Comparing the large and small arm movement samples there is a slight preference for the large arm movement. This is indicated by the negative z-scores and the small to medium effect size of 0.261.

Therefore, the arm movement has a clear impact on perceived robotic naturalness. Furthermore, robotic arm movement positively impacts perceived robotic naturalness. Different arm amplitudes and frequencies of movement are noticed by users. H_1 and H_2 were both confirmed.

5 Survey 2: Facial expressions for a social robot

This questionnaire was designed to assess the perceptions of people of the blinking pattern and the mouth shapes traits. For the first trait, an eyeblink design pattern was designed based on human eyeblink data obtained from literature from the medical field. The pattern can be described by the values in Table 3.

For the second trait, several mouth shapes were designed to convey Ekman’s seven emotions. These are displayed in Figure 1.

This questionnaire had 153 participants with a balanced gender (41/59%) and the age range 18-24 representing 45% of the total. In the eyeblink pattern section, the participants were shown a video of the robot performing the new

Table 3. Implementation details for each type of blink: Single and double. For the double blink, the contact time was also used as open time. Times in seconds.

Blink type	Sequence	Opening-closing time	Contact time	Total blink time
Single	closing-closed-opening (1x)	0.25	0.35	0.85
Double	closing-closed-opening-open (2x)	0.15	0.175	1.3

(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Fig. 1. Sketches of mouth shapes for the robot and the corresponding Ekman emotion. (a) Disgust, (b) Anger, (c) Anger, (d) Surprise, (e) Contempt, (f) Sadness, (g) Fear, (h) Happiness, (i) Surprise, (j) Contempt

blink pattern and were asked three questions. In the mouth shapes section the participants were shown images of the robot using the different mouth shapes and were requested to make it correspond to one of Ekman's seven emotions.

Regarding Casper's eye blinking, the questionnaire contained the questions (i) "How [natural] was Casper's blinking?", (ii) "What's your opinion on Casper's blinking speed?", (iii) "Did Casper always blink in the same way?".

For the Casper mouth shapes a single question was used, "What is Casper feeling?".

The second question was meant to confirm that participants found the robot's blinking speed adequate since it was designed to match a person's. The third question aimed to assess if participants noticed both blink types, the single and double blinks in Table 3. The following hypotheses were tested. (H_3) The users will consider the blinking speed to be adequate, i.e. with mean, median and mode statistics of approximately three. (H_4) The users will notice the two different blink types considered. (H_5) The users will consider the eyeblink pattern to be natural, i.e. with mean, median and mode statistics higher than three.

On the mouth shape, the goal was to test hypothesis (H_6) Users will consider that the robot is feeling the emotion that was intended for a given mouth shape. Mean, median, mode, and standard deviation statistics for the first two questions on eye blinking are $(3.54, 4, 4, 0.946)$ and $(2.88, 3, 3, 0.802)$, respectively.

Most participants (52% of them) thought that blinking speed was adequate, i.e. classified speed as a three. This was an expected result since the blinking speed for the robot was modulated according to medical research performed on humans. Concerning the different blink types, almost all participants (85% of them) acknowledged that the robot did not blink in the same way, hence recognizing the different blink types. Although most of the participants thought the blinking speed to be adequate and acknowledged the different blink types, the results also show that were still some participants (40% of them) that rated the eyeblink pattern naturalness with a three or less. However, 58% of the participants considered the eyeblink pattern to have a score equal or above four in terms of naturalness.

5.1 Sparsity measures

In the context of Social Robotics sparsity measures are not commonly used to analyse experiment data. The frequent use of questionnaires do acquire perception data originates sequences that can be organized in several forms. For example, (i) the answers in each sample can be concatenated to analyse possible bias, or (ii) answers to a specific question of all samples can be concatenated to identify trends among the population (relative to that answer), or (iii) all answers of all samples can be concatenated. The sparsity in such sequences can be an indicator of how answers are distributed.

A sparse dataset can be defined as one in which a small number of coefficients contain a large proportion of the energy [11]. The concept of sparsity is used in many engineering domains, including machine learning and signal processing, but it is also employed in other fields such as economics, for example, to evaluate income inequality [11], [10]. In the context of the present thesis, sparsity coefficients are proposed to analyze the concordance of participants' answers when faced with a categorical scale.

The performance of 16 different measures of sparsity, based on their fulfillment of six properties, is presented in [11]. Of these, only the Gini Index (GI) and pq-mean (with $p < 1$, $q > 1$) fulfill all six properties. The Hoyer index does not fulfill the Cloning property[3]. The Hoyer index is used since the Cloning property is not considered relevant for the proposed application.

5.2 Emotive mouth shapes

In this section the sparsity measures above are used to quantify the concordance of participants' answers when presented with the categorical scale of Ekman's emotions. Sparsity measures are presented as an alternative to regular descriptive statistics, such as median, mode and standard deviation, which are not appropriate since the data is categorical and the messages' order in the x-axis is random.

Table 4 shows the Gini, pq-mean ($p = 1$, $q = 3$) and Hoyer indexes produced by the emotive mouth shape questions.

Table 4. MBot's emotion for different mouth shapes. The values must be interpreted accounting for the range of values allowed for each index.

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)
Gini	0.53	0.398	0.407	0.788	0.717	0.697	0.59	0.835	0.835	0.639
pq-means	-0.606	-0.859	-0.643	-0.310	-0.337	-0.397	-0.5	-0.284	-0.284	-0.387
Hoyer	0.471	0.351	0.362	0.921	0.862	0.782	0.591	0.975	0.975	0.763

[3] If two populations have identical wealth distributions, the sparsity of wealth in one population will remain the same when the two are combined [11].

In what concerns the "Facial expressions for a social robot" questionnaire, H_3 was confirmed. The design of a robotic eyeblink pattern based on human eyeblink information seems to be appropriate. The mean, median, and mode values obtained for the question on Casper's blinking speed are around 3, meaning that users think that it was neither too slow nor too fast. H_4 was confirmed as approximately 85% of participants noticed that the robot did not always blink in the same way. This suggests that participants noticed the two blink types. H_5 was confirmed. Median and mode values for the question on the blinking naturalness were equal to four, however, the mean value was very close to three. This suggests that, although the robot blinked at an adequate speed and with an irregular pattern, it is hard to consider a robot blinking natural. Based on these results and on [9], small circular mouth shapes are accurately recognized as surprise. Additionally, bigger but still round mouth shapes are also recognized by most as surprise, however, they seem to also be interpreted as portraying more negative emotions such as fear. Also, downward concave curves are accurately interpreted as sadness and upward concave curves are accurately interpreted as happiness.

To accurately portray happiness, mouth shape 4h may be used by the MBot. To accurately portray surprise, mouth shapes 4d and 4i may be used by the MBot. To accurately portray sadness, mouth shapes 4e and 4f may be used by the MBot. To accurately portray contempt, mouth shape 4j may be used by the MBot. No mouth shapes are suggested to accurately represent fear, disgust and anger.

H_6 was only confirmed for mouth shapes 4d, 4f, 4h, 4i and 4j. When the robot was using these mouth shapes the users interpreted that the robot was feeling the intended emotion with relative agreement, i.e., $GI > 0.6$, pq-mean < -0.4 and Hoyer > 0.7. Otherwise, for mouth shapes 4a, 4b, 4c, 4e, and 4g the hypothesis was not confirmed.

6 Survey 3: Movement for a social robot

This questionnaire was designed to assess users' perceptions about trajectory smoothness and head movement.

For the first trait, the local planner's configuration parameters were adjusted so that the robot drew straighter paths rather than S-shaped ones resulting in a navigation that was perceived as more natural.

For the second trait, the robot used a leg detector to acquire a person's location and "look" at the person through the actuation of the neck motor. The other objective of the questionnaire was to assess how much the robot could rotate its head so that it is not considered eerie. Two values were tested for the maximum amplitude of rotation of the robot's head: 60° and 90°.

The questionnaire had 124 participants, with fully balanced gender (50/50%) and the age range 18-24 representing the largest portion of the population (32%). In the questionnaire the participants were shown two videos of the robot navigating between two goals in a mobile robotics lab environment. In the first video, the robot was using the new configuration of the local planner (generating straighter paths) and in the second it was using the manufacturer's configuration (generating more S-shaped paths). After each video participants were asked to rate the naturalness of the robot's movement. A final question assesses whether or not users had found a difference between the movements.

The participants were also shown two videos of the MBot passing by a person standing still. In the first video, the robot looks for a longer time at the man since it can rotate its head up to 90°. In the second video, the robot can rotate its head up to 60°, which causes it to look at the man for a smaller period. After viewing the videos, the participants were asked to answer two questions. Test hypotheses for the smooth navigation trait were as follows. (H_7) Users will consider that the robot's navigation style with the new local planner parametrization values is more natural than the robot's navigation style with the old parametrization. (H_8) Users will notice a difference between the navigation styles in the two videos.

Test hypothesis for the head movement (H_9) users will consider the robot to be more familiar with the person in video 2 in which the maximum neck RoM is $6o°$. (H_{10}) Users will notice a difference between the head movements in the two videos.

The mean, median, mode, and standard deviation statistics obtained are $(2.83, 3, 2, 1.17)$ for the manufacturer's parametrization and $(3.73, 4, 4, 1.01)$ for the new parametrization. The mean, median, mode and sample distributions show that there is a user preference for the improved navigation style. Moreover, 85% of the answers noticed a difference in navigation style between the two parametrizations.

The Wilcoxon signed ranks and Sign tests are shown in Table 5. The null hypothesis was that there is no statistically relevant difference between the two compared samples, i.e. there is no difference in perceived robotic navigation naturalness between the two navigation styles. $\alpha = 0.05$, a commonly accepted value for the level of risk [7]; The comparisons with Wilcoxon and Sign tests are statistically significant with p-values below the established α, which means that the null hypothesis can be rejected.

Table 5. The z-scores were computed based on positive ranks for both tests, thus negative values indicate that the second rank in the calculation is higher than the first one.

	Manufact. - Improved navigation	
	Wilcoxon signed rank	Sign test
z-score	-6.635	-6.076
p-value	< 0.001	< 0.001
ES	0.596	0.546

From Table 5, there were statistically significant differences between the conditions, confirmed by both tests. This suggests that the sample distributions are statistically different from one another. The results indicate that participants noticed a difference between the two navigation styles. The aforementioned preference for the improved navigation style over the manufacturer's is confirmed by the z-scores and ESs values. There are large ES values, i.e. bigger than 0.5, with corresponding highly negative z-scores. These z-scores were computed based on positive ranks, hence the ratings given to the improved navigation style were significantly higher than the ratings given to the manufacturer's one.

The results obtained for the two trait-specific questions for the look at people trait with 69% agreeing that in video 1 Casper was more familiar with the person, and 87% recognizing differences in the head movement between the two videos.

The conclusions from the "Movements for a social robot" questionnaire can be summed up as follows. H_7 was confirmed. Both descriptive and nonparametric inferential statistics used, indicate that users find the new navigation style, which favors straight paths, more natural than the manufacturer's, which favors S-shaped ones (even when not faced with obstacles). H_8 was confirmed. Most questionnaire participants (85% of them) found a difference in the robot's movement between the two videos. H_9 was not confirmed. Most users found that the robot was more familiar with the person in video 1 in which the maximum neck RoM is 90°. It is inconclusive whether the 60° of the 90° RoM should be used to increase robotic anthropomorphism; H_{10} was confirmed since 87% of the participants considered that the robot's head movement was different in the two videos. However, as previously discussed, this may not be due to the different RoMs but rather the different head movements that were performed by the robot when it wasn't looking at the person.

7 Survey 4: Perception of nonverbal sounds

The questionnaire on the sounds for a social robot was designed to assess users' perceptions of (meaningful) nonverbal sounds. R2D2-like sounds[4] were used. meant to convey simple messages that would make sense for a robot to use during interactions. The chosen messages were "Goodbye", "Hello", "Low battery", "No" and "Yes".

The questionnaire had 116 participants, with a balanced gender (59/41%) and 18-24 and 25-34 being the most represented age ranges (33% and 20%, respectively).

The participants were shown videos including the sounds and asked what the robot was saying. The options provided were the five messages that were intended for the sounds. The questionnaire included a single question, namely "On the videos with sound (a)...(f), what is Casper saying?", repeated a total of six times.

The sounds were displayed to the participants without any context or cues, hence it is expected that the participants do not fully agree about their meanings.

Regarding the sounds trait, the hypothesis is, (H_{11}) Users will consider that the robot saying the intended messages for a given sound.

[4] Credits: https://github.com/koide3/ros2d2

Table 6 shows the Gini, pq-mean ($p = 1$, $q = 3$) and Hoyer indexes corresponding to the samples produced by the six “What is Casper saying?” questions. Similarly to the mouth shape and Ekman’s emotions, sparsity measures are presented as an alternative to regular descriptive statistics which are not appropriate since the data is categorical and the messages’ order in the x-axis is random.

Table 6. Meaning of sounds. The sparsity values must be interpreted taking into account the total ranges computed for the sample size.

	(a) “Low battery”	(b) “Hello”	(c) “No”	(d) “Goodbye”	(e) “Hello”	(f) “Yes”
Gini	0.617	0.579	0.579	0.290	0.369	0.414
pq-means	-0.460	-0.494	-0.494	-0.799	-0.76	-0.692
Hoyer	0.741	0.676	0.676	0.214	0.307	0.369

The results show that H_{11} was only confirmed for sounds 6a, 6b, and 6c. Upon hearing these sounds the users interpreted the intended message with relative agreement, i.e. GI > 0.57, pq-mean < -0.5 and Hoyer > 0.67. Otherwise, for sounds 6d, 6e, and 6f the hypothesis was not confirmed.

8 Conclusions

The paper presented a detailed study on the design of basic social behaviours for a specific robot[5]. The methodology used is applicable to other studies as well. The use of sparsity measures when faced with a categorical scale is also proposed. Additionally, the study provides useful guidance for the design of anthropomorphic behaviors for social robots.

The results show that users are able to discriminate small differences in robotic behaviours. The parametrization of navigation functions was shown to have effect in the social skills of the robot. This raises the interesting possibility of considering an optimization problem over the space of parameters of the navigation system which is to be covered in future work together with real world experiments.

Acknowledgments

This work was partially supported by project LARSyS-FCT Project UIDB/50009/2020.

References

1. H. Admoni and B. Scassellati. Social eye gaze in human-robot interaction: A review. *Journal of Human-Robot Interaction*, 6(1):25—63, May 2017.

[5] The videos used in the questionnaires are available from https://sites.google.-com/view/joao-silva-sequeira/exploring-behaviours-for-social-robots

2. K. Baraka, P. Alves-Oliveira, and T. Ribeiro. An Extended Framework for Characterizing Social Robots. In C. Jost, B. Le Pévédic, T. Belpaeme, C. Bethel, N. Crook D. Chrysostomou, M. Grandgeorge, and N. Mirnig, editors, *Human-Robot Interaction: Evaluation Methods and Their Standardization*, Springer Series on Bio- and Neurosystems, pages 21—-64. Springer International Publishing, 2020.
3. I. Brace. *Questionnaire Design: How to Plan, Structure and Write Survey Material for Effective Market Research*. Kogan Page Publishers, April 2018.
4. J.J. Cabibihan, W.C. So, and S. Pramanik. Human-Recognizable Robotic Gestures. *IEEE Transactions on Autonomous Mental Development*, 4(4):305—-314, December 2012.
5. A. Castro-González, J. Alcocer-Luna, M. Malfaz, F. Alonso-Martín, and M.A. Salichs. Evaluation of Artificial Mouths in Social Robots. *IEEE Transactions on Human-Machine Systems*, 48(4):369—-379, August 2018.
6. J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. New York: Routledge, 2nd ed. edition, July 1988.
7. G.W. Corder and D.I. Foreman. *Nonparametric Statistics: A Step-by-Step Approach*. John Wiley & Sons, May 2014.
8. M.J. Doughty. Consideration of Three Types of Spontaneous Eyeblink Activity in Normal Humans: During Reading and Video Display Terminal Use in Primary Gaze, and while in Conversation. *Optometry and Vision Science*, 78(10):712–, October 2001.
9. A. Giambattista, L. Teixeira, H. Ayanoglu, M. Saraiva, and E. Duarte. Expression of Emotions by a Service Robot: A Pilot Study. In A. Marcus, editor, *Design, User Experience, and Usability: Technological Contexts*, Lecture Notes in Computer Science, pages 328–336. Springer International Publishing, 2016.
10. C. Gini. Measurement of Inequality of Incomes. *The Economic Journal*, 31(121):124—125, March 1921.
11. N. Hurley and S. Rickard. Comparing Measures of Sparsity. *IEEE Transactions on Information Theory*, 55(10):4723—4741, October 2009.
12. R. Jones. Communication in the Real World, September 2016.
13. H. Lehmann, A. Roncone, U. Pattacini, and G. Metta. Physiologically Inspired Blinking Behavior for a Humanoid Robot. In A. Agah, J.J. Cabibihan, A.M. Howard, M.A. Salichs, and H. He, editors, *Social Robotics*, Lecture Notes in Computer Science, pages 83—93. Springer International Publishing, 2016.
14. J.D. Rodriguez, K.J. Lane, G.W. Ousler, E. Angjeli, L.M. Smith, and M.B. Abelson. Blink: Characteristics, Controls, and Relation to Dry Eyes. *Current Eye Research*, 43(1), January 2018.
15. J. Romkes and K. Bracht-Schweizer. The effects of walking speed on upper body kinematics during gait in healthy subjects. *Gait & Posture*, 54:304–310, May 2017.
16. G.M. Sullivan and R. Feinn. Using Effect Size - or Why the p Value Is Not Enough. *Journal of Graduate Medical Education*, 4(3):279–282, September 2012.
17. J.L. Tracy and D. Randles. Four Models of Basic Emotions: A Review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review*, 3(4):397—-405, October 2011.
18. M. Wiggers and C.F. van Lieshout. Development of recognition of emotions: Children's reliance on situational and facial expressive cues, 1985.
19. M. Zimmerman, S. Bagchi, J. Marvel, and V. Nguyen. An Analysis of Metrics and Methods in Research from Human-Robot Interaction Conferences 2015-2021. In *Procs, 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2022)*, pages 644–648, March 2022.