Subjective Experience of Interacting with a Social Robot at a Danish Airport

Unpublished Conference Paper for SEMCON at Aalborg University, 2017

Andreas Kornmaaler Hansen, Emil Bonnerup, Juliane Nilsson, Lucca Julie Nellemann, and Sara Nielsen
School of Information and Communication Technology
Aalborg University, Aalborg, Denmark
Email: 17gr782@es.aau.dk

***Abstract*— This paper investigates the subjective experience of interacting with a social robot at Aalborg Airport (AAL) by conducting an ecological field study where 23 important attributes in Human-Robot Interaction (HRI) were elicited. Scales were developed from these attributes and tested at AAL. In both tests travellers were recruited by a remote controlled *Double* robot, which had an iPad with an interface asking if it may help the travellers with wayfinding at AAL. When the subjects had chosen the desired location they were asked to follow the robot. In the first test the robot led them to a semi-structured interview about their experience and in the second test to an experimenter who asked them to rate their experience on the developed scales. After the first test the observations and the subjects' statements were interpreted and coded using an affinity diagram. 567 affinity notes were sorted by a bottom-up procedure into 10 categories from which the 23 scale questions and scales were developed. The scales were used in a second test at AAL, where 43 subjects rated the robot and the interaction on the 23 scales. The ratings were analysed with Principal Component Analysis (PCA), which showed both positive and negative correlations within each of the three groups: *Robot's height*, *distance*, and *direction*.**

I. Introduction

This study originates from a social robot research project at Aalborg University with the aim of developing and implementing robots in a variety of contexts. This raises technological, normative, and empirical questions as to how these robots are to behave and in which settings they might be useful. Several studies have already shown that both utilitarian and hedonic attributes influence the way users experience the interaction with social robots [1], [2]. Utilitarian attributes involves the utility, practicality, and usability of the robot, whereas hedonic attributes reflect the user experience, enjoyment, and acceptance while interacting with the robot [1]. Some gender and cultural differences are found to affect the interaction with a social robot. These are the willingness to accept robots [3], and the intimate distance, which is smaller for southern Europeans and Japanese compared to Americans and northern Europeans [4]. How one experience Human-Robot Interaction (HRI) also depend on one's gender, with females being less willing to accept and interact with robots than males, who also perceive the robots as more human-like compared to how females perceive robots [1].

It is unknown whether these results generally apply to Danish travellers. People have different experiences and might articulate them differently depending on both their culture and the context. It is important to elicit attributes based on the user's own words and terms.

This study aims to elicit important attributes in HRI at an airport and develop relevant scales based on these attributes. Furthermore the study aims to analyse if some attributes have an influence on others. The developed scales might be used by robot designers for specific contexts and potentially for tailored user experience evaluations.

II. Method - Elicitation of Attributes

The first test was conducted on Danish travellers who interacted with a social robot in a natural setting. The test was conducted at Aalborg Airport (AAL) after the security check and before the travellers reached the shopping and dining area. The travellers who interacted with the robot were asked to participate in a semi-structured interview about their first impressions while being observed during both the interaction and the interview. The *Wizard of Oz* method was used.

A. Materials

For the test a *Double* robot from Double Robotics Inc, [5], with an iPad Air 2, [6], was used. Based on a small pretest it was decided to change the head mount so that the iPad is angled upwards, to make the robot appear more welcoming. The modified *Double* robot is shown on Figure 1 and Figure 2. The *Double* robot was connected to a laptop via Wi-Fi connection and its own software. The *Double* application allows a web page to be superimposed on the iPad screen while the controller is still able to control the robot from the laptop. The web page presented on the iPad was a wireframe developed in Marvel, (www.marvelapp.com), aiming to depict the potential usage of the robot as a wayfinding tool at AAL. The entire wireframe was formulated in Danish.

B. Subject Recruitment

30 subjects from the age of 8 to 62 years (M=37.9, SD=17.1) distributed among 16 females and 14 males participated. The subjects were recruited by the robot itself, which was remote controlled by a present robot controller, to provide a more ecological and undisturbed interaction between robot and subject. The robot approached potential subjects and the wireframe on the iPad asked them if it might help them find their way around AAL and presented a "Yes/No" option.

Fig. 1: *Double*'s front.

Fig. 2: *Double*'s profile.

If ”No” was pressed, the robot wished the traveller a pleasent journey and left. If ”Yes” was pressed, the subjects were presented with four wayfinding options: Food, Shopping, Toilets, or Gate information, as shown on Figure 3. The subjects were then kindly asked to follow the robot after they had chosen their preferred option. The robot then led the subjects to an interviewer who shortly informed them of the study and received verbal consent to record audio before the semi-structured interview. In total 18 interviews were conducted of which 11 were on single travellers and seven were on groups of travellers.

Fig. 3: Options presented to the subjects after having pressed ”Yes” and been informed of the robot being part of a research project.

C. Semi-structured Interview

The interview was a two part semi-structured interview. The first part consisted of probing the subjects for their first impression and experience of interacting with the robot in regard to their thoughts about the robot itself and what they think other travellers might think about the interaction. In addition to these conversation topics the subjects were asked about their opinion regarding the robots approach, usefulness, and reliability. Afterwards, the subjects were asked about previous experiences and problems at airports, in order to gather potential use cases where the robot might be helpful.

The second part consisted of asking specific questions relating to the robots physical characteristics such as speed, height, distance, movements, appearance, and approach. These questions were asked because these attributes have been found to affect the experience of HRI [4].

The two parts were conducted in continuation of each other and the questions in the second part were only asked if the subjects had not previously answered them spontaneously.

D. Roles

Five researchers were present during the test at AAL. One controlled the *Double* robot, one conducted the interview, and the remaining three observed the travellers as they interacted with or walked past the robot. Since the robot controller had no way to know what happened on the screen, one of the observers signaled to the robot controller when to start leading the subjects to towards the shopping area.

E. Data Processing

The interviews were transcribed and coded along with observations into affinity notes. The purpose was to create an affinity diagram, which brings insights and issues into a hierarchical diagram based on subjects' statements and behaviour [7]. This affinity diagram is pivotal in eliciting the attributes that affect HRI, and thereafter in creating the scales to be used for further work. To include more of the gathered insights, it was decided to use both the spontaneous answers from the conversation topics and the answers from the specific questions in the affinity diagram. The two types of answers are not differentiated.

III. Results - Elicitation of attributes

In total 567 affinity notes were sorted into 10 overall categories with individual subcategories. The 10 superordinate categories were appearance, trust, behaviour, approach, problems with the touch screen, avoidance of interaction, personal interest, positivity towards the robot, usefulness, and tech-experience.

Based on the 10 categories, attributes were elicited according to the criterion of a) being an influencing attribute and b) the possibility of formulating the attributes as a scale question. The field study was conducted on Danish speaking test subjects, and the attributes are therefore listed in both English and Danish. The scale questions are all presented on a *Visual Analogue Scale* (VAS) with closed anchor points and are either bi- or unipolar. If a scale is bipolar a midpoint will be marked either with or without a label. A bipolar scale without a label will be noted with *No label*, whereas an unipolar scale, which does not contain a midpoint, will be noted with a -.

In the following the 23 derived scales (noted *S*), along with the specific scale question (SQ), will be presented in corresponding order as presented for the test subjects in the second test and sorted by each page. An example of each set of scales is given, to illustrate how they were presented in the test.

A. Page 1

SQ01: How do you think the screen on the robot reacted?
SQ02: How did you experience the robot?
SQ03: How was it to use the robot?
SQ04: How did you experience the movements of the robot?
See Table I and Figure 4 for how the scales were presented.

TABLE I: Scale labels page 1

S	Left label	Midpoint	Right label
1	Extremely bad (Ekstremt dårligt)	No label	Extremely well (Ekstremt godt)
2	Extremely unwelcoming (Ekstremt afvisende)	No label	Extremely welcoming (Ekstremt imødekommende)
3	Extremely difficult (Ekstremt svært)	No label	Extremely easy (Ekstremt nemt)
4	Extremely wild (Ekstremt vilde)	No label	Extremely calm (Ekstremt rolige)

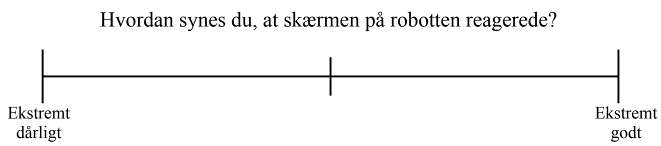


Fig. 4: Example of a bipolar scale relevant for the scale question: *How do you think the screen on the robot reacted?*, rated from *Extremely bad* to *Extremely well*.

B. Page 2

SQ05: I think that the robot stopped...
SQ06: I think that the speed of the robot is...
SQ07: I think that the height of the robot is...
See Table II and Figure 5 for how the scales were presented on page 2.

TABLE II: Scale labels page 2

S	Left label	Midpoint	Right label
5	Way too close (Alt for tæt på)	No label	Way too far (Alt for langt fra)
6	Way too slow (Alt for langsom)	Fine (Fin)	Way too fast (Alt for hurtig)
7	Way too low (Alt for lav)	Fine (Fin)	Way too high (Alt for høj)

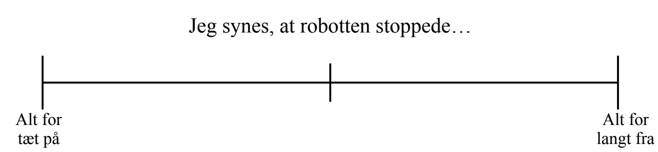


Fig. 5: Example of a bipolar scale relevant for the scale question: *I think that the robot stopped...*, rated from *Way too close* to *Way too far*.

C. Page 3 and 4

SQ08: I feel that the robot can help me
SQ09: I think that the robot was obstructing me
SQ10: I feel safe around the robot
SQ11: The robot startled me
SQ12: I like to be served by the robot
SQ13: I relied on the robot to lead me to the location I chose
See Table III and Figure 6 for how the scales were presented on page 3 and 4.

TABLE III: Scale labels page 3 and 4

S	Left label	Midpoint	Right label
8-13	Completely disagree (Helt uenig)	No label	Completely agree (Helt enig)

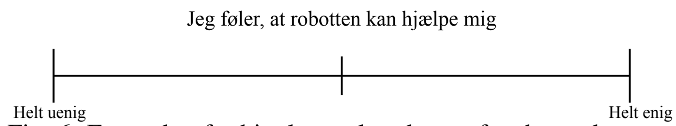


Fig. 6: Example of a bipolar scale relevant for the scale question: *I feel that the robot can help me*, rated from *Completely disagree* to *Completely agree*.

D. Page 5

SQ14: How personal did you experience the help of the robot?
SQ15: How surprised were you by the robot's approach?
For each of the two scale questions the chosen labels are listed in Table IV. Also see Figure 7 for an example of the scale used.

TABLE IV: Scale labels page 5

S	Left label	Midpoint	Right label
14	Not at all personal (Slet ikke personlig)	-	Extremely personal (Ekstremt personlig)
15	Not at all surprised (Slet ikke overrasket)	-	Extremely surprised (Ekstremt overrasket)

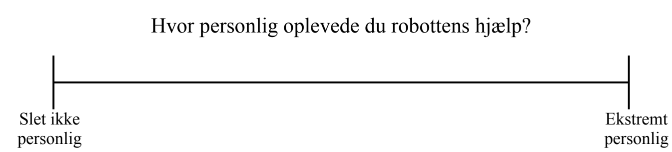


Fig. 7: Example of an unipolar scale relevant for the scale question: *How personal did you experience the robots help?*, rated from *Not at all personal* to *Extremely personal*.

E. Page 6

SQ: What do you think about the robot?
This question covers scales 16-19. The scale used was similar to the one shown in Figure 7. See Table V for the labels used.

TABLE V: Scale labels page 6

S	Left label	Midpoint	Right label
16	Not at all annoying (Slet ikke irriterende)	-	Extremely annoying (Ekstremt irriterende)
17	Not at all elegant (Slet ikke elegant)	-	Extremely elegant (Ekstremt elegant)
18	Not at all exciting (Slet ikke spændende)	-	Extremely exciting (Ekstremt spændende)
19	Not at all cute (Slet ikke sød)	-	Extremely cute (Ekstremt sød)

F. Page 7

Q: What else do you think about the robot?
This question covers scales 20-23. The scale used was similar the one shown in Figure 7. See Table VI for the labels used.

TABLE VI: Scale labels page 7

S	Left label	Midpoint	Right label
20	Not at all cool (Slet ikke sej)	-	Extremely cool (Ekstremt sej)
21	Not at all intrusive (Slet ikke anmassende)	-	Extremely intrusive (Ekstremt anmassende)
22	Not at all funny (Slet ikke sjov)	-	Extremely funny (Ekstremt fun)
23	Not at all human (Slet ikke menneskelig)	-	Extremely human (Ekstremt menneskelig)

Based on the affinity diagram a 24th attribute was derived, this attribute will not be presented along side the aforementioned scales but will be included in a separate demographic page as it does not directly concern the robot. The attribute is formulated in the scale question: *How fond of technology are you?*, which will be evaluated on an unipolar scale, similar to the other unipolar scales, with anchor points: *Not at all fond* (slet ikke glad) and *Extremely fond* (ekstremt glad).

When comparing the attributes for HRI found in this study with attributes for HRI from previous conducted studies on social robots [1], [2], [3], [8], [9], [10], [11], attributes such as distance, anthropomorphism, height, speed, movement, trust, enjoyment, technological knowledge, and usefulness reoccur. New attributes were found compared with previous mentioned studies, these are: Elegance, cuteness, coolness, startling, excitement, welcoming, obstructive, whether the robot help is seen as personal, and whether the encounter with the robot was surprising. However, these might be measured indirectly in the aforementioned studies.

Accordring to [12] *social distance* has an effect on HRI. The subjects in AAL did not mention it specifically, which might be due to *power distance*, where the subjects feel in control as they were the dominant part and the robot was the subordinate in the interaction. Furthermore this could also be due to the given task being a cooperative task, as the robot's purpose was to help the subject to a specific location of their choosing. *Task distance* in cooperative tasks might incline subjects to act more friendly, intimate, and involved in the interaction with the robot, which was supported by all the positive comments from the subjects.

IV. Part II - Scale Testing

This section describes the second part of the study where the scales were tested at AAL. The purpose was to get the travellers to rate their experience with the robot on the developed scales.

V. Method - Scale Testing

The test was conducted over the course of two days and in the same way as the first test described in section II. However, in this test instead of the robot leading them directly to an interviewer, it led them in the direction of their chosen location. After a short distance, an experimenter politely stopped them and informed them about the ongoing study. This was done to give the subjects a more natural experience, where the robot appeared to lead them in the right direction. After they accepted to participate, they were led to a PC nearby to rate their experience on the developed scales. The experimenter was ready to take notes while they rated. The order in which the scales were presented is described in section III. The physical variables of the robot were varied over the course of the test. This included the angle of approach, the robot's height (and speed), and the distance to the subjects. As a consequence of the study being ecologically performed in an actual airport it was not possible to precisely control the distance to the subjects or the angle of approach. These had to rely on subjective judgement from the researchers. Height was measured using a measuring tape. The robot height was set to either of the following five heights: 118 cm, 123.5 cm, 129 cm, 140 cm, or 151 cm. The first and last mentioned heights are dictated by the robot's minimum and maximum height, whereas 129 cm was chosen according to what the subjects in previous study mentioned as fine, which roughly corresponds to human elbow height according to the mean height of the Danish population.

A. Materials

The same materials were used as those used in the first test described in section II along with the same *Double* and tilted head mount. The same wireframe was used except with a few adjustments. Additionally, software was developed in Processing 3.3.6 (www.processing.org) to be able to collect data from the scales. The program was presented on a *Microsoft Surface Pro (5)* with a wireless mouse. The 24th scale was presented on a sheet of paper along with a few other questions to collect demographic information.

B. Scale Program

The scale program consisted of 23 scales distributed on seven pages. The number of scales presented on a single page were maximum four and minimum two. The subjects were instructed to set a marker representing their response on the scales using the provided mouse. The scales were organised to be as consistent on each page as possible e.g. the same type of scales with a midpoint were presented on one page and internally the scales were organised such that two similar attributes did not appear right after each other.

C. Subject Recruitment

Over the course of two days 43 subjects participated. They were aged from 10 to 72 (M=40.1, SD=13.4) and distributed among 16 females and 27 males. Subjects reported that they travel between 1 and 100 times per year (M=15.3, SD=18.1). The subjects were again recruited by the robot with the same wireframe as in the previous test.

D. Roles

Four researchers were present during the test at AAL. One controlled the *Double* robot, one instructed the subjects to answer the scales, one observed when the robot should start leading the subjects towards the shopping area and one noted the physical variables of the robot such as height, direction of approach, and distance to the subjects. As with the previous test the robot controller had no way to know what happened on the screen, one of the observers signaled to the robot controller when to start leading the subjects to towards the shopping area.

E. Data Processing

Data was gathered in Processing 3.3.6 and for each subject data were saved in a .csv format and analysed using MATLAB, Excel, and RStudio. Data were analysed using boxplots and Principal Component Analysis (PCA), which were further analysed in order to compare correlated attributes. The PCAs were conducted on robot's height, distance, and direction separately in order to assess how the subjects responded to the physical changes of the robot.

VI. Results - Scale Testing

Due to a programming error the dataset contains several zero-answers which could happen when subjects did not answer on a scale. It was decided to remove the missing answers. Two criteria were used to determine when to exclude some of the zero-points that occurred because of a non-answered scale. 1) If all scale ratings presented on one of the pages are zero, it was probably because subjects skipped a page and 2) If there is a strong tendency for high ratings on a scale and a single zero-answer seems unlikely, it was denoted as a missing answer.

The different ratings on the scales in the second test are visualised on Figure 8.

From Figure 8 it is notable that there exists much greater variation in some scales compared to others. For some SQs, a big part of the scale is used (1, 4, 8, 12, 14, 19), where for others a smaller part of the scale is used (2, 3, 5, 7, 9, 11, 13, 16). Furthermore some of the scale ratings appears normal distributed (1, 2, 6, 17, 20, 21) and others appears skewed (8, 9, 10, 11, 13, 18). In some SQs the data accumulates around midpoints or anchor points (5, 7, 9, 10, 11, 13, 16). SQ05 and SQ07 are centralised around the midpoint, which for SQ07 can relate to the label *fine* being too broadly describing. The closed anchor points used on the scales can induce accumulation of data points around these. Also the use of bipolar scales poses the risk of picking two words which are not each others opposite.

When looking at gender differences it seems that women rated SQ04, SQ11, and SQ21 lower than men. It suggests that females experienced the movements of the robot more calm compared to males, but also that they find the robot less intrusive and startling.

When conducting a PCA on all the scale results it showed that only 53.1 % of the total variance is explained when using the first three dimensions, which is why PCA relating to different groups such as the robot's height, distance, and direction are conducted in order to reduce the dimensions and to be able to explain more of the variance. Figure 9a shows a Scree plot from PCA relating to the robot's height, where 94.1 % of the variance is explained using three dimensions. Figure 9b shows the Biplot relating to the robot's height. Similar Biplots were made for direction and distance. The attributes which loads the most on the components are:

Height: SQ19 appears to contribute the most to PC1. SQ01 the most to PC2. SQ06 the most to PC3.

Distance: SQ01 appears to contribute the most to PC1 along with SQ15. SQ10 contributes most to PC2.

Direction: SQ10 contributes most to PC1. SQ22 and

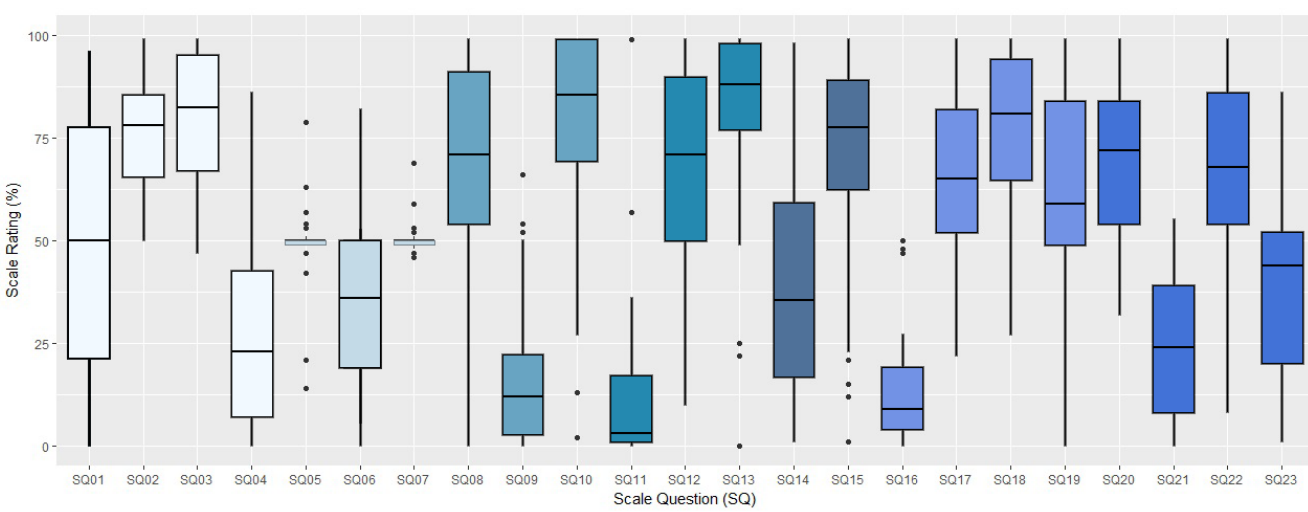
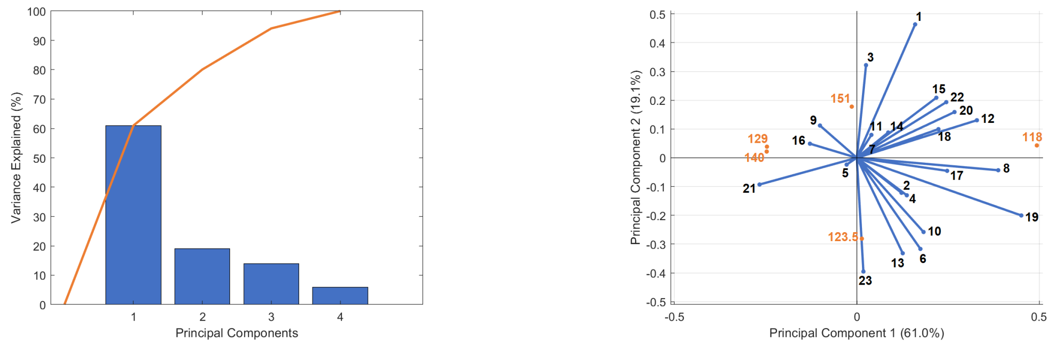


Fig. 8: Boxplot showing the 23 attributes. The boxplot shows the median, where the box is ranging from 25-75 % of the data, and the whiskers from 0-25 % and 75-100 % of the data. Each colour represents a new page of scales.



(a) Scree Plot of Robot Height

(b) Biplot of Robot Height

Fig. 9: **(a):** Scree plot showing the connection between the number of Principal Components and Variance Explained (%): PC1 (61 %), PC2 (19.1 %), PC3(14 %), PC4(5.9 %). **(b):** Biplot showing how the different attributes contributes to components and which attributes that correlates. The black numbers denotes SQ and the red to the different heights in cm.

SQ23 contributes almost equally to PC2. SQ13 and SQ20 contributes almost equally to PC3.

2D and 3D Biplots from the different PCAs shows positive and negative correlations between different attributes. These correlations are presented in Table VII.

TABLE VII: Correlations from PCA

PCA	Positive correlation	Negative correlation
Height	SQ08 + SQ17	SQ02 + SQ09
	SQ10 + SQ13	SQ04 + SQ12
	SQ12 + SQ18	SQ12 + SQ21
	SQ14 + SQ15	SQ16 + SQ19
	SQ20 + SQ22	SQ18 + SQ21
Distance	SQ01 + SQ12	SQ02 + SQ09
	SQ07 + SQ17	SQ05 + SQ21
	SQ08 + SQ21	SQ10 + SQ13
	SQ10 + SQ22	SQ13 + SQ22
		SQ14 + SQ16
		SQ19 + SQ20
Direction	SQ05 + SQ07	SQ01 + SQ12
	SQ08 + SQ10	SQ06 + SQ23
	SQ09 + SQ14	SQ09 + SQ10
	SQ18 + SQ20	SQ10 + SQ14
		SQ13 + SQ21

To investigate these correlations further, ratings from correlating SQs are compared, as seen in Figure 10. Even though correlations occur when doing PCA on groupings, these plots of comparison do not take groupings into consideration. This might be why some correlations are not found when comparing SQs, even though the Biplots showed a correlation. From Figure 10 it can be seen that the robot is perceived more exciting when subjects like to be served by the robot. Attributes that have a clear correlation from this kind of comparison are shown in Table VIII.

TABLE VIII: Correlations from graphs

Positive correlation	Negative correlation
SQ04 + SQ09	SQ02 + SQ09
SQ08 + SQ10	SQ09 + SQ10
SQ08 + SQ17	SQ12 + SQ21
SQ10 + SQ13	SQ13 + SQ21
SQ12 + SQ18	SQ18 + SQ21
SQ18 + SQ20	
SQ20 + SQ22	

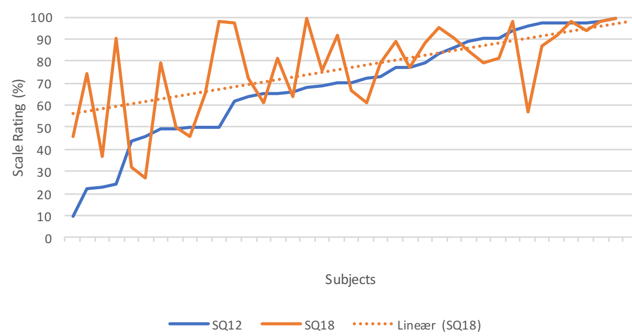


Fig. 10: Comparison between ratings on SQ12 and SQ18 based on 41 subjects. Two were removed due to incomplete datasets.

The physical variables can be compared with SQs as well as the demographic parameters in the same manner as comparing SQs with each other, to see how these variables and parameters affect the ratings of the attributes. Figure 11 were SQ06 is compared with the robot's height shows that when the robot's height increases the robot is perceived as being less annoying. When comparing SQ06 with age, the younger a subject is, the more inclined one is to rate the robot as moving too slow. Also, a small negative correlation was noted in relation to how exciting and funny the robot seemed. The older the subjects were, the less exciting and funny they rated the robot. Within height the smaller the robot was the faster and more wild it was rated. This was expected, as the robot automatically moves more slowly when the height increases and faster when the height decreases. Additionally it was rated more cute, elegant, and trustworthy, regarding to wayfinding, when it was small.

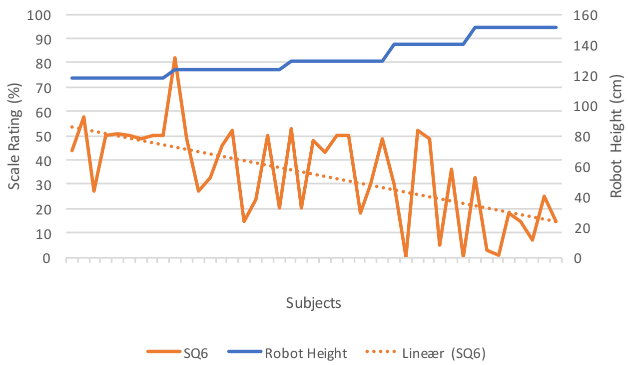


Fig. 11: Comparison between the robot's five heights (cm) and ratings (%) on SQ06 based on all subjects.

VII. Discussion

A. Equipment

Due to the Marvel wireframe running inside the *Double* application, the touch screen was less responsive than expected. This resulted in some travellers not wanting further interaction with the robot. It also affected the experience for those who chose to interact with it anyway.

The scaling software written for the second test had some response issues as well, which ultimately led to incomplete datasets because the program often failed to notice mouse-clicks. This is not ideal for a scale which relies on an immediate impulse response without much thought. This have almost definitely changed some subjects' responses in either direction of the scale due to them clicking in another location on the scale than initially chosen.

B. Interview

The interview in the first test was supposed to be very open and allow the subjects to talk freely. In reality, however, it ended up being a bit more structured, with the subjects directly answering the questions they were asked and waiting for the next one. This introduced some interviewer bias, since the predetermined questions dragged the conversation in a certain direction, and meant subjects talked a lot about a specific aspect of the interaction, because they were asked directly about it.

C. Treatment of data

In the first test spontaneous answers, answers to specific questions, and observations were all analysed together in an affinity diagram. It could be argued that these answers are two different categories, where only the spontaneous answers are the subjects' own words and areas of interest. The subjects did not provide as many spontaneous answers as hoped, hence all answers and observations were analysed together to gain more insight.

If a scale was not answered in the second test the software returned 0, which also was a valid scale response for the subjects. After looking through the data it was decided to remove some of the zeros, which seemed to stem from the unresponsive software bug. This might have led to real zero-responses being excluded from the dataset and for missing answers being included.

D. First impression

The results of this study were based entirely on the subjects' first impression of the robot. The long term experience of interacting with the robot remains unknown. Several subjects interacted with the robot, because it was new and exciting and made them curious, which might be explained by the novelty effect. Their user experience and focus points might change relative to their familiarity with the robot and therefore different attributes might be elicited when conducting a study not only regarding first impression.

E. Context

Some of the attributes which were elicited might only appear because the study was conducted at AAL, where the *Double* robot helped with specific tasks. These attributes are considered to be important for travellers at AAL and can be used when evaluating a social robot in this specific context. However, these are not attributes which can easily be adjusted by a designer such as how the subjects perceive the robot (elegant, cute, welcoming, etc.). Though this might be true, it is found that the robot was rated more cute, elegant, and trustworthy, regarding to wayfinding, when it was small. On the basis of these correlations the subjective experience might be improved by changing physical, easily adjustable characteristics such as height.

F. Labels and scale questions

It appears that there was some confusion regarding some of the scales e.g. where SQ05 was misinterpreted. The question was meant as to how close the robot stopped in relation to the subject, but some subjects thought it was meant in relation to their chosen location which they obviously could not answer. A few also commented on the attributes funny and cute. Some subjects expressed that the robot was funny, but not humorous funny and therefore had trouble rating the scale question because of this duality. Another subject commented that she refused to rate the robot as cute, because she regarded it as a human quality and that she refused to humanise the robot.

For the scale rating to SQ02, none of the subjects rated under 50 %. Half of the scale, regarding unwelcoming, might therefore be redundant. Instead an unipolar scale regarding welcoming could be developed and this would give a wider range to answer on the scale to the attribute welcoming.

The low variance and centering around the midpoint on some of the scales might be that the subjects did not find the scale questions relevant or did not understand them and therefore chose a neutral rating. It might also be that the selected stimuli was not varied enough. This could be reinforced by the midpoints showing the subjects where the middle is. If the ratings are constantly located at the same point, they do not contain much information about how different variables and parameters affect the experience of interacting with the robot.

G. Redundancy

Some of the attributes might be redundant if they correlate with other attributes having similar overall characteristics. This relates to attributes such as cool and funny, which might be redundant when also asking subjects about how cute and exciting they found the robot. This is due to the correlations, but also the word funny, which can be widely used, e.g. as humorous funny or weird funny. Attributes like how safe subjects felt using the robot could be combined with how much they relied on the robot leading them to the chosen location under a shared trust attribute. Asking subjects how trustworthy they found the robot and how much they felt it could help them could then give an impression on the three attributes by only asking two questions. Furthermore, attributes such as SQ12 (like being served by the robot) might be redundant when asking subjects how exciting and intrusive they found the robot.

VIII. Conclusion

The research conducted in this study revealed 10 main categories relating to HRI. Attributes were found for each of these 10 categories which describe the user experience of interacting with a social robot at Aalborg Airport. The observations and the 30 subjects' statements were interpreted and coded using an affinity diagram. 567 affinity notes were sorted by a bottom-up procedure into 10 categories which revolved around appearance, trust, behaviour, approach, problems with the touch screen, avoidance of interaction, personal interest, positivity towards the robot, usefulness, and tech-experience.

New attributes within the 10 categories were discovered compared to previous mentioned studies: Elegance, cuteness, coolness, startling, excitement, welcoming, obstructive, whether the robot help is seen as personal, and whether the encounter with the robot was surprising. Besides the newly found attributes some attributes were found to match previously found attributes. These attributes regards distance, anthropomorphism, height, speed, movement, trust, and usefulness.

23 attributes were elicited and scales were developed from these. After the second test, 10 instances of correlations were found for both height and distance and 9 for direction based on PCA with different groupings. 12 instances of correlation were found when comparing attributes that correlates when conducting PCA.

Based on the results in this study, it is suggested that the scale questions: *I relied on the robot to lead me to the location I chose* and *I feel safe around the robot* are combined to one attribute named: Trust. Also funny and cool are described by exciting and cute and thus seem redundant. Finally, the scale question: *I like to be served by the robot* is also described by intrusive and exciting, thus also seem redundant.

The different types of scales used in this study could explain the difference in variance between the scales. The closed endpoints might affect how much the results accumulate at the ends. The label *fine*, the use of midpoints, and scale questions such as *I think the robot stopped...* should be reconsidered to improve the user's understanding of the scale and minimize misunderstandings.

Acknowledgment

We would like to thank Karl Damkjær Hansen, postdoc at Aalborg University for giving professional insights and feedback, helping with technical support, and lending the *Double* robot to the authors. We would also like to thank Professor Dorte Hammershøi for supervising the study. Lastly we would like to thank Aalborg Airport for providing access to the wanted user group and facilities by letting the authors conduct their field studies at the airport.

References

[1] M. M. de Graaf and S. B. Allouch, "Exploring influencing variables for the acceptance of social robots," Robotics and Autonomous Systems, vol. 61, no. 12, pp. 1476 – 1486, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0921889013001334

[2] M. M. de Graaf, S. B. Allouch, and T. Klamer, "Sharing a life with harvey: Exploring the acceptance of and relationship-building with a social robot," Computers in Human Behavior, vol. 43, no. Supplement C, pp. 1 – 14, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747563214005536

[3] J. E. Katz, D. Halpern, and E. T. Crocker, In the Company of Robots: Views of Acceptability of Robots in Social Settings, J. Vincent, S. Taipale, B. Sapio, G. Lugano, and L. Fortunati, Eds. Springer, 2015.

[4] K. Dautenhahn, M. Walters, S. Woods, K. L. Koay, C. L. Nehaniv, E. A. Sisbot, R. Alami, and T. Siméon, "How may i serve you? a robot companion approaching a seated person in a helping context," pp. 172–179, 2006.

[5] D. Robotics. (2017) Work from anywhere. [Online]. Available: https://www.doublerobotics.com

[6] Apple. (2017, 11) ipad air - technical specifications. [Online]. Available: https://support.apple.com/kb/SP692?viewlocale=en_US&locale=da_DK

[7] J. B. Wendell, K. Holtzblatt, and S. Wood, Rapid Contextual Design. Elsevier LTD, Oxford, 2005. [Online]. Available: http://www.ebook.de/de/product/3268220/jessamyn_burns_wendell_karen_holtzblatt_shelley_wood_rapid_contextual_design.html

[8] D. Halpern and J. E. Katz, "Close but not stuck: Understanding social distance in human-robot interaction through a computer mediation approach." vol. 1, pp. 17–34, 2013.

[9] V. Bogicevic, M. Bujisic, A. Bilgihan, W. Yang, and C. Cobanoglu, "The impact of traveler-focused airport techonology on traveler satisfaction." pp. 351–361, 2017.

[10] E. Pacchierotti, H. I. Christensen, and P. Jensfelt, "Human-robot embodied interaction in hallway settings: a pilot user study," pp. 164–171, 2005.

[11] M. Shiomi, K. Shinozawa, Y. Nakagawa, T. Miyashita, T. Sakamoto, T. Terakubo, H. Ishiguro, and N. Hagita, "Recommendation effects of a social robot for advertisement-use context in a shopping mall," vol. 5, pp. 251–262, 2013.

[12] Y. Kim and B. Mutlu, "How social distance shapes human-robot interaction," Int. J. Human-Computer Studies, vol. 72, pp. 783–795, 2014.