HRI Users' Studies in the Context of the SciRoc Challenge Some Insights on Gender-Based Differences

Lun Wang
Sapienza University of Rome, Italy
wang@diag.uniroma1.it

Luca Iocchi
Sapienza University of Rome, Italy
iocchi@diag.uniroma1.it

Andrea Marrella
Sapienza University of Rome, Italy
marrella@diag.uniroma1.it

Daniele Nardi
Sapienza University of Rome, Italy
nardi@diag.uniroma1.it

ABSTRACT

In this paper, we present the outcomes of the first user study designed and evaluated in the context of the Smart City Robotics Challenge (SciRoc Challenge). The study presented in this paper has the main novelty of having been devised and implemented in a realistic environment: a robot competition where robot tasks were developed by participant teams, robots were fully autonomous, and user questionnaires were part of the competition score.

Specifically, our study was performed over a scenario configured to instruct a robot to take an elevator of a shopping mall asking for customers support. Leveraging the dedicated questionnaire designed for the tested scenario, we validated the experimental hypothesis if *user perception of robots' behaviour may be influenced by the user's gender*. In the end, we discuss the results of our study.

CCS CONCEPTS

• **Computer systems organization** → **Robotics**; • **Human-centered computing** → **User studies**.

KEYWORDS

HRI Evaluation, User study, Robot competitions, SciRoc Challenge

ACM Reference Format:
Lun Wang, Luca Iocchi, Andrea Marrella, and Daniele Nardi. 2020. HRI Users' Studies in the Context of the SciRoc Challenge Some Insights on Gender-Based Differences . In *Proceedings of the 8th International Conference on Human-Agent Interaction (HAI '20), November 10–13, 2020, Virtual Event, NSW, Australia.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3406499.3418763

1 INTRODUCTION

In this paper, we present some outcomes of the first user study designed and evaluated in the context of the Smart City Robotics Challenge (SciRoc Challenge[1]) in the field of Human-Robot Interaction (HRI). SciRoc Challenge is a repeatable and general-purpose test method (benchmark) developed for HRI teaming performance evaluation investigating users' attitudes using HRI methodology.

Previous evaluation studies that performed HRI analysis in dedicated lab and field tests, often by employing Wizard-of-Oz methods [7], we concretely enacted an experimental study to investigate users' attitudes in a specific HRI scenario involving fully autonomous robots. Moreover, differently from other competitions (e.g., RoboCup@Home [4]), external users were involved to assess the teams performance through a specific questionnaire and such assessment contributed to the competition score. The considered HRI task has the following features: i) *realistic and dynamic social environment*; ii) *representative* sample of *users* selected from the crowd by SciRoc organization; iii) robots configured to *act autonomously*, without the need of any external guidance. To this end, we exploited the robot competition organized for the first SciRoc Challenge, held in the shopping mall of Milton Keynes (UK)[2], in September 2019. The key objective of SciRoc is to study how robots will integrate in the (smart) cities of the future as physical agents living in them. The SciRoc challenge has been divided into a series of *episodes*, each consisting of a task addressing specific research issues. To perform our user study, we focused on Episode E4 "Take the elevator" that contains several elements for social HRI. In E4, the robot must take an elevator of MK crowded together with customers to reach a service located in another floor. The robot can enter/exit the elevator at the right floor in the presence of people nearby and/or inside and can interact with the customers in spoken language. The robot is not supposed to push buttons, but it can ask the people around to do it. The implementation of the robotic task was provided by the participant teams and the robots executed the task in a fully autonomous mode.

According to the rules of E4, four users – randomly selected by SciRoc organizers – were involved to join the task performance.

In this paper, we focus on gender-based differences and validate the following experimental hypothesis: *user perception of robots' behaviour may be influenced by the user's gender*. Exploiting the data collected with the questionnaire designed for the tested scenario[9], we found that: User perception of robots' behaviour is influenced by user's gender in relation to *Perceived Reponsiveness*, *Perceived Interactivenss* and *Perceived Naturalness*.

[1] https://sciroc.eu/

[2] https://www.centremk.com/

HAI '20, November 10–13, 2020, Virtual Event, NSW, Australia

ACM ISBN 978-1-4503-8054-6/20/11.
https://doi.org/10.1145/3406499.3418763

2 SETTING OF THE USER STUDY

Teams and Robots: five teams have entered to E4, employing five different robot variants: *UC3M*, *Gentlebots*, *HEARTS*, *eNTiTy*, and *LASR*.

Scenario and Role of Users in E4: according to the rules defined for E4, the robot encounters two persons while moving towards the elevator. Person with *role A* stands in a pre-defined location not far from the elevator. S/he can observe the robot, but is not interested in interacting with it. Person with *role B* actively moves towards the robot willing to interact with it. Once arrived in front of the elevator, the robot encounters two other persons. Both persons take the elevator with the robot. At this point, the robot interacts randomly with one of them asking to push the button for the floor it wants to reach. The two persons that take the elevator are instructed to reach a specific floor, which can be different (or the same) than the one assigned to the robot. The two persons in the elevator play the following roles: Person with *role C1* always gets off before the robot; Person with *role C2* gets off together with (or after than) the robot. Note that when a floor is reached, the robot interacts with one of the persons in the elevator asking her/him whether the floor reached is the right one for it.

Runs of experiment: the SciRoc challenge lasted 4 days. In total, 10 runs of episode E4 were scheduled. 9 general runs were performed in the first 3 days, while the final run was performed in the last day. The duration of the task was around 10 minutes. In each general run, all the five teams have performed the task foreseen in E4 according to run/trial schedule. In the final run, only the four teams with highest scores have performed the task.

Participating Users: to collect users' feedback avoiding any kind of "experimental noise", we relied on the *mixed-model factorial design*, which includes both between-subjects and within-subjects components [1]. Specifically, the study involved a total of 40 users. The same 4 users participated only to one of the 10 runs. In any run, 5 different teams/robots (within-subject) performed the test according to the run schedule. User's gender (between-subject) was declared by users, before the starting of any run.

Questionnaire and Experimental Hypothesis: at the end of any run, the participating users filled a dedicated questionnaire built ad-hoc for this episode [9]. The questionnaire has been thought to specifically keep track of 17 behavioural aspects related to: *(i) social behavior of robot*, *(ii) proxemics between human and robot*, and *(iii) collaboration with robot*. The scores assigned in the scale range from: *Absolutely No=1* to *Absolutely Yes=5*. If compared with the original questionnaire [9], we decided to convert the only negative behavior, "Perceived Strangeness", into its "positive version", i.e., "Perceived Naturalness "(see details in https://bit.ly/3jf1jYM). Here, we address the following experimental hypothesis: *The behaviour perceived of the robot is influenced by user's gender.*

Data collection: we collected 196 questionnaires overall, of which 78 were considered as invalid (i.e., not filled at all because of a failed test in a run, which was a circumstance happened especially in the first day of the competition), while 118 were considered as valid. Gender distribution was 23 male and 17 female users.

Reliability of the questionnaire: we calculated the Cronbach's alpha coefficient (α) for all of the three macro categories of the questionnaire, obtaining the following results: α of *Social Behavior of robot* = 0.907; α of *Proxemics between human and robot* = 0.921; α of *Collaboration with robot* = 0.83. According to [3], which discusses cut-off values for reliability indices, values of α coefficient greater than 0.8 indicate a reliability of the adopted scale among very good and excellent.

3 GENDER ANALYSIS

Several HRI studies have investigated how human users perceive the interaction with social robots, including the analysis of the role played by human gender in the perception of robots' behaviour in HRI [2, 5, 6, 8]. Even if in the research literature (the above works are just a representative sample) there is evidence that males and females view robots differently, from the way robots are conceptualized, to the way humans respond when they interact with them, the fact is that all the empirical studies supporting these results were mainly conducted in *controlled environments* with carefully *controlled laboratory conditions*, often with a *non-representative set of users*, or by employing Wizard-of-Oz methods [7].

To explore the possibility of performing HRI user studies in the context of robotic ompetitions, we investigated the impact of gender on users' perception in the context of E4 of the SciRoc Challenge. We first completed the missing data using mean imputation method, and then conducted Mixed-ANOVA to check how male and female users perceived differently the behaviour of the robot. We found no *interaction effect* between with-in subject factor (i.e., teams) and between-subject factor (i.e., gender), meaning that the impact of between-subject factor does not depend on the level of with-in subject factor. However, we found highly significant difference of *main effect* among the with-in subject factor, meaning that the overall effect over with-in subject effects is statistically significant, but this finding will not be further discussed in this paper.

For the following items related to *Social Behavior* of the robot: *Perceived Responsiveness (p=0.02)*, *Perceived Interactiveness (p=0.03)* and *Perceived Naturalness (p=0.019)*, we found significant differences between female and male users, meaning that female users perceived the robot's behaviour more positively than male users. No other significant difference of between factor has been found in this analysis study. As a consequence, we can partially confirm the validity of our hypothesis: Only the social behaviors of robots, i.e., *Perceived Responsiveness*, *Perceived Interactiveness* and *Perceived Naturalness* are influenced by users' gender. Consequently, the designers of social robots should make sure that the interaction style of the robot fits the user's gender and the user individual attributes.

4 CONCLUDING REMARKS

We present an experimental scenario for users' studies of HRI implemented through a robotic competition, and validate the experimental hypothesis on gender-based differences. The results of gender analysis have confirmed that user's gender does affect the social experience with a robot in the context of the SciRoc challenge. We are currently further developing our analysis of the data collected in the context of the SciRoc challenge, in particular, we aim at investigating if users' role in E4 had some impact on their perception of robots' behavior.

Acknowledgements. This work has been supported by the H2020 project SciRoc (GA 780086) and the AI4EU project (GA 825619).

REFERENCES

[1] Cindy L Bethel and Robin R Murphy. 2010. Review of human studies methods in HRI and recommendations. *International Journal of Social Robotics* 2, 4 (2010), 347–359.

[2] Friederike Eyssel, Laura De Ruiter, Dieta Kuchenbrandt, Simon Bobinger, and Frank Hegel. 2012. If you sound like me, you must be more human': On the interplay of robot and user features on human-robot acceptance and anthropomorphism. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE.

[3] S George and L Mallery. 2003. Alfa de Cronbach y consistencia interna de los ítems de un instrumento de medida. *Revista de estudios Interdisciplinarios en Ciencias Sociales* 3, 16 (2003), 3–9.

[4] Luca Iocchi, Dirk Holz, Javier Ruiz-del-Solar, Komei Sugiura, and Tijn van der Zant. 2015. RoboCup@Home: Analysis and results of evolving competitions for domestic and service robots. *Artificial Intelligence* 229 (2015), 258–281.

[5] Tatsuya Nomura. 2017. Robots and gender. *Gender and the Genome* 1, 1 (2017).

[6] Tatsuya Nomura and Satoru Takagi. 2011. Exploring effects of educational backgrounds and gender in human-robot interaction. In *2011 International conference on user science and engineering (i-user)*. IEEE.

[7] Laurel D Riek. 2012. Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines. *Journal of Human-Robot Interaction* 1, 1 (2012).

[8] Mikey Siegel, Cynthia Breazeal, and Michael I Norton. 2009. Persuasive robotics: The influence of robot gender on human behavior. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE.

[9] Lun Wang, Luca Iocchi, Andrea Marrella, and Daniele Nardi. 2019. Developing a Questionnaire to Evaluate Customers' Perception in the Smart City Robotic Challenge. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1–6.