Real-Life Experiment Metrics for Evaluating Human-Robot Collaborative Navigation Tasks
Ely Repiso, Anaís Garrell, Alberto Sanfeliu

To cite this version:
Ely Repiso, Anaís Garrell, Alberto Sanfeliu. Real-Life Experiment Metrics for Evaluating Human-Robot Collaborative Navigation Tasks. IEEE RO-MAN 2023, Aug 2023, Busan (Corée), South Korea. hal-04206085

HAL Id: hal-04206085
https://hal.laas.fr/hal-04206085
Submitted on 15 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Real-Life Experiment Metrics for Evaluating Human-Robot Collaborative Navigation Tasks

Ely Repiso¹ Anaïs Garrell² Alberto Sanfeliu²

Abstract—As robots move from laboratories and industries to the real world, they must develop new abilities to collaborate with humans in various aspects, including human-robot collaborative navigation (HRCN) tasks. Then, it is required to develop general methodologies to evaluate these robots’ behaviors. These methodologies should incorporate objective and subjective measurements. Objective measurements for evaluating a robot’s behavior while navigating with others can be accomplished using social distances in conjunction with task characteristics, people-robot relationships, and physical space. Additionally, the objective evaluation of the task must consider human behavior, which is influenced by changes and the structure of their environment. Subjective evaluations of robot’s behaviors can be conducted using surveys that address various aspects of robot usability. This includes people’s perceptions of their interaction during their collaborative task with the robot, focusing on aspects such as sociability, comfort, and task-intelligence. Moreover, the communicative interaction between the agents (people and robots) involved in the collaborative task should also be evaluated. Therefore, this paper presents a comprehensive methodology for objectively and subjectively evaluating HRCN tasks.

I. INTRODUCTION

If we plan to have robotics partners in the future, these partners need to be embedded with human-like navigation behaviors. Also, these behaviors need not only to focus on accomplishing the task satisfactorily. These behaviors must include those robots move predictably and socially to increase the number of potential users for these robots and the satisfaction of using the robots [1]–[3], as well as people’s trust in robots [4], and people’s comfort [5], along with people’s perception of safety [6].

Thinking on this objective, we need to evaluate any HRCN (Examples in Fig. 1) in a way that includes the same social norms that humans use. For example, by using works like the proxemics’ rules of Hall [7] and other studies [3], [8]–[10]. Proxemic rules are related to how humans use their surrounding space. Suppose we focus only on human-human social interactions. In that case, this use of the space is defined as several interpersonal distances to socially interact between humans depending on their level of familiarity as defined by Hall [7].

This use of the space can be translated to Human-Robot Social Interactions and, more concretely, to Human-Robot Collaborative Formations while navigating together or other types of formations to perform different tasks. For example, this work [3] includes several examples of other types of proxemics. For example, when a person is looking at a bulletin board, there are specific unwritten rules that humans respect, such as not passing between that person and the bulletin board that the person is looking at. The proxemic rules include these rules because they are related to physical distances around people. However, in this case, they refer to interactions between humans and objects, not social interactions between humans, that the robot should consider in its behavior, just like the humans do.

Many recent works in Human-robot collaborative navigation use proxemics [11]–[23]. Then, it is essential to have some guidelines to develop metrics to evaluate the robot behavior using proxemics. In addition, we can use proxemics and the following guidelines in more situations than only human-robot collaborative navigation, as several works outside this field use proxemics [20], [24]–[39]. For example, we can use them with a robot serving food at a table, Fig. {3 and 4}-left.

Furthermore, we want these robots to be helpful to the general population. In that case, we must evaluate the robot’s

Fig. 1: Different types of HRCN tasks where authors have tested the objective and subjective metrics presented. Up-left: Robot Side-by-Side Accompaniment. Up-right: Robot V-Form Accompaniment. Middle: Robot Accompaniment Plus Approaching. Bottom: Robot Approaching.
be placed will change taking into account the formation of said task (explained in Sec. II-B). Based on this, the robot’s personal space can be defined and located within the formation in the best area for the robot to perform the task. Additionally, if the robot’s velocity is taken into consideration, the area should be adjusted to include a safety margin of free space to ensure the robot has enough distance to stop before colliding with any obstacle. The robot’s appearance can transform this physical area into a subjective area, which will increase to make the person feel comfortable during the interaction. This fact can be similar to the relation with the robot. For people who feel familiar with the robot, their personal distances between them and the robot can be smaller, and for unfamiliar people, these distances should be greater [53]. In addition, this area can be similar to a person’s personal space considering her personality. These last cases are more subjective than robot size. Then, there is still much work to be done to know these distances. Examples of different cases of a robot’s personal area to account for its appearance and velocity can be seen in Fig. 2, as well as how people’s personal space can change depending on their personality.

Several works in the literature explore the physical space around people depending on several aspects [3], for example, different shapes of personal space, etc. Then, we recommend that the readers explore the state-of-art works and find the best person and robot’s personal space for their implementation. Also, we have explored these distances for different velocities and positions between robot and person in our previous work [47], and we confirmed through experiments these findings [16], [54].

**B. Consider the robot task in proxemics**

In order to tailor the metrics for a specific task, it is necessary to have an understanding of the physical formation that individuals should adopt when carrying out the task, the social relation between the members of the group, as well as their preferred personal space. Then, these proxemics include several areas with different meanings of interaction, and we need some rules to create these different areas. These areas are: the area where people can notice the interaction with the robot (Interaction area) in Sec. II-B.1, the best area where the robot should be located to perform in a good way the task (the best area to perform the task) in Sec. II-B.2, and the areas where the robot should not enter (forbidden areas) in Sec. II-B.3.

1) **Definition of the interaction area:** The robot should stay inside this area to allow people to notice that they are socially interacting. This area delimits the area of interaction between humans and robots, and it is from the human point
of view. Then, it is centered on the human position. Outside this area, people should not feel they are interacting with the robot, which corresponds to one of the two types of forbidden areas in red. However, it is not the best area for the robot to perform the task. So within this area, the value of the robot’s performance will be half of any scale. The forbidden areas will be described in Sec. II-B.3, and the best area to perform the task will be described in Sec. II-B.2.

For the interaction between one robot and one person, we can look at Hall [7] studies for the social distances between two people. We focus on these studies because some researchers agree that proxemic rules between a person and a robot should be similar to the ones among people [55], and we have followed this idea among all our previous works. Then, the area best suited for people’s social interactions with robots is the area of social distances, as we currently do not have an intimate or personal relationship with robots. In addition, these types of robots are called social robots because they are designed to maintain social interaction with us. Therefore, it is logical that through real-life experiments from our previous works [47], [54], we have found that the area that best adapts to these interactions is the social area defined by Hall [7].

In cases of one-person and one-robot interaction (or interactions with more than one person but without other obstacles or people between the robot and the person), this area of social distances is the best one to delimit the interaction between people and robots. For example, this is the case of the robot’s one-person accompaniment, the robot’s group accompaniment when the robot is at the center of the formation that can interact equally with both people, the robot’s waiter for the blue person, and the approaching where the triangle formation allows the same robot’s interaction with both people, see Fig. 3.

In addition, the size of this area should be enlarged depending on the number of group members, the robot’s position inside this group, and the environment’s elements. Then, this area should increase if there are other humans or objects between the robot and the person. For example, this is the case of the robot waiter for the different people at the table and the robot’s accompaniment of a group of people when the robot is at the side of the formation. In the case of the waiter, depending on the person’s position on the table, this distance is smaller or larger with respect to the robot’s

2) Definition of the best area to perform the task: The best task achievement closely relates to the formation of the group of people to develop this task, the robot’s physical position inside this group, the robot’s physical space (Sec. II-A).

Regarding the physical formation to perform a task, we should know this specific formation for the case of people interactions. For example, in the case of group accompaniment, we can have two different formations to do the accompaniment: side-by-side and V-formation, as shown in Fig. 5. We have extracted these formations from studies about group people dynamics that study these people formations [56]–[58]. If more than one robot position within the group is possible, we must consider them to develop different metrics for all these cases. For example, in these two different group formations for two people accompaniment, there are two different metrics to consider depending on the two different robot’s positions inside the group formation, at the lateral or the center [51], [52], see Fig. 5. As we have noticed in Fig. 5, the best robot position changes depending on these two factors: the formation to perform the task and the position of the robot inside the group formation.

Furthermore, a previous task can coerce this best area to perform the task. For example, it is the case of combining a people accompaniment with approaching a second person. In this case, the robot can select two possible positions to perform a triangle formation with both people, and the best one should be selected to have a more natural robot behavior.
Then, the robot should select the position of the triangle formation nearest to its previous position on one of the sides of the person it accompanies. You can see this behavior in the two left images of Fig. 6. Several previous researchers used this formation for people interaction similarly to us [3], [59], [60]. Also, this formation is sometimes generalized for more than three people as an O-shape formation.

Finally, the reader should consider the robot’s physical area to set up the radius that delimits the best area to perform the task. Also, this area will be more realistic if you consider the robot’s personal space. For the robot case, this space is the security distance that it needs to stop in case that any object or person interferes in its path. Regarding our robot, it is 0.3 cm from its sensors, which are located at 0.5 m of its center. It is to say, the minimum area for this best area needs to be 1.3 m around the center of the robot position. This area is drawn in green in all the figures from this point until the end of this paper.

Furthermore, the reader should notice that these metrics consider the best position for humans, but we can not control them. However, to design the people’s positions within the formation, we can expect people to behave similarly to how they would do with other people in that situation. We can expect this behavior because it is the most similar interaction that people know. Therefore, the robot will have the best performance value if it is located in these green areas that represent the best area to perform the task for the robot. In the case of the scale from 0 to 1, it corresponds to the value of 1.

3) Definition of the forbidden areas: These metrics must consider the worst areas where the robot can not be located. When the robot is inside these areas, its performance is 0, and in some cases, we also need several security measures to prevent the robot will enter there in any case. For example, with a security system that stops the robot when any object enters inside its radius of security, 0.3 m from its sensors. Two possible cases need to be considered to find these forbidden areas:

First, the robot, in any case, should be allowed to invade any personal space of any human. Then, forbidden areas should be located surrounding humans. Also, the form of these areas can be different depending on if the human is stopping or walking, and should include the human velocity. In all our previous works, we have included human velocity using the prediction of human movement. Then, we propagate static people’s forbidden areas around them using their velocity (Fig. 7-left). Consequently, our forbidden areas around humans do not include the velocity because we include it in the propagation. However, suppose other researchers do not use people’s prediction and propagation of their movement. In that case, the people’s velocity should be included directly in the shape of these forbidden areas, similar to the case of the robot in Fig. 2. In addition, these areas can include the level of familiarity that the person has towards the robot. For example, extroverted people that like robots should prefer near distances between them, and introverted people or afraid of robots should prefer higher distances between them, which also means more or less personal space for the person, Fig. 2 and Fig. 7-right.

Second, if the robot goes out of the area of social distance, the person can not notice that it is interacting with her. So then, this situation should be avoided by the robot. Outside the area of social distances, there is another forbidden area for the robot. However, in this case, it does not need any security measures to prevent the robot from entering there because it will not harm anyone.

C. Consider the environment in proxemics

These metrics should consider the group’s environment because the people groups’ formations change depending on
Consider person velocity in Personal Space  

Distances of Accompaniment Dependent of Personality

Fig. 7: Left: we show how the personal space of the person is propagated due to its velocity and the uncertainty of the person’s movement while the time increases. Right: we show how the distance of accompaniment can increase due to the accompanied person’s personality.

Fig. 8: Best robot’s area. It depends on the formation changes due to the environment because the group of people formations should change their structure to avoid obstacles together.

Furthermore, the position of these areas depends simultaneously on the behaviors of the people interacting with the robot, for example, in the case of the side-by-side while approaching a person where there is an obstacle near the approached person. The movements and best positions of the robot to accompany the person and finally to approach the other person depend on these two people’s behaviors. For example, suppose they allow enough space for the robot to move more naturally. In that case, these areas will be different that in the case of these two people prefer to make the minimum effort and induce the robot to surround the accompanied person to interact with the approached person. See Fig. 9.

So then, researchers must use methods that compute dynamic metrics that will be modified automatically, considering the task, the position of all the group members during the task performance, the environment’s configuration, and the behavior of the people interacting with the robot. There are more examples in [49], [52].

III. SUBJECTIVE METRICS

We can use the theory of USUS questionnaires [40] to evaluate the robot behavior in a subjective way considering the user’s feelings about their experience while they perform the tasks together. We have selected to evaluate the robot’s sociability, comfortableness and “task-intelligence”, and the group’s communicative interaction. Our proposed questions can be seen in Table I. These questions can be used in any Human-Robot Collaborative task because these are presented in a general way and can also be adapted to other different tasks than navigation tasks. However, the researchers need to customize these questions for the concrete number of agents (humans and robots) involved and the tasks (accompaniment, approaching, remove together the objects from a table to a container, searching for objects or people in an environment, or other) because non-experts in robotics tend to lose focus on the task they have just completed with the robot if generic questions are used. This fact has been pointed out during several previous tests, and the presented questions have demonstrated their utility in conducting user studies in several works on HRCN [16], [51], [52], [54].

IV. EXPERIMENTAL EVALUATION

The presented objective and subjective metrics were employed in over 20.000 simulations and 600 real-life experiments to evaluate the 5 different types of robot’s collaborative navigation [16], [51], [52], [54], [61], including 6 user studies. In this paper, we provide a summarized overview of all the results, which serve as experimental evidence for the effectiveness and usefulness of these metrics. The formulas for these metrics, along with two additional formulas to assess distance and angle failures, can be found in the author’s PhD thesis [62]. Furthermore, these documents will be made available on the author’s website1 in the coming months, or interested readers can directly request them. Our chosen maximum velocity for the robots is 1.2 m/s, which corresponds to the typical walking speed of humans. We

1https://elyrepiso.wordpress.com/
the method is compared with the teleoperation to assess if also the group’s communicative interaction as Tab. III shows. Sociability, and intelligence perceived by the volunteers and really represent what they are supposed to measure. In our to the accuracy of a measure, it is to say, whether the results approach, iterating as necessary to ensure that the questions similar conditions. For validity, we employed a test-retest context refers to the ability to reproduce results under valid. In all cases with a $p-value > 0.05$, our hypothesis that both methods were considered equal by the volunteers was confirmed. However, we did find a statistically significant difference in the case highlighted in light red. Analyzing the mean values of robot comfort for the ASP-VG vs. ASP-SG, we observed a higher level of comfort reported for ASP-SG. Based on the volunteers’ comments, we deduced that for people were more comfortable if they can see and feel the robot near them (ASP-SG case).

V. Conclusions

This paper provides a comprehensive set of guidelines for researchers seeking to develop objective and subjective metrics to evaluate a wide range of human-robot collaborative navigation (HRCN) tasks, such as solo navigation, guidance, accompaniment, approaching and so on. The objective metrics outlined in the paper take into account various social factors identified by Hall [7] and transform them to include the formation of the task, the relationship between the person and the robot, the physical appearance and personality of the agents, the positions of other people and the structure of the environment. Additionally, these metrics consider people’s behavior in relation to their environment.

On the other hand, the subjective metrics are concerned with assessing people’s perceptions about the robot’s comfortableness, sociability, intelligence, and group communicative interaction while they interact together. Finally, the paper highlights the effectiveness of these metrics in evaluating different types of HRCN with potential users, as evidenced by previous works conducted by the authors.

The guidelines presented in the paper offer a solid foundation and a source of inspiration for other researchers who wish to develop metrics to evaluate their experiments in the field of HRCN, as well as other types of human-robot collaborative tasks. By adhering to these guidelines, researchers can ensure that their metrics are comprehensive, accurate, and reliable, thereby advancing the state of the art in human-robot collaboration.

ETHICS APPROVAL

The Ethics Review Board of Universitat Politècnica de Catalunya approved the study, with title: Navegació Robot-Humano Colaborativa en entornos con personas y predicción de movimiento humano. The Ethics Review Board decision number is: 2021-11 UPC. Then, informed consent was obtained from all individual participants included in all studies.

REFERENCES

TABLE II: Results of all experiments about the mean and standard deviation of the Area performance metrics extracted from the proxemic metrics of this paper. Notice this value is always between 0 and 1, as said previously. For the approaching line intersection, real-life experiments were not done because the optimal intersection outperforms this behavior.

<table>
<thead>
<tr>
<th>Simulation experiments: Area metric extracted from the proxemic metrics</th>
<th>One Person Side-by-Side</th>
<th>Two Person Side-by-Side</th>
<th>Two Person V-form</th>
<th>Approaching with Splines</th>
<th>Approaching using line intersection</th>
<th>Approaching using optimal intersection</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.8639</td>
<td>0.7813</td>
<td>0.7582</td>
<td>0.9719</td>
<td>0.6121</td>
<td>0.8163</td>
</tr>
<tr>
<td>std</td>
<td>(± 0.0293)</td>
<td>(± 0.0976)</td>
<td>(± 0.0685)</td>
<td>(± 0.0722)</td>
<td>(± 0.0086)</td>
<td>(± 0.0111)</td>
</tr>
<tr>
<td>Real-life experiments: Area metric extracted from the proxemic metrics</td>
<td>mean</td>
<td>0.7764</td>
<td>0.8049</td>
<td>0.7089</td>
<td>0.9599</td>
<td>—</td>
</tr>
<tr>
<td>std</td>
<td>(± 0.0478)</td>
<td>(± 0.1373)</td>
<td>(± 0.0651)</td>
<td>(± 0.0538)</td>
<td>—</td>
<td>(± 0.0168)</td>
</tr>
</tbody>
</table>

TABLE III: All user studies results using the questioners of the current paper

<table>
<thead>
<tr>
<th>Simulation experiments: Area metric extracted from the proxemic metrics</th>
<th>ASP-SG vs Teleop</th>
<th>ASP-VG vs Teleop</th>
<th>ASP-VG vs ASP-SG</th>
<th>ASP-VG lateral vs central</th>
<th>ASP-SG lateral vs central</th>
<th>Accompaniment plus approaching</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cronbach’s alpha</td>
<td>0.71</td>
<td>0.75</td>
<td>0.7</td>
<td>0.82</td>
<td>0.82</td>
<td>0.72</td>
</tr>
<tr>
<td>ANOVAs tests plus Pairwise comparison with Bonferroni</td>
<td>p-value=0.2</td>
<td>p-value=0.2</td>
<td>p-value=0.02</td>
<td>p-value=0.2</td>
<td>p-value=0.91</td>
<td>p-value=0.01</td>
</tr>
<tr>
<td>Robot’s Comfortableness</td>
<td>p-value=0.5</td>
<td>p-value=0.9</td>
<td>p-value=0.2</td>
<td>p-value=0.2</td>
<td>p-value=0.41</td>
<td>p-value=0.1</td>
</tr>
<tr>
<td>Robot’s Sociability</td>
<td>Not Measured</td>
<td>Not Measured</td>
<td>p-value=0.2</td>
<td>p-value=0.85</td>
<td>p-value=0.8</td>
<td>p-value=0.2</td>
</tr>
<tr>
<td>Robot’s Intelligence</td>
<td>Not Measured</td>
<td>Not Measured</td>
<td>p-value=0.54</td>
<td>p-value=0.81</td>
<td>p-value=0.01</td>
<td>Not Measured</td>
</tr>
</tbody>
</table>


