Increasing perceived value between human and robots - Measuring legibility in human aware navigation

Christina Lichtenthäler, Tamara Lorenz, Michael Karg, Alexandra Kirsch

Increasing Perceived Value Between Human and Robots - Measuring Legibility in Human Aware Navigation

Christina Lichtenthäler*, Tamara Lorenz† ‡, Michael Karg*, Alexandra Kirsch*

***Abstract*— Robots will more and more enter our daily life. In order to increase their acceptance it is necessary that their movements and behavior are predictable. With our present experiment we assess the acceptance of autonomous robots in human working and living environments. As a specific indicator we define legibility as an important prerequisite for user acceptance. In a simulator study participants rated the navigation behavior of a robot with regard to several aspects of legibility. Results show that Human Aware Navigation is a method to increase the perceived value of robot navigation behavior.**

I. INTRODUCTION

In the near future not only vacuum cleaner robots will live in our houses, robots will also help us doing the housework. However, for a successful human-robot interaction the user acceptance of a robotic system plays an important role. Only if humans feel comfortable with a robot it will be accepted and used. Thus, it is necessary to adapt the robots behavior which is controlled by its underlying algorithms. Our main objective is therefore to develop algorithms causing accepted behavior. We have to evaluate these algorithms with respect to acceptance and usefulness. In marketing research we find evaluation methods dealing with the acceptance, usefulness and value of products. For this reason we derived our evaluation method from marketing approaches. Here, the term *perceived value* describes the subjective value attached to a product, which is not necessarily connected to its objective worth or price. It rather reflects the degree to which user needs are satisfied by the product [1]. The perceived value consists of different properties: perceived safety, reliability, comfort, usefulness and confidence. In the domain of human-robot interaction all these factors depend in part on the ability to infer the actions of the robot. Imagine a robot performing some household chore, such as preparing a meal. The robot fulfills its duty, but manipulates objects with sudden, unpredictable movements, rushes through the kitchen with rapid changes of direction or ignores obvious errors like a pot not being placed properly on the stove. Even if this robot will eventually serve the meal, a person might not have enough confidence to leave this robot alone, because its actions are not understood nor predictable. This reduces the perceived usefulness of a household robot: if the presence of a person is necessary (or is felt to be necessary), the person might just as well do the job alone.

∗Institute for Advanced Study, Technische Universität München
†Experimental Psychology, Ludwig-Maximilians-University München
‡Automatic Control Engineering (LSR), Technische Universität München

We therefore assume that an important factor for determining the perceived value of a robot is the *legibility* of its actions, which we define as follows:

Definition 1.1 (Legibility): Robot behavior is legible*, if a human can infer the next actions, goals and intentions of the robot with high accuracy and confidence.*

Legibility is a valid measure with and without human interaction because it is unlikely that robots and users will interact permanently and even when the robot acts independently, a human observer should feel confident. In order to develop algorithms causing accepted behavior we compare algorithms with respect to their legibility. We present the design and the results of an experiment measuring the legibility of different navigation methods. The purpose of the experiment is to compare human reactions towards perceived robotic behavior caused by different navigation methods in dynamic situations. This experiment is intended to be a prestudy before we carry out a similar experiment in a real live scenario with a real robot. We want to answer the question if the behavior of the robot caused by an algorithm or method is legible and furthermore which of the navigation methods causes the most legible behavior. We assume that the most legible algorithm is the best algorithm regarding user acceptance as well as to a high perceived value. The used framework to measure legibility of the robotic behavior is taken from Lichtenthäler et al. [2]. The framework describes an experimental setup to measure legibility by gathering explicit and implicit feedback from participants observing the robot's behavior. In our experiment we only collect explicit feedback with a questionnaire.

In the following we first state how our approach fits the context of related research (section II). Afterwards we introduce our experimental design (section III) and present the results (section V), which are discussed in section VI. Finally we conclude and provide an outlook for future work.

II. RELATED WORK

There is only a small number of contributions in the literature dealing with the legibility respectively the perceived value of robot behavior. Nevertheless there is the wide field of intention recognition [3]. Contrary to the task to predict the intentions of an agent whether it is a human or a robot we want to measure if the robot's behavior is predictable. Only in Takayama et al. [4] we find a comparable experiment towards the legibility (in their work called readability) of the robot behavior. They present a simulation-based experiment to verify the readability of robot behavior and found support

for their hypothesis that the readability is influenced by showing forethought and goal-oriented reactions.

In the experiment at hand we want to compare different navigation algorithms. Dautenhahn et al. [5] accomplished an experiment to compare three different approaching strategies. The objective of their experiment was to find preferences for approaching directions. They found that most subjects preferred to be approached from either the left or right side, with a small overall preference for a right approach by the robot. With the results at hand they developed approaching methods taking into account the preferred approaching directions.

Weiss et al. [6] concentrate on the social acceptance which is a important part of the perceived value as we mentioned it before. They developed a framework to assess the social acceptance for an outdoor robot based on questionnaires, showing that a variation of a breaching experiment is a reasonable method to evaluate robot behavior in real life situations.

III. EXPERIMENTAL DESIGN

The experiment is designed to address the question of how legible different navigation algorithms are to a human observer and how the resulting robot behavior influences the perceived value.

A. Navigation Methods

In our study we compare two variations of the human aware navigation designed by Kruse et al. [7] with two variations of the state of the art navigation method taken from the ROS navigation stack (www.ros.org/wiki/navigation). In this section we give a brief introduction to the navigation methods and the differences between them. All four navigation methods consist of a global and a local planner (see Fig. 1). The global planners we use generates the complete path to a goal using an A* algorithm. The local planner is seeded by the global plan and generates velocity commands to control the robot. Differences between the global planners result from the cost function the A* algorithm uses to compute the shortest path whereas the local planner differs in the method to determine velocity commands. In the following we describe the different global planner cost functions and the different local planners.

a) Move Base Global Planner: The cost function of the move base global planner (MB) is based on a 3D voxel grid. An obstacle causes infinite costs with descending costs in its surrounding to propagate them from obstacles out to a user-specified radius. For further informations see [8] and http://www.ros.org/wiki/navfn.

b) Human Aware Global Planner: The cost function of the human aware global planner (HA) takes the human, his/her motions and all obstacles into account. In addition to the infinite costs of obstacles it increases the costs around a human and differentiates between a standing and a moving person:

- Moving: higher costs in front of the human to avoid moving in his/her direction of motion.

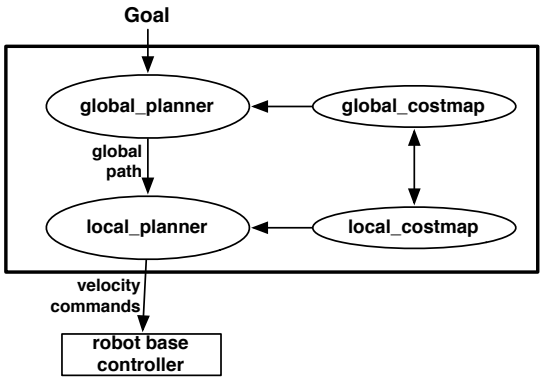


Fig. 1: Concept of navigation methods (www.ros.org/wiki/move_base)

- Standing: higher costs behind the human to avoid moving behind his/her back.

Those "social costs" should ensures a comfortable navigation behavior. Further informations can be found in [7].

c) Dynamic Window Approach: The dynamic window approach (DWA) is a real-time collision avoidance strategy developed by Fox et al. [9] (www.ros.org/wiki/dwa_local_planner). The DWA computes local controls by first determining a target trajectory in position or velocity space (usually a circular arc or other simple curve), then inverting the robots dynamics to find the desired velocity commands that will produce that trajectory [10].

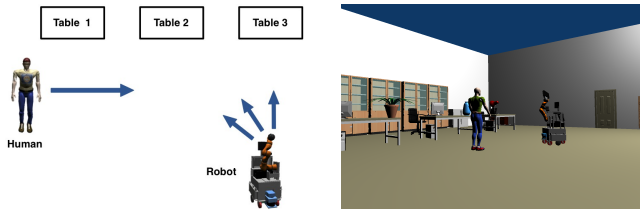
d) Waypoint Follower Local Planner: The waypoint follower (WF) is based on the human aware global planner. This local planner projects its own motion and the human motion into the future and selects a speed that avoids predictable collisions. The human motion is predicted linearly assuming constant speed and direction, while the robot motion is predicted using the global path returned by the human aware global planner.

e) Trajectory Planner: The trajectory planner (TP) by Gerkey et. al [10] (www.ros.org/wiki/base_local_planner) is based on a different paradigm then DWA. Instead of searching for feasible velocity commands to a trajectory the TP samples possible velocity commands and simulates the resulting trajectory. The TP algorithm chooses the best simulated trajectory by taking obstacle avoidance, goal distance and distance to the optimal path into account.

f) Navigation Methods: For our study we use the following four navigation methods.

- **MB-DWA:** The move base global planner with the dynamic window local planner.
- **MB-TP:** The move base global planner with the trajectory planner as local planner.
- **HA-WF:** The human aware global planner with the waypoint follower local planner.
- **HA-TP:** The human aware global planner with the trajectory planner as local planner.

The MB-DWA method is a state-of-the-art navigation method used by plenty of research groups. We compared the navi-



(a) Experimental setup (b) Simulator view.

Fig. 2: Design of the navigation study

Fig. 3: Example of an crash with the human in the video clip.

gation method MB-DWA with the human aware navigation method HA-WF, to find out which concept is the best. Additionally we want to compare the different global planners MB and HA when comparing HA-TP against MB-TP. The comparison of HA-TP against HA-WF and MB-DWA against MB-TP shows us which is the best local planner with regard to legibility on the one side and perceived value in more global terms.

B. Participants

We recruited 16 participants with the average age of 26.6 years - thereof 3 women and 13 men. Eight participants have regular contact to robots, three from time to time and five have rarely or no contact to robots.

C. Experimental Setup

For this experiment we have recorded short movie sequences using the MORSE simulator [11] with a human crossing the robot's path in an office environment (see Fig. 2b). Providing a third person perspective we generate the feeling of a noninvolved observer. Woods et al. [12] have convincingly argued that videotaped trials are a feasible approach for pilot studies like this one. Our virtual environment shows three tables in an office scenario (see Fig. 2a). In order to simplify the nomination of the tables we put three different objects with different colors (green plant, blue vase, red lamp) on each of the three tables. Our robot has to deliver a folder to one of them while a person is crossing its way as shown in Fig. 2a. The robot and the human are starting from a fixed point in each of the clips..

In order to avoid the potentially directing effect of the robot's gaze behavior we choose a robot without any artificial head. This allows us to concentrate on the legibility of the robot's navigation behavior.

a) Conditions: With three tables as possible goal positions of the robot and four navigation methods, we tested $(3 \times 4) = 12$ different observation tasks. Each observation task was displayed as a video clip once per participant in random order. In three of the 12 clips you can see the robot crashing into the human (see Fig. 3). This behavior is caused by the local planning algorithms which are not taking the human into account (MB-DWA, MB-TP, HA-TP).

b) Procedure: Participants were presented with three videos to familiarize with the virtual environment. Here we showed how the robot would move towards each table without the presence of a human crossing the scene. After this the experiment started. In each video, explicit feedback was asked at two points: (1) at short distance between the crossing human and the robot. Here we ask the participant to judge which table the robot is aiming at and if it will change its direction to do so. Additionally participants were asked to rate their confidence on this judgment. Then the video was started again and participants were able to observe the actual behavior of the robot and its whole way towards the intended table. After the robot reached the intended table the video was stopped again (2). Participants were asked to tell whether the robots actual behavior was expected and if not to rate how surprising this was. Additionally we asked for ratings on the perceived safety, comfort, and reliability of the robot. All ratings were to be done on a five-point Likert scale on which '1' stands for no (i.e. confidence, comfort,...) at all, '5' for very e.g. confident.

IV. DATA ANALYSIS

Data Analysis was performed in Excel and SPSS. Due to the non-parametric character of our data, differences of frequencies were analyzed with Pearsons Chi-Square tests and assessed ratings with a Friedmans ANOVA. For post hoc analysis of the gathered results we applied Wilcoxon tests. A Bonferroni correction was applied and so all effects for post hoc tests are reported at a 0.0167 level of significance.

V. RESULTS

When the video was stopped for the first time, participants were asked to predict the goal of the robot. Results of correct prediction are shown in Fig. 4. Most correct answers were

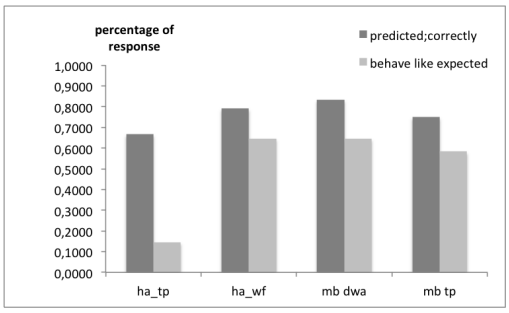


Fig. 4: Frequency of correct prediction and expected behavior in %.

given in navigation scenarios where MB-DWA (83.33%) and HA-WF (79.17%) were applied compared to MB-TP (75.00%) and HA-TP (66.67%), Pearsons Chi-Square tests

did not reveal any significant association between type of planner and number of correct responses, all $p > 0.05$.

The average confidence rating ranged from $M_{HA-TP} = 3.57$ to $M_{MB-TP} = 3.85$, $\chi^2_F(3) = 2.98$, $p > 0.05$, indicating that all navigation methods equally allowed for confidence rating above average.

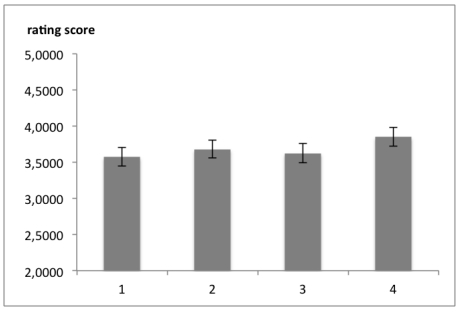


Fig. 5: Mean and standard error of the confidence factor rated on a five point Likert scale.

After participants were able to watch the actual behavior of the robot they were asked if this matched their expectations. Results of this question are shown in Fig. 4. In 50.52% of the trials expectation was met. No difference in expectations was found between MB-DWA (64.58%) and HA-WF (64.58%) or MB-TP (58.33%), both $p > 0.05$. However, expectations when HA-TP (14.58%) was applied differed significantly from expectations during observation of HA-WF, $\chi^2 = 25.09$, $p < 0.001$, and MB-TP, $\chi^2 = 19.83$, $p < 0.001$. This indicates that the robots performance when navigated by HA-TP is really not legible.

For the 49.48% of trials in which the robot behavior did not meet the expectations, we asked participants to rate their surprise (see Fig. 6). The ratings differed significantly between the navigation methods, $\chi^2_F(3) = 8.91$, $p < .05$. However, after correction post hoc tests only reveal marginal difference between HA-WF ($Mdn = 3.00$) and HA-TP ($Mdn = 4.00$), $T = 13.00$, $p = 0.04$, and no significant difference between HA-WF and MB-DWA ($Mdn = 3.25$) or MB-TP ($Mdn = 4.00$), both $p > 0.0167$.

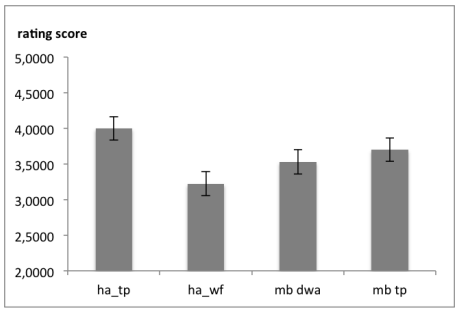


Fig. 6: Mean and standard error of the surprise factor rated on a five point Likert scale.

Regarding the rating of safety (see Fig. 7) there was a significant differences between navigation methods, $\chi^2_F(3) = 13.41$, $p < 0.01$. Post hoc tests reveal that safety was

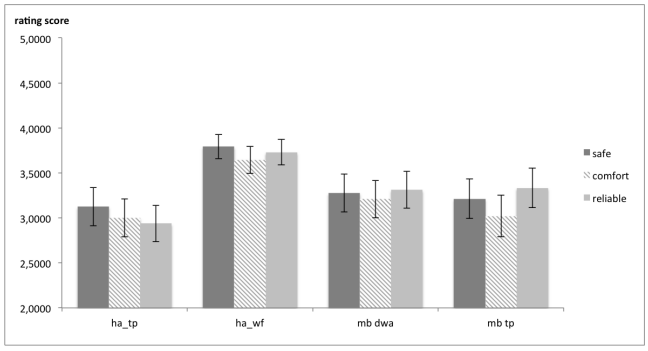


Fig. 7: Mean and standard error of the safety, comfort and reliability factor rated on a 5 point Likert scale.

perceived higher with the HA-WF ($Mdn = 3.83$) compared to HA-TP ($Mdn = 3.33$), $T = 11.00$, $p < 0.01$, and also marginally higher compared to MB-TP ($Mdn = 3.33$), $T = 24.00$, $p = 0.02$. No significant difference regarding perceived safety was found between HA-WF and HA-DWA ($Mdn = 3.00$), $p > 0.0167$.

Also the rating of comfort (see Fig. 7) showed a significant difference between navigation methods, $\chi^2_F(3) = 7.97$, $p < 0.05$. Here HA-WF ($Mdn = 3.33$) was perceived as more comfortable than HA-TP ($Mdn = 2.83$), $T = 9.50$, $p < 0.01$ and marginally more comfortable than MB-TP ($Mdn = 3.00$), $T = 14.50$, $p = .017$. Again no difference was found between HA-WF and HA-DWA ($Mdn = 2.83$), $p > 0.05$.

The reliability of navigation methods (see Fig. 7) also differed significantly, $\chi^2_F(3) = 8.83$, $p < 0.05$. Post hoc Wilcoxon tests showed that HA-WF ($Mdn = 3.67$) was perceived more reliable than HA-TP ($Mdn = 2.67$), $T = 6.00$, $p < 0.01$. No difference in reliability was found for HA-WF compared to MB-DWA ($Mdn = 3.17$) nor MB-TP ($Mdn = 3.33$), both $p > 0.0167$.

As mentioned before we defined the perceived value as consisting of safety, reliability, comfort and legibility. As we found no validated calculation method for the perceived value in the literature we computed the perceived value of each navigation method as a cumulative measure. We summed up the the safety, comfort, reliability factor and the legibility, which is composed of the results of correct prediction, expected behavior and the confidence factor. All factor measures, like safety were converted to % data.

$$\begin{aligned} pv &= (\text{safety} + \text{comfor}t + \text{reliability} + \text{prediction} \\ &+ \text{behave} + \text{confidence})/6. \end{aligned}$$

The results are shown in Figure 8. Numerically, the perceived value was highest for HA-WF (73.45%) whereas MB-DWA (69,39%) and MB-TP (66.95%) scored badly. However HA-TP (55.66%) is worse than all other navigation methods.

VI. DISCUSSION

With the study at hand we evaluate four different robot navigation methods in a simulated environment. A human observer had to watch video clips from a third-person perspective in which a robot was moving to an unknown

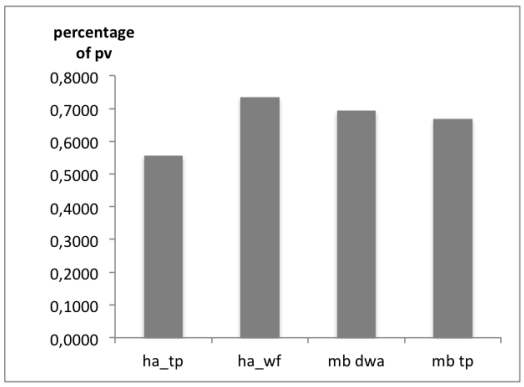


Fig. 8: Calculated perceived value.

target while a human was crossing its path. The video was stopped before the human and the robot crossed paths and the observer had to predict the future goal of the robot. Results of comparing legibility in the form of correct goal prediction between navigation methods did not show any significant differences. This is also mirrored in the reported level of confidence. Participants were equally confident in rating all navigation methods which means that all ratings and judgments on the different navigation methods were performed with comparable quality. Nevertheless, when assessing the numerical differences between numbers of correct prediction we found that the legibility of the robots behavior was worse for TP-local planners - independent of the applied global planners. While HA-WF and MB-DWA resulted in more correct predictions no striking differences between them were observed. The second judgment for accessing legibility was requested after participants had watched the video to the end and were thus able to see the whole performance of the robot. Expectations regarding the robot's behavior were met in over 60% of the trials applying HA-WF and MB-DWA respectively. Also during navigation by MB-TP expectations were met in more than half of the trials. However it is striking how few trials, namely less than 20%, were met during application of HA-TP. One possible explanation for this finding is that if one decides to apply human aware navigation in the global context, it is also essential to apply human aware local planning strategies. It almost seems as if a global human aware planner is useless regarding predictability of behavior when it is combined with a non human aware local planner. This also goes in line with better results for MB-TP. Here both global and local planner have no human aware navigation strategies which might indeed be a strategy that is attributed to a robot being a machine. Predictions and expectations regarding a robot behavior might be guided by that impression and higher ratings for HA-WF and MB-DWA might be a hint towards the fact that human aware strategies (HA-WF) or other dynamic approaches (MB-DWA) result in a higher legibility. In order to gain information from the trials in which expectations were not met, we asked participants to also rate their surprise about the actual behavior of the robot. Proofing the results from the analysis of met expectations, we found that the robots behavior during HA-TP caused highest surprise ratings. HA-WF caused lowest surprise rates and was also not significantly different from MB-DWA, the latter resulted in numerically higher ratings. One main difference between HA-WF and MB-DWA is that MB-DWA does include the possibility of crashes with the human because the human is not treated as a moving obstacle. One of the MB-DWA videos actually showed this possibility what potentially caused higher surprise rates. The same reason can of course also add to the lower legibility of the TP- local planner and cause higher surprise ratings during observation of MB-TP and HA-TP. In other words, this means that navigation with HA-WF was least surprising and thus best legible because crashes were never possible. One of our main goals in the study at hand was, to increase the perceived value of a robot behavior by increasing the legibility of its navigation. In order to access the perceived value we asked participants to rate the perceived safety of the robot, the comfort they would feel when interacting with it and the perceived reliability of its behavior. All three measures were rated best for HA-WF with the most striking difference to HA-TP. This proofs again that both local and global planner should be human aware by showing that navigation when combining a human aware global planner with a non human aware local planner is not perceived as safe in behavior nor reliable and does not result in a comfortable interaction. Analyzing safety and comfort ratings, we also found that HA-WF was rated safer and more comfortable compared to MB-TP. Regarding perceived safety one has to keep in mind that HA-WF doesnt allow for crashes at all which is most probably the reason for higher ratings. Regarding comfort it is assumed that people feel more comfortable when the responding system acts familiar. The human aware navigation includes human collision avoidance strategies in form of social costs [6] which might cause higher comfort ratings compared to merely technical collision avoidance. Nevertheless no difference in reliability was observed between HA-WF and MB-TP or MB-DWA. Nowadays people are used to interact with technical systems on a daily basis and we experience that those systems gain robustness with exponential velocity. It is thus not surprising that also systems following a more technical behavior are perceived as reliable. Finally, we calculated a score for the perceived value by summing up its components. Our results show that HA-WF resulted in the highest perceived value, again reflecting the previously described findings.

VII. CONCLUSIONS AND FUTURE WORK

A. Conclusions

To conclude we presented the design and the results of a human-robot experiment where we compared different navigation methods. We tested different combinations of global and local planners with and without human awareness. Taken together, HA-WF had a higher legibility compared to both approaches with TP local planners and a comparable legibility with the state-of-the-art non-human aware planner MB-DWA. Regarding the perceived value, HA-WF resulted

in higher perceived safety and increased feeling of comfort compared to HA-TP and MB-TP as well as a higher attribution of reliability compared to MB-TP. In summary we can therefore assume that a high legibility measured in correct predictions, met expectations and surprise leads to a high perceived value regarding perceived safety, felt comfort and attributed reliability during the interaction with a robot. Both HA-WF and MB-DWA allow for a high legibility. But if it comes to real world applications one should consider that only HA-WF does not allow collisions while navigating in the close surrounding of humans. Thus, minimizing the surprise rates has proven to be a reliable predictor, which led to the highest perceived value score. Furthermore with the results at hand we can state that the choice of the local planner is an important factor regarding the perceived value. The results showed that the TP local planner was achieving the worst results in comparison with the DWA and the WF local planners. The combination of the human aware global planner HA with the human aware global planner WF results in the best navigation method regarding the perceived value.

B. Future Work

This works shows us the way to more legible navigation methods, which we will explore further in the future. Furthermore we will carry out a similar study with a real robot in the future. We plan to measure explicit and implicit feedback as described in [2] with a focus on surprise detection during human action in order to verify our results in a real world environment.

VIII. ACKNOWLEDGMENTS

With the support of the Technische Universität München - Institute for Advanced Study and Cluster for Excellence Cognition for Technical Systems, funded by the German Excellence Initiative.

REFERENCES

[1] V. A. Zeithaml, "Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence," *The Journal of Marketing*, vol. 52, no. 3, 1988. [Online]. Available: http://dx.doi.org/10.2307/1251446

[2] "Towards a legibility metric: How to measure the perceived value of a robot," in *International Conference on Social Robotics, ICSR 2011, Work-In-Progress-Track*.

[3] F. Sadri, "Logic-based approaches to intention recognition," *Handbook of Research on Ambient Intelligence: Trends and Perspectives*, 2010. [Online]. Available: http://www.igi-global.com/viewtitlesample.aspx?id=54666

[4] L. Takayama, D. Dooley, and W. Ju, "Expressing thought: improving robot readability with animation principles," in *Proceedings of the 6th international conference on Human-robot interaction*, ser. HRI '11. New York, NY, USA: ACM, 2011, pp. 69–76. [Online]. Available: http://doi.acm.org/10.1145/1957656.1957674

[5] K. Dautenhahn, M. Walters, S. Woods, K. Koay, C. Nehaniv, A. Sisbot, R. Alami, and T. Siméon, "How may I serve you?: a robot companion approaching a seated person in a helping context," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, no. April 2005. ACM, 2006, pp. 172–179. [Online]. Available: http://dl.acm.org/citation.cfm?id=1121272

[6] A. Weiss, R. Bernhaupt, M. Tscheligi, D. Wollherr, K. Kuhnlenz, and M. Buss, "A methodological variation for acceptance evaluation of human-robot interaction in public places," in *RO-MAN 2008: Proccededings of the 17th IEEE International Symposium on Robot and Human Interactive Communication*, 2008.

[7] T. Kruse, A. Kirsch, E. A. Sisbot, and R. Alami, "Exploiting human cooperation in human-centered robot navigation," in *IEEE International Symposium in Robot and Human Interactive Communication (Ro-Man)*, 2010.

[8] E. Marder-Eppstein, E. Berger, T. Foote, B. Gerkey, and K. Konolige, "The office marathon: Robust navigation in an indoor office environment," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 300–307. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5509725

[9] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *Robotics & Automation Magazine, IEEE*, vol. 4, no. 1, pp. 23–33, 1997. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=580977

[10] B. Gerkey and K. Konolige, "Planning and control in unstructured terrain," in *ICRA Workshop on Path Planning on Costmaps*, 2008. [Online]. Available: http://pub1.willowgarage.com/apubdb_html/files_upload/8.pdf

[11] G. Echeverria, N. Lassabe, A. Degroote, and S. Lemaignan, "Modular Open Robots Simulation Engine: MORSE," in *Proceedings of the 2011 IEEE International Conference on Robotics and Automation*, 2011.

[12] S. Woods, M. Walters, K. L. Koay, and K. Dautenhahn, "Comparing human robot interaction scenarios using live and video based methods: Towards a novel methodological approach," in *Proceedings of The 9th International Workshop on Advanced Motion Control (AMC'06*, 2006, pp. 750–755.