

Incorporating Gaze into Social Navigation

Justin Hart, Reuth Mirsky, Xuesu Xiao, Peter Stone

Paper #8

Abstract—Most current approaches to social navigation focus on the trajectory and position of participants in the interaction. Our current work on the topic focuses on integrating gaze into social navigation, both to cue nearby pedestrians as to the intended trajectory of the robot and to enable the robot to read the intentions of nearby pedestrians. This paper documents a series of experiments in our laboratory investigating the role of gaze in social navigation.

I. INTRODUCTION

As mobile robots move into human-populated environments, such as homes, offices, and businesses, they must be able to negotiate the problem of navigating in spaces that they share with people. This development has given rise to research on the problem of social navigation [3, 15]. Among other goals, researchers in this area wish to improve the comfort and safety of people who must share space with robots, to make robots more interpretable to people as they navigate, and to enable robots to make progress on tasking where they may otherwise be impeded by nearby pedestrians blocking their path [2, 25]. It is also worth noting that the study of social navigation has not been limited to the domain of robotics. Significant research has been performed in virtual reality simulations or 3D-rendered game models [17, 24], and the Social Force Model — which has been leveraged in robotics — has its origins in multi-agent crowd simulations [9, 22].

A significant majority of the work on the task of social navigation in robots has focused on the position and trajectory of people with respect to the robot [4, 10, 20, 23]. A smaller collection of work has focused on cuing nearby pedestrians as to the intentions of the robot. Methods have included the addition of turn signals to the robot, as well as projection mapping arrows onto the floor in front of the robot; both indicating the robot’s intended trajectory [1, 6, 14, 19]. Research in our group has instead focused on leveraging gaze as a social cue, both to indicate the intended trajectory of the robot [8] and to interpret the intentions of nearby people [11]. This short paper discusses the evolution of our thinking on this problem based on the outcomes of several experiments in an ongoing series of studies that we are performing and concludes with a discussion of challenges we have identified. For a comprehensive review of approaches for handling interactions in the context of social navigation, we refer the reader to Mirsky et al. [16].

The first work in this series of experiments is by Fernandez et al. [6] who attached LEDs to the frame of a BWIBot robot [13], and used the LEDs in a fashion similar to a turn signal. In a test in which people pass robots heading head-on towards them in a hallway, LEDs were only successful in preventing a human and a robot from blocking each other’s paths when the

person has previously seen the turn signal being used by the robot (thus revealing the signal’s meaning). This result caused us to look to the use of gaze as a social cue to coordinate hallway-passing behavior, with the hypothesis that gaze will be easily interpreted correctly by people.

Gaze is an important indicator of where a person is about to move. Norman [18] speculated that bicycle riders avoid collisions with pedestrians by reading their gaze. Nummenmaa et al. [19] present a study in which a virtual agent (3D-rendered on a computer monitor) walks towards the study participant. The participant must choose whether to pass the agent on its left or right using keyboard commands, and the virtual agent indicates its intention by looking to its left or right. Unhelkar et al. [27] present a study in which head pose is used to determine which target a pedestrian will walk toward. Khambhaitia et al. [12] present a motion planner which coordinates the head motion of a robot to the path that the robot will take 4 seconds in the future, and asked participants in a video survey to determine the robot’s intended path as it approaches a T-intersection.

In our lab we have investigated the use of gaze in social navigation. Hart et al. [8] present a human study in which researchers acting as pedestrians in a busy hallway vary their gaze patterns to be either congruent with the direction that they intend to walk, counter to that direction, or absent (by looking down at a cell phone). The results demonstrate that pedestrians passing these researchers in a hallway are more likely to collide with them when their gaze is counter to the direction that they intend to walk in. In the same paper, Hart et al. [8] update the experimental setup from Fernandez et al. [6] to compare a gaze cue presented on a 3D-rendered virtual agent head mounted to the robot’s chassis to the use of LEDs; finding that the gaze cue is more effective than the LEDs in preventing the robot and participants in the study from blocking each other’s paths.

The first two of these robot studies surrounds the idea of the robot socially cuing its intentions to nearby pedestrians, but does not explore the idea of the robot’s behavior responding to social cues made by the people it interacts with [7, 26]. The most recent study on this topic by our group, performed by Holman et al. [11], approaches this problem from the perspective of enabling the robot to respond to human gaze. Participants are placed in a virtual environment, using wireless virtual reality equipment with an embedded eye tracker, and instructed to walk to one of five targets, in a similar fashion to the experimental design described in the work by Unhelkar et al. [27]. The study’s findings indicate that gaze can be used as an early cue indicating the target of a participant’s motion.



Fig. 1: The hallway on the left is the one used in the human-robot interaction studies by Fernandez et al. [6] and Hart et al. [8]. The one on the right is the one used in the human field study by Hart et al. [8].

We plan to leverage these results in future work to enable a robot to coordinate its motion to that of nearby pedestrians.

II. GAZE, NAVIGATION, AND HALLWAY PASSING

In this section we provide further details on our experiments on gaze and social navigation.

A. LEDs and Passive Demonstrations

Fernandez et al [6] present a study in which a robot navigates a hallway (Figure 1 (left)) and signals the side that it intends to pass a human participant using a strip of LEDs which act as a turn signal. The robot’s navigation algorithm treats the hallway as being divided into three traffic lanes through which it may navigate. Both the robot and the pedestrian start on the middle lane at opposite ends of a hallway. The robot signals that it is about to change its lane by blinking the LED light strip on the side of its chassis matching that of the direction of the lane that it intends to shift into. The LEDs are configured similarly to Figure 2 (left), which is adapted from [8]. It should be noted that there is an important difference between the appearance of the robot with LEDs in [6] and [8]. In Fernandez et al. [6], the robot has a monitor attached to its top in the LED condition, but with no face rendered on it, and mounted facing the back of the robot. This is a design choice on the BWIBot used to launch the robot’s software. In the Hart et al. [8] study, the monitor is removed because it was noted by the researchers that study participants would sometimes pause to observe the contents of the monitor, which is only the output of the ROS nodes driving the robot, and not intended as part of the interaction.

The robot’s navigation algorithm models passing a person in a hallway as a problem over three traffic lanes at three distances, as in Figure 3. If the person and the robot are both in the middle lane, then the robot has the option of

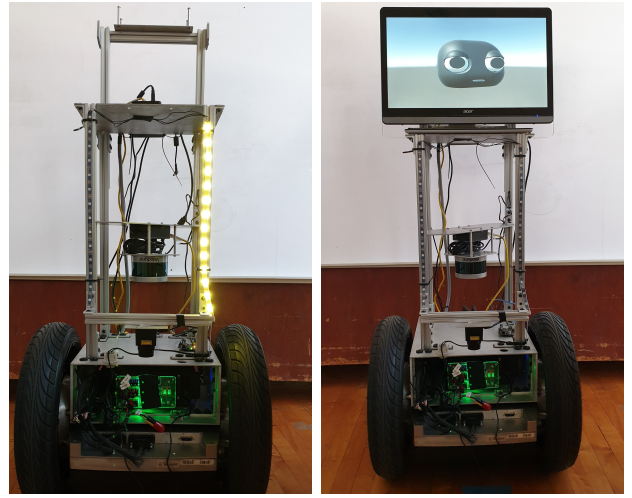


Fig. 2: The two conditions in our human-robot hallway experiment: the LED signal (left) and the gaze signal (right).

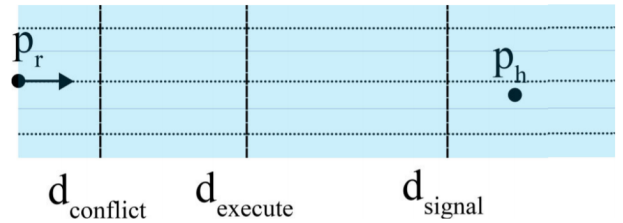


Fig. 3: A diagram of the hallway, its lanes, and the distance thresholds at which the robot signals its intention to change lanes ($d_{conflict}$), executes a lane change ($d_{execute}$), and is determined to potentially be in conflict with a person in its path (d_{signal}). The position of the robot is marked p_r and the position of the person is marked p_h .

passing the person on the left or the right, by shifting into the corresponding lane. The distances in this model are: d_{signal} , the distance at which the robot will begin to signal its intention to the pedestrian; $d_{execute}$, the distance at which the robot will begin to shift into the left or right lane; and $d_{conflict}$, at which the robot stops its motion and does not attempt to pass. If the robot and the person are in opposite lanes when they pass each other, the robot will not stop during the interaction.

In addition to these parameters, Fernandez et al. [6] introduced the concept of a “passive demonstration,” which is a sort of training episode in which the study participant is not informed that they are being trained, but wherein the robot demonstrates the signal by simply using it in front of the participant before it is relevant to their interaction. In this case, the robot moves into the right lane, using the turn signal, at the very start of crossing the hallway, then moves back to the middle prior to passing the participant. Upon coming within distance d_{signal} the robot again signals, now moving into the left lane, when passing the participant.

The study follows an inter-participant design, in which each participant traverses the hallway exactly once. The distances, d_{signal} , $d_{execute}$, and $d_{conflict}$ are set to 7 meters, 2.75 meters,

and 1 meter, respectively, and the robot always moves into the left lane when passing the pedestrian. These values are chosen based on pilot study data, indicating that pedestrians are likely pass on the right, and that $d_{execute}$ is at the last possible distance change lanes. This set of distances is chosen to assure that participants who successfully pass the robot do so based on the signal, rather than the robot’s motion. The study is set up as a 2X2 experiment where the controlled variables are whether or not the LED is used, and whether or not the robot performs a passive demonstration. The main measure is whether a participant and the robot experience a “conflict,” in which they come too close to each other. The results show that the passive demonstration condition with the LED turn signal significantly outperforms other conditions (no demonstration, no LED: 100% conflict; no demonstration, LED: 90% conflict; demonstration, no LED: 70% conflict; demonstration, LED: 20% conflict). A one-way ANOVA shows a significant main effect ($F(3, 36) = 9.913, p < 0.001$) and all pairwise post-hoc tests based on Least Squares Difference (LSD) contrasting against the “demonstration, LED” condition are significant at $p < 0.01$.

B. Gaze in Purely Human Navigation Environments

The results from the previous experiment encouraged us to search for a more naturalistic cue that people would be able to pick up on without having to observe passive demonstrations. Following previous work that looked at potential cues such as body rotation, trajectory estimation, and gaze [21, 27], we conducted a human study where we tested the viability of gaze as an intentional cue in purely human navigation [8].

In this study, a researcher navigates the hallway depicted in Figure 1 (right) and looks either in the direction in which they intend to go (*Congruent gaze*), opposite to this direction (*Incongruent gaze*), or at a mobile phone to deprive other pedestrians from leveraging their gaze (*No gaze*). The primary metric is whether the researcher comes into conflict with other pedestrians, defined as bumping into them, brushing against them, or quickly shifting to get out of each other’s way.

A total of 220 interactions were observed with 60 congruent gaze interactions, 85 incongruent gaze interactions, and 75 no gaze interactions. The mean percentage of conflicts by condition are: congruent gaze, 15%; incongruent gaze, 48%; and no gaze 28%. A one-way ANOVA shows a significant main effect ($F_{2,217} = 5.02, p = 0.007$). Post-hoc tests of pairwise mean differences using the Bonferroni criteria show significant differences between congruent gaze and the other two conditions (congruent vs. incongruent $md = 0.221, p = 0.017$; congruent vs. no gaze $md = 0.191, p = 0.033$).

These results highlight the importance of gaze as a naturalistic cue that assists people to process the navigational goal of other pedestrians around them, and adapt their own trajectory accordingly. This outcome has motivated our subsequent studies on how gaze can be leveraged both to convey the robot’s navigational goal and to infer the human navigational goal.

C. Conveying the Robot’s Navigational Intention

Returning to the hallway used in the passive demonstration experiment [6], we hypothesized that the use of gaze-like cue is more readily interpretable than the LED signal.

We designed a gaze cue using a 3D-rendered version of the Maki 3D-printable robot head.¹ The virtual head is displayed on a 21.5 inch monitor mounted to the front of the robot. When signaling, the robot turns its head 16.5° and remains in this pose, as shown in Figure 2 (right). The experimental design repeats the experimental setup from Fernandez et al. [6], contrasting the gaze signal against the LED signal, but omitting the test of passive demonstrations. Because of hardware changes, d_{signal} is reduced to 4m.

With 11 participants in the LED condition and 16 in the gaze condition, participants in the gaze condition successfully infer the robot’s goal 50% of the time, while none of the participants in the LED condition infer the goal of the robot. Important to note in the interpretation of the results from this experiment is that we expect a conflict 100% of the time unless the robot’s cue (either LED or gaze) are correctly interpreted. This is because the robot moving into the left-hand lane, which is against the convention in North America, is expected to result in conflict 100% of the time. Comparing the performance of the gaze signal against the LED signal demonstrates its ability to impact people’s navigational choices, and that people more easily interpret the gaze cue than the LED turn signal.

D. Inferring the Pedestrian’s Navigational Intention

The former set of experiments have demonstrated that a robot using a gaze-based head-turn cue can signal its intention to a person. The most recent experiment in our laboratory on the topic of social navigation is an attempt to make inroads on the inverse of that task — having the robot react to a person’s gaze in order to get out of the way — by making predictions of human walking motions based on gaze.

Holman et al [11] present a study, inspired by the experiment in Unhelkar et al. [27], in which participants wear a virtual reality headset with an embedded eye tracker and walk through a simulated room towards a goal that they are instructed to reach. For each trial, participants are first instructed to walk along a straight path towards position “A”, a target 1m directly in front of their starting position. Upon reaching position “A,” participants proceed to one of five goals placed 4m in front of the participant’s starting position, labeled 1–5, and placed 1m horizontally apart from each other. The purpose of navigating to position “A” before Goals 1–5 is to avoid conflating the effects of beginning to walk with the measured effects of the study.

A total of 7 participants (6 male, 1 female) ranging in age from 19–31 (mean 22.7) participated in this study. Study participation was limited to researchers working in the laboratory, as this study was conducted during the COVID-19 pandemic.

A multivariate Gaussian time series prediction algorithm was trained on subsets of the data collected during the pedes-

¹<https://www.hello-robo.com/>

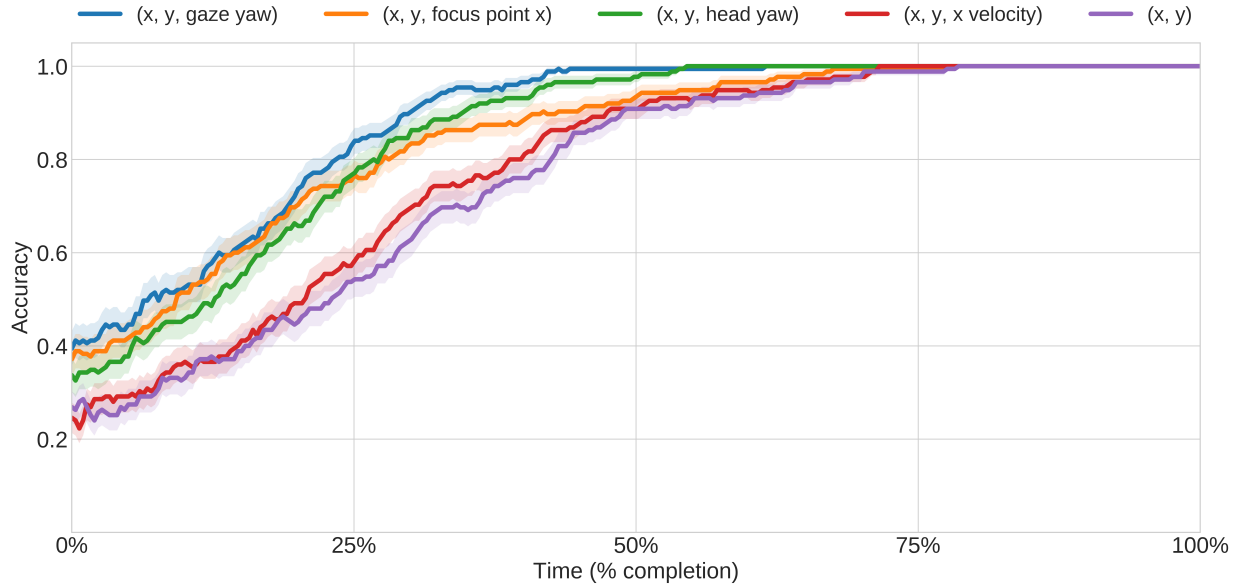


Fig. 4: Cross-validated accuracy of the multivariate Gaussian time series model over percent completion in time. Cross validation is computed with respect to a single participant over a model trained over all other participants, then computed as the mean when this procedure is repeated for all participants. The shaded region represents one standard deviation from the mean cross-validated accuracy.

trian’s journey, and is used to predict the final goal of their path. To extract meaningful results from the small amount of data collected, the accuracy of this model is tested via cross-validation. Results can be seen in Figure 4. The top line in Figure 4 indicates the performance of gaze yaw plus the position of the participant in indicating the participant’s final navigational goal, showing that this model predicts their motion goal earlier than all other tested cues. While this study is limited, in the fact that it only predicts motion toward a discrete goal, it represents an inroad towards reading gaze for social navigation.

III. CONCLUSION

This paper presents some of our efforts towards the design of a robot that can navigate in a social context while conveying its intention using a gaze-based social cue and reading the gaze of nearby pedestrians. We are currently in the process of designing a system which moves these cues beyond the confines of the hallway which we constructed for these experiments and into the real world. Here, we list a few facets of social navigation that we believe also bear further study.

- 1) **Up-Close Deconflicting Interactions:** While most human interactions when navigating in a crowd are seamless, there are still cases where there is a conflict and the joint navigation needs to be mediated. In such cases, pedestrians make eye contact or even use verbal communication to resolve the navigational conflict [19, 27, 8]. These behaviors that are either instinctive or socially learned by people, and will need to be incorporated into a social robot’s behavior.
- 2) **Context Understanding:** Navigation in a familiar place like one’s home will results with very different gaze

and motion patterns than navigation in an open mall or in a hospital. For example, an early sociological study showed that people tend to move in small groups rather than alone, but that the group size distribution highly depends on context [5]. In order to be able to leverage the gaze of pedestrians, the robot should be aware of the context of the interaction.

- 3) **Cultural Differences:** Different countries have different social norms when navigating in a crowd. It was common in our human study (see Hart et al. [8] for details) for people to shift to the right in order to avoid other pedestrians. However in other locations, people might shift to the left or to the direction that they are already more oriented towards.

As we continue to develop systems for social navigation, we expect to be able to handle a richer set of features handling a wider variety of situations and conforming to factors such as external context and cultural norms.

ACKNOWLEDGMENTS

This work has taken place in the Learning Agents Research Group (LARG) at UT Austin. LARG research is supported in part by NSF (CPS-1739964, IIS-1724157, NRI-1925082), ONR (N00014-18-2243), FLI (RFP2-000), ARO (W911NF-19-2-0333), DARPA, Lockheed Martin, GM, and Bosch. Peter Stone serves as the Executive Director of Sony AI America and receives financial compensation for this work. The terms of this arrangement have been reviewed and approved by the University of Texas at Austin in accordance with its policy on objectivity in research.

REFERENCES

- [1] Kim Baraka and Manuela M. Veloso. Mobile service robot state revealing through expressive lights: Formalism, design, and evaluation. *International Journal of Social Robotics*, 10(1):65–92, Jan 2018.
- [2] Aniket Bera, Tanmay Randhavane, Rohan Prinja, and Dinesh Manocha. Sociosense: Robot navigation amongst pedestrians with social and psychological constraints. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7018–7025, Vancouver, BC, Canada, September 2017.
- [3] Konstantinos Charalampous, Ioannis Kostavelis, and Antonios Gasteratos. Recent trends in social aware robot navigation: A survey. *Robotics and Autonomous Systems*, 93:85–104, April 2017.
- [4] Yu Fan Chen, Michael Everett, Miao Liu, and Jonathan P How. Socially aware motion planning with deep reinforcement learning. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1343–1350, Vancouver, BC, Canada, September 2017.
- [5] James S Coleman and John James. The equilibrium size distribution of freely-forming groups. *Sociometry*, 24(1): 36–45, March 1961.
- [6] Rolando Fernandez, Nathan John, Sean Kirmani, Justin Hart, Jivko Sinapov, and Peter Stone. Passive demonstrations of light-based robot signals for improved human interpretability. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 234–239, Nanjing, China, August 2018.
- [7] Rachel Gockley, Jodi Forlizzi, and Reid Simmons. Natural person-following behavior for social robots. In *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction (HRI)*, pages 17–24, Arlington, VA, USA, March 2007.
- [8] Justin Hart, Reuth Mirsky, Xuesu Xiao, Stone Tejeda, Bonny Mahajan, Jamin Goo, Kathryn Baldauf, Sydney Owen, and Peter Stone. Using human-inspired signals to disambiguate navigational intentions. In *Proceedings of the International Conference on Social Robotics (ICSR)*, pages 320–331, Golden, Colorado, USA, November 2020.
- [9] Dirk Helbing and Peter Molnar. Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282–4286, May 1995.
- [10] Peter Henry, Christian Vollmer, Brian Ferris, and Dieter Fox. Learning to navigate through crowded environments. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 981–986, Anchorage, Alaska, USA, May 2010.
- [11] Blake Holman, Abrar Anwar, Akash Singh, Mauricio Tec, Justin Hart, and Peter Stone. Watch where you’re going! gaze and head orientation as predictors for social robot navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 6183–6190. IEEE, 2021.
- [12] Harmish Khambhaita, Jorge Rios-Martinez, and Rachid Alami. Head-body motion coordination for human aware robot navigation. In *Proceedings of the International workshop on Human-Friendly Robotics (HFR 2016)*, page 8, Genoa, Italy, September 2016.
- [13] Piyush Khandelwal, Shiqi Zhang, Jivko Sinapov, Matteo Leonetti, Jesse Thomason, Fangkai Yang, Ilaria Gori, Maxwell Svetlik, Priyanka Khante, Vladimir Lifschitz, et al. BWIBots: A platform for bridging the gap between ai and human–robot interaction research. *The International Journal of Robotics Research (IJRR)*, 36 (5-7):635–659, February 2017.
- [14] Ryo Kitagawa, Yuyi Liu, and Takayuki Kanda. Human-inspired motion planning for omni-directional social robots. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 34–42, Boulder, CO, USA, 2021.
- [15] Thibault Kruse, Amit Kumar Pandey, Rachid Alami, and Alexandra Kirsch. Human-aware robot navigation: A survey. *Robotics and Autonomous Systems*, 61(12):1726–1743, December 2013.
- [16] Reuth Mirsky, Xuesu Xiao, Justin W. Hart, and Peter Stone. Prevention and resolution of conflicts in social navigation - a survey. *CoRR*, abs/2106.12113, 2021. URL <https://arxiv.org/abs/2106.12113>.
- [17] Soraia Raupp Musse and Daniel Thalmann. A model of human crowd behavior: Group inter-relationship and collision detection analysis. In *Computer Animation and Simulation: Proceedings of the Eurographics Workshop*, pages 39–51, Budapest, Hungary, September 1997.
- [18] Don Norman. *The design of future things*. Basic books, 2009.
- [19] Lauri Nummenmaa, Jukka Hyönä, and Jari K Hietanen. I’ll walk this way: Eyes reveal the direction of locomotion and make passersby look and go the other way. *Psychological Science*, 20(12):1454–1458, December 2009.
- [20] Billy Okal and Kai O Arras. Learning socially normative robot navigation behaviors with bayesian inverse reinforcement learning. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 2889–2895, Stockholm, Sweden, May 2016.
- [21] A. E. Patla, A. Adkin, and T. Ballard. Online steering: coordination and control of body center of mass, head and body reorientation. *Experimental Brain Research*, 129(4):629–634, Dec 1999.
- [22] Photchara Ratsamee, Yasushi Mae, Kenichi Ohara, Tomohito Takubo, and Tatsuo Arai. Modified social force model with face pose for human collision avoidance. In *Proceedings of the ACM/IEEE international conference on Human-Robot Interaction (HRI)*, pages 215–216, Boston, MA, USA, March 2012.
- [23] Emrah Akin Sisbot, Luis F Marin-Urias, Rachid Alami, and Thierry Simeon. A human aware mobile robot

motion planner. *IEEE Transactions on Robotics (T-RO)*, 23(5):874–883, October 2007.

- [24] Johannes Strassner and Marion Langer. Virtual humans with personalized perception and dynamic levels of knowledge. *Computer Animation and Virtual Worlds*, 16(3–4):331–342, September 2005.
- [25] Leila Takayama, Doug Dooley, and Wendy Ju. Expressing thought: improving robot readability with animation principles. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 69–76, Lausanne, Switzerland, March 2011.
- [26] Yusuke Tamura, Tomohiro Fukuzawa, and Hajime Asama. Smooth collision avoidance in human-robot coexisting environment. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3887–3892, Taipei, Taiwan, October 2010.
- [27] Vaibhav V Unhelkar, Claudia Pérez-D’Arpino, Leia Stirling, and Julie A Shah. Human-robot co-navigation using anticipatory indicators of human walking motion. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 6183–6190, Seattle, WA, USA, May 2015.