The Oz of Wizard: Simulating the Human for Interaction Research

Aaron Steinfeld
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213
412-268-6346
steinfeld@cmu.edu

Odest Chadwicke Jenkins
Dept. Of Computer Science
Brown University
Providence, RI, USA 02912-1910
401-863-7600
cjenkins@cs.brown.edu

Brian Scassellati
Yale University
New Haven, CT 06520
203-432-1219
scaz@cs.yale.edu

ABSTRACT
The Wizard of Oz experiment method has a long tradition of acceptance and use within the field of human-robot interaction. The community has traditionally downplayed the importance of interaction evaluations run with the inverse model: the human simulated to evaluate robot behavior, or “Oz of Wizard”. We argue that such studies play an important role in the field of human-robot interaction. We differentiate between methodologically rigorous human modeling and placeholder simulations using simplified human models. Guidelines are proposed for when Oz of Wizard results should be considered acceptable. This paper also describes a framework for describing the various permutations of Wizard and Oz states.

Categories and Subject Descriptors
H.5.2 [**INFORMATION INTERFACES AND PRESENTATION (e.g., HCI)**]: User Interfaces – *Evaluation/methodology, Theory and methods.*

General Terms
Measurement, Performance, Design, Experimentation, Human Factors, Theory.

Keywords
Wizard of Oz, human-robot interaction, evaluation, interaction.

1. INTRODUCTION
1.1 Position
The Wizard of Oz approach, where an experimenter secretly fills in for a piece of technology while a participant conducts a task [1], is a well established and accepted method for evaluating human behavior and performance when using a hypothetical technology or system capability. Technical publication of work utilizing this method does not trigger skepticism and doubt during peer review nor do questions rise regarding whether such work belongs within the domain of human-robot interaction (HRI). Using an inclusive and interdisciplinary model of HRI, we posit there is a place for cutting edge technology research that supports and enables further research on the human aspects of HRI. We envision synergistic feedback between these to forms of HRI research, where human studies evaluate the viability of technologies, both current and future, and enabling technologies research makes these ideas tangible while exploring new mechanisms. Therefore, we propose that an inverse methodology, the “Oz of Wizard”, is a valid approach to evaluating enabling technologies research that supports or enhances human-robot interaction.

Furthermore, we argue that the methodological rigor found within the field of human modeling should not be uniformly required when simulating human input for HRI research. Unlike human-computer interaction with deterministically controlled digital environments, in robotics there are often external and physical factors, namely uncertainty of various forms, which prohibit utilization of precise models. In certain cases, we feel that high quality research on technological advances in the domain of human-robot interaction do not always require precise human simulation.

1.2 History
Research using the Wizard of Oz technique is widespread within human-robot interaction. For example, several papers at recent HRI conferences have utilized the method (e.g., [2-4]). An interesting nuance is that use with robots has led to a largely unnoticed modification to the original concept.

Human-computer interaction and experimental psychology studies using Wizard of Oz have largely focused on just technology function, regardless of the environment. However, the field of HRI has extended the methodology due to the nature of robotics. Within HRI this method also includes the influence of the environment. An experimenter may drive a robot through a cluttered scene, thereby simulating basic mobility, path planning, and perception. The subsequent behavior will not be the same as if the robot were moving through a sparsely populated room. Robot behavior is simulated with respect to how it interacts with the environment.

This is an important distinction. In the past, the environment only influenced the participant, not the technology. In robotics, the environment can effect both the robot and the human. In fact, it is quite realistic to expect interactions to occur where the influence of the environment does not independently effect the robot and the human. However, HRI Wizard of Oz experiments inherently assume that the environment affects the robot and human independently.

HRI'09, March 11–13, 2009, La Jolla, California, USA.

1.3 Need

Quite often, the prime motivation for not wanting to bring in actual human participants is directly related to time and logistics. Many groups are explicitly interested in advancing the science of algorithms, embodiments, and mechanisms needed for human-robot interaction. Maintaining a rapid pace of exploration and/or development is not always possible if each step is expected by their peers to be thoroughly tested with human participants. Inexperience with human experiments amplifies logistical barriers. Likewise, human participation can be expensive – especially for interactions that may consume long periods of time and/or have more than minimal risk to the participant.

The obvious argument is to use a computerized human model to simulate human input. This is an accepted practice (e.g., [5]) and is based on previously vetted experimental research on human interaction, cognition, and perception. However, human models of all types have limitations that can prevent a human-robot interaction researcher from contributing to the field within a reasonable period of time. These barriers include, but are not limited to: (a) access to models that are proprietary or sparsely published, (b) expertise with specialized modeling approaches, (c) expensive specialized software, and (d) training on the fundamental science behind the models.

A good example is research on the repeatability and reliability of a robot component. If a specialist in mechanical hands wanted to quantify how robust a robot hand gripping algorithm was, it would be cost prohibitive to use a wide array of human participants to shake hands with a robot thousands of times. They may also lack access to an adequate database describing the myriad of human hand motions for each of the millions of desired permutations of size, shape, and motion. The experimenter is caught in a no-win situation. They cannot bring in the quantity and variety of human participants expected by the field and they cannot utilize a precise digital human model. Their use of a simplified model for human hand sizes, shapes, and motions would be considered a weakness during peer review.

Likewise, safety and equipment limitations can be barriers to human involvement. The experimenter may lack a safe and appropriate manifestation of a robot hand. This may be due to only having access to a hand with a reduced number of fingers or a hand that could easily break bones. Institutional Review Board requirements can exacerbate this situation dramatically. There is a real risk of harm and the latency due to review can take two to three times longer than some technology development cycles. Again, inexperience with human experiments can lead to even longer delays.

1.4 Why the Inverse Matters

In all of these cases, the researcher is conducting human-robot interaction research and should be considered on comparable footing to those focused on human behavior. Human-robot interaction is not just human behavior when exposed to robots; the behavior of the robot when exposed to a human, even a simulated human, is also a valid topic within HRI. By using simplified models of human behavior, researchers can test variability and/or feasibility in technologies that produce and enable robot behavior.

It is important to note that our argument is *not* that it is acceptable to use simplified human models in all cases. We are merely stating that it is reasonable to take such an approach for cutting edge research on human-robot interaction technology when certain barriers are present. We propose the following guidelines for when to use simplistic human models:

- The risk to human participants is high, or
- Utilizing human participants is logistically infeasible

combined with:

- Access/availability of precise human models is poor, if at all

We also endorse use of simplified human models during early iterations in advance of experiments or more precise human modeling. However, we recommend only reporting such preliminary research in publication to demonstrate how subsequent algorithms and robots are worthy of more precise human modeling and/or experiments with human participants.

2. FRAMEWORK

We work from the notion that humans and robots are components within a human-robot system [6, 7]. In such systems, the behavior of human(s) and robots(s) are coupled together and receive feedback through some system dynamics, typically a physical environment. As a function, the overall behavior of the system is caused by the behavior of the robots, humans, and the influences of their environment:

System behavior = f(robots, humans, environment) (1)

with the shorthand: b = system behavior, r = robot(s), h = human(s), and e = the environment. The behavior of each component within the environment can then be expressed as:

wizard = robot, as influenced by the environment

oz = human, as influenced by the environment

As mentioned earlier, HRI Wizard of Oz experiments assume that the environment effects the robot and human independently. This allows subsequent expression of simulating either the Wizard or the Oz as:

Wizard of Oz: $b = f(r, h, e) \approx f(h, e) \approx w(o)$ (2)

Oz of Wizard: $b = f(r, h,e) \approx f(r, e) \approx o(w)$ (3)

It is helpful to think of these as two sides of the same coin. Wizard of Oz controls robot behavior, making it a dependent variable of human behavior. Such studies focus on human behavior (as an independent variable) through the function of overall system behavior given exposure to robot behavior (as a dependent variable). In contrast, Oz of Wizard does the inverse where human behavior is controlled in some manner to focus on robot behavior as the independent variable. In other words, this case is the study of robot behavior as a function of overall system behavior given exposure to simplified human behavior.

This, of course, also permits exploration of a variety of combinations besides just the Wizard of Oz and the Oz of Wizard. Figure 1 summarizes these in the context of how close to reality the Wizard and the Oz are within the evaluation. To some degree, these combinations could also be considered a starting point for defining the types of research within the HRI community.

2.1 Wizard of Oz

This is the traditional model where robot behavior is simulated, usually by an experimenter [1]. As previously stated, this approach is widely accepted by the field [2-4]. The evaluation captures the

Figure 1. Wizard/Oz Combinations

influence of human behavior in the environment but does not measure actual robot behavior, as influenced by the environment. Realistic system behavior can be estimated but may not be realized until years of technology advances have occurred.

2.2 Oz of Wizard

Like its predecessor, such experiments can estimate realistic system interaction. However, the risk of error is directly tied to the simulation of the human. Human involvement is simulated through detailed models or through controlled approximations, depending on what barriers and resources are present. The latter may be sufficient for gross estimation of system behavior. This, in turn, can inform the next technology research iteration.

Oz of Wizard primarily applies to enabling technologies for HRI research. This category includes, but is not limited to, vision and learning algorithms, robot platforms with novel integration, and architectures for cognitive modeling. Work in this area often comes from or has a strong intersection with other technology-focused research areas, such as computer vision, machine learning, and artificial intelligence as well as broader robotics research. Evaluation for such enabling technologies must demonstrate feasibility for advancing human-robot interaction in terms of both the validity of the proposed technology and its suitability to enable new or better modalities for interaction.

Technological validity can often come from metrics used by the intersecting research community. For example, such metrics can be based on Receiver Operating Characteristic Curves [8, 9] for recognition in vision or speech interfaces, mean squared error from ground truth for prediction [10] and classification [11] learning from demonstration, and properties from autonomously learned POMDP models [12, 13]. However, satisfying technological metrics alone does not constitute an HRI contribution. For example, precise human pose tracking from video [14] can be done in a manner that, in theory, supports human-robot interaction, but is computationally too expensive and time consuming for applicability in the near future. Thus, it is critical for proposed Oz of Wizard papers describe a path to feasibility where fundamental assumptions and limitations are clearly stated and can be overcome in leading to systems suitable for use in experimental HRI.

In the area of cognitive modeling, work by Trafton, et al [5] is an excellent example of Oz of Wizard with moderately precise human modeling. This work used ACT-R to emulate the through processes of a young child for learning the game of Hide and Seek. While only one child was observed, not enough for a valid population sample, the level of modeling is not simplistic. The resulting HRI evaluation focuses on the success of the robot to interact with the human during a game. The work is clearly focused on HRI technology advancement.

2.3 Oz with Wizard

When human participants are introduced into an Oz of Wizard evaluation, but not measured precisely, the combination can be called Oz with Wizard. We use "with" to express that the measurement of Oz is not precise and or not measured at all. Oz merely accompanies the Wizard[1]. Examples of evaluations in this model are measurements of robot reliability and performance during actual use. Such evaluations have validity on robot behavior in a realistic environment but lack clarity on Oz.

Srinivasa, et al [15] describe a robotic home assistant which was demonstrated around humans during a number of events, including a large lab open house. Robot behavior was clearly affected by the presence of humans but the evaluation is strictly focused on robot-centric metrics. For example, the authors report a failure rate of roughly 20 out of 220 when the robot attempted to move a mug to the dishwasher.

2.4 Wizard with Oz

Many studies in the field of HRI fall into the next category, Wizard with Oz. In this case, measurement of robot behavior, as influenced by a realistic environment, is neglected in favor of close

[1] We elected to use "with" after some debate. We wish to emphasize the notion that this term implies being present, but on a lesser footing. I.e., "the dog was with its owner".

measurement of the Oz. Examples include, but are not limited to, quantitative laboratory studies with real robots and context assessments preceding robot deployment.

In a study by Humphrey and Adams [16], twenty-four participants were recruited to test various compass visualizations for remote operation of a mobile robot. The authors used a simulated robot and environment but did not control the robot behind the scenes. The evaluation largely measured the human component of the system through metrics including preference, situation awareness, and workload. System behavior was examined with task performance measures.

Yamaoka, et al [17] explored the question of how close should a robot get to a user during social interaction. This study exposed participants to a real robot during a simulated retail sales event and obtained ratings on user comfort and robot likeability. In this case, human behavior and the whole system is being measured but there is still a large portion of Wizard being simulated due to the environmental constraints placed on the experiment. Specifically, the study was limited to a 3x3 m area with pristine conditions (e.g., no additional customers, retail noise, clutter, etc).

Likewise, this category includes research where real robot behavior is tested in a simulated environment. For example, Hoffman and Breazeal's [18] work on anticipation algorithms collected data on human perception of robot behavior from 32 human participants. The experiment was run using a simulated robot in a simulated environment but the robot behaviors were real.

Context assessments preceding robot development and deployment are included in this category since the researcher is extensively measuring not just the human's expectations and perception of robot involvement but also the environment and tasks that will directly affect robot behavior. Examples of this include, but are not limited to, ethnographic studies (e.g., [19]) and surveys of human expectations (e.g., [20]).

2.5 Wizard and Oz

When both the Wizard and the Oz in an evaluation are real and tested in the envisioned environment, the researcher has full representation of both Wizard and Oz. This is the preferred method for evaluating human-robot interaction and is manifested as a real-world experiment where increasing levels of environment realism leads to a greater distance from the origin. The whole system is being influenced by the actual environment and no simulated behaviors are required. The assumption that the environment effects the robot and human independently can also be relaxed and system behavior can be directly measured, rather than estimated.

The work by Scholtz, et al [21] is a good example of Wizard and Oz. This study involved eleven teams competing in an urban search and rescue (USAR) competition. Robots were deployed in a physical environment explicitly designed to emulate challenges typically encountered by robots within an USAR setting. While this was not a real-life field test it does capture a great deal of environmental realism. The competitive nature of the event also raises the human's stress levels above a typical laboratory experiment. The authors also collected data on the robot, human, and system, thus leading to evaluation of both Wizard and Oz.

Field operational tests are the ideal. Such evaluations are admittedly resource intensive but they permit a strong feedback loop between technology development and evaluation. Successful execution of a field test requires mature technology that works for the user. Besides issues of abandonment, reliability, and acceptance, technology maturation is also driven by the noise present in real-life tasks and environments. This was especially evident in Casper and Murphy's [22] assessment of HRI during a live USAR deployment. The seventeen findings from post-hoc analysis detail a broad collection of issues related to HRI, ranging from robot sensors to operator fatigue, group interaction, acceptance, and the impact of the environment.

As with before, this category is not limited to quantitative evaluations. Mutlu and Forlizzi [23] conducted an ethnographic analysis of a service robot deployment within a hospital intermittently over 15 months. This research captured details about HRI for an actual robot product within a real environment and workflow. The result is a comprehensive assessment of the system as a whole.

2.6 Wizard nor Oz

This is the case where all aspects of the system are simulated (neither Wizard nor Oz are real). This is the least desirable approach in that there is limited basis in reality for all of the components. Good work can still be accomplished in this category but the onus on authors is heavier. Assessment of scientific advancement can be more challenging if human, robot, and environment models are not precise or grounded in empirical data already collected in other experiments.

While intelligent transportation is not traditionally viewed as HRI, we point to the work done by Krishnan, et al [24] on a simulated rear-end collision-warning system as an example of good work in this category. The team compiled a model using inputs drawn from a wide range of data to provide a highly realistic prediction of system performance along various design parameters. Inputs included data drawn from literature on human response time, braking rates, traffic mix, and vehicle mass. The team even acquired traffic speed data from a local municipality.

3. SELECTING A TECHNIQUE

In arguing for the acceptance of this range of methodologies, it is important to discuss the selection process in designing an appropriate experiment. In some cases, choices will be dictated by the availability of either technology or appropriate safety and feasibility constraints (as discussed in Section 1.3). However, in many other cases, multiple methodological approaches will be feasible. In these cases, researchers today often make these decisions based on expediency and cost. We argue that the choice of a methodology should rather be guided by an informed decision that weighs the costs of a study with the potential applicability of the proposed research. We motivate this discussion with examples drawn from our own work, as the decision process behind methodological choices can only be assumed from most published work in HRI.

3.1 A Clear Example of Wizard of Oz

In recent work, Bainbridge, et al [25] were interested in understanding the impact that embodiment has on how a user will respond to potentially difficult or unusual requests from a robot. Their application domain was socially assistive robotics [26], in which, for example, a robot might encourage a stroke victim to perform a series of difficult rehabilitation exercises. The focus here was on the human user's response, not on the development of unique technology or an autonomous capability of the robot, and on typical adults without disability, thus providing a baseline for future work.

To study the effect of embodiment, they placed subjects in an office environment and asked them to follow the directions of a robot on where to move piles of books. In some cases, the books were moved to a shelf (a typical task) while in others the robot indicated that expensive textbooks should be thrown into the trash (an unusual task). The robot would either be present in the room or displayed live on video feed on a large flat-panel display. A great deal of effort was expended in providing appropriate controls to mitigate the differences between a 2-D projection and a 3-D figure, between a system that made noise in the room and a remote system that broadcast audio, and the interactivity that might be present on either system.

After careful thought, a Wizard of Oz methodology was selected. This allowed for precise control of the interactivity and reliability of the robot (either physical or virtual) while maintaining a focus on human responses. This methodology was costly; more than 60 subjects were recruited, and data recording, coding, and analysis required more than three months. In the end, the study provided evidence that humans were more willing to perform the unusual task with a real robot than with a virtual character. Although the robot design and the task design do not match the application domain, the results still offer a strong argument as to why robots might be more valuable in assistive technology than virtual characters (even characters with the fidelity of a live broadcast).

3.2 Strictly Oz of Wizard

In work by Jenkins et al. [9, 27, 28], the objective was to develop state estimation systems, from monocular vision, to enable socially interactive robots to perceive non-verbal human cues, namely human pose and gestures. By enabling perception of non-verbal cues, humans could interact with robots more of a peer-to-peer manner, where less direct control of the robot is necessary. While the purpose of this work was to support HRI, this effort was clearly Oz of Wizard in using simplified models of human subjects, such as limitations on the type of movements that could be performed and the user's rough body proportions. These assumptions were made to demonstrate the feasibility of the state estimation systems through an experimental prototype. Further, this work emphasized aspects that would allow the proposed methods to run fast enough for online and onboard computation, and thus within plausibility for application to HRI in the foreseeable future.

The Oz of Wizard approach has been used to make additional advances on this enabling technology. Recent work indicates that sensing, not necessarily algorithms for computational perception, play more of a role in estimating non-verbal cues [28, 29]. This work replaced the monocular color camera with an infrared-based depth camera. As a result, a system capable of interactively following a person and recognizing gestures was created that worked with common "off-the-shelf" algorithms for perception, such as Support Vector Machines for person detection and Hidden Markov Models for gesture recognition.

3.3 Mixing Wizard and Oz

We seek in this section to provide an example that both mixes some of the combinations of Wizard and Oz components and one which changes over time as an experiment becomes more mature.

In a series of studies, Gold and Scassellati [30-32] developed a computational model of language acquisition that allowed a robot to learn the meaning of pronouns (such as "I", "you", "here", and "there") by listening to conversations between competent adult speakers. The goal of this work was both to provide an algorithm for acquiring these words from real-world discourse and to provide a possible model of how human children perform this word-learning task.

The first published work in this line of research [30] focused on the technology development for an algorithm that could learn that "I" referred to the speaker and "you" referred to the addressee when other fixed words in the vocabulary (e.g., "ball") were known and could be correctly identified within a scene. While the primary effort was on the technology development, it was important to establish that the algorithm could be successful on the kind of utterances that people generate. Because only a proof-of-concept was required, the authors elected to use an often-used database of parent-to-child utterenaces [34] as the input to their algorithm. From the transcripts of these conversations, the authors manually provided the system with information about the environment that matched what they could be inferred from the conversation (including, for example, who had the ball). In this respect, although the human utterances were taken from actual parent-child conversations, the system used only a simplification of real human-human conversations and with a limited environment. In the classification presented here, this study used an Oz with Wizard method, as the technology was novel but the human interaction simulated.

As the technology was refined, [31, 33, 35] the nature of the experiments shifted away from studies with pre-canned human data and toward experiments that allowed the robot to learn directly from overheard conversations in the real world. For example, in [33] the robot was able to learn the meanings of the words "I", "you", "am" and "are" from listening to an exchange between two students playing catch in front of the robot. Most interesting from this study was that the robot was able to learn these words successfully only when it overheard conversations between two other people, not from conversations involving only one user speaking directly to the robot. (When speaking only to the robot, it is difficult to determine what the word "you" really means, as it always refers to the robot!) This finding matches results from developmental psychology in which second-born children learn pronouns more quickly than first-born children, presumably because there are more conversations for the second child to overhear (analysis of this is presented in [35]). While the sampling employed only a few pairs of subjects recorded for only a few minutes at a time, the results demonstrated novel technology deployment on real user interactions in the real world, thus qualifying it as Wizard and Oz. The final results from this study were both a technological advance in machine perception and a greater understanding of the nature of human input that allows for language learning.

4. DISCUSSION

As stated above, the acceptance of the Wizard of Oz model within the HRI community suggests that the inverse model, Oz of Wizard, should also be viewed as an appropriate HRI methodology. The simulation of human behavior, as influenced by the environment, is a powerful approach for advancing research on the technology side of HRI and should not be downplayed by the community.

We acknowledge there is risk to endorsing Oz of Wizard and Wizard nor Oz methodologies with simplified human models. Solid and constructive peer review combined with evaluation of the work in the spirit of the proposed guidelines can foster both a successful model for reporting results and greater understanding and mutual benefit within the entire field of HRI research. Such understanding is critical in establishing a common foundation to move the HRI community forward. The risk engendered by the assumption that human behavior follows some simplified behavioral model is no greater than the risk engendered by the assumption that future technology development can be accurately predicted.

We suggest that researchers using simplified human models self-assess their work before submitting such work for publication. If it is hard to justify such models using the proposed guidelines, then we strongly recommend incorporation of more Oz into the work prior to submission. This can be through: (a) obtaining better human models, (b) utilization of an Oz with Wizard approach, or (c) moving fully into Wizard and Oz.

Through definition of Wizard of Oz and Oz of Wizard methodologies, we also see a manner in which to assess the quality of other types of HRI research. The fundamental aspect of this framework is the influence of the environment and the researcher's use of simulation through software, physical setting, and/or task. This is a logical factor to take into account when categorizing HRI research since robotics is "in the world".

In closing, the complex nature of robotics is what makes HRI different and exciting when compared to human-computer interaction and related fields. This added complexity comes from the inherently broader set of disciplines required for successful deployment of HRI in the real world. It is important for the HRI research community to accept this interdisciplinary nature as a valued asset rather than a weakness. As such, we must be open and accepting of quality research done on the interaction technology side of the human-robot system.

5. ACKNOWLEDGMENTS

Parts of this work are based on work supported by the National Science Foundation under Grants No. CBET-0742350, SES-0835767 and EEEC-540865.

6. REFERENCES

[1] Kelley, J. F. 1984. An iterative design methodology for user-friendly natural language office information applications. *ACM Trans. Inf. Syst.* 2, 1 (Jan. 1984), 26-41. DOI= http://doi.acm.org/10.1145/357417.357420

[2] Shiomi, M., Kanda, T., Koizumi, S., Ishiguro, H., and Hagita, N. 2007. Group attention control for communication robots with wizard of OZ approach. In *Proceedings of the ACM/IEEE international Conference on Human-Robot interaction* (Arlington, Virginia, USA, March 10 - 12, 2007). HRI '07. ACM, New York, NY, 121-128. DOI= http://doi.acm.org/10.1145/1228716.1228733

[3] Kahn, P. H., Freier, N. G., Kanda, T., Ishiguro, H., Ruckert, J. H., Severson, R. L., and Kane, S. K. 2008. Design patterns for sociality in human-robot interaction. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (Amsterdam, The Netherlands, March 12 - 15, 2008). HRI '08. ACM, New York, NY, 97-104. DOI= http://doi.acm.org/10.1145/1349822.1349836

[4] Shiomi, M., Sakamoto, D., Takayuki, K., Ishi, C. T., Ishiguro, H., and Hagita, N. 2008. A semi-autonomous communication robot: a field trial at a train station. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (Amsterdam, The Netherlands, March 12 - 15, 2008). HRI '08. ACM, New York, NY, 303-310. DOI= http://doi.acm.org/10.1145/1349822.1349862

[5] Trafton, J. G., Schultz, A. C., Perznowski, D., Bugajska, M. D., Adams, W., Cassimatis, N. L., and Brock, D. P. 2006. Children and robots learning to play hide and seek. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction* (Salt Lake City, Utah, USA, March 02 - 03, 2006). HRI '06. ACM, New York, NY, 242-249. DOI= http://doi.acm.org/10.1145/1121241.1121283

[6] Granda, T., Kirkpatrick, M., Julien, T., and Peterson, L. 1990. The evolutionary role of humans in the human-robot system. In *Proc. Human Factors Society 34th Annual Meeting*. 664--668.

[7] Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., and Goodrich, M. 2006. Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction* (Salt Lake City, Utah, USA, March 02 - 03, 2006). HRI '06. ACM, New York, NY, 33-40. DOI= http://doi.acm.org/10.1145/1121241.1121249

[8] Forsyth, D. and Ponce, J. 2003. *Computer Vision A Modern Approach*, Prentice Hall.

[9] Jenkins, O., González, G., and Loper, M. 2007. Interactive human pose and action recognition using dynamical motion primitives. *International Journal of Humanoid Robotics*, 4(2):365-385.

[10] Grollman, D. H., and Jenkins, O. C. 2008. Sparse incremental learning for interactive robot control policy estimation. In *International Conference on Robotics and Automation* (ICRA, Pasadena, CA, USA, May 2008).

[11] Chernova, S. and Veloso, M. 2008. Multi-thresholded approach to demonstration selection for interactive robot learning. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (Amsterdam, The Netherlands, March 12 - 15, 2008). HRI '08. ACM, New York, NY, 225-232. DOI= http://doi.acm.org/10.1145/1349822.1349852

[12] Doshi, F. and Roy, N. 2007. Efficient model learning for dialog management. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction* (Arlington, Virginia, USA, March 10 - 12, 2007). HRI '07. ACM, New York, NY, 65-72. DOI= http://doi.acm.org/10.1145/1228716.1228726

[13] Schmidt-Rohr, S. R., Knoop, S., Lösch, M., and Dillmann, R. 2008. Reasoning for a multi-modal service robot considering uncertainty in human-robot interaction. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (Amsterdam, The Netherlands, March 12 - 15, 2008). HRI '08. ACM, New York, NY, 249-254. DOI= http://doi.acm.org/10.1145/1349822.1349855

[14] Vondrak, M., Sigal, L., and Jenkins, O. 2008. Physical simulation for probabilistic motion tracking. In *Computer Vision and Pattern Recognition* (CVPR, Anchorage, AK, USA, Jun 2008).

[15] Srinivasa, S., Ferguson, D., Vande Weghe, M., Diankov, R., Berenson, D., Helfrich, C. and Strasdat, H. 2008. The Robotic Busboy: steps towards developing a mobile robotic home assistant. In *International Conference on Intelligent Autonomous Systems*.

[16] Humphrey, C. M. and Adams, J. A. 2008. Compass visualizations for human-robotic interaction. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (Amsterdam, The Netherlands, March 12 - 15, 2008). HRI '08. ACM, New York, NY, 49-56. DOI= http://doi.acm.org/10.1145/1349822.1349830

[17] Yamaoka, F., Kanda, T., Ishiguro, H., and Hagita, N. 2008. How close?: model of proximity control for information-presenting robots. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (Amsterdam, The Netherlands, March 12 - 15, 2008). HRI '08. ACM, New York, NY, 137-144. DOI= http://doi.acm.org/10.1145/1349822.1349841

[18] Hoffman, G. and Breazeal, C. 2007. Effects of anticipatory action on human-robot teamwork efficiency, fluency, and perception of team. In *Proc. of the ACM/IEEE International Conference on Human-Robot Interaction* (Arlington, Virginia, USA, March 10 - 12, 2007). ACM, New York, NY, 1-8. DOI= http://doi.acm.org/10.1145/1228716.1228718

[19] Gemperle, F., DiSalvo, C., Forlizzi, J., and Yonkers, W. 2003. The Hug: a new form for communication. In *Proceedings of the 2003 Conference on Designing For User Experiences* (San Francisco, California, June 06 - 07, 2003). DUX '03. ACM, New York, NY, 1-4. DOI= http://doi.acm.org/10.1145/997078.997103

[20] Takayama, L., Ju, W., and Nass, C. 2008. Beyond dirty, dangerous and dull: what everyday people think robots should do. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (Amsterdam, The Netherlands, March 12 - 15, 2008). HRI '08. ACM, New York, NY, 25-32. DOI= http://doi.acm.org/10.1145/1349822.1349827

[21] Scholtz, J., Young, J., Drury, J., and Yanco, H. 2004. Evaluation of human-robot interaction awareness in search and rescue. In *Proc. IEEE International Conference on Robotics and Automation*.

[22] Casper, J., and Murphy, R. 2003. Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center. *IEEE Transactions on Systems, Man and Cybernetics B*, 33.

[23] Mutlu, B. and Forlizzi, J. 2008. Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (Amsterdam, The Netherlands, March 12 - 15, 2008). HRI '08. ACM, New York, NY, 287-294. DOI= http://doi.acm.org/10.1145/1349822.1349860

[24] Krishnan, H., Gibb, S., Steinfeld, A., and Shladover, S. E. 2001. Rear-end collision-warning system: Design and evaluation via simulation, *Transportation Research Record - Journal of the Transportation Research Board* 1759, 52-60.

[25] Bainbridge, W., Hart, J., Kim, E. and Scassellati, B. 2008. The effect of presence on human-robot interaction. In *IEEE International Symposium on Robot and Human Interactive Communication* (Munich, Germany).

[26] Tapus, A., Mataric, M. and Scassellati, B. 2007. The grand challenges in socially assistive robotics. *IEEE Robotics and Automation Magazine.* Vol. 4, No. 1. p. 35-427.

[27] Jenkins, O. C., González, G., and Loper, M. M. 2007. Tracking human motion and actions for interactive robots. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction* (Arlington, Virginia, USA, March 10 - 12, 2007). HRI '07. ACM, New York, NY, 365-372. DOI= http://doi.acm.org/10.1145/1228716.1228765

[28] Koenig, N., Chernova, S., Jones, C., Loper, M., and Jenkins, O. 2008. Hands-free interaction for human-robot teams. In ICRA 2008 *Workshop on Social Interaction with Intelligent Indoor Robots* (Pasadena, CA, USA, May 2008).

[29] Koenig, N., Chernova, S., Jones, C., Loper, M., and Jenkins, O. 2008. Hands-free human-robot interaction. In *HRI Caught on Film 2, Proc. of the 3rd ACM/IEEE International Conference on Human Robot Interaction*. ACM, New York, NY, 383-388.

[30] Gold, K. and Scassellati, B. 2006. Using context and sensory data to learn first and second person pronouns. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction* (Salt Lake City, Utah, USA, March 02 - 03, 2006). HRI '06. ACM, New York, NY, 110-117. DOI= http://doi.acm.org/10.1145/1121241.1121262

[31] Gold K. and Scassellati, B. 2006. Grounded pronoun learning and pronoun reversal. In *5th International Conference on Development and Learning* (ICDL-06, Bloomington, IN).

[32] Gold, K., Doniec, M., and Scassellati, B. 2007. Learning grounded semantics with word trees: prepositions and pronouns. In *Proceedings of the 6th IEEE International Conference on Development and Learning* (ICDL, London, England, July 2007).

[33] Gold, K. and Scassellati, B. 2007. A robot that uses existing vocabulary to infer non-visual word meanings from observation. In *Proceedings of the Twenty-Second Annual Meeting of the Association for the Advancement of Artificial Intelligence* (AAAI. Vancouver, BC, Canada, August, 2007).

[34] MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk, 3rd ed*. Mahwah, NJ: Lawrence Erlbaum Associates.

[35] Gold, K. (2008). Using Sentence Context and Implicit Contrast to Learn Sensor-Grounded Meanings for Relational and Deictic Words: The TWIG System. Yale University, May 2008.