Robots predicting the Interruptibility of Humans

Siddhartha Banerjee and Sonia Chernova
Georgia Institute of Technology
{siddhartha.banerjee,chernova}@gatech.edu

***Abstract*—In this paper, we outline our research into the appropriate timing of robot interruptions of humans. This research is motivated by the rising popularity of collocated human robot task execution and the resulting need for robots to be mindful of human cognitive limitations. We draw inspiration from existing work in Psychology, Human Factors Research, HCI, Ubiquitous Computing, and build upon preliminary work in Robotics to present a paradigm for interruption that seeks to mitigate deleterious effects of unchecked robot interruptions.**

I. Introduction

Deployment of service robots is most likely to happen in scenarios where humans and robots will be working together. Collaborative scenarios will require humans and robots to work on the same task, or be working on different tasks, within the same environment. We focus on the latter scenario and state that the goal of effective deployment of service robots is to minimize the performance penalty of the group members in their respective tasks. In particular, we examine the problem of user *interruptibility*.

Many co-robot tasks will require the robot to interrupt the human either to signal task completion, to request a new task, or to notify of an error. Psychology literature has shown that interruptions at the wrong time have the potential to be disruptive and introduce a performance penalty on the human [9]; meanwhile, in the right context, interruptions sometimes have the potential to be beneficial [23]. Thus, we hypothesize that the ability to reason about the iterruptibility of a person can have significant impact on human-robot interaction and team effectiveness.

In our own day-to-day experience, humans are quite accurate and remarkably consistent at judging the interruptibility of others. Consider for example the situations depicted in Table I. It is easy for most readers to gauge the likely interruptibility of a human from simple contextual cues in the pictures. In real time interactions, humans determine the interruptibility of others based on contextual cues gathered from multiple sources of information, including audio, vision, location, time, and past experience.

Our main goal, therefore, is to enable autonomous mobile service robots become capable of providing similar interruptibility judgments as those of humans. Crucially, we desire that the robot use no knowledge other than what is derived from its on-board sensors (RGB-D camera, laser scanner) and past experience to provide this judgment. More specifically, our work seeks to make the following contributions:

1) To create a *model of interruptibility* for a human in a given scene that can serve as a reliable proxy for cognitive load.
2) To create an autonomous system capable of *making appropriate interruptions* given the model of interruptibility in the current context.

Within the scope of this paper, we focus on the first of the above goals, for which we build upon the framework proposed by Chiang et al. [5]. The authors characterize the decision process for interruption as a Human-Aware Markov Decision Process (HAMDP), within which decisions are made based on the state of a relevant human. This state is determined at any point of time by a combination of environmental factors—the human's orientation relative to the robot, their focus of attention, their voice activity—and an inferred Theory-of-Mind-like variable for which the value encompasses past observations of the human. The authors use this state variable in a reinforcement learning based approach to learn human preferences to interruption methods. We propose to extend the HAMDP formulation to learn the appropriate time to interrupt a human. In this paper we outline our approach; as a next step we will conduct user study evaluations to establish the applicability of the above models.

II. Related Works

The characteristics of interruptions and their effects on the human psyche has been the focus of research in the field of Psychology and Human Factors Research since the early twentieth century. In particular, Gillie and Broadbent [9] and Speier et al. [23] expound on the cognitive effects of interruptions on humans—both works identify appropriate timing of interruptions as a key factor in the nature of these cognitive effects. Meanwhile, recent work by Rivera [17] characterizes the factors considered by humans, particularly nurses in a hospital, during the decision making process for interruptions.

The fields of Human-Computer Interaction (HCI) and Ubiquitous Computing (Ubicomp) have also contributed to the study of interruptibility by building upon the body of psychological literature available. Some researchers have classified interruption types when studying their effects on humans [14, 19], while others have focused on characterizing the environment [10, 7], and still others have tried to define structures on the interruptee's task to determine an acceptable time to interrupt [1, 11]. We found that determining the context of the interruption was an important factor for all the interruption research in HCI and Ubicomp; thus our work draws inspiration from this realization to incorporate a focus on interruption context.

Within the context of robotics, researchers working on companion robots [18], shopping mall assistants [20, 12, 22], receptionists and bartenders [15, 2, 8], and other conversational agents [3, 16, 4] have all dealt with the problem of interrupting humans. In particular, research by Saulnier [21] and Chiang et al. [5] has focused on the problem of generating appropriate interruptions for humans: the second of our goals outlined above.

III. Approach

Our current work focuses on the first of the proposed research questions, to create a *model of interruptibility* for a human that can serve as a reliable proxy for cognitive load. We wish to use this model to make a decision on when to interrupt such that we cause the least penalty in pre-interruption task performance for the interruptee.

Humans use multiple audio, visual, and contextual cues when making this decision to interrupt another person. Rivera [17] found that nurses at a hospital weighed the following factors during the decision making process:

1) Level of comfort with the interruptee
2) Reaction of the interruptee to past interruptions
3) Level of busyness projected by the nonverbal behavioural cues of the interruptee
4) Role and position of the interruptee within a nursing unit
5) Current task of the interruptee, as determined through audio and visual cues
6) Urgency and nature of the interrupting task
7) Experience based knowledge of the consequences of the interruption to the interrupted task, the interrupting task, the interrupter, and the interruptee

While the decision parameter of level of comfort, *parameter* 1, can be hard to influence or quantify within robotics, obtaining the other parameters is a focus of active research in the community. Work on task and action recognition [13], and the field of socially intelligent robotics [24], is particularly relevant for obtaining these other parameters. Therefore, we draw inspiration from this work to build upon research done by Chiang et al. [5].

As outlined above, Chiang et al. [5] characterized the interruption decision problem as a *Human Aware Markov Decision Process* with a state formulation that represented the affect of a human interruptee as well as the robot's belief about the human from past observations. In order to derive this state, the authors focused on three nonverbal behaviour cues from the human—visual focus of attention, pose of the body, and voice activity detection. These cues loosely correspond to the decision *parameters* 2 *&* 3 used by the nurses. While the ultimate aim of the authors in this work was to use these features to determine personalized interruption behaviours for a person, we believe the same features can be used to also determine when to interrupt a person.

In Table I, we present snapshots from video streams collected by our mobile robot as it moves around an office building environment. The depicted snapshots are taken from the same area of the building — a small kitchen and eating

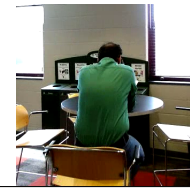
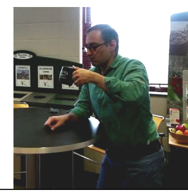
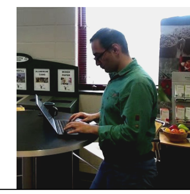
Independent Variable	**Interruptible**	**Not Interruptible**
Human's Pose		
Human's Attention Allocation		
Objects in the Scene		

TABLE I: Snapshots from video streams depicting our lab colleague under different conditions of the independent variable. The scenes have been sorted into interruptible or not based on the authors' judgment.

(a) (b)

Fig. 1: Individual video frames analyzed using the STRANDS Perception Pipeline [6] to detect and orient a person. Here we show outputs indicating the orientation of the face (a) and the pose of a person in a scene (b).

area. The images showcase some of the cues that communicate the interruptibility of a person. Direct eye contact is one of the early cues of social engagement and serves as a significant factor in determining interruptibility. However, even when no eye contact is being made (five out of the six example images), secondary factors, such as pose, location and objects of interest, can provide additional information. For example, in the bottom row of Table I, understanding the context, in this case the activity the person is engaged in, provides interruptibility cues — someone drinking coffee is typically far more interruptible than someone answering their email.

Within the context of our work we are in the process of developing a real-time interruptibility prediction model. This research effort is still in its early stages; we have collected a data set of individual users in natural human environments, and Figure 1 presents example features extracted based on that data. The video segments have been annotated by us on a six-point scale of interruptibility based on our own judgment. User studies in the near future will focus on the problem of garnering accurate annotations for training our models.

A key extension our approach provides over prior work is the use of semantically grounded contextual reasoning based on recognition of objects in the environment. Our hypothesis is that decisions made with contextual cues can lead to markedly better interruptibility estimates. At this point we are hard coding the semantics of recognized objects in the scene, but the automatic creation of these semantics is also the focus of active research in our laboratory.

Once the person and object detection pipelines are tested, and the methods of gathering supervision are complete, we seek to validate our approach in an extended real world deployment within an office building environment.

References

[1] Piotr D. Adamczyk and Brian P. Bailey. If not now, when? In *Proc. of the Conf. on Human factors in Computing Systems '04*, pages 271–278, New York, New York, USA, 2004. ACM Press. ISBN 1581137028. doi: 10.1145/985692.985727.

[2] Dan Bohus and Eric Horvitz. Dialog in the open world. In *Proc. of the Int. Conf. on Multimodal Interfaces '09*, page 31, New York, New York, USA, 2009. ACM Press. ISBN 9781605587721. doi: 10.1145/1647314.1647323.

[3] Dan Bohus and Eric Horvitz. Decisions about turns in multiparty conversation. In *Proceedings of the 13th international conference on multimodal interfaces - ICMI '11*, page 153, New York, New York, USA, 2011. ACM Press. ISBN 9781450306416. doi: 10.1145/2070481.2070507.

[4] Crystal Chao. *Timing Multimodal Turn-Taking in Human-Robot Cooperative Activity*. PhD thesis, Georgia Institute of Technology, 2015.

[5] Yi-Shiu Chiang, Ting-Sheng Chu, Chung Dial Lim, Tung-Yen Wu, Shih-Huan Tseng, and Li-Chen Fu. Personalizing robot behavior for interruption in social human-robot interaction. In *2014 IEEE International Workshop on Advanced Robotics and its Social Impacts*, pages 44–49. IEEE, sep 2014. ISBN 978-1-4799-6968-5. doi: 10.1109/ARSO.2014.7020978.

[6] Christian Dondrup, Nicola Bellotto, Ferdian Jovan, and Marc Hanheide. Real-time multisensor people tracking for human-robot spatial interaction. In *Workshop on Machine Learning for Social Robotics at Int. Conf. on Robotics and Automation*. ICRA/IEEE, 2015.

[7] James Fogarty, Scott E. Hudson, Christopher G. Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny C. Lee, and Jie Yang. Predicting human interruptibility with sensors. *ACM Transactions on Computer-Human Interaction*, 12(1):119–146, mar 2005. ISSN 10730516. doi: 10.1145/1057237.1057243.

[8] Mary Ellen Foster, Andre Gaschler, and Manuel Giuliani. How Can I Help You? Comparing Engagement Classification Strategies for a Robot Bartender. In *Proc. of the 15th ACM on Int. Conf. on Multimodal Interaction '13*, pages 255–262, New York, New York, USA, 2013. ACM Press. ISBN 9781450321297. doi: 10.1145/2522848.2522879.

[9] Tony Gillie and Donald Broadbent. What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychological Research*, 50(4):243–250, apr 1989. ISSN 0340-0727. doi: 10.1007/BF00309260.

[10] Eric Horvitz and Johnson Apacible. Learning and reasoning about interruption. In *Proceedings of the 5th international conference on Multimodal interfaces - ICMI '03*, page 20, New York, New York, USA, 2003. ACM Press. ISBN 1581136218. doi: 10.1145/958432.958440.

[11] Shamsi T. Iqbal and Brian P. Bailey. Leveraging characteristics of task structure to predict the cost of interruption. In *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*, page 741, New York, New York, USA, 2006. ACM Press. ISBN 1595933727. doi: 10.1145/1124772.1124882.

[12] Yusuke Kato, Takayuki Kanda, and Hiroshi Ishiguro. May I help you? In *Proc. of the 10th Annual ACM/IEEE Int. Conf. on Human-Robot Interaction '15*, pages 35–42, New York, New York, USA, 2015. ACM Press. ISBN 9781450328838. doi: 10.1145/2696454.2696463.

[13] Yu Kong and Yun Fu. Introduction. In *Human Activity Recognition and Prediction*, pages 1–22. Springer International Publishing, Cham, 2016. doi: 10.1007/978-3-319-27004-3_1.

[14] Daniel McFarlane and Kara Latorella. The Scope and Importance of Human Interruption in Human-Computer Interaction Design. *Human-Computer Interaction*, 17(1):1–61, mar 2002. ISSN 0737-0024. doi: 10.1207/S15327051HCI1701_1.

[15] Marek P. Michalowski, S. Sabanovic, and Reid Simmons. A spatial model of engagement for a social robot. In *9th IEEE Int. Workshop on Advanced Motion Control, 2006.*, pages 762–767. IEEE, 2006. ISBN 0-7803-9511-1. doi: 10.1109/AMC.2006.1631755.

[16] Tomislav Pejsa, Dan Bohus, Michael F. Cohen, Chit W. Saw, James Mahoney, and Eric Horvitz. Natural Communication about Uncertainties in Situated Interaction. In *Proceedings of the 16th International Conference on Multimodal Interaction - ICMI '14*, pages 283–290, New York, New York, USA, 2014. ACM Press. ISBN 9781450328852. doi: 10.1145/2663204.2663249.

[17] A. Joy Rivera. A socio-technical systems approach to studying interruptions: Understanding the interrupter's perspective. *Applied Ergonomics*, 45(3):747–756, may 2014. ISSN 00036870. doi: 10.1016/j.apergo.2013.08.009.

[18] Stephanie Rosenthal, Manuela M. Veloso, and Anind K. Dey. Is Someone in this Office Available to Help Me? *Journal of Intelligent & Robotic Systems*, 66(1-2):205–221, apr 2012. ISSN 0921-0296. doi: 10.1007/s10846-011-9610-4.

[19] Nadine Sarter. Multimodal Support for Interruption Management: Models, Empirical Findings, and Design Recommendations. *Proceedings of the IEEE*, 101(9):2105–2112, sep 2013. ISSN 0018-9219. doi: 10.1109/JPROC.2013.2245852.

[20] Satoru Satake, Takayuki Kanda, Dylan F. Glas, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. How to approach humans? In *Proc. of the 4th ACM/IEEE Int. Conf. on Human Robot Interaction '09*, page 109, New York, New York, USA, 2009. ACM Press. ISBN 9781605584041. doi: 10.1145/1514095.1514117.

[21] Paul Saulnier. Exploring Socially Appropriate Nonverbal Robotic Interruption, 2011.

[22] Chao Shi, Masahiro Shiomi, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. Measuring Communication Participation to Initiate Conversation in HumanRobot Interaction. *International Journal of Social Robotics*, 7(5):889–910, nov 2015. ISSN 1875-4791. doi: 10.1007/s12369-015-0285-z.

[23] Cheri Speier, Joseph S Valacich, and Iris Vessey. The effects of task interruption and information presentation on individual decision making. In *Proceedings of the eighteenth international conference on Information systems*, pages 21–36. Association for Information Systems, 1997.

[24] Zerrin Yumak and Nadia Magnenat-Thalmann. Multimodal and Multi-party Social Interactions. pages 275–298. 2016. doi: 10.1007/978-3-319-19947-4_13.