Foundations and Trends® in
Robotics
5:4

A Survey of Methods for Safe Human-Robot Interaction

Przemyslaw A. Lasota, Terrence Fong
and Julie A. Shah

now
the essence of knowledge

A Survey of Methods for Safe Human-Robot Interaction

Przemyslaw A. Lasota
Massachusetts Institute of Technology, USA
plasota@mit.edu

Terrence Fong
NASA Ames Research Center, USA
terry.fong@nasa.gov

Julie A. Shah
Massachusetts Institute of Technology, USA
julie_a_shah@csail.mit.edu

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Robotics

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

P. A. Lasota, T. Fong and J. A. Shah. *A Survey of Methods for Safe Human-Robot Interaction*. Foundations and Trends® in Robotics, vol. 5, no. 4, pp. 261–349, 2014.

This Foundations and Trends® issue was typeset in LaTeX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-68083-278-5

A Survey of Methods for Safe Human-Robot Interaction

Foundations and Trends® in Robotics
Volume 5, Issue 4, 2014

Editorial Board

Editorial Scope

Topics

Foundations and Trends® in Robotics publishes survey and tutorial articles in the following topics:

- Mathematical modelling
- Kinematics
- Dynamics
- Estimation methods
- Artificial intelligence in robotics
- Software systems and architectures
- Sensors and estimation
- Planning and control
- Human-robot interaction
- Industrial robotics
- Service robotics

Information for Librarians

Foundations and Trends® in Robotics, 2014, Volume 5, 4 issues. ISSN paper version 1935-8253. ISSN online version 1935-8261. Also available as a combined paper and online subscription.

Foundations and Trends® in Robotics
Vol. 5, No. 4 (2014) 261–349

DOI: 10.1561/2300000052

now
the essence of knowledge

A Survey of Methods for Safe Human-Robot Interaction

Przemyslaw A. Lasota
Massachusetts Institute of Technology,
USA
plasota@mit.edu

Terrence Fong
NASA Ames Research Center,
USA
terry.fong@nasa.gov

Julie A. Shah
Massachusetts Institute of Technology,
USA
julie_a_shah@csail.mit.edu

Contents

Abstract

Ensuring human safety is one of the most important considerations within the field of human-robot interaction (HRI). This does not simply involve preventing collisions between humans and robots operating within a shared space; we must consider all possible ways in which harm could come to a person, ranging from physical contact to adverse psychological effects resulting from unpleasant or dangerous interaction. In this work, we define what safe HRI entails and present a survey of potential methods of ensuring safety during HRI. We classify this collection of work into four major categories: safety through control, motion planning, prediction, and consideration of psychological factors. We discuss recent work in each major category, identify various sub-categories and discuss how these methods can be utilized to improve HRI safety. We then discuss gaps in the current literature and suggest future directions for additional work. By creating an organized categorization of the field, we hope to support future research and the development of new technologies for safe HRI, as well as facilitate the use of these techniques by researchers within the HRI community.

P. A. Lasota, T. Fong and J. A. Shah. *A Survey of Methods for Safe Human-Robot Interaction*. Foundations and Trends® in Robotics, vol. 5, no. 4, pp. 261–349, 2014.
DOI: 10.1561/2300000052.

Abstract

Ensuring human safety is one of the most important considerations within the field of human-robot interaction (HRI). This does not simply involve preventing collisions between humans and robots operating within a shared space; we must consider all possible ways in which harm could come to a person, ranging from physical contact to adverse psychological effects resulting from unpleasant or dangerous interaction. In this work, we define what safe HRI entails and present a survey of potential methods of ensuring safety during HRI. We classify this collection of work into four major categories: safety through control, motion planning, prediction, and consideration of psychological factors. We discuss recent work in each major category, identify various sub-categories and discuss how these methods can be utilized to improve HRI safety. We then discuss gaps in the current literature and suggest future directions for additional work. By creating an organized categorization of the field, we hope to support future research and the development of new technologies for safe HRI, as well as facilitate the use of these techniques by researchers within the HRI community.

P. A. Lasota, T. Fong and J. A. Shah. *A Survey of Methods for Safe Human-Robot Interaction*. Foundations and Trends® in Robotics, vol. 5, no. 4, pp. 261–349, 2014. DOI: 10.1561/2300000052.

1

Introduction

Human-robot interaction — collaboration, communication, and cooperation between humans and robots — is a rapidly growing area of robotics research. From introducing robotic co-workers into factories (Unhelkar et al., 2014; Gleeson et al., 2013; Knight, 2013), to providing in-home robot helpers (Graf et al., 2004), to developing robotic assistants for astronauts on-board the International Space Station (ISS) (Fong et al., 2013; Diftler et al., 2011; Bualat et al., 2015), there are a wide variety of beneficial applications for HRI. Whether this interaction involves an industrial robot, mobile manipulator, free-flyer, or even a self-driving car or wheelchair, one should always approach the development of HRI platforms and technologies from a safety-focused perspective. The successful advancement of HRI depends upon safety being a top priority and an integral component of any HRI application. In order to understand how to tackle the challenging problem of ensuring safety in HRI, it is necessary to clearly define what safe HRI entails and what has been accomplished thus far in terms of standardizing safety metrics and methods, and survey the current literature to identify areas that warrant further research and development.

1.1 Defining Safety in HRI

In order to ensure safe HRI, it is necessary to first understand what constitutes safety and its various components. In 1942, science fiction writer Isaac Asimov proposed three "Laws of Robotics," the first of which states: "A robot may not injure a human being or, through inaction, allow a human being to come to harm" (Asimov, 1942). Inspired by Asimov's definition, we can identify two distinct ways in which a robot could inflict harm on a human being.

The first is through direct physical contact. In simple terms, in order for HRI to be safe, no unintentional or unwanted contact can occur between the human and robot. Furthermore, if physical contact is required for a given task (or strict prevention of physical contact is neither possible nor practical) the forces exerted upon the human must remain below thresholds for physical discomfort or injury. We define this form of safety in HRI as **physical safety**.

Preventing physical harm alone, however, does not necessarily translate to stress-free and comfortable interaction. Consider, for example, a hypothetical manufacturing scenario in which a robot uses a sharp cutting implement to perform a task in proximity to human workers, but is programmed to stop if a human gets too close. While direct physical harm is prevented through careful programming, this type of interaction can be stressful for humans. Importantly, psychological discomfort or stress can also be induced by a robot's appearance, embodiment, gaze, speech, posture, and other attributes (Mumm and Mutlu, 2011; Butler and Agah, 2001).

Stress can have serious negative effects on health (McEwen, 1993), which makes stressful HRI a potential source of harm. Furthermore, psychological discomfort caused by any of the other aforementioned factors, as well as robotic violation of social conventions and norms during interaction, can also have serious negative effects on humans over time. We define the prevention of this type of indirect, psychological harm as maintaining **psychological safety**. It is important to note that psychological harm, in contrast with physical harm, is not limited to proximal interaction, as it can also be sustained through distal interaction via a remote interface.

As HRI can be applied in a multitude of domains, we apply a broad definition of the term "robot" in the context of this work. Although the individual works described in this survey are generally presented in the context of interaction with one type of robot in a specific domain, the methods for safety in HRI we present in the following sections are domain independent and relevant to a wide array of robot types, such as manipulator arms, drones, personal robots, and self-driving cars.

1.2 Safety Standards and Criteria

The development of guidelines and requirements in the form of international safety standards represents an important effort toward ensuring safety during human-robot interaction. The International Organization for Standardization (ISO) has been working toward releasing documents that specify how best to maintain safety during interaction between humans and industrial robots. The first step in this process was the release of the ISO 10218 document entitled "Robots and robotic devices – Safety requirements for industrial robots," which is composed of two parts: "Robots" and "Robot systems and integration" (International Organization for Standardization, 2011a,b). The ISO 10218 outlines some potential methods of safe collaborative manipulation — for example, *speed and separation monitoring* and *power and force limiting* — as well as relevant safety requirements.

The technical specification accompanying this document is the ISO/TS 15066 (entitled "Robots and robotic devices – Collaborative robots") (International Organization for Standardization, 2016). This technical specification provides additional information and details about how to achieve the requirements established by ISO 10218. It includes quantitative biomechanical limits, such as allowable peak forces or pressures for various parts of the body, as well as equations for speed and separation monitoring. In support of the development of the ISO technical specification, organizations including the National Institute of Standards and Technology (NIST) collaborated with ISO to develop protocols and metrics that would allow for characterization of the effectiveness of a robot's safety methods (National Institute of Standards and Technology, 2013).

The safety criteria mentioned above were developed in part through study of human-robot collisions. Recent experiments have incorporated collisions between robots and instrumented crash-test dummies, both in simulation (Oberer and Schraft, 2007) and using actual physical hardware (Haddadin et al., 2007, 2009). Other research has incorporated crash tests involving simulated human tissue, such as abdominal samples collected from pigs (Haddadin et al., 2012). Work with actual human-robot collisions has also been conducted to classify pain (Povse et al., 2010) and injury thresholds (Fraunhofer IFF, 2013), as well as to investigate the effectiveness of control strategies (Haddadin et al., 2008). Various injury prevention criteria for HRI have resulted from these works (Jung-Jun Park and Jae-Bok Song, 2009; Oberer and Schraft, 2007; Haddadin et al., 2012). Importantly, the findings are discussed in relation to the ISO standard regulations, providing feedback for their further refinement and improvement. (Haddadin (2013) have presented a detailed discussion of the limitations of the current standards and proposed improvements.) By combining the efforts of academic and industrial research groups and standardization organizations, more suitable and relevant standards and metrics can be developed and introduced in subsequent revisions of the ISO standards.

While the development of the aforementioned international safety standards represents a crucial first step toward improving HRI safety, it is important to note that these standards are being developed specifically for industrial applications. Although many of the principles would likely transfer to other types of robots and applications, the standards' scope is too narrow to fully address other uses, such as robotic tour guides or assistants for the elderly. We therefore must look beyond these industrial standards in order to identify all the pertinent aspects of safe HRI and the various possible safety methods that could be employed to address them.

1.3 Goals and Scope

The main goal of this work is to organize and summarize the large body of research related to facilitation of safe human-robot interaction. This survey describes the strategies and methods that have been developed thus far, organizes them into subcategories, characterizes relationships between the strategies, and identifies potential gaps in the existing knowledge that warrant further research.

1.3.1 Method

As there is a vast amount of work that could be applied to safe HRI, it was imperative to select a cohesive and meaningful subset of research. We conducted a survey to identify the various *methods* that could be utilized to make HRI safe. This is in contrast to other work, such as that of Vasic and Billard (2013), who partially outlined these possibilities but organized the paper according to application and focused on other aspects of safety, such as potential sources of danger and liability.

Also, we chose to focus our survey on recent research. A survey on safety in HRI by Pervez and Ryu (2008) covered much of the earlier work conducted within the field; this review mostly discusses research that had been published since that survey. Additionally, our survey focuses on the safety aspects of proximal HRI, and so we do not consider, for example, safety concerns during remote operation. Furthermore, we chose to focus this survey on interaction with robots acting as independent entities, and so we do not consider the regime of interaction with wearable robots, such as exoskeletons or orthotics. (We direct the reader interested in the latter topic to recent works in both industrial and medical applications (Kolakowsky-Hayner et al., 2013; O'Sullivan et al., 2015; Zeilig et al., 2012).) This survey also does not focus on the psychological safety aspects of interacting with social robots and the potential impact such robots can have when emulating human personality traits or social behaviors. (The reader interested in these aspects should consult works relating social psychology to robotics, such as papers by Young et al. (2008) and Fong et al. (2003).)

For the present work, we chose not to focus on robot hardware development as a potential method of ensuring safety in HRI. In recent years, robotics manufacturers have become increasingly involved in the development of robots designed specifically for proximal HRI. (Examples of such robots include the ABB YuMi (ABB, 2015), the RethinkRobotics Baxter and Sawyer (Robotics, 2015a,b), and the KUKA LBR (KUKA, 2015).) There has also been a significant amount of work in hardware development for safe HRI within the academic community, and the technologies used by these new robots are often a product of this research. This includes work on new actuators designed to be human-safe, such as series elastic actuators (SEA) (Pratt and Williamson, 1995), variable impedance actuators (VIA) (Vanderborght et al., 2013), distributed macro-mini actuation (Zinn et al., 2004), or external hardware, such as robot skins (Hoshi and Shinoda, 2006). (We direct the reader interested in compliant actuator designs to the review by Ham et al. (2009).)

Defining the scope of our work as outlined above, our selection process focused on papers published between 2008 and 2015 from the conference proceedings of the ACM/IEEE International Conference on Human-robot Interaction (HRI), IEEE International Conference on Robotics and Automation (ICRA), Robotics: Science and Systems (RSS), the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), the IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob), and the International Conference on Advanced Robotics (ICAR), as well as journal articles published in the *International Journal of Robotics Research* (IJRR), the *Journal of Mechanical Science and Technology* (JMST), the *IEEE Transactions on Robotics* (T-RO), the *IEEE Transactions on Automation Science and Engineering* (T-ASE), and the *Journal of Robotic Systems.*

We first grouped papers according to theme; common keywords among papers within each theme were then used as further search criteria. We focused our final selection on publications with higher impact factors and according to the selectivity of the publication venue. We relaxed these constraints if a topic associated with a keyword was underrepresented or the work was published within the last 3 years. We also recursively investigated works cited by the collected papers to identify additional potential sources. The resulting collection was then organized into the following main themes: safety through control, planning, prediction, and consideration of psychological factors, as depicted in Figure 1.1.