Available online at www.sciencedirect.com
ScienceDirect
Reliability Engineering and System Safety 93 (2008) 1280–1286
RELIABILITY ENGINEERING & SYSTEM SAFETY
www.elsevier.com/locate/ress

Short communication

Analytical basis for evaluating the effect of unplanned interventions on the effectiveness of a human–robot system

Julie A. Shah[a,*], Joseph H. Saleh[b], Jeffrey A. Hoffman[a]

[a]Department of Aeronautics and Astronautics, 77 Massachusetts Avenue, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[b]School of Aerospace Engineering, 270 Ferst Drive, Georgia Institute of Technology, Atlanta GA 30332, USA

Received 11 April 2006; received in revised form 11 June 2007; accepted 12 June 2007
Available online 5 July 2007

Abstract

Increasing prevalence of human–robot systems in a variety of applications raises the question of how to design these systems to best leverage the capabilities of humans and robots. In this paper, we address the relationships between reliability, productivity, and risk to humans from human–robot systems operating in a hostile environment. Objectives for maximizing the effectiveness of a human–robot system are presented, which capture these coupled relationships, and reliability parameters are proposed to characterize unplanned interventions between a human and robot. The reliability metrics defined here take on an expanded meaning in which the underlying concept of failure in traditional reliability analysis is replaced by the notion of intervention. In the context of human–robotic systems, an intervention is not only driven by component failures, but includes many other factors that can make a robotic agent to request or a human agent to provide intervention, as we argue in this paper. The effect of unplanned interventions on the effectiveness of human–robot systems is then investigated analytically using traditional reliability analysis. Finally, we discuss the implications of these analytical trends on the design and evaluation of human–robot systems.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Human–robot interaction; Multi-agent system; System reliability; Risk to humans; Mean time between intervention

1. Introduction and motivation

Human–robot systems are being increasingly considered, and used in a number of military operations, civilian search, and rescue operations, and are proposed as an integral part of future space missions to the Moon and Mars [1–3]. The increased relevance of human–robot systems raises the issue of how to optimally (and reliably) design these systems to best leverage the varied capabilities of humans and robots. The question of optimality in turn raises the question of what metrics to use in order to guide the design, and evaluate the performance, of human–robot systems. Unfortunately, an analytical framework of common metrics does not currently exist to compare the performance of different human–robot systems. Formulating such a framework is challenging in part because techniques do not exist to incorporate the effect of human–robot interaction into methods for evaluating metrics such as productivity, reliability, and risk to humans [4]. The operational environment of human–robot systems is often hostile to the human agents and the metrics of productivity, reliability, and risk to humans are strongly coupled.

This paper addresses the relationships between productivity, reliability, and risk to humans for human–robot systems operating in a hostile environment. Objectives for maximizing the effectiveness of a human–robot system are presented, which capture these coupled relationships, and parameters are proposed to characterize unplanned interventions between a human and robot (Section 2). The effect of unplanned interventions on measures of effectiveness is discussed qualitatively (Section 3). Next, the effect of unplanned interventions is formulated analytically in terms of the reliability parameters defined by the authors (Section 4). The potential implications of this preliminary analysis on the design and evaluation of human–robot systems are then discussed (Section 5).

*Corresponding author. Tel.: +1 617 253 2448; fax: +1 617 253 0147.
E-mail address: arnoldj@mit.edu (J.A. Shah).

0951-8320/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ress.2007.06.007

2. Metrics and definitions

In designing human–robot systems to operate in a hostile environment, two objectives ought to be considered:

(1) To maximize the amount of useful work that human agents can do within a time window (human productivity).
(2) To minimize the risk to human agents. In many situations, this entails minimizing the time human agents spend in the hostile environment (exposure).

The details of the specific task, the mission-level timeline, and the environment in which the task is to be performed determine which of these measures is most appropriate. Minimizing exposure may be the primary objective in many terrestrial applications involving few tasks, dangerous and unpredictable environments, and low penalties for entering and exiting the operational environment. For example, urban search and rescue teams may minimize the risk to rescuers by using robotic agents to search buildings in danger of structural collapse and identify victims before sending in rescuers, even though this strategy may not maximize the productivity of each rescuer. In contrast, maximizing human productivity may also be the primary objective in situations where there is some penalty for entering and exiting the operating environment. For example, astronauts preparing for an extravehicular activity (EVA) on the Space Shuttle and International Space Station are required to breathe pure oxygen for up to a few hours before the EVA to avoid decompression sickness. With this penalty, mission planners may prefer to maximize the astronauts' productivity during the span of a single EVA rather than require the astronauts to perform many short EVAs—even if many short EVAs would minimize the time astronauts spend in the hostile environment. While both maximizing human productivity and minimizing exposure are important considerations for human space flight, maximizing human productivity is also likely to be the dominant objective for situations in which astronauts have other useful work to do in the extravehicular environment. For example, robotic agents may be deployed to reduce the amount of time humans would spend working on a particular task. Astronauts would then use the extra time to begin working on other tasks, thereby increasing the human productivity during an EVA. The motivation for the following discussions is to formulate an analytical basis for investigating the effect that human–robot interaction has on the objectives to maximize human productivity and minimize exposure.

This preliminary analysis investigates the effect of unplanned interventions, a specific type of human–robot interaction, on each of these objectives. The following definitions are presented for the purpose of this analysis:

An *intervention* is defined as a robotic agent receiving unplanned assistance from a human agent.

Mean time between interventions (MTBI) is the mean time that a human–robot system operates nominally (human and robotic agents are not engaging in an intervention). This is defined as

$$\mathrm{MTBI} = \frac{t_{\text{btw-int}}}{n_{\text{int}}}, \tag{1}$$

where $t_{\text{btw-int}}$ is the cumulative time that the system is not engaging in unplanned interventions requiring a human agent and n_{int} is the number of unplanned interventions requiring a human agent.[1]

MTBI is analogous to the mean time between failures (MTBF) as defined in the IEEE Standard [5].[2] However, while MTBF refers to component or system reliabilities, MTBI takes on a broader meaning. MTBI is a function of:

(1) The environment (and uncertainty in the environment) that the system or robotic agent is operating in.
(2) The autonomy of the system or robotic agent, in this case defined as the ability to accommodate variations in the environment in pursuit of its goals [6] (in this case, without human intervention).
(3) The inherent component or system reliabilities.

For example, the MTBI for a rover operating in a boulder-strewn field may be dependent on the size and distribution of boulders, and also the rover's ability to autonomously navigate among boulders. The inherent reliabilities of the components utilized by the rover while navigating will also affect MTBI. In defining MTBI we have renegotiated the meaning of MTBF to better capture properties of a human–robot system in the same spirit that [7] defined a maintenance-free operating period to better analyze and predict the properties of aerospace systems.

Mean time completing intervening (MTCI) is the mean duration of interventions. This is defined as

$$\mathrm{MTCI} = \frac{t_{\text{int}}}{n_{\text{int}}}, \tag{2}$$

where t_{int} is the cumulative time spent by human agents engaging in unplanned interventions and n_{int} is the number of interventions requiring a human agent.

MTCI is analogous to the mean time to repair (MTTR) as defined in the IEEE Standard [5].[3] MTCI is a function of many variables, including (but not limited to)

- the nature of the failure or problem requiring an intervention;

[1] By this definition, MTBI is a measure of the mean time to an intervention plus the mean time to complete the intervention.

[2] MTBF is defined in IEEE Std. 493-1997 as: The mean exposure time between consecutive failures of a component. It can be estimated by dividing the exposure time by the number of failures in that period, provided that a sufficient number of failures has occurred in that period.

[3] MTTR is defined in IEEE Std. 493-1997 as: The mean time to repair or replace a failed component. It can be estimated by dividing the summation of repair times by the number of repairs, and, therefore, it is practically the average repair time.

- the design parameters describing cognitive abilities and interaction among agents including
  - the amount of information agents are able to gather about the nature of the failure prior to and during the intervention
  - the amount and type of information that can be transferred between a robotic and human agent;
- the physical distance between a robotic and human agent;
- the lag in communications between a robotic agent and human agent;
- the available resources and tools.

The goal of this work is to build on these objectives and parameters to analytically describe the effect of interventions on the effectiveness of human–robot systems, and explore potential implications for the design and evaluation of human–robot systems. Next, the effect of interventions on system effectiveness is discussed qualitatively to form the basis for an analytical discussion.

3. Qualitative discussion of the effect of interventions

In this section, the concept of a mission timeline is used to qualitatively discuss the effect of interventions. A nominal mission timeline in which the human–robot system performs a collaborative task without unplanned interventions is shown in Fig. 1a. We assume that the human and robot begin working on the collaborative task at the same time $t = 0$. The mission time, t_(max.mission), is the maximum time window that the human–robot system has to perform a specific task, and represents time constraints associated with humans working in a hostile environment. Consider for example an astronaut performing a spacewalk or a scuba diver on a dive. In these cases, the maximum time window is dictated by the amount of life support consumables (e.g. oxygen) the human agents can carry with them. This is the same maximum time window referred to in the objective to maximize human productivity in Section 2. The nominal amount of time for the human–robot system to perform a specific task is labeled t_(task) in Fig. 1a. The nominal amount of time required for the human agents to fulfill their part of the specific task is labeled t_(human.finish). The time remaining once the human agents fulfill their part of the specific task is the time available for the humans to do other work, either within or outside the hostile operating environment.

In this paper, robotic agents that do not require interventions are referred to as "reliable", while agents that do require interventions are referred to as "unreliable."

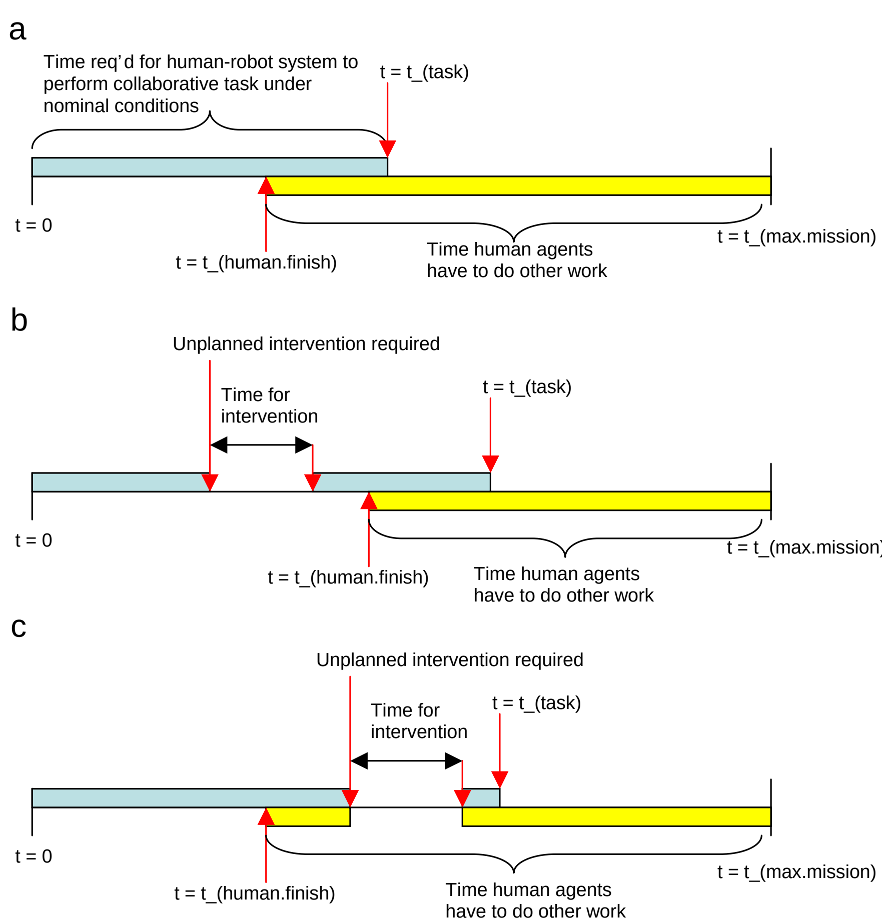

Fig. 1. Effect of interventions on mission-level timeline.

These qualifiers, reliable and unreliable, are obviously not used in their traditional sense, but they take on an expanded meaning in which the underlying concept of failure (or time to failure) is replaced by the notion of intervention (or time to intervention). In the context of human–robotic systems, an *intervention* is not only driven by component failures, as discussed in Section 2.

The case where robotic agents are "unreliable" and require unplanned interventions is depicted in Fig. 1b,c. Interventions while human agents are still fulfilling their part of the task increase t_(human.finish) and t_(task), as shown in Fig. 1b. Unplanned interventions after t_(human.finish) lead to a situation in which human agents may be required to remain in the hostile operational environment and attend to these unplanned interventions. If human agents do not remain in the operational environment and a time penalty is incurred for repeatedly returning to the operational environment, the time required to respond and attend to interventions increases. This situation is depicted in Fig. 1c. Each of these situations would significantly increase exposure and decrease the time available for human agents to do other work.

However, imagine that the human–robot system was "reliable"—in the sense that the robots did not often run into problems requiring intervention. In this case, the human agents would have a choice: once they finish their primary task, they could remain in the operational environment and begin working on other tasks. This would increase human productivity. Or, once the human agents finish their primary task, they could return to the safe environment secure in the knowledge that the robots will continue to work without requiring interventions. This would minimize the exposure of the human agents. In other words, a "reliable" human–robot system provides the option of maximizing the effectiveness of the human–robot system by either maximizing human productivity or minimizing exposure. In the next section, these relationships are quantified using the metrics and definitions presented in Section 2.

4. Analytic formulation of the effect of interventions

In the following analyses, the effect of interventions on the objective of maximizing human productivity is explored by expressing the time available for humans to do other work as a function of MTBI and MTCI. In addition, the effect of interventions on the second objective discussed in Section 2, namely minimizing human exposure to the hostile environment, is explored by expressing the probability of intervention after the t_(human.finish) as a function of MTBI and MTCI.

4.1. Effect of MBTI and MTCI on time for other work

Under nominal conditions (no interventions required), humans will complete their task in t_(human.finish), as shown in Fig. 1a. The remaining time, assuming the window of operation in the hostile environment t_(max.mission) and

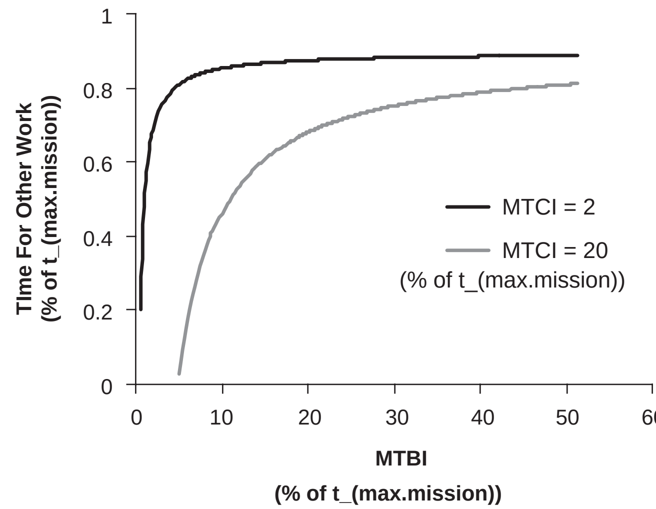

Fig. 2. Time for other work as a function of MTBI and MTCI.

nominal conditions, is given by

$$t_{\text{other}} = t_{\text{max.mission}} - t_{\text{human.finish}}. \tag{3}$$

As the robotic agents start requiring interventions, the time available for humans to do other work decreases, and thus limits the productivity of the human agents. This section explores how the time available for other work varies as a function of the MTBI required by the robotic agents, and with the expected duration of interventions (MTCI).

A general expression of the time available for humans to do other work is presented in Eq. (4). This expression is normalized by the nominal amount of time (without interventions) for the human–robot system to perform a specific task. Time for other work is given by

$$\begin{aligned}\frac{t_{\text{other}}}{t_{\text{task}}} &= 1 - \frac{t_{\text{human.finish}}}{t_{\text{task}}} - \frac{t_{\text{human.finish}}}{t_{\text{task}}\text{MTBI}}\text{MTCI} \\ &\quad - \frac{(t_{\text{task}} - t_{\text{human.finish}})}{t_{\text{task}}\text{MTBI}}\text{MTCI},\end{aligned} \tag{4}$$

where t_(human.finish) is the nominal amount of time (with no interventions) required for the human agents to fulfill their part of the specific task and t_(task) is the nominal amount of time for the human–robot system to perform a specific task. The first two terms in Eq. (4) represent the time for other work with no interventions. The third term accounts for the total time spent engaging in interventions before t_(human.finish), as depicted in Fig. 1b. The last term accounts for the total time spent engaging in interventions after t_(human.finish), as presented in Fig. 1c.[4]

Fig. 2 shows the amount of time available for human agents for other work as a function of MTBI and MTCI.

The figure shows that as MTBI increases, the time for other work initially increases sharply and then

[4]Eq. (4) also assumes that the MTBI is measured from the end of one intervention to the start of the next intervention, and that two interventions cannot occur during the same time.

plateaus. Increasing MTCI by an order of magnitude decreases the time for other work and softens the transition between the initial increase and plateau as a function of MTBI.

This analysis assumed that MTCI and MTBI are constant throughout the mission duration, t_(task) = 20% of t_(max.mission), and t_(human.finish) = 10% of t_(max.mission). The curves represent the relationships between MTBI (ranging from 0% to 60%) and the specific MTCI (2% or 20%) for which the time for other work is positive. The relationship between MTBI and MTCI, which yield a positive time for other work is given by

$$\frac{\text{MTCI}}{\text{MTBI}} \leqslant \frac{t_{\text{max.mission}} - t_{\text{human.finish}}}{t_{\text{task}}}. \tag{5}$$

4.2. Effect of MTBI and MTCI on the probability of intervention

Once the human agents finish their part of the task, they may choose to exit the hostile environment as soon as possible to minimize exposure. Under nominal conditions, humans complete their part of the task in t_(human.finish), and the robotic agents finish the task without requiring interventions as shown in Fig. 1a. In this case, the human agents may leave the hostile environment directly after t_(human.finish). However, if the robotic agents are likely to require interventions after t_(human.finish), the human agents may instead choose to remain in the hostile environment for a certain amount of time such that the probability of intervention past this point is within a specified threshold. This section explores the probability of intervention after t_(human.finish) as a function of MTBI required by the robotic agents, and the expected duration of interventions (MTCI).

The probability of intervention after human agents finish their part of the task is described using a Poisson distribution to model the occurrence of initiating events. In this case, the initiating event is an intervention. The probability that at least one intervention will be required between when the human agents finish their part of the task and when the task is complete is given by

$$F(t) = 1 - \exp\left[-\int_{t\_\text{human,tot}}^{t\_\text{task,tot}} h(s)\,\mathrm{d}s\right], \tag{6}$$

where $t_{\text{human,tot}}$ is the time required (including interventions) for the human agents to perform their part of the task and $t_{\text{task,tot}}$ is the total time (including interventions) to perform the task. Also, $h(s) = 1/\text{MTBI}$ is constant, and MTCI is constant.

The resulting expression for the probability that at least one intervention will be required after t_(human.finish) as a function of MTBI, MTCI, t_(task), and t_(human.finish) is

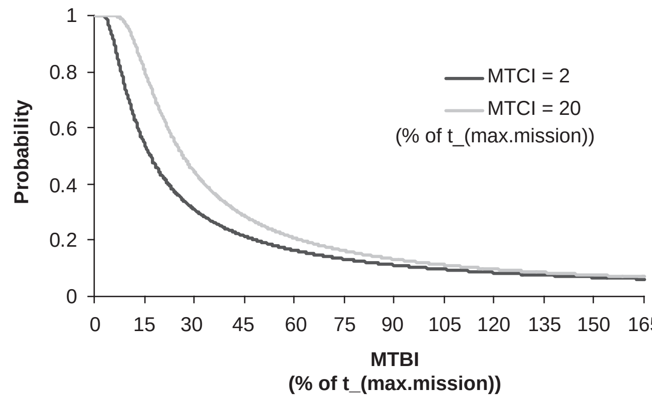

Fig. 3. Probability of intervention as a function of MTBI and MTCI.

given by

$$F(t) = 1 - \exp\left[-\frac{t_{\text{task}}}{\text{MTBI}} + \frac{t_{\text{human.finish}}}{\text{MTBI}} - \frac{t_{\text{task}} - t_{\text{human.finish}}}{\text{MTBI}^2}\,\text{MTCI}\right], \tag{7}$$

where t_(human.finish) is the nominal amount of time (with no interventions) required for the human agents to fulfill their part of the specific task, and t_(task) is the nominal amount of time for the human–robot system to perform a specific task.

Fig. 3 shows the probability that an unplanned intervention is required after the human agents have finished their part of the task as a function of MTBI and MTCI.

The figure shows that as MTBI increases, the probability of intervention decreases. Increasing MTCI by an order of magnitude shifts the curve and increases the probability of intervention for a given MTBI.

This analysis assumed that both MTBI and MTCI are expressed as a percentage of the total mission time, t_(task) = 20% of t_(max.mission), and t_(human.finish) = 10% of t_(max.mission).

5. Discussion

Preliminary analysis of the trends presented in Section 4 yields interesting insights for design and evaluation of human–robot systems. Fig. 2 indicates that the time available for other work is sensitive to both MTBI and MTCI. The sensitivities of time available for other work to changes in MTBI and MTCI are shown in Figs. 4 and 5, and are, respectively, described by[5]

$$s_{\text{MTCI}} = \left|\frac{\partial t_{\text{other}}}{\partial \text{MTCI}}\right| = \frac{t_{\text{task}}}{\text{MTBI}} \tag{8}$$

[5]As noted in Section 2, MTBI is a measure of the mean time to an intervention plus the MTCI. This equation assumes that MTCI is small compared with the mean time to an intervention.

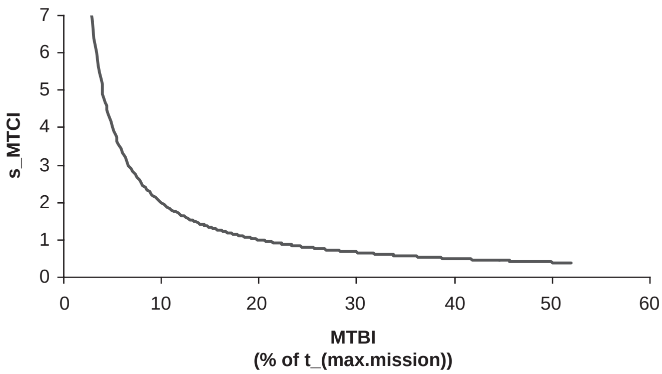

Fig. 4. Sensitivity of time available for other work to MTCI.

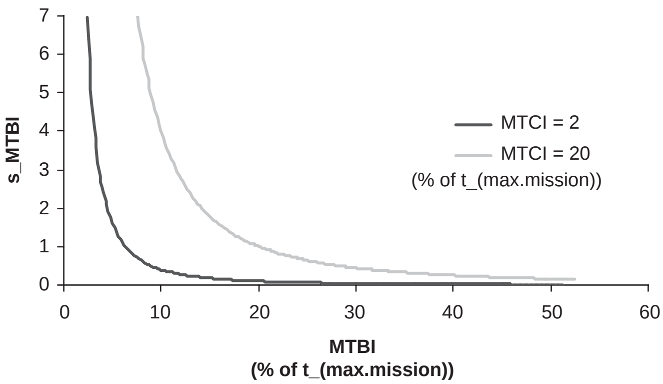

Fig. 5. Sensitivity of time available for other work to MTBI.

and

$$s_{\mathrm{MTBI}} = \left|\frac{\partial t_{\mathrm{other}}}{\partial \mathrm{MTBI}}\right| = \frac{(\mathrm{MTCI})t_{\mathrm{task}}}{\mathrm{MTBI}^2}. \quad (9)$$

In Figs. 4 and 5, t_{task}, MTBI, and MTCI are expressed as a percentage of the maximum mission time, and $t_{\mathrm{task}} = 20\%$ of the maximum mission time.

Increases in MTCI reduce the time available for human agents to do other work. However, this analysis shows that the sensitivity of time available for other work to MTCI is not a function of MTCI, and the sensitivities to MTCI quickly decreases as MTBI is increased. In other words, as the frequency of interventions decreases, the objective of maximizing human productivity becomes less sensitive to the duration of interventions. Also, increases in MTBI result in increased time available for other work, and the sensitivity to MTBI increases as MTCI increases. In other words, increases in the duration of interventions result in greater sensitivity to the frequency of interventions. This suggests that a designer may be able to compensate for large or uncertain MTCI and achieve increases in time available to do other work with modest increases in MTBI.

Interestingly, Fig. 3 indicates that the probability of robotic agents requiring an intervention after the astronauts have finished their part of the task is primarily a function of MTBI. The sensitivity of probability of intervention to changes in MTBI and MTCI are shown

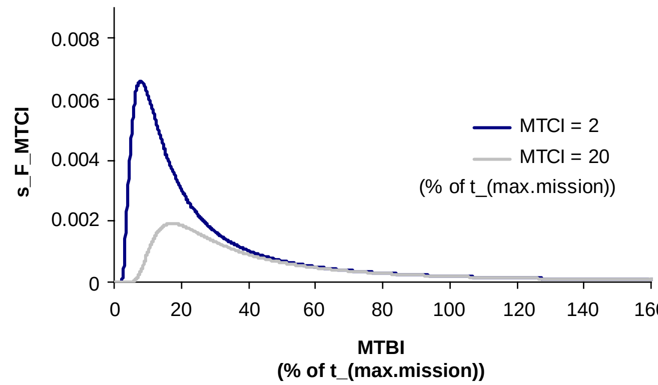

Fig. 6. Sensitivity of probability of intervention to MTCI.

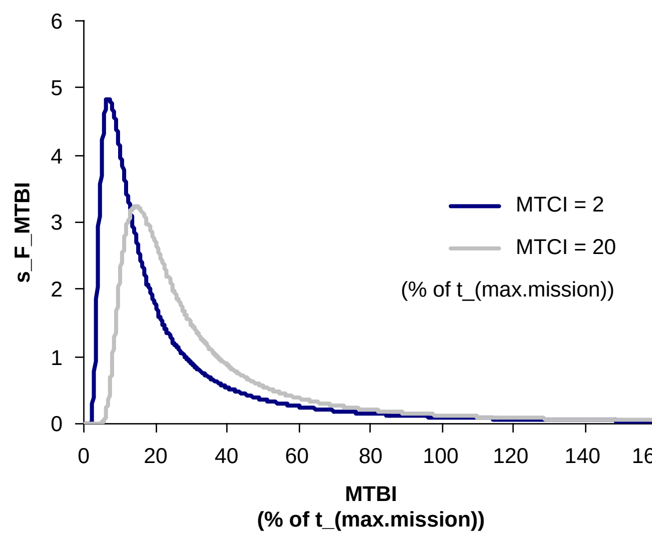

Fig. 7. Sensitivity of probability of intervention to MTBI.

in Figs. 6 and 7, and are, respectively, described by[6]

$$s_{F\_\mathrm{MTCI}} = \left|\frac{\partial F(t)}{\partial \mathrm{MTCI}}\right| = \frac{t_{\mathrm{task}} - t_{\mathrm{human.finish}}}{\mathrm{MTBI}^2} \exp\left[-\int_{t\_\mathrm{human,tot}}^{t\_\mathrm{task,tot}} h(s)\mathrm{d}s\right] \quad (10)$$

and

$$s_{F\_\mathrm{MTBI}} = \left|\frac{\partial F(t)}{\partial \mathrm{MTBI}}\right| = \left(\frac{t_{\mathrm{task}}}{\mathrm{MTBI}^2} - \frac{t_{\mathrm{human.finish}}}{\mathrm{MTBI}^2} + 2\frac{t_{\mathrm{task}} - t_{\mathrm{human.finish}}}{\mathrm{MTBI}^3}\mathrm{MTCI}\right) \times \exp\left[-\int_{t\_\mathrm{human,tot}}^{t\_\mathrm{task,tot}} h(s)\mathrm{d}s\right], \quad (11)$$

where t_{task} is the nominal amount of time (with no interventions) for the human–robot system to perform the task, $t_{\mathrm{human.finish}}$ is the nominal amount of time (with no interventions) required for the human agents to fulfill their part of the specific task, $t_{\mathrm{human,tot}}$ is the time required

[6]As noted in Section 2, MTBI is a measure of the mean time to an intervention plus the MTCI. This equation assumes that MTCI is small compared with the mean time to an intervention.

(including interventions) for the human agents to perform their part of the task, $t_{task,tot}$ is the total time (including interventions) to perform the task, and $h(s) = 1/\mathrm{MTBI}$.

In Figs. 6 and 7, t_{task}, $t_{human.finish}$ MTBI, and MTCI are expressed as a percentage of the maximum mission time, $t_{task} = 20\%$, and $t_{human.finish} = 10\%$ of the maximum mission time.

This sensitivity analysis shows that the probability of intervention is nearly three orders of magnitude more sensitive to MTBI than MTCI. This suggests that MTBI is the primary driver, and unknown or uncertain MTCI may not significantly impact the design of a system to minimize the probability of intervention.

The meanings of MTBI and MTCI in the context of this analysis have implications for human–robot system design and evaluation. As discussed previously, MTBI is a function of

(1) The environment (and uncertainty in the environment) that the system or robotic agent is operating in.
(2) The autonomy of the system or robotic agent, in this case defined as the ability to accommodate variations in the environment in pursuit of its goals [6] (in this case, without human intervention).
(3) The inherent component or system reliabilities.

MTCI is a function of many variables, including (but not limited to)

- the nature of the failure or problem requiring an intervention;
- the design parameters describing cognitive abilities and interaction among agents including
  - the amount of information agents are able to gather about the nature of the failure prior to and during the intervention
  - the amount and type of information that can be transferred between a robotic and human agent;
- the physical distance between a robotic and human agent;
- the lag in communications between a robotic agent and human agent;
- the available resources and tools.

This preliminary analysis suggests that a designer may be able to greatly impact both measures of the effectiveness of a human–robot system by increasing MTBI, despite the likely variability and unpredictability of MTCI. In particular, the objective of maximizing human productivity becomes less sensitive to MTCI as MTBI increases. In addition, the objective of minimizing exposure is significantly more sensitive to changes in MTBI than MTCI. These are encouraging results since MTBI is primarily a function of parameters that designers may influence, such as agent autonomy and component and system reliabilities.

These trends also have interesting implications for experiments aimed at evaluating the effectiveness of human–robot systems. Accurately characterizing MTCI through experimentation may not be necessary to formulate reasonable evaluations of the effectiveness of a human–robot system. This is fortunate since MTCI is dependent on a host of different factors and is likely to be difficult to accurately quantify. Another approach would be to characterize MTBI through experimentation with the factors that result in interventions, and conduct a sensitivity analysis to various MTCI.

References

[1] Carleson J, Murphy R. Human–robot interaction in rescue robotics. IEEE Trans Syst Man Cybern 2004;34(2):138–53.
[2] Burke JL, Murphy RR, Rogers E, Lumelsky VJ, Scholtz J. Final report for the DARPA/NSF interdisciplinary study on human–robot interaction. IEEE Trans Syst Man Cybern Pt C 2004; 34(2):103–12.
[3] Schenker P.S., Huntsberger T.L., Pirjanian P., Baumgartner E.T. Planetary rover developments supporting mars exploration, sample return and future human–robotic colonization. In: Proceedings of the 10th Conference on Advanced Robotics, 2001. p. 31–47.
[4] Shah J, Saleh J, Hoffman J. Review and synthesis of considerations in architecting heterogeneous teams of humans and robots for space exploration. IEEE Trans Syst Man Cybernet Pt C 2007;37(5).
[5] Institute of Electrical and Inc. Electronics Engineers. IEEE recommended practice for the design of reliable industrial and commercial power systems. IEEE Std. 493–1997, 1998. IEEE Gold Book.
[6] Thrun S. Toward a framework for human–robot interaction. Human–Comput Interact 2004;19(1–2):9–24.
[7] Kumar U, Knezevic J, Crocker J. Maintenance free operating period—an alternative measure to MTBF and failure rate for specifying reliability? Reliab Eng System Saf 1999;64(1):127–31.