When to Interrupt: A Comparative Analysis of Interruption Timings within Collaborative Communication Tasks

Nia Peters[1], Griffin Romigh[2], George Bradley[3], and Bhiksha Raj[1]

[1] Carnegie Mellon Universtiy 5000 Forbes Ave. Pittsburgh, PA 15213 U.S.A
[2] Air Force Research Laboratory, Wright-Patterson Air Force Base, Dayton, OH 45433
[3] Walden University 1000 S Washington Ave Minneapolis, MN 55401

{nbradley}@ece.cmu.edu, {griffin.romigh}@us.af.mil, {george.bradley}@waldenu.edu, {bhiksha}@cs.cmu.edu

Abstract. This study seeks to determine if it is necessary for a software agent to monitor the communication channel between a human operator and human collaborators to effectively detect appropriate times to convey information or "interrupt" the operator in a collaborative communication task. The study explores the outcome of overall task performance and task time of completion (TOC) at various delivery times of periphery task interruptions. A collaborative, goal-oriented task is simulated via a dual-task where an operator participates in the primary collaborative communication task and a secondary keeping track task. User performance at various interruption timings: random, fixed, and human-determined (HD) are evaluated to determine whether an intelligent form of interrupting users is less disruptive and benefits users' overall interaction. There is a significant difference in task performance when HD interruptions are delivered in comparison with random and fixed timed interruption. There is a 54% overall accuracy for task performance using HD interruptions compared to 33% for fixed interruptions and 38% for random interruptions. These results are promising and provide some indication that monitoring a communication channel or adding intelligence to the interaction can be useful for the exchange.

Keywords: Interruption management · Collaborative Communication · Turn-taking · Human-human-computer interactions

1 Introduction

The primary focus of this work is on exploring when to disseminate system-mediated interruptions within collaborative multi-tasks in auditory environments. In this work, an interruption *is the intention of a 3rd party to convey new information to a member participating in a collaborative communication task, which aids in one task but disrupts another*. Imagine a military operation in which an Unmanned Aerial Vehicle (UAV) operator is communicating with team members on the ground. In this operation, an example of a collaborative communication task between the operator and the teammates is a task in which the UAV operator and human teammates are coordinating the location of a particular ground object based on the teammates' perspective as well as the UAV's location and perspective. In conjunction with this task, the UAV operator may also be tasked with keeping track of various UAV state changes.

Figure 1 illustrates a generic representation of this specific scenario for the focused collaborative, communication task where an operator is communicating with human teammates and a machine teammate simultaneously.

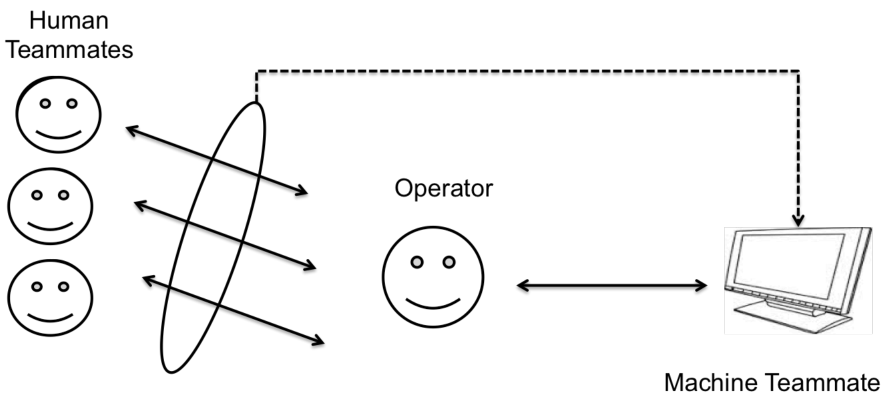

Fig. 1. Collaborative Communication & Monitoring Task

In Figure 1 notice there is bidirectional communication between the operator and human teammates as well as the operator and her machine teammate, but the human teammates are oblivious to the communication between the operator and the machine teammate. Here you have the collaborative communication between the operator and human teammates and another task between the operator and machine teammate. In this dual-task scenario, we would like to determine whether the machine teammate should monitor the communication channels between the operator and the human teammates before it communicates information to the operator. If monitoring the communication channel prior to an interruption is beneficial, when is the most optimal time to interrupt the operator that is the least invasive to the overall exchange and results in the operator performing all tasks quickly and effectively.

There is empirical research dedicated to manipulating time on the delivery [1], [2], [3] of system-mediated interruptions [4] in multi-task environments [5]. There is also literature that explores immediate interruption or notification dissemination [6], [7], [8] within dual-task scenarios. Studies have shown that delivering interruptions at random times can result in a decline in performance on primary tasks [1], [6], [9], [8], [10]. Additionally, studies have illustrated that interrupting users engaged in tasks has a considerable negative impact on task completion time [11], [6], [2], [9], [12], [3].

It is imperative that system interruptions are delivered at moments of minimal cognitive workload [13] to avoid information overload and maintain task efficiency. Miyata and Norman [14] argue that delivering notifications at moments of lower mental workload would lower the cost of interruptions. Coordinating interruptions within a dialogue is an important factor in maintaining efficient communication and preventing information overload [15].

Much of the current literature is focused on one user engaged in a primary task interrupted by a peripheral task. This study differs in that the primary task is a collaborative task between two or more users and the secondary task is presented to one of the collaborating users. This work aims to determine if monitoring a collaborative communication task will improve overall task performance and time of complete (TOC) within collaborative, goal-oriented exchanges. In this study *monitoring* refers to listening to a communication exchange between two users prior to distributing an interruption. The monitoring method used in this report is referred to as human-determined (HD) interruptions where a human interrupter listens to or monitors a collaborative communication channel between two users and decides on appropriate times of interruption. The hope is that humans use cues while monitoring a communication channel prior to interrupting that could be less disruptive to the exchange between collaborating participants. If humans monitoring the communication channel prior to interruptions are indeed valuable for the overall exchange, this could help justify designing an interruption system that mimics optimal human interruption cues in collaborative exchanges. In this study, there is a significant increase in operator task performance when the communication channel is monitored and human interruptions are disseminated in comparison to random and fixed times interruptions.

The rest of this paper is arranged as follows: Section 2 illustrates the methods used in this work. In Section 3 the evaluation process is explained. Section 4 presents the results, which are further discussed in Section 5. Finally, Section 6 is the conclusion.

2 Methods

A collaborative, goal-oriented task is simulated via a dual-task of a dyadic collaborative communication scenario (primary task) in conjunction with a keeping track task (secondary task). Data are collected through a series of exercises undertaken by a team of three individuals: an interrupter, an operator, and the operator's teammate. In the primary task, the operator and teammate perform a collaborative communication task in an effort to accomplish a common goal similar to that illustrated between the UAV operator and his human teammates. In the secondary task, the operator simultaneously performs a keeping track task. Within this task the operator must keep track of various UAV states where the state information is disseminated via the experimenter sending interruptions at varying interruption timings: fixed, random, and human determined (HD). Finally, human performance on the secondary task and the time of competition (TOC) on the primary task are evaluated to determine the implications of interrupting at different times. The assumption is that a human monitoring the communication channel (HD interruptions) will result in interruptions sent at a less disruptive time allowing the operator to complete the primary task at a faster rate and improve secondary task performance.

- Research Question I: Is there a difference in the time of completion (TOC) for the primary task for different interruption timings?
  - H_{01}: There is no difference in the TOC for the primary task for different interruption timings.
  - H_1: There is a difference in the TOC for the primary task for different interruption timings.

- Research Question II: Is there a difference in the performance of the secondary task for different interruption timings?
  - H_{02}: There is no difference in the secondary task performance for different interruption timings.
  - H_2: There is a difference in the secondary task performance for different interruption timings.

2.1 Data

The goal of the data collection is to characterize how different interruption strategies affect task performance in a team communication task. The overall data collection consists of 30 teams engaged in a collaborative task. In the data collection, a team is defined as two users participating in a communication task with one team member acting as an operator. If there is a team consisting of User1 and User2, User1 as the operator is one team and User2 as the operator is another team. This setup is used because the evaluation is done with respect to the operator, not the team. All teams are chosen randomly from a total of 23 participants. From the 23 participants, 2 are selected randomly to participate as human interrupters in addition to one person not participating in the collaborative communication task. In total 3 human interrupters were used in this data collection.

2.2 Tangram Task

The Tangram Task is a collaborative communication task where users communicate over a push-to-talk communication channel to rearrange Tangram shapes on their individual screens until the shapes in the corresponding color columns correspond correctly. One user has an interface similar to Figure 2. The other user has the exact same interface except the columns and objects in the columns are rearranged.

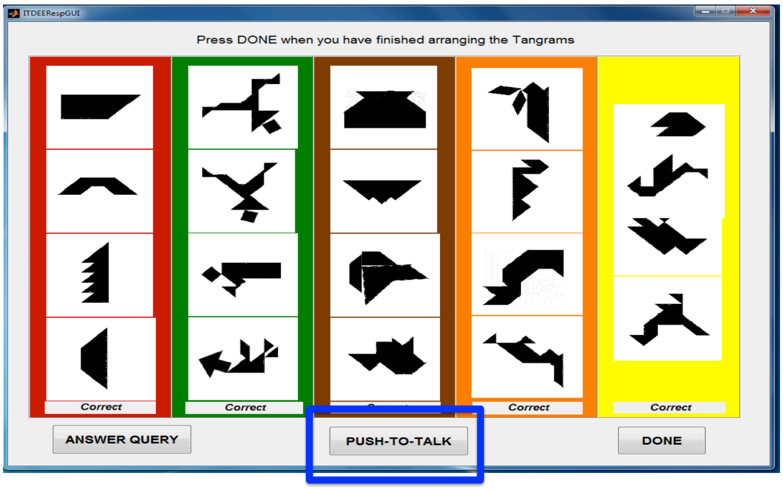

Fig. 2. Interface of Tangram Task

Two member teams are instructed to speak across a push-to-talk communication channel and collaborate with one another to arrange their Tangram shapes so that the shapes in the color column correctly correspond. Each team has an unlimited time to complete this task, but participants are instructed to complete the task as quickly as possible without jeopardizing task performance. Since one of the team members is the operator and simultaneously completing the keeping track task, there are three types of human-human communication interactions during this task between the team members:

- The non-operator describes the Tangram shapes and asks for feedback from their partner
- The operator describes the Tangram shapes and asks for feedback from their partner
- Both of the teammates describe the Tangram shapes and request feedback

This experimental task is inspired by the task described in [16] and used because it is a close simulation of the collaborative communication task described in the afore-mentioned UAV scenario.

2.3 Keeping Track Task

In the keeping track task, the operator is interrupted by the experimenter with updates about various objects and their attributes and then queried about the state of a particular object while simultaneously performing the Tangram task. Random, fixed, and human-determined (HD) interruptions (updates and queries) are communicated to the operator at various timings:

- **Fixed Interruptions** - interruptions (updates and queries) automatically disseminated every 9 seconds
- **Random Interruptions** - interruptions automatically disseminated randomly between 5 -13 seconds
- **Human Determined** - interruptions disseminated based on a human interrupter listening to the collaborative communication channel and determining based on the activity on the channel when to execute an interruption.

Human interrupters were instructed to send interruptions during times when they think their interruptions would be the least disruptive to the overall communication task. Human determined interruptions are being used to determine if an intelligent monitoring mechanism is necessary for interruptions.

Users must respond to queries via a push-to-talk communication mechanism. The operator was instructed to respond to the most recent state update of an object prior to a query. Multiple updates are sent to the operator followed by a query pertaining to information presented in the updates. The interruptions are presented as *blocks* of sizes 3-5 where 3, 4, or 5 updates were presented before a query. For example, an interruption block of size 3 presents 3 updates before a query. The query corresponds to information within the current block. The *lag* is how far a query is from its corresponding update. Below is an example of an Update/Query block:

{Update I}: Hawk-88' LOCATION is Point Bravo
{Update II}: Raven-3's FUEL-LEVEL is 30%
{Update III}: Falcom-11's ALTITUDE is 1900 ft.
{Fourth[Query I}: What is Raven-3's current FUEL-LEVEL?

This example illustrates an update/query block with a block size of 3 and a lag of 2 meaning there were 3 state updates before the user is queried and the attribute in question is two positions away from the query. All interruptions are communicated via a synthesized human voice. The keeping track task is motivated by [17].

3 Evaluation

The outcome from the data collection is used to determine if the operator performs the Tangram task faster when a human is listening in on the task and interrupting in comparison to automatically sent interruptions. The data is also evaluated to determine if the operator is able to perform better on the keeping track task when a human interrupts in comparison to fixed and randomly disseminated interruptions. Using the results from the data collection, we evaluate two metrics: the primary task time of completion (TOC) and the secondary task performance.

As described in Section 2.2, the primary task or the Tangram task required a two-member team to communicate and complete a collaborative task. Note that one of the participants in the team is the designated operator and simultaneously performing the keeping track task. Since we would like to see which interruption mechanism: random, fixed, or HD is least disruptive and results in a faster completion time over the primary task, we extract the TOC or how long it takes a given operator to complete the Tangram task with his partner. For each interruption-timing category, we average over all of the operators' TOC (in minutes) for the primary task.

Additionally, as described in Section 2.3, the secondary task or the keeping track task required the designated operator of a team to respond to queries when presented with a block of updates about UAV objects and their varying states. The experimenter annotated the results of this task. The percent correct served as the secondary task performance metric. The percent correct was calculated as the number of queries answered correctly divided by the total number of queries presented to a given operator. For each interruption-timing category, we average over all the operators' percent correct on the secondary task.

4 Results

A one-way analysis of variance (ANOVA) was conducted to evaluate the relationship between interruption timings and time of completion (TOC) on the primary task. The independent variable, interruption-timing factor included three levels: random, fixed, and human-determined (HD). The dependent variable was the average time it took the operators to complete the Tangram task (TOC) in minutes. The ANOVA was not significant at the 0.05 level, $F(2,174) = 0.96$, $p = 0.39$. The 95% confidence intervals for the differences, as well at the means and standard deviations for the three interruption timing groups are reported in Table 1.

Table 1. Primary Task TOC (minutes) Results

Interruptions	Mean	Std. Error	95% Confidence	
			Lower	Upper
Fixed	3.32	0.19	2.94	3.70
Random	3.45	0.19	3.07	3.83
HD	3.087	0.19	2.70	3.46

Although on average operators completed the primary task in less time when HD interruptions were administered, the TOC averages are close across interruption timings and there is no significant difference in the completion times. Here we can accept the H_{01} null hypothesis.

A one-way ANOVA was conducted to evaluate the relationship between interruption timings and user performance on the secondary task. The independent variable, the interruption-timing factor, included three levels: random, fixed, and HD. The dependent variable was the percent correct of queries operators answered correctly on the keeping track task. The ANOVA was significant at the 0.05 level, $F(2,174) = 7.52$, $p = 0.001$. The strength of the relationship between the interruption timings and the percent correct on the secondary task, as assessed by the η^2, was average, with the interruption-timing factor accounting for 8% of the variance of the dependent variable.

Table 2. Secondary Task Performance (Percent Correct) Results

Interruptions	Mean	Std. Error	95% Confidence	
			Lower	Upper
Fixed	33.47	4.05	25.47	41.47
Random	38.38	4.05	30.38	46.38
HD	54.70	4.05	46.70	62.70

Follow-up tests were conducted to evaluate pairwise differences among the means. The Levene's Test results in a significance of 0.45, supporting our assumption of homogenous variances among the three groups. We then conducted post hoc comparisons using the Tukey HSD, a test that assumes equal variances among the three groups. As illustrated in Table 3, there was a significant difference in the means between the HD interruptions and fixed timed interruptions as well as a significant difference in the HD interruptions and the randomly timed interruptions. The 95% confidence intervals for the pairwise differences, as well at the means and standard deviations for the three interruption timing groups are reported in Table 2 and 3.

Table 3. Task Performance (Percent Correct) Comparison across Interruptions

		Mean Diff.	Std. Error	Sig.
HD	Fixed	21.23	5.73	0.001
	Random	16.32	5.73	0.014

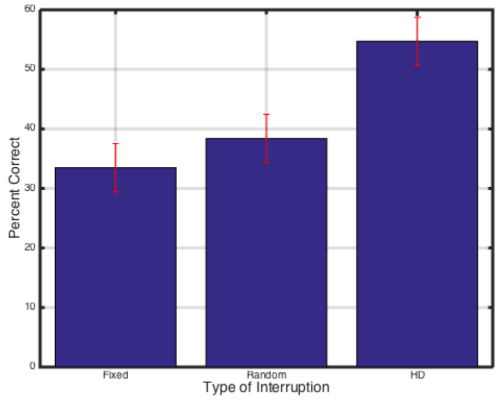

Fig. 3. Secondary Task Performance results

As summarized in Table 2 and 3 and Figure 3, the average accuracy for the primary task of all 30 teams is 54.7% for Human Determined interruptions, 38.4% for randomly timed interruptions, and 33.5% for fixed-timed interruptions. These results indicate that there is a significant difference in user performance when the operator is interrupted with HD interruptions. These results suggest that we can reject the H_{02} null hypothesis.

5 Discussion

The aforementioned results illustrate that monitoring the communication channel can improve task performance since the interruptions are probably being allocated as points that are less disruptive. This does not necessarily result in a lower time of completion for the primary task. This work did not directly evaluate the influence of interruptions on the TOC of a task [11], [6], [2], [9], [12], [3], but focused more on the how the TOC of a task was affected by various interruptions timings. From these results, various interruption-timing disseminations do not have an impact on the time of completion of a task.

The current results do however corroborate current literature that has shown that delivering interruptions as random times can result in a decline in performance on tasks [1], [6], [9], [8], [10]. The current work augments this claim by not only showing that randomly timed interruptions can be detrimental to task performance but in comparison HD interruptions are an improved mechanism of disseminating interruptions. There is no difference illustrated in this work between fixed-timed and randomly-times interruptions. Human-determined interruptions on the other hand significantly improve task performance over both of these alternative mechanisms.

In future work, we would like to use these results to justify the necessity of an intelligent interruption management system that can be used in collaborative, goal-oriented tasks. It was illustrated that the user task performance was degraded when fixed and random interruptions were sent. If monitoring the communication channel before interruptions indeed provide some improvement in the overall task performance within collaborative tasks, then there is justification for proposing an intelligent mechanism that also monitors a communication channel before interruptions. The next step in this work is developing a baseline model of what cues to monitor within a communication channel before an interruption. We could begin by exploring human interruptions (in this work stated as human determined interruptions) to determine if there are optimal human cues that could be integrated into an interruption management model. The long-term goal of this work is to design an intelligent interruption agent whose performance exceeds the performance of a fixed, random, and human interruption performance in collaborative, goal-oriented tasks.

6 Conclusion

In conclusion, in this work a collaborative communication and keeping track task to simulate a collaborative goal-oriented interaction between an operator and his human and machine teammates was developed. In evaluating the time of completion and performance on an experimental dual-task, the results suggest that although there is no significant differences in the effect interruption timings have on the time of completion of a task, the performance of the task in influenced by the time at which interruptions are disseminated. These results show that there could be some benefit to monitoring a task prior to sending interruptions in a collaborative communication task, that supports how well the task is performed.

The benefit of fixed timed and randomly timed interruptions is that they can be integrated into collaborative tasks between users and disseminate interruptions automatically. If the benefits of automatically sent interruptions could be merged with the benefits of a monitoring system, there is the potential for developing an intelligent interruption management system in collaborative communication task environments. Such a system would be beneficial in any interaction in which two or more users are collaborating to accomplish a common goal and there is peripheral information coming in that supports the current exchange or an additional task.

References

1. Brian P Bailey and Joseph A Konstan, “On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state,” *Computers in human behavior*, vol. 22, no. 4, pp. 685– 708, 2006.
2. Mary Czerwinski, Edward Cutrell, and Eric Horvitz, “Instant messaging: Effects of relevance and timing,” in *People and computers XIV: Proceedings of HCI*. British Computer Society, 2000, vol. 2, pp. 71–76.
3. Christopher A Monk, Deborah A Boehm-Davis, and J Gregory Trafton, “The attentional costs of interrupting task performance at various stages,” in *Proceedings of*

the Human Factors and Ergonomics Society Annual Meeting. SAGE Publications, 2002, vol. 46, pp. 1824– 1828.

4. D Scott McCrickard, Christa M Chewar, Jacob P Somervell, and Ali Ndiwalana, “A model for notification systems evaluation assessing user goals for multitasking activity,” *ACM Transactions on Computer- Human Interaction (TOCHI)*, vol. 10, no. 4, pp. 312– 338, 2003.
5. Daniel C McFarlane and Kara A Latorella, “The scope and importance of human interruption in human- computer interaction design,” *Human-Computer Inter- action*, vol. 17, no. 1, pp. 1–61, 2002.
6. Mary Czerwinski, Edward Cutrell, and Eric Horvitz, “Instant messaging and interruption: Influence of task type on performance,” in *OZCHI 2000 conference proceedings*, 2000, vol. 356, pp. 361–367.
7. Laura Dabbish and Robert E Kraut, “Controlling interruptions: awareness displays and social motivation for coordination,” in *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. ACM, 2004, pp. 182–191.
8. Kara A Latorella, *Investigating interruptions: Implications for flightdeck performance*, 1996.
9. John G Kreifeldt and ME McCarthy, “Interruption as a test of the user-computer interface,” in *JPL Proc. of the 17th Ann. Conf. on Manual Control p 655-667(SEE N 82-13665 04-54)*, 1981.
10. Joshua S Rubinstein, David E Meyer, and Jeffrey E Evans, “Executive control of cognitive processes in task switching.,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 27, no. 4, pp. 763, 2001.
11. Edward Cutrell, Mary Czerwinski, and Eric Horvitz, “Notification, disruption, and memory: Effects of messaging interruptions on memory and performance,” 2001.
12. Daniel McFarlane. "Coordinating the interruption of people in human-computer interaction." *Human-computer interaction, INTERACT*. Vol. 99. 1999.
13. Brian P Bailey and Shamsi T Iqbal, “Understanding changes in mental workload during execution of goal- directed tasks and its application for interruption management,” *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 14, no. 4, pp. 21, 2008.
14. Yoshiro Miyata and Donald A Norman, “Psychological issues in support of multiple activities,” *User centered system design: New perspectives on human-computer interaction*, pp. 265–284, 1986.
15. Tammar Shrot, Avi Rosenfeld, Jennifer Golbeck, and Sarit Kraus, “Crisp: an interruption management algorithm based on collaborative filtering,” in *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2014, pp. 3035–3044.
16. Clark, Herbert H., and Deanna Wilkes-Gibbs. “Referring as a collaborative process.” Cognition 22.1, 1986, 1-39
17. Venturincv, Michael. "Interference and information organization in keeping track of continually changing information." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39.4. 1997, 532-539.