Inferring Collaboration Strategies and Usability from Remote Observations in a Spaceflight Analog Environment

Jessica J. Marquez, Ph.D.
Shivang Shelat
Jimin Zheng
John A. Karasinski, Ph.D

Human-Computer Interaction Group
NASA Ames Research Center

Introduction

- It is difficult to remotely evaluate astronaut software interactions.
- Usual design practices in HCI must be coupled with naturalistic observations [1].
    - To be accurate, users must be “astronaut-like” and in an extreme, isolated environments.
- Voice logs in spaceflight analogs provide a method for unobtrusive data collection.
    - Useful for studying collaborative work with a software tool.
    - Helpful for ensuring usability in high fidelity environments.

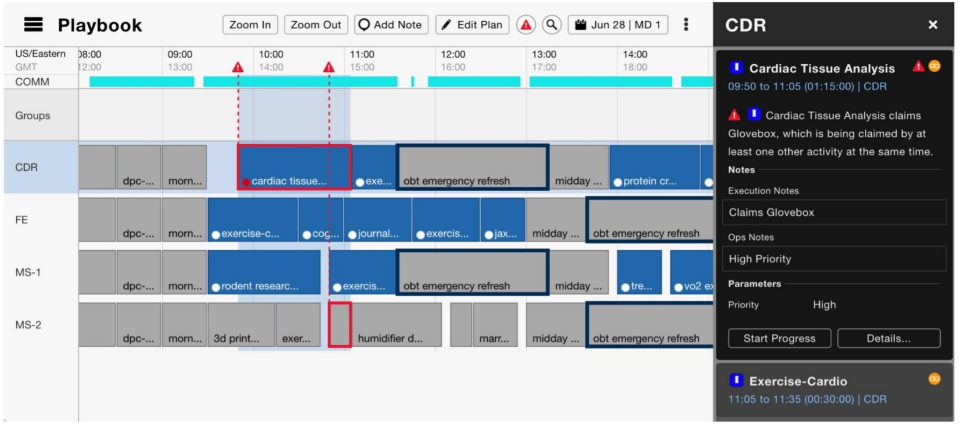
HERA Campaign 6 (C6)

- HERA is an isolation analog located at NASA Johnson Space Center (JSC) that simulates future long duration exploration missions.
    - For each of C6's four missions, four astronaut-like crew members live in the habitat for 45 days.
    - Crew only interacts with their family, friends, and mission control virtually.

An exterior view of the Human Exploration Research Analog (HERA) [2].

Present Work

- Studying conversations among crew members in HERA C6 to...
  - **Provide insights into crew collaboration** while using Playbook, a mission planning and scheduling software tool for astronauts to build their own timelines.



A screenshot of Playbook's user interface. Schedule violations are highlighted so the user can make adjustments until a timeline is feasible for the whole crew [3, 4].

What is Self-Scheduling?

Self-scheduling is the ability for an astronaut to autonomously manipulate their own spaceflight schedule.

Crew Procedure

Team Preference Meeting **(TPM)**	Self-Scheduling Session **(SS)**	Execution
An open discussion on timeline preferences, e.g. “lots of free time in the afternoon”, “hygiene periods in the morning”.	Independent scheduling by the assigned crew leader, ideally integrating team preferences into a feasible schedule. Then, this scheduler fills out a **workload** questionnaire.	Crew executes the self-scheduled timeline.

Audio Recordings

- In Campaign 6, all crew wore Philips Audio Recorders (DVT4010) throughout the day.
- Identified TPM and SS timeframes, locating and trimming relevant clips of audio.
- Audio was transcribed using Whisper, an automatic transcription software created by OpenAI [5].
    - Errors in transcripts were manually fixed.

Categorizing Crew Interactions

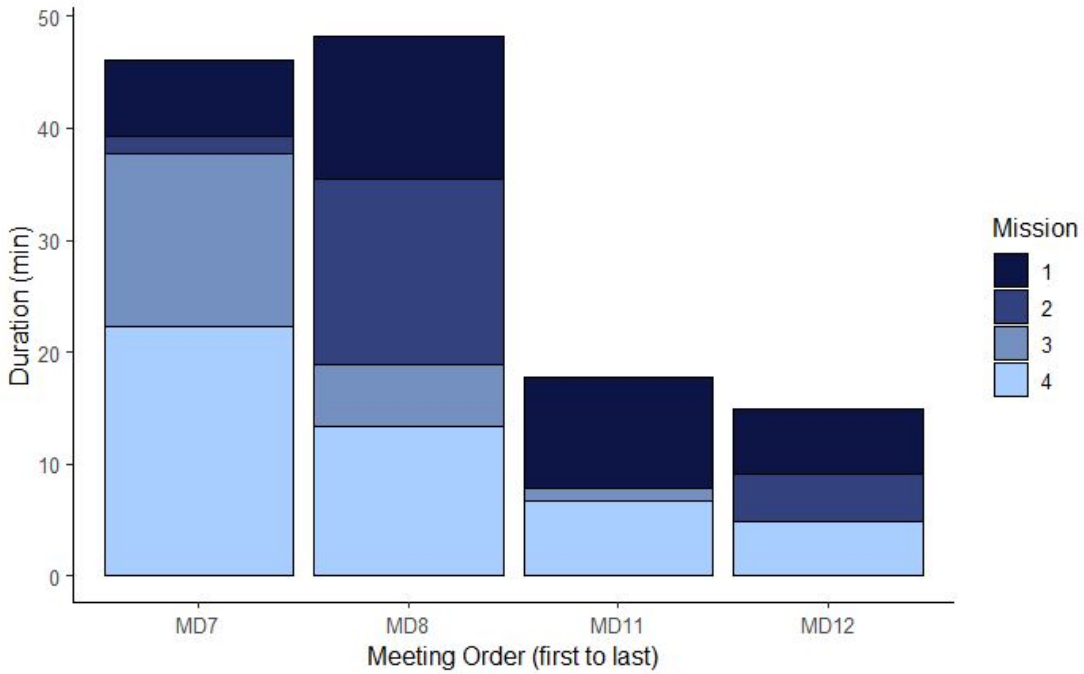
Label	Category	Description	Example
CRTimeline	Collaboration Regarding Timeline	Discussion on timeline content and preferences	*“I personally like the questionnaires stacked together... knock ’em all out.”*
CRTask	Collaboration Regarding Task	Discussion on the nature of the assigned task at hand (either TPM or SS)	*“We can talk about what our preferences are.”*
CRPlaybook	Collaboration Regarding Playbook	Discussion on how to use the tool or navigate the interface	*“But where do you see the tasks?” “They’re right there on add-to-plan.”*
OT	Off-Topic	Jokes, tangents, or unrelated topics	*“Have you watched the latest season of that show?"*

Counts of Interactions

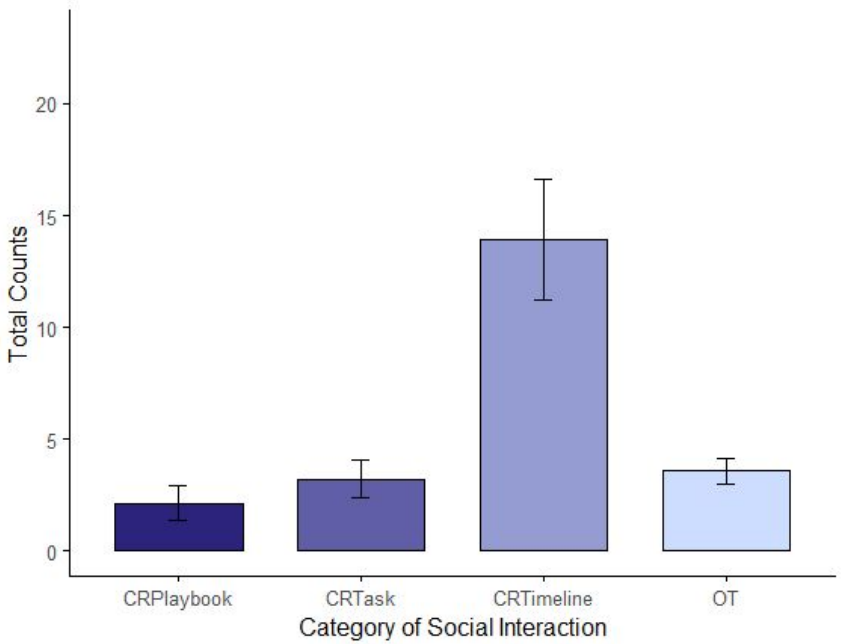
- Two independent raters counted crew interactions in the transcripts and tallied them according to their respective categories.
    - Interrater reliability was excellent, ICC = .91, 95% CI [.83, .95]
    - Counts were then averaged between raters.
- Our dependent variable of interest is **total counts** during a TPM or SS dedicated to a category of discussion.

Finding 1: TPM durations decrease

- Over the course of the mission, crews spend less time discussing preferences.
- Preferences might be mainly captured by 1 to 2 meetings.
- Conversation about preferences may have shifted to having them while completing self-scheduling activity.



Finding 2: TPMs are mostly on-task

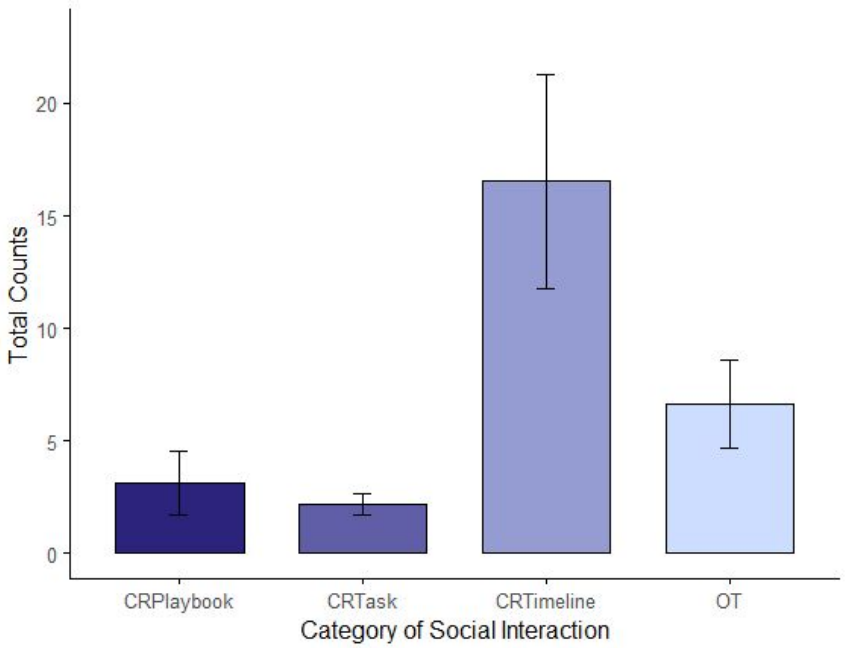


Counts during TPMs dedicated to different topics of discussion. Two TPMs were skipped and are not included in this plot (n = 14). Error bars show standard error.

- Lots of collaboration during TPMs revolved around **timeline preferences**, not Playbook features, the nature of the task at hand, or unrelated topics.
- This suggests that...
  - Task instructions are clear.
  - Crew members take advantage of the time to understand everyone's preferences before creating the team's schedule.

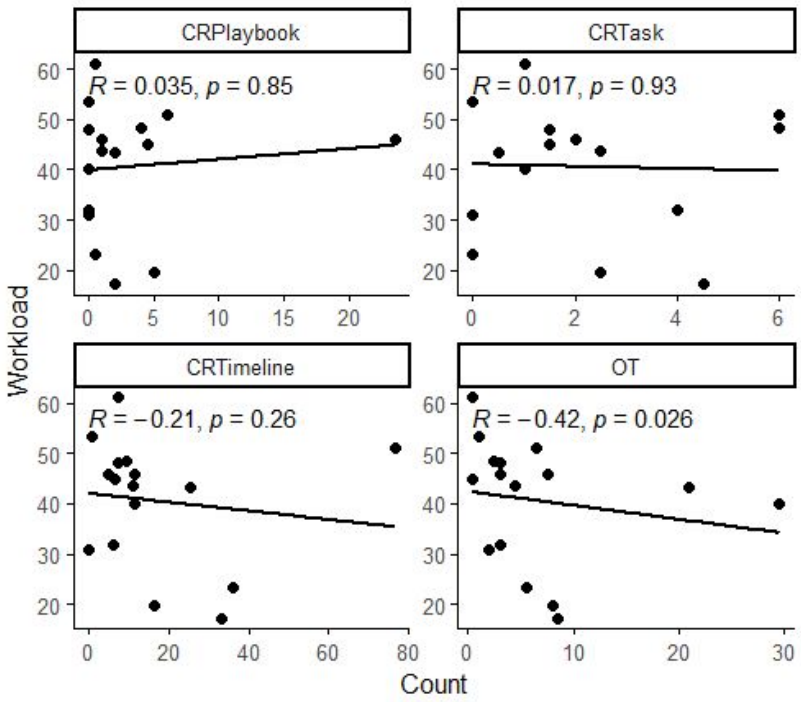
Finding 3: SS involves unexpected collaboration

- We asked each crew members to complete self-scheduling activity independently (i.e., this was not meant as a team activity).
- We did not expect crew to have conversations with each other during self-scheduling.
  - Surprisingly, they ***did***, mostly discussing timeline content and preferences.
- This behavior greatly varied across crews, as illustrated by the large standard error bar on CRTimeline.



Counts during SS dedicated to different topics of discussion (n = 16). Error bars show standard error.

Finding 4: Chatting during SS correlates with lower workload



Kendall's correlation tests between counts of interactions and weighted NASA-TLX scores during SS.

- Greater counts of off-topic conversation correlate with lower workload reports (NASA-TLX) directly after self-scheduling.
- This suggests that crews chat more when the self-scheduling task incurs low workload.
    - Off-topic chatter could be considered a secondary task.

Summary

- Recorded conversations provided, for the first time, an unobtrusive glimpse into behaviors during self-scheduling in a spaceflight analog environment.
- Conversation analysis did not identify any usability issues but did show that the self-scheduling task is more of a collaborative task than previously thought.
    - Crew collaboration happens during self-scheduling with Playbook.
    - Crew members communicate preferences during not just Team Preferences Meetings, but Self-Scheduling sessions too.
- Conversation analysis may lead to viable proxy measures of workload.
    - This is particularly valuable in a spaceflight analog environment where participant survey compliance may be challenging.

References

1. Clinkenbeard, M. (2018). Multimodal conversation analysis and usability studies: exploring human-technology interactions in multiparty contexts. *Communication Design Quarterly Review*, *6*(2), 103-113.
2. Brief, J. (2015, April 30). *HERA External View*. NASA. https://www.nasa.gov/image-feature/human-exploration-research-analog-hera-external-view
3. Marquez, J. J., Shelat, S., & Karasinski, J. (2022). Promoting Crew Autonomy in a Human Spaceflight Earth Analog Mission through Self-Scheduling. *ASCEND 2022*, 4263.
4. Marquez, J. J., Pyrzak, G., Hashemi, S., McMillin, K., & Medwid, J. (2013). Supporting real-time operations and execution through timeline and scheduling aids. In *43rd International conference on environmental systems* (p. 3519).
5. Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). Robust speech recognition via large-scale weak supervision. *arXiv preprint arXiv:2212.04356*.