

Off-Topic Chatter as an Unobtrusive Assessment of Workload

Katherine E. Homer¹, Shivang Shelat², John A. Karasinski¹, and
Jessica J. Marquez¹

¹NASA Ames Research Center, ²San Jose State University, Moffett Field, CA

Background

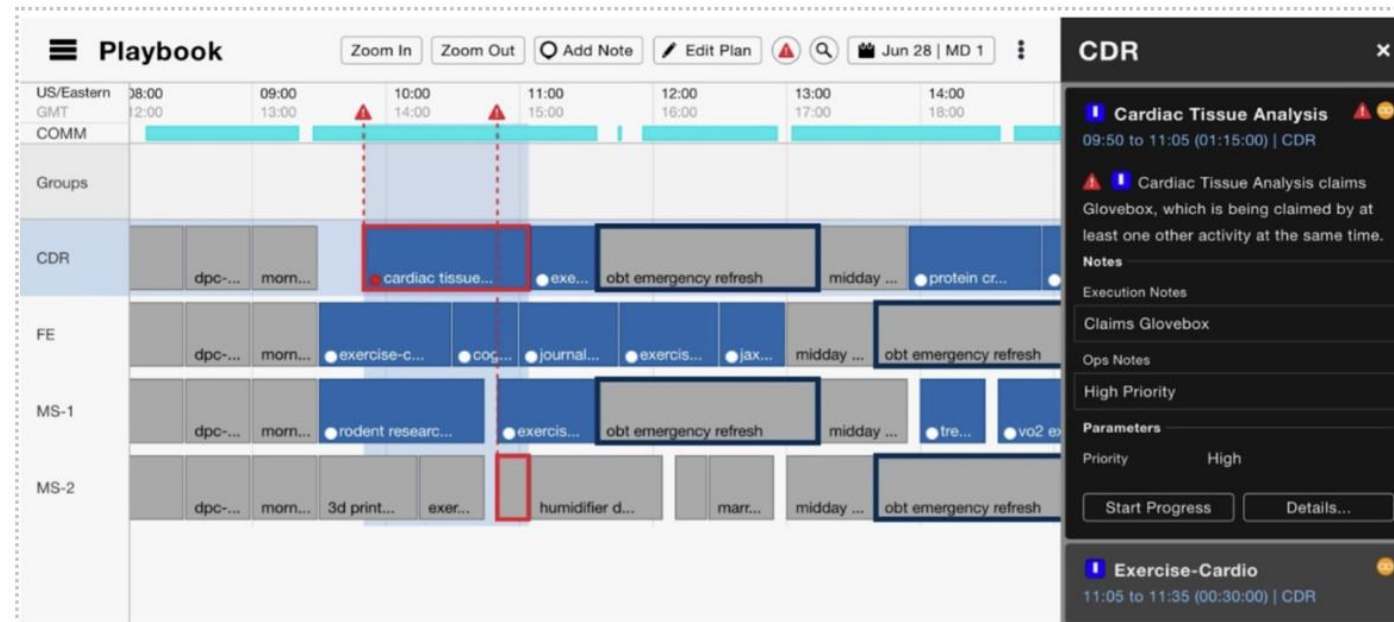
- Long Duration Exploration Missions (LDEMs) have significant new challenges due to their extreme environment, communication delays, and high safety risk.
- HCI analysis of space exploration technology must be conducted in representative environments to ensure usability is not impacted by certain factors.
- Analogs such as HERA and CHAPEA attempt to simulate aspects of LDEMs to provide more study opportunities but these continue to have their own limitations.
- Prior work by Shelat et al. suggested that off-topic chatter during collaborative scheduling is negatively associated with workload.

Workload

- Surveys – NASA-TLX and Bedford Scale
- Performance – primary task
- Performance – secondary task
- Physiological measurements – HR, EDA, respiration, eye-tracking etc.
- Speech-based vocal biomarkers – vocal acoustics and speech signals
- Conversation analysis – linguistic features

Playbook

Playbook is a mission planning and scheduling software tool which allows novice planners to organize activities and create timelines.



Screenshot of Playbook's user interface with schedule violations highlighted so the user can move activities until a timeline is viable for the whole crew.

Study Overview

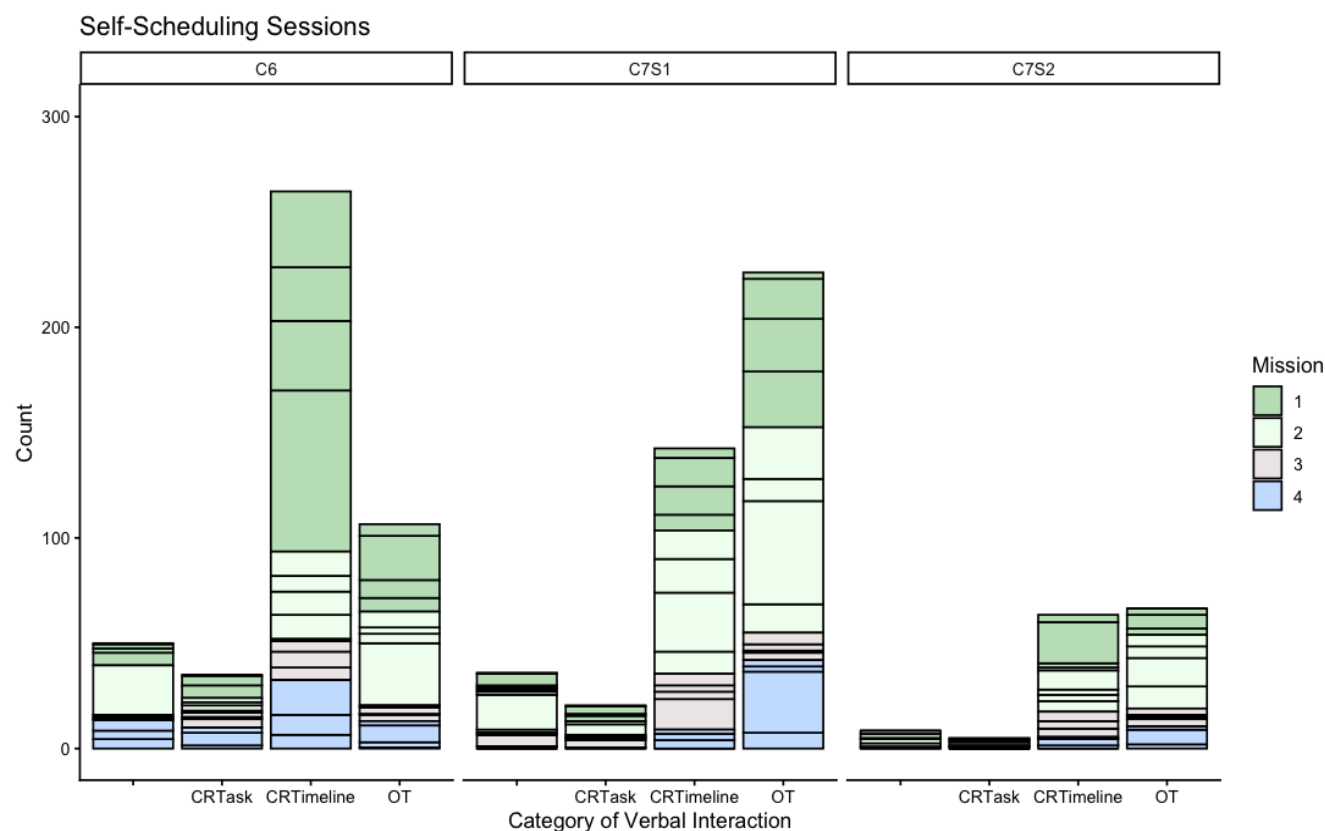
- Playbook was used as the scheduling software in NASA HERA Campaigns 6 and 7 (C6 and C7).
- Increasing crew autonomy phases
 - No autonomy – all timelines were fixed and could not be edited by the crew
 - Limited autonomy – the crew were allowed to reschedule many activities
 - High autonomy – crew self-scheduling was entirely initiated by the crew without external constraints.
- During the Limited autonomy phase, crew members were asked to complete self-scheduling sessions (SS) where one person, the planner, was responsible for scheduling one of the mission days.
 - 1 SS per person (4) in C6
 - 2 SS per person (8) in C7

Study Overview

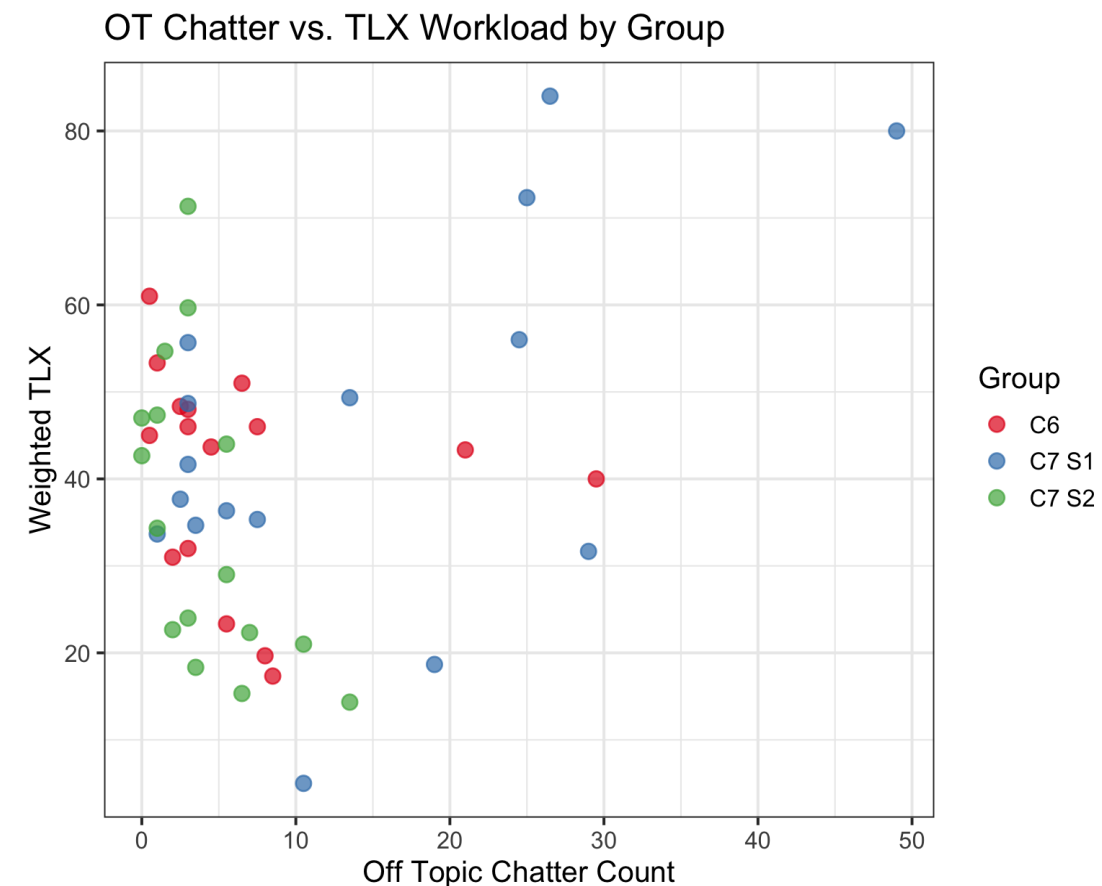
- Voice recordings of these SS were transcribed using Whisper and corrected with manual review.
- Two independent reviewers then categorized the crews' conversation into one of four categories.

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	"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 the add-to-plan."
OT	Off-Topic	Jokes, tangents, or unrelated topics	"Have you watched the latest season of that show?"

Results C6 and C7



Counts of observed conversation types during SS sessions in C6, C7 S1 and C7 S2 (n = 48 total).



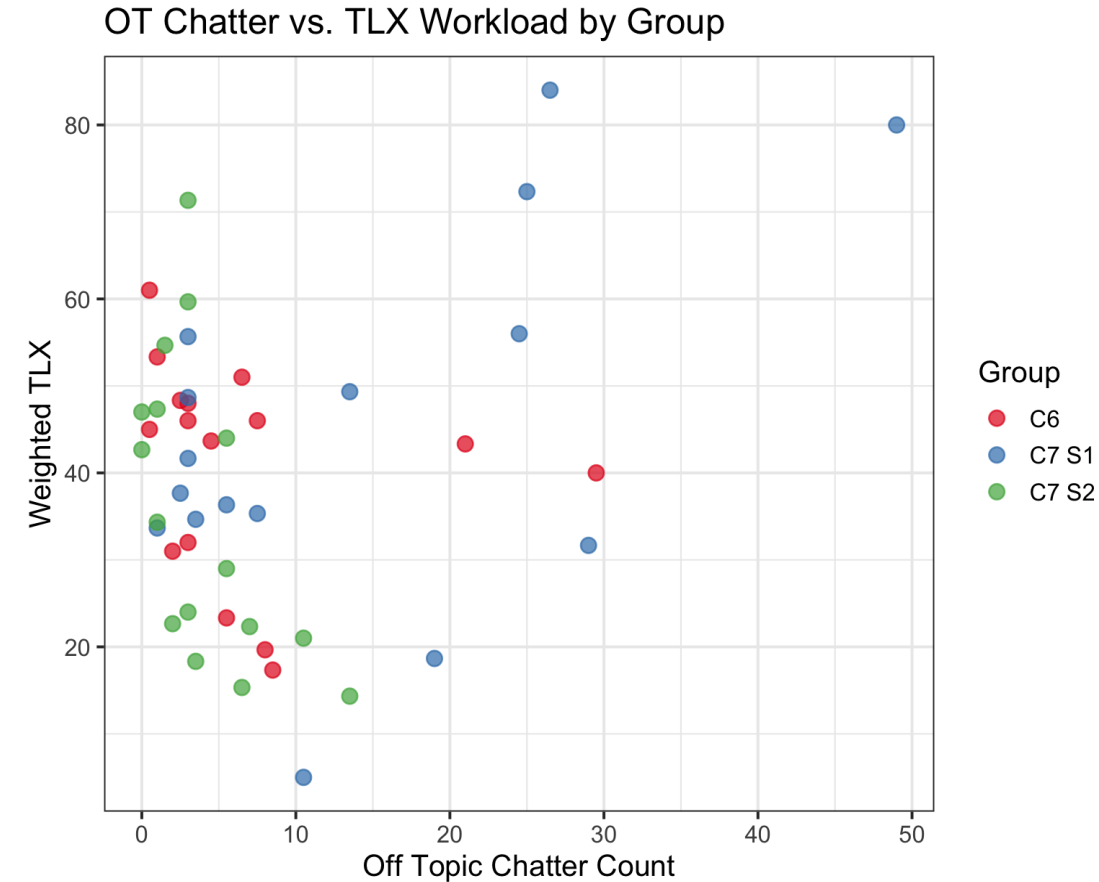
OT chatter count and workload for each SS in HERA C6 and C7.

Results - Outliers

- IQR $\rightarrow$ 6 outliers when $OT > 22.5$
- Rosner with $k = 8 \rightarrow$ 8 outliers
- Rosner with $k = 9 \rightarrow$ 8 outliers ✓
- Z-score $\rightarrow$ 7 outliers

Two outliers occurred in C6, five in C7.

Upon review, three outliers were removed resulting in $n = 16$ for C6, $n = 13$ for C7 S1, and $n = 16$ for C7 S2.



OT chatter count and workload for each SS in HERA C6 and C7.

Results - Statistics

Session	NASA-TLX Dimension	τ	p	p-fdr
C6	Weighted TLX	-0.42	0.027	-
	Mental Demand	-0.24	0.217	0.614
	Physical Demand	0.04	0.834	0.927
	Temporal Demand	-0.16	0.409	0.614
	Performance	0.02	0.927	0.927
	Effort	-0.32	0.093	0.559
	Frustration	-0.19	0.328	0.614
C7 S1	Weighted TLX	-0.09	0.667	-
	Mental Demand	-0.05	0.803	0.950
	Physical Demand	-0.25	0.289	0.868
	Temporal Demand	-0.08	0.709	0.950
	Performance	-0.03	0.900	0.950
	Effort	-0.24	0.264	0.868
	Frustration	0.01	0.950	0.950
C7 S2	Weighted TLX	-0.44	0.018	-
	Mental Demand	-0.13	0.493	0.493
	Physical Demand	-0.18	0.388	0.465
	Temporal Demand	-0.42	0.028	0.085
	Performance	-0.29	0.156	0.234
	Effort	-0.38	0.045	0.090
	Frustration	-0.56	0.004	0.023

Kendall's rank correlations were conducted for each of the C6, C7 S1, and C7 S2 sessions separately.

A significant negative monotonic relationship was found between workload and OT chatter in the SS C6 and C7 S2 sessions.

Prior to FDR correction, a weak monotonic correlation was found between effort and OT-chatter in C6. In C7 S2 significant negative correlations were found between OT chatter and the workload dimensions of temporal demand, effort, and frustration.

After correction, the only correlation to remain significant was that between frustration and OT chatter in C7 S2.

Discussion

- In both HERA C6 and HERA C7 S2, increased OT chatter was monotonically correlated with decreased workload. This did not hold for HERA C7 S1, indicating this could be an inconsistent or temporally dynamic correlation.
- While not consistent, the finding from C7 S2 further supports the previous analysis conducted by Shelat et al. which first identified the relationship between workload and OT conversation during scheduling in C6.
- In C7 S2, temporal demand, frustration, and effort all had negative monotonic correlations with OT chatter prior to FDR correction.
- Definitive conclusions on which dimensions of workload drive the relationship are not yet possible because of low power caused by the large number of tests run and the small sample sizes typical of analog research.

Discussion - Causality

Future work should focus on the causal direction of the relationship between workload and OT chatter:

1. Decreased workload leads to an increase in OT chatter
2. Increased OT chatter decreases feelings of workload
3. A third variable (e.g., task difficulty, crew mood, or time of day) drives both decreased workload and OT chatter

We speculate Option 1 is the most likely because prior research indicates that verbal communication during complex tasks generally decreases task performance and increases workload.

Casner and Schooler found that when pilots utilized automation, their workload decreased and they reported a higher percentage of task-unrelated thoughts, or mind-wandering. This supports Option 1: since low workload increases off-task thoughts, low workload also likely increases OT chatter.

Discussion - Limitations

- Audio recordings were cut to approximately the correct start and end time of the SS, but the process was manual and reviewers observed that there were some transcripts with OT conversation prior to the start of the SS and/or after the SS had ended.
- There were SS in which the crew was multitasking (e.g., simultaneously completing other surveys or experiments) while the planner developed the schedule.
- Our transcripts did not differentiate between crew members so any communication between crew, even if it did not include the planner, was incorporated into our analysis.
- Despite utilizing all conversation, only the planner of the SS filled out the NASA-TLX.

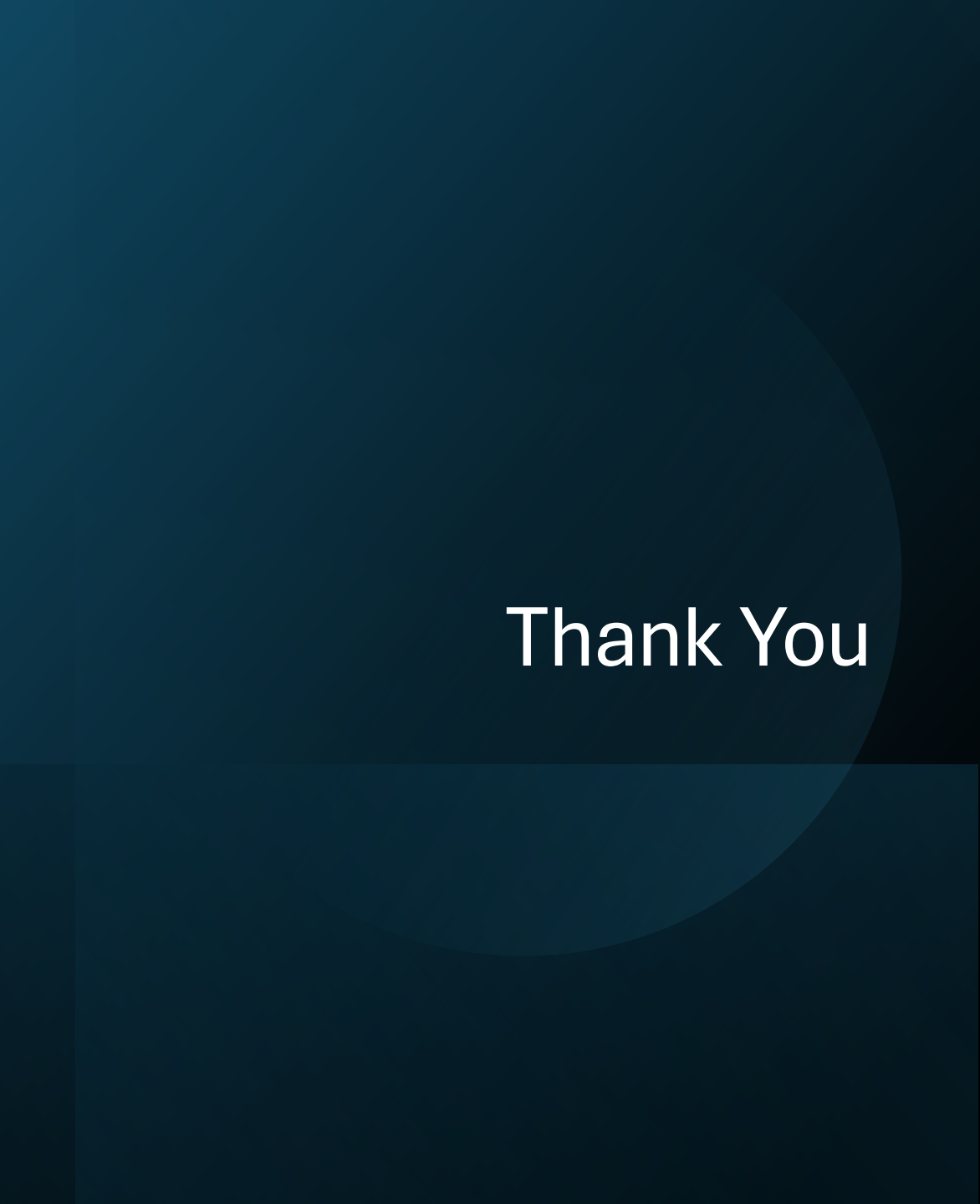
Conclusions

This research supports a likely negative correlation between OT chatter and workload.

As such, tracking OT chatter could be a valuable tool to unobtrusively monitor workload levels of astronauts during collaborative HCI in real time.

This method:

- Requires no additional or cumbersome hardware
- Does not distract astronauts from their tasks - which may impact performance or safety
- Does not require additional crew time to complete surveys – which impacts overall mission efficiency
- Has the benefit of assessing workload in team environments during collaborative HCI tasks



Thank You

Acknowledgments: This study was funded by NASA (WBS #44494.02.01.16.01).

The authors have no competing interests to declare that are relevant to the content of this study.