Formative Usability Assessment of an Augmented Reality Head-Mounted Display for Astronaut Extravehicular Activities

Annlyle J. Diokno, Benjamin Rubin, Yining Zhang, Israel Cantu, Cameron Huang, Titan Chen, Xiaoxuan "Alicia" Cheng, Mert Culcu, Hecun "Hailey" Liu, Sumodha Pokhrel, Sahitha Vuddagiri, Alexandra Xu, Yining "Elena" Zhang, Jing Chen, & Raudel Avila

PROJECT OVERVIEW

Extravehicular activities (EVAs) are vulnerable to cognitive overload, error, and task-switching costs. Tasks include:

Ingress and Egress: Entering and exiting the lunar lander safely

Geological Sampling: Scanning and documenting lunar samples

Navigation: Traversing unfamiliar, low-light terrain

These tasks are performed in **low-light, unfamiliar terrain.**

Astronauts must complete the tasks within **tight timeframes** due to limited consumable resources, such as oxygen and battery life.

AR Usability Considerations:

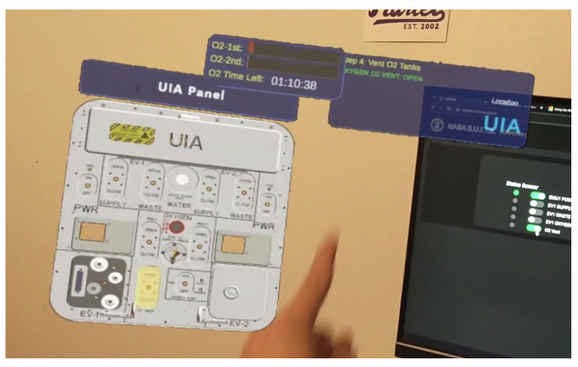
Iterative, goal-oriented usability testing is essential to ensure task efficiency, safety, and user satisfaction (Weiss, 2024).

Standard questionnaires like the SUS should be combined with **observational analyses** and **post-test inquiries** to uncover usability issues (Weiss, 2024).

OBJECTIVES

To conduct a **formative usability assessment** to iteratively refine an **AR interface** that supports astronauts through **EVA operations.**

VISUAL ENVIRONMENT



METHODOLOGY

A **four-phase** approach:

1 PREPARATION

- Obtained IRB approval to ensure the highest ethical standards

2 DESIGN & FORMATIVE EVALUATION

- Hierarchical task analysis (HTA)
- Wireframe (Version 1)
- Formative Evaluation Methods:
  1) Cognitive Walkthrough
  2) User Failure Modes & Effects Analysis (uFMEA)

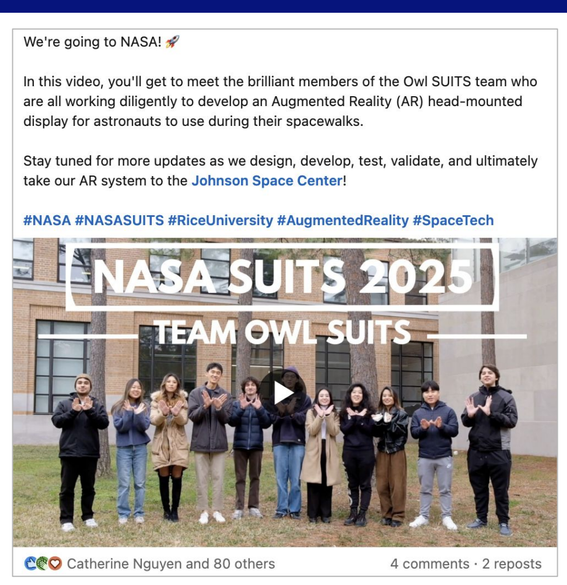
3 ITERATIVE DESIGN

- Refined the prototype by addressing the identified user errors from the formative evaluation (Version 2)
- Heuristic Evaluation on Version 2

4 USABILITY TEST ($N = 5$)

- ***Qualitative:*** Think-aloud, post-test inquiry
- ***Quantitative*** (per ISO 9241-11, 1998):
  - Efficiency: Task time
  - Effectiveness: NASA-TLX (Hart & Staveland, 1988)
  - Satisfaction: Modified SUS with ARS (Bangor et al., 2008, 2009)

OUTREACH



LinkedIn Post

Our post has generated 81 likes and 4,312 impressions on LinkedIn. According to the post analytics, it's received a total of 1,474 views and over 6 hours of watch time.

RESULTS

Qualitative Results (Themes):

1) Visibility and readability challenges

"My vision is so bad in the dark."

"It's kind of hard to see the paper [mock UIA]."

2) Poor system responsiveness

"It won't let me click."

"Is there an easier way to tap it?"

3) Appreciation for color-coded cues

"I like that it highlighted what to press."

"Instead of having to search for small text, I just look for the color."

Quantitative Results:

- *Efficiency*: Task times ranged between 250 to 449 seconds
- *Effectiveness*: NASA-TLX scores ranged from 27 to 61
- *User Satisfaction*: SUS scores ranged from 47.5 to 72.5 and ARS scores ranged from 4 to 6

Metric	Task Time (s)	NASA-TLX	SUS	ARS
Mean (*SD*)	320.6 (*75.4*)	40 (*13.4*)	58 (*11.0*)	5.2 (*.8*)

CONCLUSION

- Users found color-coded cues helpful, but poor light visibility hindered readability
- Unresponsive tapping and poor feedback disrupted task flow and caused confusion
- Low satisfaction and observed usability issues signal a need for design refinement

References

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, *24*(6), 574–594. https://doi.org/10.1080/10447310802205776

Bangor, A., Kortum, P., & Miller, J. A. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies, 4*(3), 114–123.

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology. 3*. 77-101. https://doi.org/10.1191/1478088706qp063oa

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 139–183. https://doi.org/10.1016/s0166-4115(08)62386-9

ISO. (1998). 9241-11. *Ergonomic requirements for office work with visual display terminals (VDTs)—Part 11: Guidance on usability*. Geneva: The International Organization for Standardization.

Weiss, H. (2024). *Evaluation of Augmented Reality and Wearable Sensors to Assess Neurovestibular and Sensorimotor Performance in Astronauts for Extravehicular Activity Readiness* (Doctoral dissertation)