Fiscal Year:	FY 2022	**Task Last Updated:**	FY 01/05/2022
PI Name:	Selva, Daniel Ph.D.		
Project Title:	HCAAM VNSCOR: Virtual Assistant for Spacecraft Anomaly Treatment During Long Duration Exploration Missions		
Division Name:	Human Research		
Program/Discipline:			
Program/Discipline--Element/Subdiscipline:			
Joint Agency Name:		**TechPort:**	Yes
Human Research Program Elements:	(1) **HFBP**:Human Factors & Behavioral Performance (IRP Rev H)		
Human Research Program Risks:	(1) **HSIA**:Risk of Adverse Outcomes Due to Inadequate Human Systems Integration Architecture		
Space Biology Element:	None		
Space Biology Cross-Element Discipline:	None		
Space Biology Special Category:	None		
PI Email:	dselva@tamu.edu	**Fax:**	FY
PI Organization Type:	UNIVERSITY	**Phone:**	607-255-6351
Organization Name:	Texas A&M University		
PI Address 1:	Aerospace Engineering Department		
PI Address 2:	701 Ross St 3141 TAMU		
PI Web Page:			
City:	College Station	**State:**	TX
Zip Code:	77843-0001	**Congressional District:**	17
Comments:			
Project Type:	GROUND	**Solicitation / Funding Source:**	2017-2018 HERO 80JSC017N0001-BPBA Topics in Biological, Physiological, and Behavioral Adaptations to Spaceflight. Appendix C
Start Date:	03/06/2019	**End Date:**	03/05/2023
No. of Post Docs:	1	**No. of PhD Degrees:**	0
No. of PhD Candidates:	2	**No. of Master' Degrees:**	0
No. of Master's Candidates:	0	**No. of Bachelor's Degrees:**	0
No. of Bachelor's Candidates:	2	**Monitoring Center:**	NASA JSC
Contact Monitor:	Whitmire, Alexandra	**Contact Phone:**	
Contact Email:	alexandra.m.whitmire@nasa.gov		
Flight Program:			
Flight Assignment:			
Key Personnel Changes/Previous PI:			
COI Name (Institution):	Diaz Artiles, Ana Ph.D. (Texas A&M Engineering Experiment Station) Dunbar, Bonnie Ph.D. (Texas A&M Engineering Experiment Station) Wong, Raymond Ka Wai Ph.D. (Texas A & M, College Station)		
Grant/Contract No.:	80NSSC19K0656		
Performance Goal No.:			
Performance Goal Text:			

Task Description:	This task is part of the Human Capabilities Assessments for Autonomous Missions (HCAAM) Virtual NASA Specialized Center of Research (VNSCOR). The research objective of this proposal is to investigate the impact of using Virtual Assistants (VA) to support crewmembers in the context of anomaly treatment during Long Duration Exploration Missions (LDEM), when ground support will be limited. A VA will be developed building upon the software architecture from existing VAs developed by the Principal Investigator (PI) for similar purposes. The VA will provide support for various aspects of anomaly treatment, including detecting and diagnosing the anomaly, as well as recommending a course of action. It will also have the ability to take initiative in the dialog with the user (mixed-initiative mode), and the ability to provide explanations for its actions. The impact of the VA on performance, cognitive workload, situational awareness, and trust, will be assessed through a set of three experiments with human subjects in a laboratory environment. The first experiment will establish the baseline impact (master-slave, no explanations), and subsequent experiments will study the effect of switching to the mixed-initiative mode and adding explanations. The system will also be deployed and tested in the Human Exploration Research Analog (HERA) analog environment.
Rationale for HRP Directed Research:	
Research Impact/Earth Benefits:	This project will provide standards and guidelines that will help NASA design similar virtual assistants to support astronauts during future long duration exploration missions. Such standards and guidelines will concern both the functionality and the user interface of the virtual assistant.
Task Progress:	In Year 3 of this project, we have completed the first lab experiment, thus achieving Specific Aim 1, pending publication of the journal paper with the result. A total of 12 participants (mean age $\pm$ SD = 24.8 $\pm$ 1.9 years) were selected from an astronaut-like population. The participants were largely comprised of students at Texas A&M University, who had either already completed or were currently pursuing a graduate degree in a science, technology, engineering, mathematics (STEM) field. In addition, most participants had previous experience working with technical procedures. Each participant completed two experiment sessions: a control session without Daphne and a treatment session with Daphne. During both conditions, subjects were tasked with detecting, diagnosing, and resolving Environmental Control and Life Support Subsystem (ECLSS) anomaly scenarios. We selected two different anomaly scenario groups (5 anomaly scenarios per group) based on the perceived difficulty of resolving that particular anomaly and completing its corresponding resolution procedure. We counterbalanced participants with respect to which anomaly group they received first and in which experiment session they could use Daphne. We measured performance metrics by evaluating the number of anomalies that participants could successfully resolve with and without the use of Daphne-AT. Additional performance metrics included the time it took for each subject to solve each anomaly, whether or not the subject detected the anomaly, diagnosed the root cause, and selected the correct anomaly resolution procedure. We also recorded the number of attempts for the latter three metrics. In addition, we measured cognitive workload by employing the NASA Task Load Index (TLX) survey, and situational awareness using the Situational Awareness Rating Technique (SART). Finally, the participants' trust in autonomous systems was measured using Jian's trust scale. The summary of the results is as follows. For performance, a significant difference (p=0.002) was found in the number of anomalies correctly resolved with Daphne (4.1$\pm$0.3) vs. without Daphne (2.7$\pm$0.4). This supports our hypothesis that a VA can improve performance in anomaly resolution. For cognitive workload, a significant difference (p=0.002) was found in the TLX score with Daphne (45.53$\pm$6.67) vs. without Daphne (65.97$\pm$4.81). This supports our hypothesis that a VA can reduce cognitive workload in anomaly resolution. For situational awareness, a marginally significant difference (p=0.052) was found in the SART score with Daphne (5.19$\pm$0.56) vs. without Daphne (4.41$\pm$0.71). This does not fully support our hypothesis that a VA can improve situational awareness in anomaly resolution. Finally, the responses to the Jian scale items showed generally high trust in automation. A much more detailed description of the experiment and the results is in the paper about to be submitted to the Human Factors and Ergonomics Society (HFES) journal. The draft paper will be provided to the NASA Human Research Program (HRP) as soon as it is submitted. This year we also completed the first HERA mission Campaign 6, Mission XXII (C6M1). Campaign 6 / C6M2 will also be complete by the time the period of performance of this report is exhausted, which gets us about halfway in meeting Specific Aim 3. Our HERA study was divided into two phases. In Phase 1, each crewmember performed two experiment sessions per week. In each session, the crewmember worked on a single anomaly scenario, either with Daphne or without. Thus, at the end of Phase 1, each subject had worked individually on 3 anomalies with Daphne and 3 without. Because of a technical issue with Daphne, one of the crewmembers ended up with only 2 anomalies with Daphne and 4 without Daphne. In Phase 2, crew members worked together as a team. They conducted two sessions per week, and worked on one anomaly scenario per session. This resulted in a total of 3 anomalies with Daphne and 3 without. Preliminary results are as follows. Concerning performance, all scenarios were successfully resolved and therefore this was not an interesting metric. We are currently looking at the time to diagnosis of the anomaly to see if there are any interesting trends there. Trends appear to indicate that in some cases the crew took more time to solve scenarios with Daphne than without, especially in Phase 2, but we attribute this to some logon issues we encountered towards the end of the mission. Our perception is that the diagnosis was very fast both with and without Daphne, suggesting that the anomaly scenarios were "too easy". The subjects did spend considerable time conducting the procedures to resolve the anomalies, but that was not the focus of our study. Concerning cognitive workload, we observed a small difference between the TLX scores with Daphne (23.42$\pm$5.17) and without Daphne (25.20$\pm$3.54) but this difference was not significant with only 4 subjects (p=0.42). It remains to be seen if the effect becomes significant with the full 16 subjects. It is also interesting to note that these scores are very low compared to the scores obtained in the laboratory experiments. This could be due to the shorter training and less experience of the participants of the lab experiments compared to the HERA crew. This is also consistent with the explanation that the scenarios were too "easy" for the HERA crew. Concerning situational awareness, we observed the opposite trend compared to the lab experiments, although in both cases the effects are not significant. In this case, SART scores were slightly lower with Daphne (20.46$\pm$3.35) than without Daphne (21.21$\pm$3.50), for a p-value of 0.41. We note that the effects were not consistent among crewmembers, and the overall result is driven by one of the 4 crewmembers, who reported a significantly higher level of situational

awareness without Daphne vs. with Daphne. We are looking in more detail into the components of the SART score (understanding, demand, supply) because the trends appear to be conflicting there too. While the literature has reported cases where using some kind of assistance results in a decrease of situational awareness (e.g., autopilots in commercial aircraft), this may be due to other factors (e.g., minor technical issues with the VA) as well. More research is needed to understand this – we hope that missions 2,3,4 will help us elucidate this issue. We also note that crewmembers mentioned in the exit debrief interviews that adding explanations would likely significantly increase situational awareness with Daphne, and this is something that we plan on testing in our lab experiments 2 and 3.

Concerning trust, the results were similar to the ones from the lab experiments, showing high trust in the VA across the board. In the exit interviews, crewmembers indicated that they trusted Daphne "right away", after the first few of her recommendations were proven to be correct.

Some of the insights we got from the exit interviews are as follows. The crew generally exhibited strong interest in using VAs for anomaly resolution, and enjoyed using Daphne. Most of them showed an interest in the social aspects of VA and mentioned that they "attributed a personality to Daphne" and "talked about her as if she were another crewmember". They all mentioned establishing trust very quickly, once and for all, thanks to Daphne "getting it right the first 3 times or so". Almost none of them found the question answering capabilities essential because they were "going for speed" and didn't feel like they needed to ask any questions. However, all of them mentioned that question answering would be very useful in cases where Daphne recommended more than one diagnosis with the same confidence level. Moreover, crewmembers expressed an interest in the more interactive diagnosis and advanced explanations capabilities we are currently developing as something that would "significantly increase the usefulness" of the tool. Finally, they all confirmed that the scenarios generally felt very easy to diagnose and adding some more complexity would make it more interesting and fun.

As for next steps, in Year 4 we will support the remaining missions of C6, thus finalizing Specific Aim 3. We plan on publishing a journal paper with the results. Otherwise, most of the activity will focus on achieving Specific Aim 2: we will finalize the development of the enhanced VA capabilities (explanations and mixed initiative) and conduct lab experiments 2 and 3.

Bibliography Type:	Description: (Last Updated: 02/23/2024)
Abstracts for Journals and Proceedings	Selva D, Josan P, Dutta P, Abbott R, Viros i Martin A, York K, Dunbar B, Wong RKW, Diaz Artiles A. "Virtual assistant for anomaly resolution In long duration exploration missions: Baseline effects on performance, cognitive workload, and situational awareness." 2022 NASA Human Research Program Investigators' Workshop, Virtual. February 7-10, 2022. Abstracts. 2022 NASA Human Research Program Investigators' Workshop, Virtual. February 7-10, 2022 (Abstract #1133-000353). , Feb-2022
Abstracts for Journals and Proceedings	Josan P, Dutta P, Abbott R, Viros A, York K, Dunbar B, Wong R, Selva D, Diaz-Artiles A. "Results from first laboratory testing of virtual assistant Daphne-AT. " 2022 NASA Human Research Program Investigators' Workshop, Virtual. February 7-10, 2022. Abstracts. 2022 NASA Human Research Program Investigators' Workshop, Virtual. February 7-10, 2022 (Abstract #1133-000351). , Feb-2022
Abstracts for Journals and Proceedings	Dutta P, Kaur PK, Viros A, York K, Abbott R, Dunbar B, Wong R, Diaz-Artiles A, Selva D. "Virtual assistants for anomaly treatment – lessons learned and path forward." 2022 NASA Human Research Program Investigators' Workshop, Virtual. February 7-10, 2022. Abstracts. 2022 NASA Human Research Program Investigators' Workshop, Virtual. February 7-10, 2022 (Abstract #1133-000410). , Feb-2022