Fiscal Year:	FY 2020	**Task Last Updated:**	FY 03/25/2020
PI Name:	Lee, John Ph.D.		
Project Title:	HCAAM VNSCOR: Conversation Analysis to Measure and Manage Trust in Virtual Assistants		
Division Name:	Human Research		
Program/Discipline:			
Program/Discipline--Element/Subdiscipline:			
Joint Agency Name:		**TechPort:**	No
Human Research Program Elements:	(1) **HFBP**:Human Factors & Behavioral Performance (IRP Rev H)		
Human Research Program Risks:	(1) **HSIA**:Risk of Adverse Outcome Due to Inadequate Human Systems Integration Architecture (IRP Rev L)		
Space Biology Element:	None		
Space Biology Cross-Element Discipline:	None		
Space Biology Special Category:	None		
PI Email:	jdlee@engr.wisc.edu	**Fax:**	FY
PI Organization Type:	UNIVERSITY	**Phone:**	608-890-3168
Organization Name:	University of Wisconsin, Madison		
PI Address 1:	Department of Industrial and Systems Engineering		
PI Address 2:	1513 University Ave		
PI Web Page:			
City:	Madison	**State:**	WI
Zip Code:	53706-1539	**Congressional District:**	2
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:	04/15/2019	**End Date:**	04/14/2023
No. of Post Docs:		**No. of PhD Degrees:**	
No. of PhD Candidates:	1	**No. of Master' Degrees:**	
No. of Master's Candidates:		**No. of Bachelor's Degrees:**	
No. of Bachelor's Candidates:		**Monitoring Center:**	NASA JSC
Contact Monitor:	Williams, Thomas	**Contact Phone:**	281-483-8773
Contact Email:	thomas.j.will1@nasa.gov		
Flight Program:			
Flight Assignment:	NOTE: End date changed per S. Huppman/HRP and NSSC information (Ed., 3/20/2020) NOTE: End date changed to 3/31/2020 per NSSC information (Ed., 1/22/2020)		
Key Personnel Changes/Previous PI:			
COI Name (Institution):	Cross, Ernest Ph.D. (NASA Johnson Space Center) McGuire, Kerry Ph.D. (NASA Johnson Space Center)		
Grant/Contract No.:	80NSSC19K0654		
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 goal of this research is to develop conversation analysis to measure and mitigate inappropriate trust in virtual assistants. These trust measurements will guide system design, particularly the multimodal interactions and mode switching, as well as how to mitigate over trust and trust recovery. We will use conversation analysis to measure trust at multiple time-scales from real-time interactions to longitudinal monitoring of trust over a long duration exploration mission.

Conversation analysis provides a promising, but relatively unexplored approach to measuring trust. We propose a conversation analysis at the micro, meso, and macro levels which includes not just the words, but also pauses and facial expressions. Specifically, at the micro-level, conversation elements include voice inflections, pauses between works and keystrokes, gaze shifts, and facial expressions. The meso-level analysis includes words exchanged during interactions with the virtual assistant along with other team interactions as they relate to the automation. At the macro level, conversational analysis considers interaction time, interaction effort, frequency of interaction, turn-taking, barging in tendency, and whether it is the person or the virtual assistant who initiates the interaction. Additionally, prior research into conversational analysis indicates there are novel ways of managing or calibrating trust through the presentation of information, e.g., manipulating the tone and cadence of the system when using speech and through facial expressions (Nass & Brave, 2005; DeSteno et al., 2012).

Due to time delays in communication, long duration exploration missions will require greater crew autonomy and greater reliance on automation. For this approach to work trust calibration needs to be engineered into the system. Trust is a critical construct that mediates how well human operators use automated systems, such as virtual assistants, that provide decision support. Trust affects people's willingness to rely on automated systems in situations that have a degree of uncertainty and risk. Trust strongly affects the effectiveness of human-agent collaboration, particularly in the willingness to accept suggestions from a virtual assistant. Knowing whether or not to trust automation can be further complicated by lack of sleep, workload, task risk, and task complexity. Moreover, as we continue to push the limits of intelligent systems and rely on them more as decision aids trust calibration (i.e., operator trust is at a level which matches the automation's capabilities) becomes essential to mission execution.

Appropriate calibration of trust requires matching the operator's trust to the virtual assistant's current capabilities. Calibration of trust is not something that can happen once, but must occur throughout the life cycle of the interaction between operator and automated system (Hoffman et al., 2009). Trust is a dynamic construct that continuously increases and decreases due to a number of factors, primary the performance of the automated system, i.e., higher performance leads to higher trust and vice versa. Although much effort focuses on creating more capable and trustworthy automation, less effort has considered the equally important consideration of creating trustable automation. Trustable automation is automation that is understandable and that naturally promotes calibrated trust. Therefore, we aim to create trustable automation by continuously measuring operators' trust unobtrusively and in real-time, and then use this measure to guide the virtual agent to employ one or more countermeasures to calibrate trust and improve human-system performance.

References

DeSteno D, Breazeal C, Frank RH, Pizarro D, Baumann J, Dickens L, Lee JJ. Detecting the trustworthiness of novel partners in economic exchange. Psychol Sci. 2012 Dec;23(12):1549-56. http://doi.org/ ; PubMed PMID: 23129062

Hoffman RR, Lee JD, Woods DD, Shadbolt N, Miller J, Bradshaw JM. The dynamics of trust in cyberdomains. IEEE Intelligent Systems. 2009 Nov-Dec;24(6):5-11. https://

Nass C, Brave S. Wired for Speech : How Voice Activates and Advances the Human-Computer Relationship. Cambridge, MA: MIT Press, 2005.

Rationale for HRP Directed Research:

Research Impact/Earth Benefits:

The outcomes of this research will make two important contributions to the overall HCAAM VNSCOR effort. First, it will promote more effective interactions and acceptance of virtual assistants. Second, it will provide new analytic techniques for understanding how people work with automated agents as team members.
Virtual assistants and other types agents enabled by artificial intelligence represent an important opportunity to extend human capabilities, but only if they are accepted and trusted appropriately. If people trust the virtual assistant too much they will rely on it in situations that exceed its capability, and if they trust it too little they will fail to engage it when it could benefit the team. One pathway towards appropriate trust is to make the virtual assistant more trustworthy: increase its technical capabilities to accommodate any situation. Another approach it to make it more trustable: communicate its capability and allow its capability to be challenged in its interactions with people. Such trustable technology requires three important advances to the state of knowledge in the field:

1. An ability to ascertain how much people currently trust the technology

2. An ability to convey uncertainty and its capability, particularly as part of conversational interactions

3. Interaction affordances that provide the opening for people to assess the capability of the assistant, particularly as part of conversational interactions.

These three advances for trustable technology require the development of new analytic techniques for understanding human interaction with automated teammates. Real-time, unobtrusive measures of trust represent a particularly valuable, but challenging measure to develop. Trust is most often measured with ratings and indirectly through people’s decision to rely on automation, which are obtrusive not diagnostic. Conversation and text-based interactions offer a promising, but unexplored way to assess trust. Text analysis has a 50-year history in domains as diverse as psycholinguistics and cognitive science, and more recently natural language processing, affective state assessment, and sentiment analysis. Building on the foundation of text analysis makes it possible for this research to immediately contribute to data analysis of previous and future studies of automation-human teaming, and to contribute to the foundation of conversational agent design.

Task Progress:	In this phase of the project, we identified and validated nine conversational indicators of trust for implementation in other VNSCOR efforts. Three indicators for each measurement level are identified. We completed IRB (Institutional Review Board) and received approval from NASA IRB and are awaiting review from Wisconsin IRB and finalize SRD (science requirements document). Other goals accomplished: acquire Total Organic Carbon Analyze (TOCA) hardware device for use in Human Exploration Research Analog (HERA); develop pre and post survey questions; complete study design for HERA study; complete eye tracking specification. For the micro-transactions that occur during the interaction with a virtual agent, we would extract vocal and physiological features from the voice-based conversation, gaze, heart rate, facial expression, and Galvanic Skin Response (GSR). Measurement #1 Speech Trust Recognition and #2 Valence/Arousal Recognition consider paralinguistic or non-verbal information, which includes the sound spectrum of the speech, apart from the actual speech content. For #1, the acoustic feature extraction task is conducted based on spectral features (e.g., Mel-frequency cepstral coefficients (MFCCs)) extracted from speakers' voices using Librosa library in Python 3.7.3. A Multilayer Perceptron (MLP) is then used to train four discrete emotions associated with trust (i.e., happy, calm, angry, and fearful). For #2, prosodic, spectral, and glottal waveforms are extracted. We adopt a three-layer model incorporating Adaptive neuro fuzzy inference systems (ANFIS) to get continuous emotion dimensions (i.e., valence and arousal) classification to identify trust indicators. For #3, we translate all physiological measures (i.e., Galvanic skin Conductance; gaze; heart rate; facial expression) extracted into sequences of letters for the subsequent data analysis by using the Symbolic Approximation (SAX). For the meso-transactions, as stated in the proposal, we would extract verbal features to show lexical indicators of trust, which include words spoken or typed by the crew members. We combine three trust features techniques (i.e., lexical analysis, topic modeling, and word embedding) with three predictive models (i.e., percent score, F&F decision tree, and lasso and xgboost) to form measurement #4, #5, & #6. We have identified a trust lexicon from a comprehensive review of trust scales. The words used in these trust scales complement the words that have previously been used to define the trust lexicon. The macro-level of indicators will be analyzed through conversation turning-taking (e.g., cooperative vs. competitive overlap) as the coordination measurement (#7). For Measurement #8, we would measure the reliance and compliance behaviors, including frequency and duration of communications with the intelligent agent and the rate of adopting its recommendations for troubleshooting. Finally, questionnaires (#9) have been identified to measure subjective trust based on a comprehensive literature review of trust scales. Through text analysis, we have explored the similarities and differences between existing 38 trust scales with 488 items. The words comprising the scales were coded with GloVe and then computed the embedding for each item and each scale. We will use the Uniform Manifold Approximation and Projection (UMAP), a visualization of the dimension-reduced embeddings. This semantic space provides an understanding of how to operationalize trust and the guidelines for the selection of the trust scale. A composition of trust scales based on domains (dispositional, history-based, and situational) and categories (automation, E-commerce, human-human) is identified, which can used as guidelines for choosing the appropriate trust surveys for this study. NASA IRB Approval did not occur until December. This delayed the team's access to the previous study data that was to be used for initial training of the trust algorithm. The team continued to develop core functionality such as the ability to translate speech to text.
Bibliography Type:	Description: (Last Updated:)