Concept of Operations Evaluation for Using Remote-Guidance Ultrasound for Exploration Spaceflight

Victor W. Hurst IV*; Sean Peterson; Kathleen Garcia; Douglas Ebert; David Ham; David Amponsah; Scott Dulchavsky

BACKGROUND: Remote-guidance (RG) techniques aboard the International Space Station (ISS) have enabled astronauts to collect diagnostic-level ultrasound (US) images. Exploration-class missions will likely require nonformally trained sonographers to operate with greater autonomy given longer communication delays (> 6 s for missions beyond the Moon) and blackouts. Training requirements for autonomous collection of US images by non-US experts are being determined.

METHODS: Novice US operators were randomly assigned to one of three groups to collect standardized US images while drawing expertise from A) RG only, B) a computer training tool only, or C) both RG and a computer training tool. Images were assessed for quality and examination duration. All operators were given a 10-min standardized generic training session in US scanning. The imaging task included: 1) bone fracture assessment in a phantom and 2) Focused Assessment with Sonography in Trauma (FAST) examination in a healthy volunteer. A human factors questionnaire was also completed.

RESULTS: Mean time for group B during FAST was shorter (20.4 vs. 22.7 min) than time for the other groups. Image quality scoring was lower than in groups A or C, but all groups produced images of acceptable diagnostic quality.

DISCUSSION: RG produces US images of higher quality than those produced with only computer-based instruction. Extended communication delays in exploration missions will eliminate the option of real-time guidance, thus requiring autonomous operation. The computer program used appears effective and could be a model for future digital US expertise banks. Terrestrially, it also provides adequate self-training and mentoring mechanisms.

KEYWORDS: spaceflight, astronauts, ultrasound, telemedicine, remote guidance.

Hurst VW IV, Peterson S, Garcia K, Ebert D, Ham D, Amponsah D, Dulchavsky S. *Concept of operations evaluation for using remote-guidance ultrasound for exploration spaceflight*. Aerosp Med Hum Perform. 2015; 86(12):1034–1038.

Numerous publications on ultrasound (US) imaging in actual and analogue spaceflight environments report reliance on remote guidance (RG) of minimally trained caregivers or research operators with no professional expertise in US.[1,2,4] High-quality US images from multiple organ systems (e.g., cardiovascular, respiratory, musculoskeletal, genitourinary) have been obtained consistently, attesting to the general maturity of the remotely guided diagnostic US capability in orbital spaceflight and in remote terrestrial settings with communication latency of up to 2 s.[7,8,14] However, increased latency in data flow and voice communication associated with greater distances is expected to impair the effectiveness of these techniques. The actual communication delay that would render RG ineffective has not been objectively established. The two-way communication delay on missions to the Moon is expected to reach ~5-6 s and the delay on missions to Mars is expected to last 10 to 45 min.[3,5,13] A simulated Mars mission with a 15-min communication delay demonstrated successful use of US by a general physician with remote assistance to diagnose appendicitis before surgery.[13] Anecdotal reports, however, are not sufficient to influence the medical requirements for exploration missions of the future.

To assess the effects of intermediate delays on a remotely guided collection of US images by nonexperts, a simulated exploration mission outpost was set up with the communication delay set at 5 s (a lunar mission scenario with additional satellite uplinks and ground segments). For longer delays without RG

*Posthumously.

From Crew Health and Research, Wyle Science, Technology and Engineering, Houston, TX, the Canadian Space Agency, Saint-Hubert, Quebec, Canada, and the Department of Surgery, Henry Ford Hospital System, Detroit, MI

This manuscript was received for review in November 2011. It was accepted for publication in August 2015.

Address correspondence to: Scott Dulchavsky, M.D., Ph.D., Department of Surgery, Henry Ford Hospital, 2799 W. Grand Blvd., Detroit, MI 48202; sdulcha1@hfhs.org.

DOI: 10.3357/AMHP.3244.2015

possibility, the same tasks were undertaken with autonomous US imaging. A subset of operators in the 5-s scenario, and all operators in the absence of RG, were assisted by a computer-based learning tool called Onboard Proficiency Enhancer-Light (OPE-L). The operators learned how to use this tool immediately before they used it in the study. This type of training, in which an operator learns how to use a specific tool or execute a specific skill immediately before applying that tool or skill, is called "just-in-time" training. The hypothesis was that the OPE-L would have a positive impact on the user's ability to collect useful US images and possibly replace RG for certain medical scenarios.

METHODS

Subjects

The procedures described in this report were reviewed and approved by the NASA Johnson Space Center Institutional Review Board. Selected for this study were 30 operator subjects with no US training or with lifetime US imaging exposure not exceeding 2 h (12 men, 18 women, ages 32–54). Seven subjects had more than 2 yr of medical education.

Software

OPE-L is a menu-driven presentation that has multimedia instructions for all steps leading to a successful capture of each target US image. The original OPE was used aboard the International Space Station (ISS) to assist astronauts in their collection of US images. It covered US basics, understanding RG language, relevant anatomy, specific imaging procedures, and imaging tips and pitfalls.[1,3,4] OPE-L contains only the target US images and an illustrative video of each of the imaging tasks [a fracture assessment of a damaged limb and a Focused Assessment with Sonography in Trauma (FAST) abdomen assessment (**Fig. 1**) used in this study].

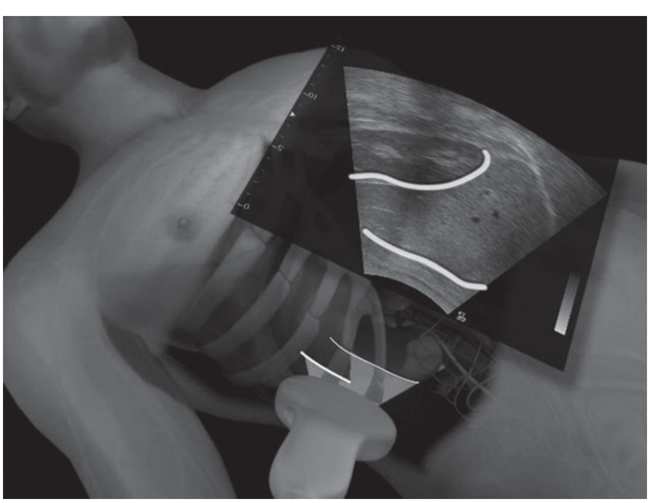

Fig. 1. OPE-L for FAST abdomen assessment. The image being collected is shown in the upper right of the video view, and the probe placement is shown in the remaining space of the view.

Procedure and Design

The subjects were randomly assigned to 3 groups of 10 subjects each for the entirety of the experiment:

- Remote Guidance (Group A). Subjects were guided by an expert (radiologist or emergency physician with US certification and experience in remote medical discourse), with a 5-s round-trip communication delay in both the audio and video. The expert received a video link from the US machine.
- Autonomous operation with OPE-L (Group B). Subjects used OPE-L to review the procedures and target images before and at any time during the task. Use of the OPE-L included, in part, the subject viewing a ~2-min video.
- Remote Guidance with OPE-L (Group C). Subjects were remotely guided in addition to using OPE-L.

All groups had a US cue card affixed to the edge of the US screen for reference purposes. The card includes a layout of the US keyboard to help the subject locate the features and settings on the device, an anatomical layout of the human body with locations specific for US probe placement, and images detailing US probe manipulation techniques (pan, rotate, tilt). A similar cue card is used for current US operations aboard the ISS.

Each subject underwent a standardized 10-min training session immediately before performing the imaging tasks. Each subject was blinded to their group assignment before their training session and none received any special focus with respect to the tools they were using during their session (for example, it was not emphasized to subjects of Group B that they needed to pay particular attention to OPE-L). The training included explanation of the tasks, basic probe and equipment manipulation techniques, familiarization with the cue card and the OPE-L computer tool, and use of RG language for this task.

The subjects then completed both experimental tasks in one session. The first task was to assess an extremity phantom (two limbs) for bone fracture and capture a total of four images of intact and fractured segments. Subjects were also asked to indicate their level of diagnostic confidence. The second task was to perform a FAST abdominal protocol on a human male subject screened to fit the demographics of the astronaut population and to collect four target images for later interpretation. Completion times were recorded for both tasks. Upon finishing both tasks, each subject completed a 22-question questionnaire to capture their perception of training effectiveness, the equipment cue card, the OPE-L computer tool, and the RG. The questionnaire also assessed the subjects' perceived level of difficulty and frustration in completing each experimental task.

Data Analysis

Image quality was rated by the same FAST-certified emergency medicine physician who was blinded to the group assignment (A, B, or C). Each of the four target images collected for each of the tasks was rated on a scale of 0 to 100%. The ratings for the four views were averaged to formulate the mean rating and then converted to a 1–4 scale (**Table I**). A ratio of the image quality (percentage) to the completion time was referred to as either

Table I. Image Quality Rating Scale for Fracture and FAST Images.

RATING	FRACTURE CRITERIA	FAST CRITERIA
4	Clearly shows fracture	Clearly shows hepatorenal interface, splenorenal interface, diaphragm, pericardial space, and perivesicular structures
3	Shows some images of the fracture but they are not clear	Shows some but not all of the above structures; shows part of the hepatorenal interface, splenorenal interface, or pericardial space; shows bladder, but not perivesicular structures
2	Shows very limited images of the fracture; hard to visualize	Shows very limited views of the above structures; not adequate for evaluation of hepatorenal interface, splenorenal interface, pericardial space; bladder only minimally visualized
1	Inadequate images; unable to interpret	Inadequate images; unable to interpret

"fracture index" or "FAST index" for the respective tasks. These indices represented image quality adjusted for the corresponding completion time.

Analysis of variance (ANOVA) with Tukey's honestly significant difference test was used to determine statistical significance of differences in the three group means for the following variables: completion time, image quality, fracture and FAST indices, difficulty index, and frustration index. This test controlled for the multiple comparisons and provided a pairwise comparison of groups. In addition, *t*-tests were used to compare the performance of subjects who had received greater than 2 yr of medical training with those who had not.

RESULTS

The area of injury for the fracture portion of the study was correctly identified by 100% of the subjects in Group C and by 90% of the subjects in both Groups A and B. The assistance provided by the RG to Groups A and C involved only anatomical positioning and not real-time confirmation of the fracture from the imagery. The self-reported confidence levels of the subjects in making a correct diagnosis of the fracture from the images collected were 94.9%, 92.5%, and 94.3% for Groups A, B, and C, respectively [$F(2,27) = 0.24$; $P = 0.79$].

Task completion times for the fracture assessment were significantly longer for Group C than for Group A (**Table II**). The differences between Group B and the other two groups were not significant. The quality of the fracture images collected was not significantly different between the three groups. Significant differences were identified between Group A and the other two groups with respect to fracture index (image quality adjusted for completion time) (**Table III**). In addition, the fracture index for Group B was not significantly different from the index for Group C. The mean task completion time for the FAST assessment was similar in all groups, but there was a trend toward quicker completion time for Group B (**Table IV**).

All groups collected FAST images of acceptable diagnostic quality; however, there was a trend with Group B having the lowest image quality and Group C having the highest. The mean FAST index (image quality adjusted for completion time) was similar for all groups [Group A: 3.7, Group B: 3.7, Group C: 4.2; $F(2,27) = 0.25$; $P = 0.78$]. Subgroup analysis revealed that the image collection performance of those with previous medical training (defined as greater than 2 yr of medical school) was significantly different from the performance of other subjects, as demonstrated by image quality [mean difference = 0.7 (0.2 – 1.3), $P = 0.01$] and FAST index [mean difference = 1.8 (0.5 – 3.0), $P = 0.01$]. Questionnaire data revealed a trend that Group B found image collection to be more difficult than Group A [$F(2,27) = 3.1$; $P = 0.06$]. The level of frustration with completing the tasks was similar for the two tasks of the experiment [$F(2,27) = 0.92$; $P = 0.41$].

DISCUSSION

Exploration-class missions to the Moon, Mars, and near-Earth asteroids are a realistic expectation that will likely materialize in the foreseeable future. The medical risk profiles and tolerance for such missions are being defined and will determine the complexity of the diagnostic and treatment capabilities.[3,6] US imaging technology will likely be a dual-purpose (research and clinical) resource of the future spaceflight health care system, similar to the current ISS configuration. This notion is supported by the many previous successes in the effective use of US in spaceflight.[3,11,15] Multiple studies have shown that minimally trained astronauts and cosmonauts aboard the ISS can be

Table II. Mean Task Completion Times for the Fracture Assessment.

GROUP	MEAN COMPLETION TIME (min)	95% CONFIDENCE INTERVAL
Group A – RG Only	7.7*	6.6-8.8
Group B – Autonomous + OPE-L	9.6	8.2-11.1
Group C – RG + OPE-L	10.7*	9.4-12.0

RG: remote guidance; OPE-L: Onboard Proficiency Enhancer-Light.
* Statistically significant difference between group means, as identified by Tukey post hoc comparisons ($P < 0.01$).

Table III. Mean Fracture Index.

GROUP	MEAN FRACTURE INDEX	95% CONFIDENCE INTERVAL
Group A – RG Only	13.5*,‡	11.5-15.5
Group B – Autonomous + OPE-L	10.6*	9.1-12.1
Group C – RG + OPE-L	9.6‡	8.3-11.0

RG: remote guidance; OPE-L: Onboard Proficiency Enhancer-Light.
* Indicates statistically significant differences between group means, as identified by Tukey post hoc comparisons ($P < 0.01$).
‡ Indicates statistically significant differences between group means, as identified by Tukey post hoc comparisons ($P < 0.01$).

Table IV. Mean Task Completion Times for the FAST Assessment.

GROUP	MEAN COMPLETION TIME (min)	95% CONFIDENCE INTERVAL
Group A – RG Only	22.7	18.9-26.4
Group B – Autonomous + OPE-L	20.4	16.6-24.2
Group C – RG + OPE-L	22.7	18.1-27.3

RG: remote guidance; OPE-L: Onboard Proficiency Enhancer-Light.

remotely guided by ground-based sonographers to collect high-quality US images of practically all feasible anatomical targets.[3,4,16] The transfer of this capability to exploration-class missions will require transition from near-real-time RG from the ground to delivery of appropriate expertise from an onboard computer-based expertise bank. The experiment reported here generally demonstrates the feasibility of such on-site information transfers.

This study was an initial assessment to answer the question of whether a longer communication delay can affect remotely guided collection of US images by non-US experts with either limited or no medical training. Choosing the 5-s communication delay experienced during lunar missions, the investigators demonstrated that this did not affect RG effectiveness seen in previous studies.[1,2,7] The image quality for the fracture assessment was at a diagnostic level for subjects receiving only RG (Group A). The image quality remained at this level when the OPE-L was added to the RG paradigm (Group C). It is understood that the difference in time between these two groups to collect their high-quality images was significant (7.7 min for Group A vs. 10.7 min for Group C); however, this difference can be explained, in part, by the actual amount of time needed to view the OPE-L video (~2 min) plus the time needed to listen to and execute supplementary instructions from the RG. Viewing of the OPE-L video may also explain the ~2-min difference in image collection times between Groups A and B; the latter group had to watch the OPE-L video before executing the task. Overall, the time differences can be considered negligible and would not have a serious effect in a medical contingency. Autonomous use of the OPE-L, that is, in the absence of RG (Group B), resulted in image quality comparable to that of the two groups that were given RG. This finding, combined with the fact that autonomous use of the OPE-L did not significantly increase the task completion time, practically proves the concept of using carefully organized information from onboard stores versus receiving information through RG. Taken together, the data support the investigators' hypothesis that use of the OPE-L or similar computer-based learning tool can replace RG in certain scenarios.

The mean task completion times for the FAST assessments were not significantly different between the three groups, nor were the mean image quality levels. It is understood that the mean image quality levels from the FAST portion of the study were lower than those in fracture assessment, yet of acceptable diagnostic quality; it must be noted that the complexity and nature of the two tasks is different and such comparison may easily be challenged.

The subjects with previous medical training produced FAST images of significantly greater quality than nonmedically trained subjects. This cohort was able to do so despite not having substantial US experience. This predictable advantage can be explained by their knowledge of anatomy in general and familiarity with cross-sectional representation of anatomy, and possibly by a more efficient RG discourse since the experts were not blinded to the professional background of the subjects. This finding is complemented by the report of Otto et al., who assessed the impact of a Mars mission-relevant communication delay of 15 min on RG techniques for diagnosing and treating a simulated case of appendicitis.[13] The research team used a caregiver who had a broad medical background but limited expertise in US and surgical procedures. In the simulated Mars research environment at Devon Island, the team demonstrated that this level of caregiver was able to be successfully guided by remote experts through US imaging procedures to provide diagnostic-quality images. Taken together, these findings suggest that a physician crewmember would be a more successful US operator regardless of US experience level and they lend support to the current NASA requirement to have a physician for lunar and planetary missions of durations greater than 210 d.[12] For missions not requiring a physician ($<$ 210 d), but still involving significant communication delays, further research is needed to determine the appropriate level of training to enable nonmedical personnel to perform diagnostic US and US-guided procedures.[17,18]

The difficulty and frustration levels associated with the tasks were similar for all three groups. Although differences were not significant, OPE-L tended to reduce the levels of perceived difficulty and frustration, suggesting that optimizing the OPE-L and allowing subjects to gain experience with computer-based procedure guidance might further improve confidence and performance. Alternatively, other modes of information delivery could be used in certain scenarios. The "virtual guidance" system developed by Martin et al. uses commercially available video glasses and audio/video guidance to collect diagnostically adequate US imagery for a given protocol.[10]

The present study demonstrates that increasing the communication delay to 5 s did not affect an RG expert's ability to guide non-US experts to collect diagnostic-quality US images. However, further increases in communication delay could impair the effectiveness of RG, especially in complex procedures requiring more than mere identification of standard and easy-to-recognize target images.

The most important result of this experiment is the success at task performance in the group of subjects operating in full autonomy and drawing necessary directions from the OPE-L computer tool. This finding shows that the potential in future spaceflight programs of adaptive multimedia tools for supporting US protocols of all levels of complexity warrants more attention, and so does their potential in limited-resource terrestrial settings.

Because of the limited scope of the experiment, the training for all subjects in the study occurred immediately before the experiment. Such timing is different from just-in-time

US training experience on the ISS,[1,2,9] yet it is possible in case of a medical emergency if appropriate computer-based resources are available. Any degradation in proficiency over time was not simulated in this study. The use of artificial limb phantoms for the fracture portion of the study substantially simplifies the task compared to actual trauma scenarios. Also, it is understood that application of the results involving RG are limited to LEO, lunar, and possibly lunar Lagrangian point missions and are not applicable to a Mars expedition with much longer communication delays. Lastly, the study was conducted in a normal-gravity environment and without attempts to reproduce any ergonomic circumstances of spaceflight.

ACKNOWLEDGMENTS

The authors thank Mary Freire De Carvalho, Ph.D., for performing the statistical analysis for this study; Ashot Sargsyan, M.D., for his support of the experiment and review of the manuscript; and Jane Krauhs, Ph.D., for her professional editing of the manuscript. The work for this study was performed in the Space Medicine Division at NASA JSC under National Space Biomedical Research Institute contract # L69030 with Wyle.

Authors and affiliations: Victor W. Hurst IV, M.S., Ph.D., Kathleen Garcia, B.S., M.S., Douglas Ebert, Ph.D., and David Ham, B.S., Crew Health and Research, Wyle Science, Technology and Engineering, Houston, TX; Sean Peterson, M.D., Canadian Space Agency, Saint-Hubert, Quebec, Canada; and David Amponsah, M.D., and Scott Dulchavsky, M.D., Ph.D., Department of Surgery, Henry Ford Hospital System, Detroit, MI.

REFERENCES

1. Chiao L, Sharipov S, Sargsyan AE, Melton S, Hamilton DR, et al. Ocular examination for trauma: clinical ultrasound aboard the International Space Station. J Trauma. 2005; 58(5):885–889.
2. Fincke EM, Padalka G, Lee D, van Holsbeeck M, Sargsyan AE, et al. Evaluation of shoulder integrity in space: first report of musculoskeletal ultrasound on the International Space Station. Radiology. 2005; 234(2): 319–322.
3. Foale CM, Kalery AY, Sargsyan AE, Hamilton DR, Melton S, et al. Diagnostic instrumentation aboard ISS: Just-in-time training for non-physician crewmembers. Aviat Space Environ Med. 2005; 76(6):594–598.
4. Hart R, Campbell MR. Digital radiography in space. Aviat Space Environ Med. 2002; 73(6):601–606.
5. Hawley S. Mission to Mars: risks, challenges, sacrifices and privileges. One astronaut's perspective. Journal of Cosmology. 2010; 12:3517–3528.
6. Jones JA, Kirkpatrick AW, Hamilton DR, Sargsyan AE, Campbell M, et al. Percutaneous bladder catheterization in microgravity. Can J Urol. 2007; 14(2):3493–3498.
7. Jones JA, Sargsyan AE, Barr YR, Melton S, Hamilton DR, et al. Diagnostic ultrasound at MACH 20: retroperitoneal and pelvic imaging in space. Ultrasound Med Biol. 2009; 35(7):1059–1067.
8. Kirkpatrick AW, Nicolaou S, Campbell MR, Sargsyan AE, Dulchavsky SA, et al. Percutaneous aspiration of fluid for management of peritonitis in space. Aviat Space Environ Med. 2002; 73(9):925–930.
9. Marshburn TH, Hadfield CA, Sargsyan AE, Garcia K, Ebert D, Dulchavsky SA. New heights in ultrasound: First report of spinal ultrasound from the International Space Station. J Emerg Med. 2014; 46(1):61–70.
10. Martin DS, Caine TL, Matz T, Lee SMC, Stenger MB, et al. Virtual guidance as a tool to obtain diagnostic ultrasound for spaceflight and remote environments. Aviat Space Environ Med. 2012; 83(10):995–1000.
11. Martin DS, South DA, Garcia KM. Ultrasound in space. Ultrasound Med Biol. 2003; 29(1):1–12.
12. NASA Technical Standard, NASA-STD-3001, NASA Space Flight Human System Standard, Volume 1, Revision A: Crew Health; in particular, Section 4.1.6.3. Washington (DC): NASA; July 2014.
13. Otto C, Comtois JM, Sargsyan AE, Dulchavsky A, Rubinfeld I, Dulchavsky SA. The Martian chronicles: remotely guided diagnosis and treatment in the Arctic Circle. Surg Endosc. 2010; 24(9):2170–2177.
14. Otto C, Hamilton DR, Levine BD, Hare C, Sargsyan AE, et al. Into thin air: extreme ultrasound on Mt. Everest. Wilderness Environ Med. 2009; 20(3):283–289.
15. Sargsyan AE, Hamilton DR, Jones JA, Melton S, Whitson PA, et al. FAST at MACH 20: clinical ultrasound aboard the International Space Station. J Trauma. 2005; 58(1):35–39.
16. Scheuring RA, Mathers CH, Jones JA, Wear ML. Musculoskeletal injuries and minor trauma in space: incidence and injury mechanisms in U.S. astronauts. Aviat Space Environ Med. 2009; 80(2):117–124.
17. Sirek AS, Garcia K, Foy M, Ebert D, Sargsyan A, et al. Doppler ultrasound of the central retinal artery in microgravity. Aviat Space Environ Med. 2014; 85(1):3–8.
18. Wagner MS, Garcia K, Martin DS. Point-of-care ultrasound in aerospace medicine: known and potential applications. Aviat Space Environ Med. 2014; 85(7):730–739.