HUMAN FACTORS ISSUES FOR TELEPRESENCE ROBOTS

James E. Dunstan
LunaCorp
Fairfax, Virginia

Abstract

Realistic and immersive interfaces to planetary data sets have until now relied on expensive computing capabilities such as those provided by high-end computers such as Silicon Graphics (SGI) machines, and have been limited to one or two senses, primarily vision. Expanding on its prior work in this area, LunaCorp is working to build scalable consumer-level hardware/software interfaces for current and future planetary data sets for both virtual exploration and real-time "actualization" of robotic exploration. In the course of building prototype hardware, LunaCorp has learned a number of lessons about how humans interact with remote locations, which is the subject of this paper.

I. Introduction and Background

LunaCorp's corporate mission is to bring the excitement of space exploration and development to the general public.* For without wide public enthusiasm and participation, the High Frontier is destined to remain the province of underfunded governmental programs, subject to the vagaries of politics. If and only if the public can be drawn into space exploration, and allowed to participate, can we open the High Frontier for expansion.

Since 1983, LunaCorp has looked for ways to allow the public to interact with space exploration in both an educational and entertaining manner. LunaCorp produced two "edutainment" CD-ROMs, "Return To The Moon,"* and "Mission: Planet Earth."* LunaCorp next developed a real-time voxel landscape engine for an Internet rover simulator, dubbed "Robots In Cyberspace."

In 1997, LunaCorp teamed with Carnegie Mellon University's Robotics Institute* to help design the Nomad robot which was sent to the Atacama Desert in Chile to demonstrate long duration autonomous navigation capabilities. Nomad itself was a full-size version of the rover LunaCorp intended to send to the Moon's equator. LunaCorp arranged for the satellite link to bring real-time telemetry back from the desert, and utilizing the six degree-of-freedom (6-DOF) motion platform system from ViRtogo,* fed the motion data from Nomad live into the platform, for the first ever transportation of motion cues between continents in real-time.

To increase its experience in combining visual, motion, and sound cues, in 1998, LunaCorp principals developed the first motion platform arcade game designed to run on a standard PC, Lunar Defense.* LunaCorp also continued to develop Internet-based robotic interfaces, including its Vote-2-Drive™ interface, to allow an unlimited number of users to interact with a real exploratory robot by defining safe paths and then allowing the user community to define the path the robot would take. LunaCorp also updated its voxel-based landscape engine as the core of its Moonroll lunar driving simulator, available for sale to science centers.

In 2000, LunaCorp completed construction of its first fully enclosed test chamber, dubbed Telepresence Experimental Portal No. 1 ("TXP-1").

* Background information on LunaCorp can be found at – http://www.lunacorp.com.

* Jerry Pournelle named "Return To The Moon" his CD-ROM of the month in Byte Magazine, July, 1995, declaring, "if you have any interest in the moon, this is for you."

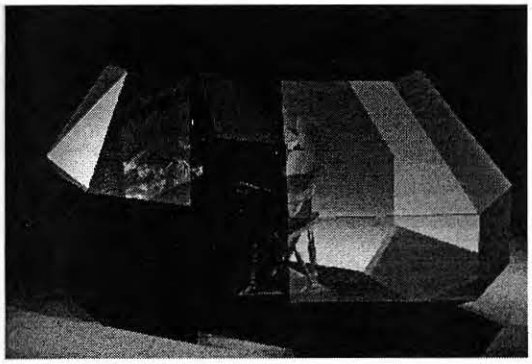
* Les Krantz of CD-ROMs Rated ranked "Mission: Planet Earth" second of ten Atlas programs reviewed with a score of 81.60 out of 100. Title is now out of print.

* http://www.frc.ri.cmu.edu/projects/lri/.

* http://www.virtogo.com.

* http://www.lunardefense.com.

A user can be closed inside TXP-1 and once there either control a simulated rover over the voxel generated landscape, or drive a scale model rover on the adjacent moonyard. Work is progressing toward the completion of TXP-2, a more stylized enclosure that also will be available to the museum and science center market by late 2001.



As LunaCorp has expanded its expertise in the past decade, its focus has remained constant – to provide the best possible environment to allow the maximum number and types of users to interact with distant exploratory robots, to literally allow users to teleport their senses to a remote location, to experience places no other humans have ever experienced. Whether the humans are highly trained technicians or casual visitors to a science center or theme park, the goal of providing the richest possible environment remains the same.

To create the most immersive environment possible, LunaCorp has come to realize that the hardware/software user interface must carefully balance ***all*** of the following cues:

1) Vision;
2) Sound;
3) Motion;
4) Temperature;
5) Tactile (force feedback input controls); and
6) Smell.

The end result is a highly immersive experience that truly teleports the senses to a different place. In the course of developing the technologies necessary to build a legitimate "planetary interface," LunaCorp has learned a great deal about the interplay between the sensory cues discussed above. The next section of this paper will discuss the "lessons learned" by LunaCorp in developing an optimal planetary interface.

II. Lessons Learned In Building Planetary Robotic Interfaces

A. Lesson One: Video Is a Pig!

In any planetary robotic mission (including missions to remote locations on planet Earth), the communications system must compete for both mass and power against other onboard systems, such as locomotion, thermal control, etc. Compromises must always be made, and ultimately, there never seems to be enough power or bandwidth to bring back as much data as the end users would like. Compounding this problem is the fact that visual images take up a huge percentage of the usable bandwidth of any robot communications system.

In order to understand the data rates required for the various senses, we will first discuss the desired minimum data rate for each sense.*

1. Video Data

As a baseline, one can look at standard NTSC resolutions as a starting point for telepresence. Most people are comfortable and accustomed to viewing video information in this format. For the purposes of this discussion, we will define NTSC to be 30 frames per second (FPS) of 640x480 pixel frames. Since we are discussing this in computer terms, we will define the color resolution to be 24 bits per pixel. A little bit of math shows the data rate:

640 x 480 == 307,200 Pixels per frame
307,200 * 30 == 9,216,000 Pixels per second
9,216,000 * 24 == 221,184,000 bits per second
or 27,648,000 bytes per second

* The author acknowledges and thanks Steve Bress of Entropy Engineering for most of the contents of this section. *See* http://www.entropyengineering.com. Steve was the major force behind the Lunar Defense video arcade game, and has been an integral part of LunaCorp's technology development team.

While 28 Megabytes per second may seem like a lot, it is well within the sustained transfer capabilities of a modern, low cost hard drive.

Using the same math, we can see that the bandwidth required for HDTV video is dramatically higher. Using the ideal case of a 1920x1050 HDTV signal and only a 30 Frame per second update rate, the numbers work out as follows:

1920x1050 == 2,016,000 Pixels per frame
2,016,000 * 30 == 60,480,000 Pixels per second
60,480,000 * 24 == 1,451,520,000 bits per second or 181,440,000 bytes per second

2. Audio Data

Until recently, CD Quality has been synonymous with High Quality Audio. In part, this is due to the simple fact that most people have never heard anything better, not that they couldn't tell the difference if they did. In any case, if it's good enough for most people, it's good enough for this discussion. Standard audio CDs use 2 channels of 16 bits per sample at 44.1 kHz.

44,100 * 16 * 2 == 1,411,200 Bits per second or 176,400 bytes per second.

In theory, this is a lot of data, but it works out to be less than one percent of the bandwidth required for the video. For surround systems, 6 channels of audio are used, increasing the required bandwidth to 529,200 bytes per second. This brings the bandwidth up to almost two percent, but is still somewhat negligible from a system design point of view. Basically, if the capacity exists for video, the extra capacity required for audio shouldn't be an undue burden.

3. Motion Data

Experience with the ViRtogo 6-DOF motion platform shows that high quality motion can be achieved with less than 500 bytes per second of information data. This is due in part to the fact the motion platform is a real world machine with physical response times. It is not just a data construct. As such, when a command is given to the machine, it will take a certain amount of time for it to respond.

The inner ear is the source of motion sensing in humans. It is also a physical device with real world response times. This keeps the maximum required data rate to provide quality motion extremely low.

4. Smell

Smell requires an even lower data rate than motion. The current generation of smell generators, have a limited palette of smells that they are capable of generating. In addition, a smell must travel across a physical distance from its generation point to the nose user. Changing the smell too quickly would lead to 'blurring' of the smell data. We estimate at this point that the maximum useable data rate for smell information will be 10 bytes per second, and that is being quite generous.

5. Temperature

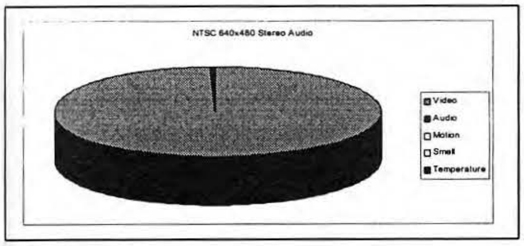
Temperature is another low speed sense. Most robotic probes have only a single temperature sensor mounted in a strategic location. In virtually all non-catastrophic occurrences, the temperature will change slowly over a period of time. To bring this experience back through telepresence is estimated to require no more than 2 bytes per second.

Even in the case of a high speed change in temperature data, the human response time to temperature variations makes higher data rates somewhat pointless.

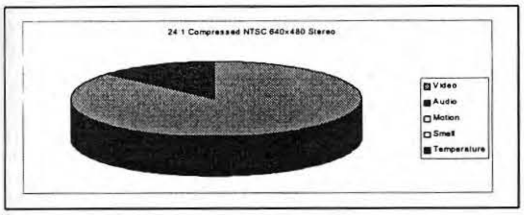
In order to bring back the digital equivalent of NTSC video at 15 frames per second uncompressed, the communications system must transmit 27 megabytes per second. In contrast, sound requires only 177 kilobytes per second, motion needs only 500 bytes per second, and smell and temperature even less:

	Raw NTSC 640x480 Stereo Audio	Percent of Total
Video	27,648	99.3621%
Audio	177	0.6361%
Motion	0.5	0.0018%
Smell	0.01	0.0000%
Temperature	0.002	0.0000%
Total	27,825.512	100.0000%

When represented graphically, the demands of the video feed is even more evident (where the light portion of the pie chart depicts the bandwidth demands of the video and the dark sliver represents all other sensory cues):



Even a highly compressed video stream (at 24:1 compression), still requires over 85% of the bandwidth of the signal returning from the rover.



The prognosis gets far worse if one wishes to bring back higher fidelity visual images such as HDTV. At the standard uncompressed HDTV standard, a planetary robot would be required to transmit 181 megabytes per second. Any way you "slice" it, the video demands eat the bandwidth of an exploratory robot alive. As a result, when trade studies are conducted for planetary missions, the other cues lose out to the demands on bandwidth of video, such that the ***only*** sensory cues that come back from most planetary robots are visual imagery.

The legitimate question must then be asked: Is the video information the only information that should be returned from a planetary robot? This leads us to our next lesson learned.

B. Lesson Two: Visual Cues Are Not As Important As They Seem

One of the key lessons we have learned at LunaCorp is that once additional sensory stimuli are added to a remote experience beyond just visual cues, the ability of the user's eyes to discern screen resolution or frame rate drops drastically. This is especially true when motion cues are introduced. With a user positioned on a motion platform system which is constantly moving in response to motion cues from the remote environment, the user's inner ear forces him or her to forget about such niceties as whether the image being transmitted is in 640x480 or 1024x768 format, or the exact frame rate of the video. When additional sensory inputs are added, such as surround sound (especially low frequency rumble), and tactile sensations such as a force feedback control mechanism, visual cues become just one of many sensory inputs rather than the dominant or exclusive interface to the remote environment.

Indeed, by dedicating even a modest percentage of the bandwidth (e.g., 15 percent) of a planetary robot to non-visual sensory inputs, this allows bandwidth demands of the visual imagery to be reduced substantially, either through a reduction in native resolution, or through the trimming of the frame rate of the video. Further, a "lossier" compression can be used in conjunction with additional sensory cues, since the resulting artifacts will be far less noticeable to a user inundated with other sensory cues.

This is not a case of "robbing Peter to pay Paul." While it is true that under such a scenario the video brought back may not be of as high quality, when this video is combined with the other sensory data, the overall experience is far more immersive than just higher video fidelity alone. Further, as discussed above, this is not a one-to-one tradeoff since the demands on bandwidth for video data far exceed the combined demands for bandwidth for the other sensory data discussed herein. Rather, an approach to bringing back remote planetary data which includes motion cues, sound, temperature data, tactile "feel" and even smell, can allow mission designers to place less demands on their communications system and still bring back an environment that is richer than if they built a more robust communications link to bring back just higher quality video.

C. Lesson Three: The Addition of Any Sensory Cues Beyond Visual and Single Channel Audio Radically Increase the Sense of Immersion for the User

Twenty-first century humans are heavily reliant on visual cues in our lives. In fact, we rely much too heavily on our eyesight to interact with our environment. Hence the high degree of eyestrain, myopia and other factors that lead so many of us to need corrective measures for our eyes.

Biologically, however, we are not "built" to rely exclusively on our eyesight. In engineering terms, we are constructed with a balanced sensory suite of receptors which combine inputs in real-time to allow us to fully interact with our environment. We do not see the mosquito flying near us, we hear the buzz. We do not see the smoke rising from the burnt toast, we smell it. We do not sense a sweltering summer day by viewing it, we feel the hot atmosphere against our skin.

Yet when we design robots to explore distant planets for us, we return only visual data, because that is how we are used to interacting with remote locations through the medium of television. Our experience at LunaCorp with the TXP-1 is that once the sensory inputs are expanded beyond just sight and sound, the user no longer is watching a remote location, he or she is ***experiencing*** that remote location. The level of immersion increases dramatically with each new sensory data type. Add motion to sight and sound, and the user is forced to feel the remote environment. Add thermal changes and the user must now begin to think about whether the remote environment is friendly or hostile. Add smell, and the user will attempt to calibrate this remote environment to places they have been before (does Mars smell more like the desert outside Bakersfield or the frozen plains of Canada)? Have a joystick "fight back" as the user tries to drive over a rock, and the user has been transported to that remote environment and fully immersed therein.

D. Lesson Four: Small Motion Cues Have a Much Greater Sensory Impact Than Large "Throws"

LunaCorp's largest experience is in melding motion cues with sight and sound. Beginning with the CMU Atacama Desert Trek in 1997, through the development of the Lunar Defense video arcade game, and more recently experience in developing the TXP-1 chamber, LunaCorp has had ample opportunity to deal with remote motion data, and translate that data to users. In the course of gaining that experience LunaCorp has found that motion is best translated in small increments at high frequency rates (>= 10 Hz).

This finding cuts against the common experience of most people who have experienced a motion ride at a theme park. There, the emphasis is on throwing people around as much as possible, and it can make for a very wild ride. The reason that motion ride designers do this, however, has less to do with either realism or sensation, but rather has everything to do with the hardware with which they are working. Most theme park motion rides use multi-person platforms, sometimes up to as many as 20 people on a platform. This simply is a matter of economics – the cost per seat goes down the more people you can put onto a platform, or so the industry believes.

The problem is that no configuration of a larger platform will allow true motion effects to be imparted to the visitor. Take the example below of an overhead view of a 15 person platform with the screen toward the top of the page.

O	O	O	O	O
X	O	Y	O	O
O	O	O	O	O

If the remote environment calls for the platform to roll left, all rider "X" will experience is a drop, not a real roll. Similarly, if the platform is to roll right, rider "X" will simply rise. In this configuration, only rider "Y", positioned dead-center, will feel the true motions imparted by the platform. Because of this, designers accentuate the motions so that even if the feeling isn't accurate, at least the platform is giving people a thrilling ride.

LunaCorp's approach has been different. Using ViRtogo's 6-DOF single person platform, the rider is always dead-center in the motion envelope. We can therefore impart actual motions to the rider without having to throw them around as much. In fact, we have discovered that faster positional changes invoke a much greater sense of motion than actual movement distance. By updating the chair position 10 times a second, the rider feels like they are moving much more than if the chair was updated twice a second, but the actual positional change was far greater. (We discovered this phenomenon late one night while working on the Lunar Defense motion code. The story involves pizza, Mountain Dew, and resulted in our own version of the Space Whoops.)

By using more frequent updates of the motion platform, this also allows us to "texture" the feel of the motion, either to make the ride more authentic, or more "Hollywood." For example, during the 1997 Atacama Desert Trek, the motion data we were getting from Chile was only coming in at 3 Hz. As a result, as Nomad pitched up to climb a ridge, there was a bit of a stair-step feel to the ride. We were able to interpolate the code to smooth this out. Also, Nomad's motion sensor was located in the center of the body of the robot rather than on any of the wheels. We knew from the video we were getting from Nomad that the robot was experiencing more motion than was being translated in the data. We could also tell from the video the overall texture of the area Nomad was traversing, and were therefore able to add a subtle rumble into the motion "mixer" to convey the feel of the terrain.

Such subtleties are not possible on large motion platforms that require large throws to thrill an audience. And as it turns out, on a price-per-seat basis, the ViRtogo 6-DOF motion platform is actually cheaper than larger platform systems, so we can win both ways with this approach.

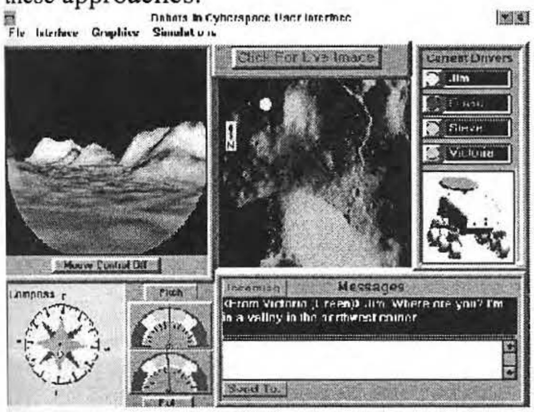
E. Lesson Five: Orientation (Compass Heading) is the Most Difficult Information to Convey to An Amateur Driver

A major goal of LunaCorp is to allow amateur drivers the opportunity to control a robot in a distant environment, including out to the Moon. In both our lunar driving simulators and scale model driving in TXP-1, we have discovered that many amateur drivers cannot get their bearings – they end up driving around in circles. This is especially true on alien landscapes such as the Moon, where there are no recognizable landmarks to guide the driver.

To date, he have built robot interfaces with a variety of tools designed to display bearing, including:

1) Analog (floating needle) compasses;
2) Digital compasses; and
3) Overhead maps with leading arrow indicators of direction.

The sample interface below shows several of these approaches.



LunaCorp currently is planning a series of experiments to test driver responses to a variety of bearing indicators to determine which provide the best information to amateur drivers and can be assimilated in the shortest time. LunaCorp views this as a key development initiative for 2001-2002, as we hone the technologies necessary to allow a wide variety of users to interact with remote robots.

III. Implications of LunaCorp's Experience for Trained Science Users

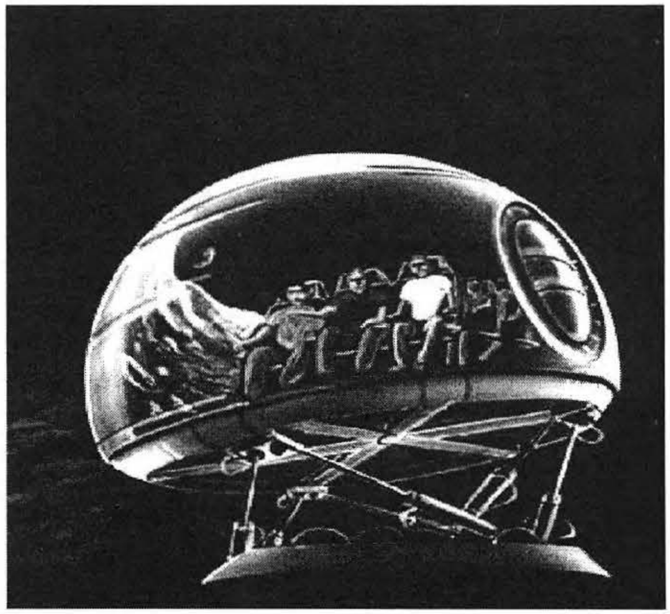
Scientists who design robots for planetary missions or experiments which will ride onboard such missions may at first dismiss this analysis as not applicable to "their" needs. They have certain data demands, and everything else is just a drain on bandwidth. This simplistic conclusion is very wrong.

First, as set forth at the outset, the fight for bandwidth on any planetary mission is intense, and visual imagery will ***always*** dominate the bandwidth. Anything that can be done to lessen the burden on the bandwidth can only aid the design of the host robot and the chances that experiment packages will be able to return the data necessary to validate their experiments.

Second, LunaCorp has experimented in reconstructing remote environments based on the video brought back. In preliminary work with both the Apollo 15 and Apollo 17 landing sites, using both still and video imagery returned, it is clear that video itself painted a very distorted picture of the true local environment. For instance, both the horizon of the Moon and any local slopes are not accurately conveyed in the visual imagery because of the lack of atmospheric distortion we take for granted on Earth and the curvature of the smaller circumference Moon. The Apollo astronauts themselves commented on this "distortion", many times becoming winded as they climbed steep slopes that looked to them to be nearly flat when they began their climbs.

Now consider a future robotic mission to the Moon (or to Mars), where the robot does not necessarily have the real-time intelligence to discern these distortions. If the controllers back on Earth order a robot to move in a certain direction based solely on what they ***think*** they see, there is a significant chance that the robot could be placed in danger. Instead, if full motion data were also being brought back with the video feed, a controller operating from a control center with a motion platform system would know the slopes of the surrounding terrain.

Similarly, for a solar powered robot, keeping in the sunlight is necessary for survival. It is possible for a robot to be partially "shaded" from the sun by a boulder or other obstruction without the straight ahead video indicating that it is operating in a shadow. Temperature gauges onboard the robot could report this instantly, however, and an operator who suddenly felt a drop in temperature within the control station (and our preliminary test show that only a few degree temperature drop is necessary to alert an operator to a temperature change) would know that the sun was being occulted. And while it is true that this information can also be conveyed visual through a series of numbers or analog gauges built into the interface, these visual cues have to compete with the rest of the imagery, whereas a blast of cold air is guaranteed to give the operator a wake up call. Finally, as in automobile operations on Earth, the introduction of a strange smell requires instant attention. The same holds true for planetary exploration. If the robot develops a problem, the introduction of smell into the control environment can be a quick signal of the need for quick actions. In short, multiple senses are always better than relying solely on visual cues.



IV. Conclusions

Finally, scientists should take seriously the use of multi-sensorial interaction with planetary robots for a simple reason – it will generate more interest in the mission. The more immersive the environment, the more people will want to interact, and the more support a mission can draw. Faced with ever diminishing resources to conduct planetary missions, scientists and engineers ***must*** find ways to draw the interest and support of the public for exploration. It is in the community's own best interest to find ways to allow the public to become part of their explorations. That is why LunaCorp was created, and we stand ready to aid the planetary exploration community in engaging public attention and translating the science in a way that even casual participants can understand and enjoy. Working together, we can build missions that sustain both real science, and real entertainment – a win-win scenario.