Performance Effects of Multi-sensory Displays in Virtual Teleoperation Environments

Paulo G. de Barros
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA, USA, 01609
+1 508-831-6617
pgb@wpi.edu

Robert W. Lindeman
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA, USA, 01609
+1 508-831-6712
gogo@wpi.edu

ABSTRACT

Multi-sensory displays provide information to users through multiple senses, not only through visuals. They can be designed for the purpose of creating a more-natural interface for users or reducing the cognitive load of a visual-only display. However, because multi-sensory displays are often application-specific, the general advantages of multi-sensory displays over visual-only displays are not yet well understood. Moreover, the optimal amount of information that can be perceived through multi-sensory displays without making them more cognitively demanding than visual-only displays is also not yet clear. Last, the effects of using redundant feedback across senses on multi-sensory displays have not been fully explored. To shed some light on these issues, this study evaluates the effects of increasing the amount of multi-sensory feedback on an interface, specifically in a virtual teleoperation context. While objective data showed that increasing the number of senses in the interface from two to three led to an improvement in performance, subjective feedback indicated that multi-sensory interfaces with redundant feedback may impose an extra cognitive burden on users.

Categories and Subject Descriptors

H.5.2 **[User Interfaces]**: Auditory (non-speech) feedback, Graphical user interfaces, Haptic I/O, Evaluation/methodology, H.5.1 **[Multimedia Information Systems]**: Artificial, augmented, and virtual realities, I.2.9 **[Robotics]**: Operator interfaces.

General Terms

Design, Performance, Measurement, Human Factors.

Keywords

Multi-sensory interfaces; robot teleoperation; virtual environment; urban search-and-rescue; visual, audio and vibro-tactile feedback.

1. INTRODUCTION

Since the creation of *Sensorama* [15] in 1962, all human senses have been used by the entertainment industry, as well as researchers in the area of Virtual Reality, as sources of information display for virtual environments (VEs). They have been evaluated in terms of their impact on user presence [35], and performance [3]. Despite that effort, few researchers have looked into integrating all senses into a single display or measuring the effect of such integration on user perception, or user efficiency and effectiveness [16]. This work evaluates the impact on user performance and cognition of multi-sensory feedback (vision, hearing and touch) in a virtual robot teleoperation search task.

Results show that a well-designed, tri-sensory display can increase user performance and reduce workload compared to a bi-sensory display. Results also show that redundant feedback is only useful if it helps user awareness of unnoticed parts of the displayed data.

The remainder of this paper is organized as follows. Section 2 reports related work. Section 3 summarizes our interface. The experiment hypotheses are detailed in Section 4, followed by a description of our study in Section 5. Section 6 summarizes the results, which are analyzed in Section 7. Last, Section 8 draws conclusions about the results and describes future areas of work.

2. RELATED WORK

Multi-sensory interface research encompasses a large variety of research areas. In the context of this work, focus will be given to Virtual Reality (VR) and Human-Robot Interaction (HRI).

Research on the integration of multiple senses in perception has shown that sense prioritization is dependent on the reliability of sensory channels [10]. Although systems providing multi-sensory stimulation have been used for some time now, studying the effects the conjunctive use of multiple senses to interact with real and virtual worlds has seldom been undertaken [4][16][19][37]. Moreover, the results obtained by individual researchers are difficult to generalize due to their task-specific nature [33].

Of all the senses, vision is by far the one that has been the most studied, with stereoscopic head mounted displays, CAVEs and powerful GPUs. Hearing has been explored for adding realism to scenes, but also to help in performing specific tasks, such as search and localization [12]. Stereoscopic, surround and bone-conduction [23] sound systems have been experimented with as audio displays with and without the use of HRTFs [11]. For touch and proprioception [27], vibro-tactile [2] and force feedback [16] have been used to signal actions [36], support interactions with virtual objects and display geo-spatial data using specialized [5][21] or mobile devices [30][31]. Multi-modal displays have also been reported to reduce user workload [3]. Contact feedback classifications for vibro-tactile devices have been proposed [22] and have even been used to guide the blind [5].

In the area of HRI, specifically urban search-and-rescue (USAR) teleoperation, interface design and implementation guidelines have yet to be standardized, although some progress has been made [8][25][34]. Interfaces for real USAR teleoperation often simply consist of keyboard, mouse, gamepads [38], and touchscreens [26] or visual displays [28].

Although current USAR teleoperation interfaces aim to improve Situation Awareness (SA) [9] and efficiency [12][28][38], little effort has been put on validating reductions in the operator cognitive load. Adding multisensory cues has been partially explored [5][6][7][32][38], and although novel visual interfaces have been evaluated [18][28], research in this field still lacks an extensive evaluation of the benefits of multi-sensorial interfaces.

Previous studies in USAR virtual robot teleoperation, vehicle driving [37] and pedestrian navigation [29] have shown that adding properly designed vibro-tactile displays to visual ones can improve navigation performance [6]. It has also been found that redundant feedback in such displays led to higher levels of SA, and increased navigation performance variability among operators [7]. Nonetheless, the reason behind such an effect is not yet well understood and could be the result of interface design issues affecting the reliability of the display multi-sensory channels [10].

With the exception of a few user studies comparing the use of audio or vibration with visual-only interfaces [11][12][16], to our knowledge, little has been done in evaluating the impact of individual components of USAR multi-sensory robot interfaces.

The current work builds on these previous results, and evaluates the effect of adding audio feedback to a bi-sensory interface (vision and touch), and the effect of redundant data presentation in multi-sensory displays. Notice that the focus of this work is on the output to the user, not the input from the user.

3. ROBOT INTERFACE

Results from previous studies suggest that vibro-tactile feedback by itself is not an optimal navigation interface. Instead, it should be used as a supplement to other interfaces [29]. In this work, three multi-sensory interfaces with increasing complexity were created by supplementing a vibro-tactile one with extra feedback.

Interface 1, the control case interface that was used as a starting point for the two other interfaces evaluated here, was designed following USAR interface guidelines and is based on the work of Nielsen [28] and de Barros & Lindeman [6]. It is composed of a visual interface (Figure 1) with a vibro-tactile belt display (Figure 2a). The visual interface fuses information as close as possible to the operator's point of focus, around the parafoveal area [19].

The visual part of Interface 1 contains a third-person view of the robot (dimensions: 0.51m × 0.46m × 0.25m), which sits on a blueprint map of the remote environment and has the video from the robot camera (60º FOV, rotating range: ±100° horiz. and ±45° vert.) presented on a rotatable panel. The blue dots on the map appear as nearby surfaces detected by robot sensors. The camera panel orientation matches the camera orientation relative to the robot. Furthermore, the robot avatar position on the map matches the remote robot position in the real-world VE. A timer with the elapsed time is shown in the top-right corner of the screen.

The vibro-tactile feedback belt (Figure 2a) is an adjustable neoprene belt with eight tactors (ruggedized eccentric DC mass motors [24]) positioned at the cardinal and intermediate compass points (forward = north). Tactor locations were adjusted for subject waist. The tactors provide the user with collision proximity feedback (CPF). The closer the robot is to colliding in the direction the tactor points, the more intense a tactor in the belt continuously vibrates, similar to the work of Cassineli [5]. The vibro-tactile feedback is only activated when the robot is within a distance $d \leq 1.25$m from an object. If an actual collision occurs in a certain direction, the tactor pointing in that direction vibrates continuously at the maximum calibrated intensity. The intensity and range values were identified as optimal in a pilot study.

Interface 2 builds upon Interface 1 and adds audio feedback. The first type of sound feedback is a stereoscopic bump sound when collisions between the virtual robot and the VE occur. The second type of sound feedback is an engine sound that increases its pitch as speed increases to give feedback about robot moving speed.

Interface 3 builds upon Interface 2 but adds extra visual feedback to the interface. A ring of eight dots is displayed on the top of the robot and mimics the current state of the vibro-tactile belt. It is an improvement over previous work on redundant displays [7]. The positioning on the belt of each tactor is associated with one of the dots in the ring and their locations match. The more intensely a tactor vibrates, the more red the dot associated with that tactor becomes (as opposed to its original color black). The second added visual feature is a speedometer positioned on the back of the robot as a redundant display for the engine sound. Table 1 summarizes the interface features that each interface contains.

For all three interfaces, the user controlled the virtual robot using a Sony PlayStation2 Dual-shock® gamepad (Figure 2b).

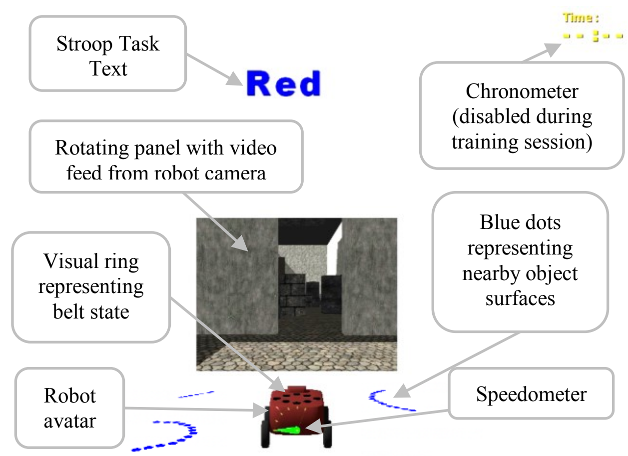

Figure 1. Visual components for all three interfaces. The visual ring and speedometer are only part of Interface 3.

Figure 2. (a) Vibro-tactile belt; (b) PlayStation® 2 controller.

Table 1: Display features for interfaces treatments.

Interface Number	Standard Visual Interface	Vibro-tactile feedback	Audio feedback	Visual ring and speedometer
1	X	X		
2	X	X	X	
3	X	X	X	X

The right thumbstick controlled robot movement using differential drive. The left thumbstick controlled camera pan-tilt [7]. The controller allowed subjects to take pictures with the robot camera.

Sound feedback was displayed through an Ion iHP03 headset. The headset was worn for all treatments. An ASUS G50V laptop was used in the study. It was positioned on top of an office table at 0.5m from the subject's eyes. The environment was run in a window with resolution of 1024×768 at a refresh rate of 17 fps.

4. HYPOTHESES

The use of vibro-tactile and enhanced interfaces has been shown to improve user performance [2][4][16][18]. Results from other previous work [6] have shown that vibro-tactile feedback can improve performance if used with a visual interface as a complementary source of collision proximity feedback (CPF) in a simple virtual teleoperation task. What is not a consensus yet among these and other studies [37], however, is whether the use of redundant feedback actually brings overall benefits.

Additionally, in another study using redundant feedback as a graphical ring [9], the results were inconclusive due to interface occlusion problems. This motivated us to improve on this interface and create a similar ring structure, but now sitting on top of the robot avatar to resolve the reported occlusion problem. With this new ring layout, it is possible that the redundant visual display benefits outweigh any potential disadvantages.

Our current study evaluates the impact on cognitive load and performance of adding redundant and complementary audio-visual displays to a control interface with vibration and visual feedback. Based on the insights collected from other previous work, our previous studies and with the interface enhancements proposed, the following two results are hypothesized:

H1. Adding redundant and complementary sound feedback to the control interface should improve performance in the search task;

H2. Adding redundant visual feedback should lead to even further improvements in performance in the search task.

5. USER STUDY

The current study was designed to confirm whether the enhancement of a visual-tactile interface with extra audio and visual information would lead to a reduction or increase in operator cognitive load and performance. We opted for a fielded interface experiment [7]. Our interface attempts to approximate what is used by researchers and experts to perform a real robot teleoperation task. This approach increases the chances of detecting the effects of multi-sensory feedback in a reasonably realistic virtual robot teleoperation context, as opposed to a lab-oriented approach, where low-complexity interfaces are tested.

5.1 Methodology

To evaluate the validity of the proposed interfaces, a search task was designed to best reproduce what happens in real USAR teleoperation situations, but in a slightly simpler manner. Subjects had to search for twelve red spheres (radius: 0.25m) in a debris-filled environment. Subjects were unaware of the total number of spheres. They were asked to find as many spheres as possible in as little time as possible and also avoid robot collisions. When the experiment was over, subjects drew sketchmaps of the VE showing the locations of the spheres found.

The experiment consisted of a within-subjects design where the search task was performed by each subject for all interface types (Table 1). The *independent variable* (*I.V.*) was the type of interface, with three possible treatments: Interface 1 (control), Interface 2 (audio-enhanced) and Interface 3 (visually-enhanced). Interface and virtual world presentation order for each subject was balanced using Latin Square to compensate for any effects within trials. The virtual worlds were built with the same size (8m x 10m), number of objects, walls and hidden spheres. They had similar complexity in terms of optimal traversal paths, traversal time, number of obstacles, and sphere levels of occlusion. The pictures taken with the robot camera (800×640) were displayed on a web page during sketchmap drawing when the search was over.

While performing the main search task, each subject also performed a secondary task, a visual Stroop task [13]. Users had to indicate whether the color of a word matched its meaning. For example, in Figure 1, the word "red" does not match its color. The words were presented periodically (every 20±~5s) for 7.5±~2.5s, disappearing after that. Users were asked to answer the Stroop task as soon as they noticed the word on screen using the gamepad. The purpose of this task was to measure user cognitive load variations due to exposure to interfaces with different levels of multi-sensory complexity. The NASA-TLX test [14] was taken after each of the interface treatments to measure user workload.

The objective *dependent variables* (*D.V.*) were the following: the time taken to complete the search task, average robot speed, the number of collisions, the number of spheres found, the number of collisions per minute, the ratio between number of collisions and path length, the number of spheres found per minute, the ratio between number of spheres found and path length, and the quality of the sketchmaps. These variables were normalized on a per-subject basis. Here is an example that explains this normalization process: if subject *A*, for a *D.V. X*, had the following results (*Interface 1*, *Interface 2*, *Interface 3*) = (10, 20, 30), these values would be converted to (10/60, 20/60, 30/60) ~ (0.17, 0.33, 0.5). The reason behind such normalization is presented in Section 6.1.

In addition to these variables, cognitive load was compared using the Stroop task results. The Stroop task objective *D.V.*s were: the percentage of incorrect responses, response time, and percentage of unanswered questions. The first two variables were analysed for three data subsets: responses to questions where color and text matched, responses to questions where color and text did not match, and all responses. These variables were also normalized. For subjective *D.V.s*, the treatment and final questionnaires compared subjects' impressions of each interface. The former was completed three times for each interface. The latter was completed once and comparatively rated all three interfaces. Subjective workload was measured using the NASA-TLX questionnaire.

The study took approximately 1.5±0.5 hours per subject. The experiment procedure steps are listed in Table 2. For each trial, the time and location of collisions were recorded. Subject gender and age, how often they used computers, played video games, used robots, used remote-controlled ground/aerial/aquatic vehicles (RCVs) and used gamepads was collected in the demographics questionnaire. For all but the first two questions, a Likert scale with four values ("daily" (1), "weekly" (2), "seldom" (3) or "never" (4)) was used. The spatial aptitude test had nine questions about associating sides of an open cube with its closed version and questions about map orientation. Subjects had strictly five minutes to complete the spatial test. The instructions page explained the experiment procedure, the task and the interface.

The training sessions used environments similar in complexity to the ones used in the real task. During training sessions (~4 min.), subjects had to find one red sphere and take a picture of it. The idea was to make subjects comfortable with the robot controls and output displays. The treatment questionnaire is summarized in Table 3. Subjective questions (3-8) were adapted from the SUS [35] and SSQ [20] questionnaires and followed a Likert scale (1-7). The final questionnaire is summarized in Table 4 and its questions 1-5 were also given on a Likert scale (1-7).

The sketchmaps were evaluated using the approach proposed by Billinghurst & Weghorst [1], but on a 1 to 5 scale. Maps were scored twice by two evaluators. The definition used for scoring map goodness is similar to the ones used in [1] and [6], that is, how well the sketched map helps in guiding one through the VE.

Table 2: Experimental procedure for one subject.

Step	Description
1	Institutional Review Board approved consent forms;
2	Demographics questionnaire;
3	Spatial aptitude test;
4	Study instructions and Q&A session;
5	User wears belt and headset. Robot interface explained;
6	Task review;
7	Training explanation and Q&A followed by training task;
8	Study task review and Q&A followed by study task;
9	During task, video and objective data is recorded;
10	Trial is over: treatment questionnaire with sketch map;
11	NASA-TLX questionnaire;
12	Five-minute break before next trial;
13	Steps 7-12 repeated for the other two interface treatments;
14	Three treatments are over: final questionnaire.

Table 3: Treatment questionnaire summary.

#	Question description
1	Report the number of spheres found;
2	Draw on a blank paper a map of the house and objects and indicate location of spheres found;
3	How difficult it was to perform the task compared to actually performing it yourself (if the remote environment was real);
4	Sense of being there in the computer generated world;
5	To what extent there were times during the experience when the computer generated world became the "reality" for you, and you almost forgot about the "real world" outside;
6	Whether the subject experienced the computer generated world more as something he saw, or somewhere he visited;
7	When navigating in the environment whether the subject felt more like driving or walking;
8	How nauseated the subject felt;
9	How dizzy the subject felt.

Table 4: Final questionnaire summary.

#	Question description
1	How difficult it was to learn;
2	How confusing it was to understand the information presented;
3	How distracting the feedback provided was;
4	How comfortable its use was;
5	How it impacted the understanding of the environment;
6	General comments about experiment.

5.2 Virtual Environment

The virtual worlds and robot interface (Figure 1) were built on the C4 game engine (www.terathon.com). According to the AAAI Rescue Robotics Competition classification, the experiment VE has difficulty level yellow. It is a single level with debris on the floor [17].

6. RESULTS

This section presents the significant results obtained in this study. Therefore, if a variable is not discussed in detail in this section, its results led to no statistically significant difference (SSD).

In order to generate the results presented here, data was processed in two ways. Continuous values were processed using a single-factor ANOVA with confidence level of $\alpha = 0.05$. This analysis was done before and after the normalization process described in 5.1. Trends had a confidence level of $\alpha = 0.1$. When a SSD among groups was found, a Tukey test (HSD, 95% confidence level) was performed to reveal the groups that differed from each other. In order to reveal such differences in more detail, data was further analyzed with ANOVA ($\alpha = 0.05$) in a pair-wise fashion.

Owing to their categorical nature, the Likert scale data obtained from the treatment and final questionnaires were processed using the Friedman test for group comparisons and the Wilcoxon Exact Signed-Rank test for pair-wise comparisons.

6.1 Demographics

A total of 18 university students participated in the experiment. Their average age was 25 years ($\sigma = 3.18$). In terms of experience levels among groups exposed to interfaces in different orders, SSDs were found for computer and RCV levels. Group 123 had more computer experience than Group 312. On the other hand, Group 312 had more RCV experience than Group 123. These differences were the main motivator for applying the data normalization explained in Section 5.1.

6.2 Subjective Measures

For the treatment questionnaires, a SSD was found for *Being there* for Interface 1 and Interface 2 (Figure 3a). The latter led to higher being there levels compared to the former ($\chi^2 = 6.28$, $p = 0.04$, *d.o.f.* = 2). Moreover, a SSD was also found for *Walking* results between Interface 2 and Interface 3 (Figure 3b). When exposed to Interface 3, moving around the computer-generated world seemed to subjects to be more like walking than when exposed to Interface 2 ($\chi^2 = 7.82$, $p = 0.02$, *d.o.f.* = 2). These results seem to support H1, but go against the claim in H2.

The final questionnaire showed interesting results, especially for Interface 2. On the one hand, a pair-wise Wilcoxon test showed Interface 2 was more difficult to use than Interface 1 ($w = 18.5$, $z = -1.75$, $p = 0.09$, $r = -0.29$, Figure 4a). On the other hand, Interface 2 was more comfortable to use than Interface 1 ($\chi^2 = 5.51$, $p = 0.06$, *d.o.f.* = 2, Figure 4b). It also more positively impacted the comprehension of the environment compared again to Interface 1 ($\chi^2 = 10.98$, $p < 0.01$, *d.o.f.* = 2, Figure 4c).

Interface understanding levels also differed (Figure 4d). Using Interface 2 and Interface 3 made it more straightforward to understand the information presented than using Interface 1 ($\chi^2 = 5.52$, $p = 0.06$, *d.o.f.* = 2). A pair-wise Wilcoxon test showed that Interface 2 had a statistically significant increase compared to Interface 1 ($w = 10.0$, $z = -2.15$, $p = 0.04$, $r = -0.36$). The same pair-wise comparison for Interface 3 and Interface 1 only showed a trend however ($w = 15.0$, $z = -1.89$, $p = 0.07$, $r = -0.31$). These results from the final questionnaire seem to support H1, but do not present any evidence in support of H2.

For the NASA-TLX questionnaire, a trend indicated that Interface 2 had a higher temporal workload score than Interface 1 (w = 37.0, z = -1.87, p = 0.06, r = -0.31, Figure 5a). This measure indicates how hurried or rushed subjects felt during the task. Subjects felt more in a rush when exposed to Interface 2. Because no difference in task time was detected among interface groups, the only other factor that could have affected subjects' rush levels would have to be related to the visual timer on screen and subjects' behavior towards it. A plausible explanation would be that subjects were able to check the timer more often to see how efficiently they were doing. This behavioral change would only be possible if the rest of the interface was less cognitively demanding. Hence, an increase in timer look-ups could have been due to a decrease in cognitive demand from the rest of the interface. If this claim is true, such a decrease would support H1.

For the NASA-TLX performance measure, a trend has indicated a lower rating for Interface 3 compared to Interface 1 (w = 103.0, z = 1.80, p = 0.08, r = 0.30, Figure 5b). This measure indicates how successful subjects felt in accomplishing the task. In other words, Interface 3 made subjects feel as if they performed worse than with Interface 1. This result goes against what was claimed in H2.

6.3 Objective Measures

For the objective measures, two variables led to relevant results. For the normalized number of collisions per minute (Figure 6a), trends were found between pairs of interfaces (1, 2) (F [2, 15] = 3.70, p = 0.06) and (1, 3) (F [2, 15] = 3.65, p = 0.06). For the normalized number of collisions per path length SSDs were found for the same pairs of interfaces (1, 2) (F [2, 15] = 4.32, p = 0.04) and (1, 3) (F [2, 15] = 4.16, p = 0.05). These results support H1.

No SSDs were obtained by the analysis of the Stroop task data, although there was a slight decrease in response time for Interface 2 and Interface 3, as can be seen in Figure 7a.

The mean, S.D. and median for the number of collisions, number of spheres found, task time, average robot speed (m/s) and map quality are shown in Table 5, but no SSD was found for these.

Table 5: The triplets (mean μ, S.D. σ, median η) for the dependent variables' non-normalized data.

D.V.	Interface 1	Interface 2	Interface 3
Cols.	(17.1, 9.9,16)	(12.8, 8.6, 11)	(14.7, 11.6, 9)
Sphs.	(8.1, 2.6, 9.0)	(7.7, 2.5, 8)	(8.2, 2.7, 8.5)
Time	(275, 112, 232)	(291, 109, 265)	(272, 93, 269)
Speed	(.56, .06, .56)	(.54, .05, .54)	(.54, .06, .54)
Map	(3.1, 1.0, 3.1)	(3.0, 1.2, 3.0)	(3.0, 1.0, 3.2)

6.4 Subject Comments

Subject comments were collected on the treatment and final questionnaires. The comments were categorized according to interface features (touch, audio, extra GUI, map, etc.) or experimental features (Stroop task, learning effects). For each category, the comments were divided into positive and negative ones. One score point was added for each comment for a feature.

There was a prevalence of positive comments directed to the audio interface. One subject stated: "Adding the audio feedback made it feel much less like a simulation and more like a real task. Hearing collisions and the motor made it feel like I was actually driving a robot." Another said, "The sound made it much easier to figure out what the robot was doing. It was clear when there was a collision." Most comments praised the collision sound, but not so much the motor sound.

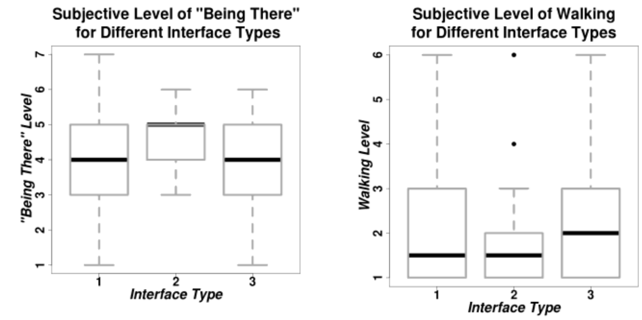

Figure 3: (a) Interface 2 increased user sense of being in the VE; (b) Interface 3 made users feel more like walking rather than driving.

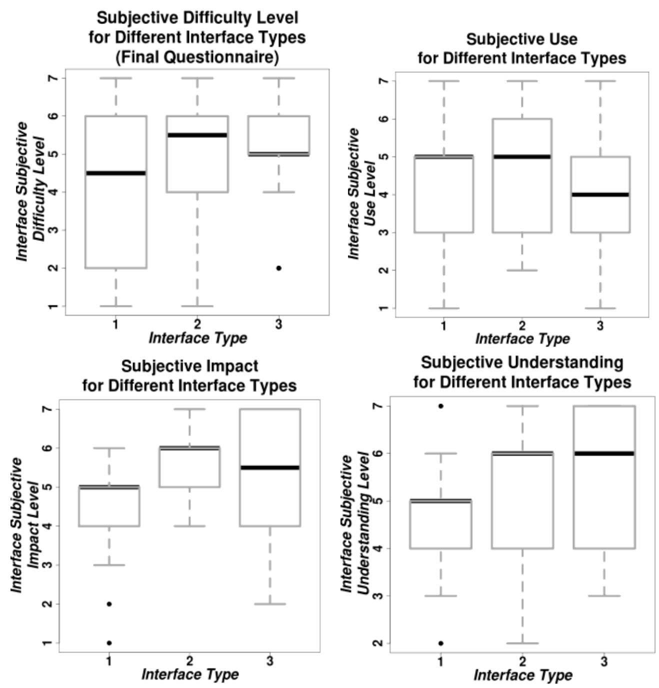

Figure 4: (a) Interface 2 was deemed more difficult to use than Interface 1, but it was also (b) more comfortable and (c) better impacted comprehension than Interface 1; (d) both Interfaces 2 and 3 helped better understand the environment than Interface 1.

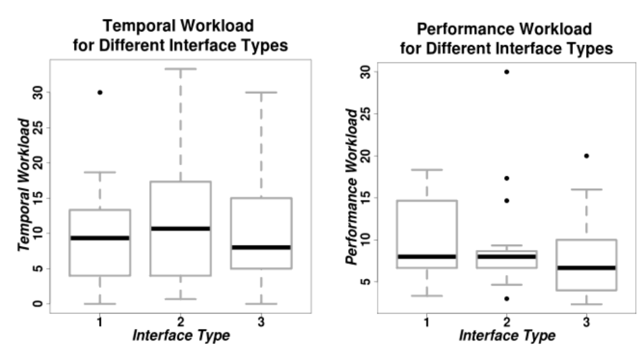

Figure 5: (a) Subjects felt significantly more rushed when using Interface 2 than with Interface 1; (b) Interface 3 caused subjects to feel as if they performed worse than Interface 1.

For the belt, it seemed that having it on all the time, even when it was evident no collision was imminent, annoyed subjects. A few subjects admitted that the belt was useful for navigation however. Many subjects seemed to ignore the belt feedback for the vast majority of the time and only used it when either a collision had already occurred or when passing through narrower places. These comments comply with the ones obtained in other studies [6].
For redundant feedback, it seemed to have distracted more than helped. One subject mentioned: "The visual speed feedback was not very useful at all, since the auditory speed feedback conveyed the idea much more effectively, so the visual speedometer became a distraction." The comments support the slight worsening in results for Interface 3 detected in Figures 3b and 7a.

Subjects' comments confirm the results obtained from subjective and objective measures supporting H1, but rejecting H2.

7. DISCUSSION

The main goal of this work was to search for answers to the question of how much one can make use of multi-sensorial displays to improve user experience and performance before an overwhelming amount of multi-sensorial information counter-balances the benefits of having such an interface. As a second goal, this work aimed at assessing the potential benefits, if any, of having redundant feedback in multi-sensory displays.

In other previous work [6], it was shown that, in the context of virtual robot teleoperation, adding touch-feedback to a visual-only interface as an aid to collision avoidance significantly improved user performance. In addition, other work [7] showed that adding redundant visual feedback for representing the same information as touch feedback could lead to a performance decrease, although the reason for that was assumed to be occlusion problems and not the fact that display of information was redundant.

Based on the interface and experiment results of these and other previous studies, our current study explored enhancing a visual-tactile interface with audio and redundant visual displays. Our enhancements over previously proposed interfaces allowed us to more accurately measure not only the impact of adding feedback to an extra human sense, but also to measure the effects of different types of redundant feedback in multi-sensory displays.

Unlike the belt feedback, which provided collision proximity feedback as the robot approached the surface of a nearby object, the collision audio display provided feedback only after a collision had occurred. This difference in feedback behavior led to an interesting result. Even though the audio feedback provided was an after-the-fact type of feedback, it led to further reductions in the number of collisions with the environment. But the audio display could not have helped reduce collisions in the same way as the touch display because of this difference in time of feedback. And the speed with which subjects moved the robot was not significantly affected by the engine sound feedback. Hence, two possible explanations for such reductions are:

1. The sound feedback made the remote VE feel more real and helped subjects become more immersed and focused on the task, leading them to perform the task with fewer collisions,
2. The sound feedback allowed subjects to better understand the relative distances between the robot and the remote VE. By experimenting with collisions a few times, subjects used sound feedback to learn what visual distance to maintain from walls to better avoid collisions from a robot camera perspective.

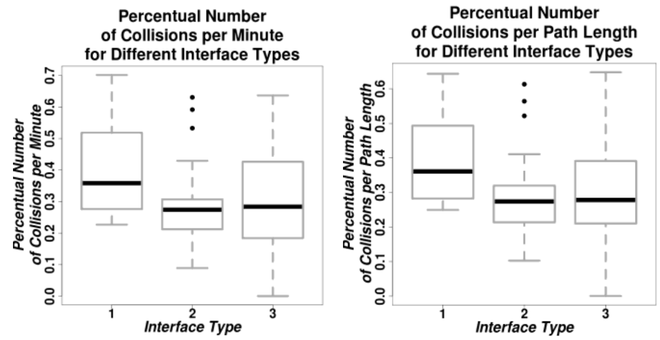

Figure 6: Both Interface 2 and Interface 3 caused a decrease in number of collisions: (a) per minute; (b) per path length.

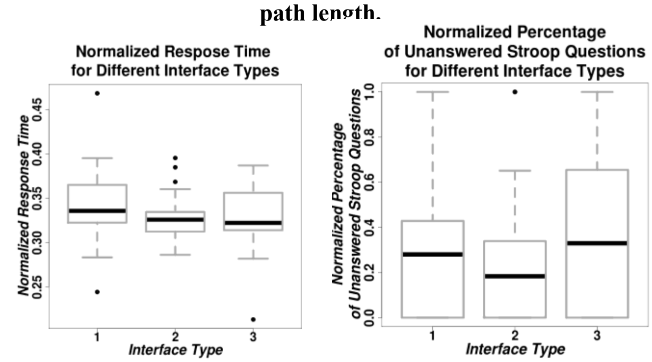

Figure 7: Stroop task results for (a) normalized response time and (b) normalized percentage of unanswered questions.

Even though both explanations matched subject feedback on the topic, we believe that the latter is a more plausible one. The distance estimation between the robot and the remote VE was not as easy to do using only the vibro-tactile feedback from the belt due to the continuous nature of the cues it provided.

Subjective feedback and objective data indicated that the engine sound did not have a major role in improving understanding of the relationship between robot and environment. Nevertheless, it was reported that this sound did improve their presence levels. Hence, the addition of the sense of hearing to the multi-sensory display improved performance and Hypothesis 1 (H1) was confirmed.

Hypothesis 2 (H2), on the other hand, was rejected. As mentioned earlier, results from similar studies on redundant feedback were inconsistent [6][37]. This work showed that redundant feedback may not always improve performance. In fact, its effect may vary depending on how the multi-sensory interface is integrated.

One explanation for the degradation in results for Interface 3 is considered here. It seems that the addition of new visual features created a new point on screen users needed to focus on. The basic visual interface (used in Interface 1 and Interface 2) already demanded a great deal of the user's attention with points of focus for: the timer on the top-right corner, the Stroop task text field, the robot camera panel and the map blueprint. Hence, adding more focus points in Interface 3 might have reduced user performance more than the amount of performance improvement that the addition of such interface features could have added.

However, would the same results be obtained if the extra visual information added was novel instead of redundant? In the case of this study, because the information displayed by the enhanced visual display was already being presented in other forms, no information was gained for most subjects, who already effectively read that same information through the vibro-tactile belt. For these subjects, the visual enhancements were either ignored or caused distraction, the latter to the detriment of their performance. Nonetheless, it would be interesting to compare the improvement results of individually using an audio-visual only interface or a visual-only interface with the speedometer and visual ring added to the current audio-visual-tactile interface.

Last, the use of the touch and audio feedback as opposed to the visual feedback for collision detection and proximity might be an indication that, when offered the same information through different multi-sensory displays, users may try to balance load among multiple senses as an attempt to reduce their overall cognitive load. Interesting though this claim may seem, the results obtained here are unable to support this notion. The verification of such a claim and the search for an answer to the question stated in the previous paragraph is the subject of future studies.

8. CONCLUSION

The main goal of this work was to give one more step towards understanding the effects on users of multi-sensory interfaces. We have explored the effects of adding audio to an existing visual-tactile interface. The context in which this exploration took place was in a virtual robot teleoperation search task in a 3D VE.

The study has shown that adding audio as the third sense to the bi-sensorial interface (visuals, touch) resulted in improvements in performance. This meant the user had not yet been cognitively overwhelmed by the control case display and could still process further multi-sensory data without detriment on performance.

This study also presented evidence indicating that displaying more data to a certain sense (vision) when it is already in high cognitive demand is detrimental to performance if the added data does not improve the user's SA of the system and environment. It remains to be seen how much of an effect the information relevance of the newly added visual data has on counter-balancing such detriment in performance. In order to measure such an effect, a new study needs to be carried out to compare the impact of a multi-sensory interface by adding more visual data that is not yet conveyed through other senses (novel data) versus adding visual data that is already conveyed through another sense (redundant data).

Redundancy could be beneficial to mitigate the fact that vision is uni-directional. A visual display could become at least partially omni- or multi-directional by adding redundant feedback through senses such as hearing and touch. The larger the number of focus points on screen, and the larger their relative distance, the higher the chances are that the user will miss some information or event. Having data redundancy spread across a multi-sensory display in a balanced, fused, non-distracting and non-obtrusive manner could reduce event misses and increase SA and comprehension.

Following the same thread of reasoning, it would be interesting to explore the validity of the following more general statement:

> *Redundant information over multiple senses brings no benefit to the user of a multi-sensorial display that already maximizes the user's omni-directional perception of relevant data.*

In other words, the more omni-directional a display is, the more data can be perceived by the user simultaneously, the smaller the chances are that changes in the data displayed are missed, and hence, the smaller the need is for providing redundant data displays. Admittedly, the study presented here barely scratches the surface of such a topic. Similar studies exploring the optimization of multi-sensorial omni-directionality must be performed and their results cross-validated for this statement be considered as plausible. Such studies should aim at complementing not only visual displays using other senses, but also complementing displays for other senses such as touch, with which it is only possible to feel as many surfaces as one's body pose can touch.

This work has provided a glimpse into the potential performance increase that multi-sensory displays can provide to 3D spatial user interaction. It has shown that multi-sensory displays can not only lead to more natural forms of information presentation but also display more information with reduced cognitive cost.

Nevertheless, the question of how complex multi-sensory displays can get is still not completely answered. Using three senses in an interface proved to be better than using only two, but what if more senses are considered? Is it possible to display data to olfactory and gustatory senses to improve displays for practical applications? Our research group aims at improving the current answers we have for these questions in future studies.

9. ACKNOWLEDGMENTS

The authors appreciate the research funding support from the Worcester Polytechnic Institute Computer Science Department.

10. REFERENCES

[1] Billinghurst, M. and Weghorst, S. 1995. The use of sketch maps to measure cognitive maps of virtual environments. *Virtual Reality Annual International Symposium*, 40-47.

[2] Blom, K.J. and Beckhaus, S. 2010. Virtual collision notification. In *Proceedings of IEEE Symposium on 3D User Interfaces*, 35-38.

[3] Bowman, D., Kruijff, E., LaViola Jr., J., Poupyrev, I. 2005. *3D User Interfaces: Theory and Practice*, parts 2 and 3, 27-310, Addison-Wesley, Boston, MA. 2005.

[4] Burke, J.L., Prewett, M.S., Gray, A.A., Yang, L. Stilson, F.R.B., Coovert, M.D., Elliot, L.R., and Redden, E. 2006. Comparing the effects of visual-auditory and visual-tactile feedback on user performance: a meta-analysis. In *Proceedings of the 8th International Conference on Multimodal Interfaces*. ACM, New York, NY, 108–117.

[5] Cassinelli, A., Reynolds, C., and Ishikawa, M. 2006. Augmenting spatial awareness with haptic radar. *Tenth International Symposium on Wearable Computers*. (Montreux, Switzerland, October 2006). ISWC'06, 61-64.

[6] de Barros, P.G., Lindeman, R.W. 2012. Poster: Comparing Vibro-tactile Feedback Modes for Collision Proximity Feedback in USAR Virtual Robot Teleoperation. *Proc. of IEEE 2012 Symposium on 3D User Interfaces*. 3DUI'12, 137-138. 2012.

[7] de Barros, P.G., Lindeman, R.W., Ward, M.O. 2011. Enhancing robot teleoperator situation awareness and performance using vibro-tactile and graphical feedback. *Proceedings of IEEE 2011 Symposium on 3D User Interfaces*. 3DUI'11, 47-54.

[8] Drury, J.L., Hestand, D., Yanco, H.A., and Scholtz, J. 2004. Design guidelines for improved human-robot interaction.

Extended Abstracts on Human Factors in Computing Systems. CHI '04, 1540.

[9] Endsley, M.R., and Garland, D.G. 2000. Theoretical underpinning of situation awareness: a critical review. *Situation Awareness Analysis and Measurement*, Lawrence Erlbaum, Mahwah, NJ.

[10] Ernst, M.O. and Bülthoff, H.H. 2004. Merging the senses into a robust percept. Trends in cognitive sciences. 8, 4 (Apr. 2004), 162–9.

[11] Gonot, A. et al. 2007. The Roles of Spatial Auditory Perception and Cognition in the Accessibility of a Game Map with a First Person View. *International Journal of Intelligent Games & Simulation*. 4, 2 (2007), 23–39.

[12] Grohn, M. et al. 2005. Comparison of Auditory, Visual, and Audiovisual Navigation in a 3D Space. *Transactions on Applied Perception*. 2, 4 (2005), 564–570.

[13] Gwizdka, J. 2010. Using Stroop Task to Assess Cognitive Load. *ECCE*. (2010), 219-222.

[14] Hart, S.G. 2006. NASA-Task Load Index (NASA-TLX): 20 Years Later. *Proc. of the Human Factors and Ergonomics Society 50th Ann. Meeting*, HFES '06, 904-908.

[15] Heilig, M. L., *Sensorama Simulator*, U. S. Patent 3,050,870, 1962.

[16] Herbst, I. and Stark, J. 2005. Comparing force magnitudes by means of vibro-tactile, auditory and visual feedback. *IEEE International Workshop on Haptic Audio Visual Environments and their Applications*. HAVE'05, 67–71.

[17] Jacoff, A., Messina, E., Weiss, B.A., Tadokoro, S., and Nakagawa, Y. 2003. Test arenas and performance metrics for urban search and rescue. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*. IROS'03. 3, 3396-3403.

[18] Johnson, C.A., Adams, J.A., Kawamura, K. 2003. Evaluation of an enhanced human-robot interface. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*. SMC'03. 900-905.

[19] Kaber, D.B., Wright, M.C., Sheik-Nainar, M.A. 2006. Investigation of multi-modal interface features for adaptive automation of a human–robot system. In *International Journal of Human-Computer Studies*, 64, 6 (Jun. 2006), 527-540.

[20] Kennedy, R.S., and Land, N.E., 1993. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *The International J. of* Aviation *Psychology*, 3, 3, 203-220.

[21] Koslover, R.L. et al. 2012. Mobile Navigation Using Haptic, Audio, and Visual Direction Cues with a Handheld Test Platform. *IEEE Transactions on Haptics*. 5, 1 (Jan. 2012), 33–38.

[22] Lindeman, R.W. 2003. Virtual contact: the continuum from purely visual to purely physical. In *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society*. HFES'03, 2103-2107.

[23] Lindeman, R.W., Noma, H., and de Barros, P.G. 2007. Hear-Through and Mic-Through Augmented Reality: Using Bone Conduction to Display Spatialized Audio. *IEEE Int'l Symp. on Mixed and Augmented Reality*, 1-4.

[24] Lindeman, R.W. and Cutler, J.R. 2003. Controller design for a wearable, near-field haptic display. In *Proceedings of the 11th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 397-403.

[25] McFarlane, D.C. and Latorella, K.A. 2002. The scope and importance of human interruption in human-computer interaction design. *Human*-Computer *Interaction*, 17, 1-61.

[26] Micire, M. et al. 2011. Hand and finger registration for multi-touch joysticks on software-based operator control units. 2011 IEEE Conference on Technologies for Practical Robot Applications (Apr. 2011), 88-93.

[27] Mine, N., Brooks Jr., F.P., Sequin, C. 1997. Moving Objects in Space: Exploiting Proprioception in Virtual-Environment Interaction, *Proc. of SIGGRAPH*, Los Angeles, CA, 19-26.

[28] Nielsen, C.W., Goodrich, M.A., and Ricks, B. 2007. Ecological interfaces for improving mobile robot teleoperation. *IEEE Transactions on Robotics*, 23, 5, 927-941.

[29] Pielot, M. and Boll, S. 2010. Tactile Wayfinder: Comparison of Tactile Waypoint Navigation with Commercial Pedestrian Navigation Systems. (2010), 76–93.

[30] Pielot, M. and Poppinga, B. 2012. PocketNavigator: Studying Tactile Navigation Systems. (2012), 3131–3139.

[31] Raisamo, R. et al. 2012. Orientation Inquiry: A New Haptic Interaction Technique for Non-visual Pedestrian Navigation. EuroHaptics Conference (2012), 139–144.

[32] Sibert, J., Cooper, J., Covington, C., Stefanovski, A., Thompson, D., and Lindeman, R.W. 2006. Vibrotactile feedback for enhanced control of urban search and rescue robots. *Proc. of the IEEE* Symposium *on Safety, Security and Rescue Robots,* Gaithersburg, MD, Aug. 2006.

[33] Sigrist, R. et al. 2013. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. Psychonomic bulletin & review. 20, 1 (Feb. 2013), 21–53.

[34] Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., and Goodrich, M. 2006. Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, 33-40.

[35] Usoh, M., Catena, E., Arman, S., and Slater, M. 2000. Using presence questionnaires in reality. *Presence-Teleoperators and* Virtual *Environments*, 9, 5, 497-503.

[36] Thullier, F. et al. 2012. Vibrotactile Pattern Recognition: A Portable Compact Tactile Matrix. IEEE Transactions on Biomedical Engineering. 59, 2 (Feb. 2012), 525–530.

[37] Van Erp, J.B.F. and Van Veen, H.A.H.C. 2004. Vibrotactile in-vehicle navigation system. Transportation Research. F, 7 (2004), 247–256.

[38] Yanco, H.A., Baker, M., Casey, R., Keyes, B. Thoren, P., Drury, J.L., Few, D., Nielsen, C., and Bruemmer, D. 2006. Analysis of human-robot interaction for urban search and rescue. In *Proceedings* of *the IEEE Symposium on Safety, Security and Rescue Robots*.