Vol. 2, No. 2, Haptics-e, October 17, 2001, http://www.haptics-e.org

Virtual Training for a Manual Assembly Task

Richard J. Adams *
Department of Electrical Engineering
University of Washington
rjadams@u.washington.edu

Daniel Klowden †
Department of Computer Science and Engineering
University of Washington
dklowden@cs.washington.edu

Blake Hannaford *
Department of Electrical Engineering
University of Washington
hannaford@ee.washington.edu

ABSTRACT

This paper describes an experiment conducted to investigate the benefits of force feedback for virtual reality training of a real task. Three groups of subjects received different levels of training before completing a manual task, the construction of a LEGO™ biplane model. One group trained on a Virtual Building Block (VBB) simulation, which emulated the real task in a virtual environment, including haptic feedback. A second group was also trained on the VBB system, but without the force feedback. The last group received no virtual reality training. Completion times were compared for these different groups in building the actual biplane model in the real world ANOVA analysis showed a significant change in performance due to training level.

1 INTRODUCTION

Virtual reality systems represent a powerful tool for training humans to perform tasks which are otherwise expensive or dangerous to duplicate in the real world. The idea is not new. Flight simulators have been used for decades to train pilots for both commercial and military aviation. These systems have advanced to a point that they are integral to both the design and the operation of modern aircraft. Virtual reality systems are also making headway into training for manned space operations. In 1993, NASA used an immersive virtual environment to train flight team members for a Hubble Space Telescope repair mission [1]. They concluded the virtual training system had a beneficial effect on crew performance during flight operations. The U.S. military is aggressively pursuing networked virtual environments for the distributed simulation of integrated combat operations [2]. This technology will allow diverse land, sea, and air elements to train together in complex scenarios involving both real and autonomous agents. The military is also interested in virtual reality systems for maintenance training. The National Guard is investigating a virtual training system for maintenance and trouble shooting tasks on the M1A1 Abrams tank, the M2A2 Bradley fighting vehicle, and the TOW II missile system [3].

All of the above examples of virtual reality training make extensive use of advanced computer graphics. Some of them incorporate audio feedback, but none provides force cues to the user. When the task to be performed involves the manual manipulation of objects, the need for haptic feedback becomes evident. In order for ‘virtual reality’ to better approximate ‘reality’ for manual tasks, the senses of touch and kinesthesia must be addressed.

Klatzky et al. (1985) [4] introduced the term "haptics" to the engineering community and studied "exploratory procedures" in which subjects identified objects through manipulation with the hands. In contrast, our study used a haptic device to simulate the creation of assemblies of a well known and visible object.

Force feedback is beginning to find its way into virtual reality training systems. In Clover et. al. [5] a PUMA 560 robot was used to simulate the control column forces of a Boeing 777 aircraft. NASA has also taken steps to add haptic information to their virtual reality simulators [6]. The Charlotte™ manipulator was used to replicate the feel of massive payloads handled by EVA crew members [7]. Some of the most exciting applications of force feedback are found in surgical simulations. Much of this research has focused on training for minimally invasive procedures [8], [9], [10].

The study described in this paper addressed a fundamental issue common to many virtual training systems. How well does training in a virtual environment transfer to a manual task in the real world? A fairly elementary task, the construction of a LEGO™ biplane model, was used to investigate the problem. A virtual mock-up of biplane assembly, incorporating both visual and haptic feedback, provided the training platform

2 APPARATUS

The VBB system consists of the Excalibur force display and virtual environment software, which emulates the behavior of LEGO™ blocks. In one hand, the user grasps the handle of the force display. In the other hand, the subject holds a two-button wireless mouse. The operator sees a 3-D graphical representation of the scene on a large monitor and feels a 3-D haptic rendering of the scene through the Excalibur display. Fig. 1 shows the VBB system.

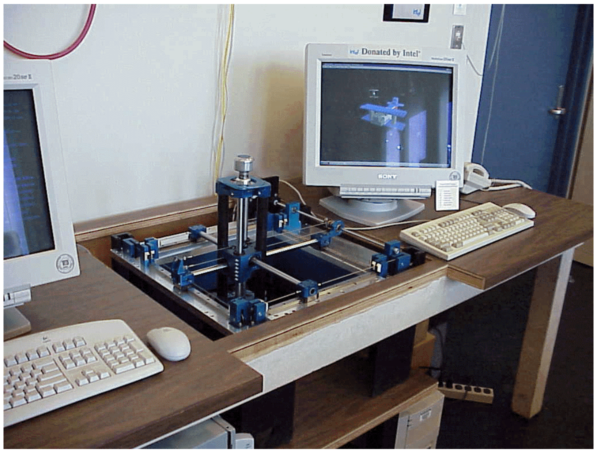

Fig. 1. The Virtual Building Block System

2.1 Excalibur Force Display

The Excalibur force display is a three degree-of-freedom Cartesian manipulator, conceived to act as a haptic interface to virtual or remote environments. Its brushless motors generate control forces through a steel cable transmission along three mutually orthogonal translational axes. The human test operator grasps a handle mounted on the end effector. Excalibur is capable of rendering peak forces of up to *200 N* and continuous forces of up to *100 N* in each axis over the workspace of $300x300x200\ mm^3$. It is controlled by a 266 MHz Pentium II™ PC executing real-time software at *1000 Hz*. Optical encoders supply measurements of resolved motor position with a resolution of *0.008* mm. Additional details on the Excalibur device can be found in [11].

2.2 Virtual Building Block Software

The VBB system simulates the behavior of a collection of LEGO™ blocks which can be manipulated by the human test operator using the 3-DOF Excalibur interface. The critical functions of device I/O, control, collision detection, and virtual environment dynamics are performed on a Pentium II™ PC at a rate of 1000 Hz. A second PC, equipped with OpenGL™ acceleration, is dedicated solely to graphics. The two computers communicate through a serial connection.

The software has two primary modes of operation: SELECT and PLACE. In SELECT mode, Excalibur motion drives a wire-frame cursor in the virtual environment. When this cursor is moved inside an individual block and the user holds down the left mouse button, the software switches to PLACE mode. In PLACE mode, Excalibur movement drives the selected block. When this block comes into contact with others in the environment, a collision takes place and the selected block is constrained. If haptic feedback is activated, the user feels the inter-block reaction forces. When the selected block is properly aligned with another in the horizontal plane, they can be ‘snapped’ together along the vertical axis. Since Excalibur is a three-axis device, only translational movements are possible. The blocks always remain orthogonally aligned with each other.

When the software is in SELECT mode, the operator can hold down the right mouse button and use the cursor to ‘cluster’ a group of blocks together. The clustered bricks can then be moved as an assembly. Keyboard commands are used to rotate the working view of the environment right, left, up, down, in, or out. The virtual model can also be flipped upside-down to easily permit work on the underside. The VBB software currently supports 10 different types of blocks, with up to 500 blocks simulated at once.

The VBB system generates compelling forces through the use of a virtual coupling network which guarantees the stability of the simulation [12]. This coupling acts as a spring-damper link between the Excalibur display and the virtual environment. A virtual coupling stiffness level of *75,000 N/m* and a damping of *90 N/(m/s)* provide a rigid and stable feel to collisions between virtual blocks and to constraints when blocks are snapped together. Details of virtual coupling network design are found in [13].

3 EXPERIMENTAL DESIGN

The purpose of the VBB training study was to assess the influence of haptic feedback on the efficacy of virtual reality training. A total of 15 subjects (engineering graduate students in their 20's) were exposed to one of three different treatments: virtual training with haptics, virtual training without haptics, and no virtual training. The subjects understood in general terms that the purpose of the experiment was to measure the performance of groups who received different forms of task training. Nearly inaudible sounds emitted by the device were the same in both virtual training conditions. The motors were active in both conditions, because gravity and friction compensation was active in both VBB treatments.

They then built a real LEGO™ biplane model five times in succession. The dependent variable was completion time for the real model. Fig. 2 shows the virtual and real LEGO™ biplane models.

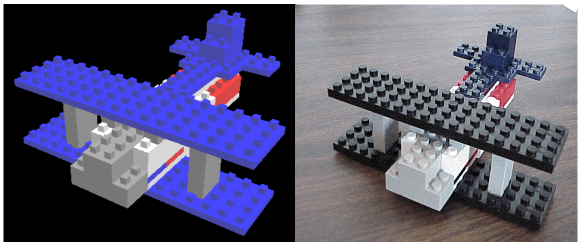

Fig. 2. Virtual and Real Biplane Models

There are two primary considerations in designing a suitable experiment for the VBB training study: skill transfer and subject aptitude. The nature of the training study precludes the use of a within-subjects design, in which each individual is subjected to all treatments. Once a subject has undergone one level of training and performed the actual biplane construction task, the skills learned will not be forgotten if the subject is then re-trained at a different level. There is a very high degree of skill transfer from one treatment to another. It is also likely this transfer will be asymmetrical, that is, skill transfer from one treatment to another will be different if the order in which the treatments are applied is switched [14]. This drives us to a between-subjects design, in which each individual is subjected to only one treatment.

The choice of a between-subject design leads to a second problem, subject aptitude. Every individual has a certain level of ability to construct LEGO™ models. One way to control for variation in individual subject aptitude is by randomly assigning a sufficient number of subjects to each treatment. Unfortunately, the number of subjects required to minimize the probability of forming unequal groups, thus biasing results, can be excessive. The higher the variance in inter-subject aptitude, the greater the required number of subjects. When it is impractical to use a large number of subjects, an alternative may be to form matched groups [14].

3.1 Matching Test

For the VBB training study, a matching test was formulated to provide a measure of each subject's aptitude for building the LEGO™ biplane model. This test consisted of having each subject build a LEGO™ model of a hydrofoil boat. The boat model, consisting of 29 pieces, was of similar complexity to the biplane model, made up of 37 pieces. The subjects (male and female engineering students in their 20's) first watched the examiner build the hydrofoil model once. They then built the model three times in succession, with a short break between each iteration. The average completion time of the three iterations was used as the variable for forming the matched groups.

The matching test was completed at least two weeks prior to administration of the actual experiment. The likelihood of significant skill transfer, in the form of increased LEGO™ assembly aptitude, from the matching test to the experimental task was considered low. If any skill transfer did take place, it should apply equally to all three treatment groups. The hydrofoil boat construction pre-test was therefore not prone to bias the results of the VBB study.

A total of 15 subjects were pre-tested. These subjects were then rank ordered from 1 to 15 according to their matching test score and grouped into 5 matched subject triads. Each of the three subjects in a triad was randomly assigned to one of the three treatments. Thus a total of 5 subjects, one from each triad, received each treatment.

3.2 Treatments

The treatments in this study were the levels of training received before building the real biplane model. Prior to the treatment, each subject watched a 4-minute video on biplane construction. This video accomplished three things. First, it was the only communication received by the subjects regarding the procedures and nature of the experiment. This allowed us to eliminate a potential source of bias due to unconscious cues from the experimenter. Second, it communicated to the subjects the way we evaluated performance (i.e., requiring perfect assembly of the model). Finally, it was the means by which we could specify the airplane design to the subjects since some internal parts are not visible on the assembled model.

- *Treatment 1* - virtual training with haptics. After completing a VBB system familiarization task, the subject watched the video on biplane construction. The subject then practiced building the biplane for 30 minutes on the VBB system *with* force feedback.
- *Treatment 2* - virtual training without haptics. After completing a VBB system familiarization task, the subject watched the video. The subject

then practiced building the biplane for 30 minutes on the VBB system *without* force feedback.

- *Treatment 3* - no virtual training. The subject watched the video and directly proceeded to the real task.

3.3 Experimental Procedure

After undergoing one of the three treatments, each subject then built the real LEGO™ biplane 5 times in succession, with a short break between iterations to allow the examiner to disassemble the model. Before beginning, the subject was instructed to complete the biplane as rapidly as possible, but to ensure the final product exactly matched the desired model. A second, completed example of the LEGO™ biplane was available to the subject for inspection during the construction task. When the biplane model was successfully completed, the subject's completion time was recorded.

4 RESULTS

Table 1 shows completion times for the real LEGO™ biplane test. Results are presented for only 14 of the original 15 subjects. One subject in the third treatment group was not able to complete the biplane experiment.

Table 1.
LEGO™ Model Completion Times in Seconds

treat. group	match. triad	iter. 1	iter. 2	iter. 3	iter. 4	iter. 5	test mean	pre-test mean
1	1	80	64	57	62	78	68	68
1	2	92	80	72	80	67	78	73
1	3	109	212	105	93	100	124	92
1	4	233	170	113	95	90	140	102
1	5	172	134	90	88	88	114	110
2	1	111	104	74	80	67	87	62
2	2	145	180	97	78	68	114	80
2	3	173	101	92	80	88	107	94
2	4	188	144	101	100	94	125	98
2	5	293	158	209	83	75	164	106
3	1	235	109	104	82	80	122	70
3	2	279	151	112	107	114	153	87
3	3	194	157	93	107	97	130	94
3	5	240	199	148	142	146	175	124

4.1 Analysis of Variance

The data in Table 1 was analyzed using a one-factor block design analysis of variance (ANOVA) model of the form,

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} . \qquad (1)$$

y_{ij} is the data (completion time) for the ith treatment (i = 1, 2, or 3) and jth matched subject triad (j = 1, 2, 3, 4, or 5). μ is the intercept, or common effect . It can be thought of as the overall test mean. τ_i is the effect of the ith treatment. β_j is the block effect of the jth triad. ε_{ij} is the error term associated with the ith treatment and jth triad. By including the block effect in the model, we account for a very large portion of inter-subject variability, which would otherwise be included in the error term.

We were interested in two hypotheses. The main hypothesis is that the effects of the three different treatments are identical,

$$H_0\text{: } \tau_1 = \tau_2 = \tau_3 \qquad (2)$$

If this hypothesis is rejected, then we can say VBB training has a significant effect on performance in the real task. The second hypothesis is that the block effects for all 5 matched subject triads are identical,

$$H_1\text{: } \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 \qquad (3)$$

If H_1 is rejected, then we can say inter-subject variability, as predicted by the matching test, is significant.

The ANOVA analysis was conducted with the aid of the SPSS™ statistical General Linear Model procedure. All hypotheses were tested at the *5%* significance level. Each iteration, presented as columns in Table 1, was treated as an independent experiment. The two hypotheses were evaluated for all 5 iterations as well as for the overall average completion time (test mean). The p-level, the smallest level of significance at which the null hypothesis can be rejected, is given in Table 2 for each of these 6 cases.

Table 2. *p*-levels for VBB Study

	iter. 1	iter. 2	iter. 3	iter. 4	iter. 5	test ave.
H_0	* .040	.656	.342	* .018	.052	* .027
H_1	.182	.336	.167	.093	.296	* .039

* - significant at the *5%* level

The ANOVA analysis indicated a significant performance difference ($p < 0.05$) due to training level for the first and forth iterations of biplane construction.

The average time also showed a significant training effect. Fig. 3 shows the progression of the mean

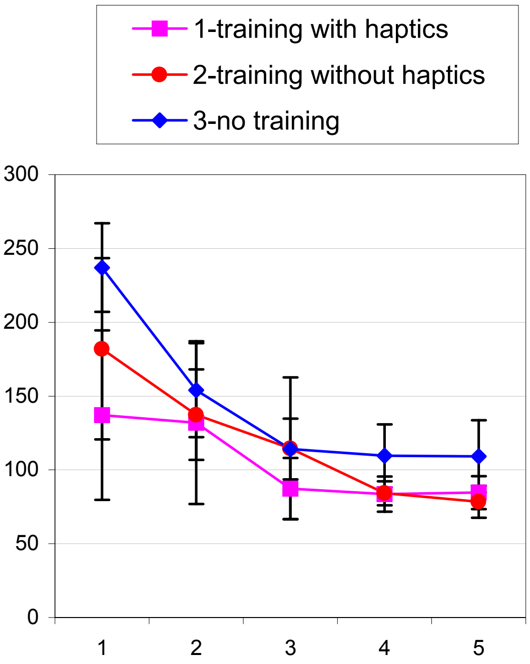

Fig. 3. Mean Biplane Completion time Versus Iteration

completion time for each treatment group with iteration number.

The p-levels for test average in Table 2 were included to highlight the success of using matched groups for this experiment. The block effect for this measure is statistically significant at the *5%* level. A quick glance at the test average column in Table 1 confirms the individual completion times for each treatment group align closely with the rank order predicted by the matching pre-test. A correlation analysis between pre-test average and actual test average, controlling for treatment level, showed a correlation of 0.84. Correlation levels on the order of 0.45 have been shown to be sufficient to justify the use of matched groups for studies with similar numbers of subjects [14]. To highlight the importance of including matched subject triad as a blocking effect in the model (1), we reran the analysis using a simple one-factor ANOVA. We previously showed treatment level for the first iteration to be significant with $p = .040$. Without the blocking term, the p-level rises to .084, incorrectly leading us to accept the null hypothesis, H_0.

4.2 Post Hoc Analysis - Tukey HSD

The results presented above only allowed us to conclude that there was a significant treatment effect for the first and fourth iterations. Additional analysis is necessary to determine the relative importance of training with or without haptic feedback. A widely accepted approach to conducting pair-wise comparisons of treatment effects is the Tukey Honest Significant Different (HSD) test [15]. Tukey HSD tests the hypothesis that two treatments are equivalent while controlling for the overall Type I error rate (the probability of incorrectly rejecting the null hypothesis). The results of Tukey HSD analysis are shown in Table 3 for the two iterations where H_0 is rejected.

Table 3.
Pair-Wise Comparisons Using Tukey HSD

hypothesis	iter 1	iter 4
$\tau_1 = \tau_3$	* .049	* .037
$\tau_2 = \tau_3$	.295	* .041
$\tau_1 = \tau_2$	.386	.997

* - significant at *5%* level

For iteration 1, training with haptics showed a significant improvement over no training at all. Completion times for those trained without haptics were not significantly different from those trained with haptics or those with no training. For iteration 4, subjects trained on the VBB system both with and without haptics performed significantly better than those with no training. Again though, the difference between those trained with haptics and those trained without is not shown significant.

The VBB training study succeeds in showing a significant benefit in the use of virtual reality training with force feedback. Results were inconclusive as to whether training with haptics improves performance versus training without force feedback. The variance in human performance (completion time) is high in comparison to mean differences between types of training. Data from this study implies at least twice the number of subjects (30) would be required to show a significant difference between treatments 1 and 2 for the first iteration of biplane construction.

4.3 Discussion

It has been suggested that human training can be broken down into three components: cognitive, perceptual, and motor demands (see for example [16]). The cognitive portion of building block training consists primarily of the construction of an internal model of the task within the user's memory. The trainee gradually learns where each particular piece fits within the final model and develops a strategy for putting the model together. During initial trials, the subject is likely to make mistakes, but in later iterations, the user works without hesitation and without faults. The motor demands of the task concern the dexterous manipulation of the pieces. The trainee must learn how to handle the pieces, move them within the environment, orient them, and connect them together. Poor command of task motor demands leads to slow progress, drops, and misaligned pieces. The perceptual aspect of training may be thought of as the glue which binds the cognitive and motor portions to force the trainee's comprehension of the environment. The blocks have certain properties which the user perceives through his or her senses. These include spatial characteristics such as shape, size, and orientation and physical properties such as rigidity, mass, and friction.

The addition of haptic feedback to a virtual reality system can contribute to all three components of training. However, the component which seems most likely to benefit from force feedback, motor demand, is actually the hardest one to address. Tasks requiring fine dexterous hand and finger movements are extremely difficult to duplicate using a haptic device. Perfectly replicating the motions, forces, and textural information which take place in the manipulation of small building blocks is far beyond the capabilities of the Excalibur (or any existing system). Motor demands for coarser tasks, such as certain industrial drilling and machining operations, are more likely to be reproduced by modern force feedback displays. It is the perceptual component of training which appears to benefit the most from the VBB system. The ability to feel blocks as they bump into each other and fit together is a powerful addition to the virtual training system. The VBB system conveys haptic information on the shape, size, and orientation of blocks which greatly enhances the users' perception of the environment. This enhanced sense of immersion allows the trainee to operate more effectively in the virtual world and more rapidly construct a cognitive model of the task.

Subjects who had the benefit of force feedback completed the biplane in the virtual world an average of twice in 30 minutes. Subjects who trained without haptics completed the virtual model an average of only one time in this period. Notice that even with haptics, it took users an average of 15 minutes to build the virtual biplane compared to less than 4 minutes for the first iteration of untrained users working on the real model.

The fact that the initial iteration is significantly affected by VBB training is not surprising. It is this first trial when untrained subjects, having only passively watched the biplane built in a 4 minute video, struggle the most to figure out how the pieces are arranged. The subjects who have had the benefit of the VBB training system had already formed an internal model of the process in their memories. If the virtual environment adequately represents the real task, this internal model assists these subjects in performing the initial biplane assembly. It is also understandable that the training effect was less pronounced in subsequent iterations. Having performed the actual task once or more, the untrained subject is much better equipped to handle the task. After enough iterations of the real task have been performed, we would expect the level at which a subject is pre-trained to become insignificant.

It was surprising to find a significant training effect in the fourth iteration of biplane construction, with all subjects having had the opportunity to build the actual model three times. Further analysis of the data suggests that although the mean differences in completion times were small at this latter stage, the variance in the data decreased faster (with iteration) than the mean. The overall standard deviation in first iteration data was 68.5. At the forth iteration, this value had decreased to 20.7. In hindsight, more iterations should have been added to verify this trend and that in steady-state, no significant difference in performance was present.

5 CONCLUSIONS

We conducted a study to assess the benefits of force feedback for virtual training in a real manual task. Three groups of subjects received different levels of virtual training for assembly of a LEGO™ biplane model: virtual training with haptics, virtual training without haptics, and no training. After training, all subjects then constructed the real LEGO™ model five times in succession. Analysis of completion times for the real task reveals that subjects trained with force feedback performed significantly better than those who received no training. Although average completion times for subjects trained with force feedback were better than those for subjects trained without, high variance in the data prevents a significant difference from being shown. A greater number of subjects will be required in any future study if this distinction is to be confirmed.

ACKNOWLEDGMENTS

We would like to thank the U.S. Air Force, for support of Captain Rick Adams during this research, Boeing Defense and Space for support of the Excalibur development, and the University of Washington Department of Statistics for consultation on statistical methods and experiment design.

REFERENCES

[1] Loftin, R.B., and Kenney, P.J., "Training the Hubble Space Telescope Flight Team," *IEEE Computer Graphics and Applications*, vol. 15, no. 5, pp. 31-37, Sep, 1995.

[2] Mastaglio, T.W., and Callahan, R., "A Large-Scale Complex Virtual Environment for Team Training," *Computer*, vol. 28, no. 7, pp. 49-56, July, 1995.

[3] McLin, D.M., and Chung, J.C., "Combining Virtual Reality and Multimedia Techniques for Effective Maintenance Training," *Proceeding of the 24th AIPR Workshop: Tools and Techniques for Modeling and Simulation, SPIE*, vol. 2645, pp. 204-210, 1996.

[4] Klatzky, R.L., Lederman, S.J. and Metzger, V.A., "Identifying Objects by Touch: An Expert System," Perception and Psychophysics, vol. 37, no. 4, pp. 299-302, 1985.

[5] Clover, C.L., Luecke , G.R., Troy, J.J., and McNeely, W.A., "Dynamic Simulations of Virtual Mechanisms with Haptic Feedback Using Industrial Robotics Equipment," *Proc. IEEE Int. Conf. Robotics and Automation*, Albuquerque, NM, pp. 3205-10, 1997.

[6] Carignan, C.R., and Akin, D.L., "Actively Controlled Mockups for EVA raining in Neutral Buoyancy," *Proceedings of IEEE Conf. on Systems, Man, and Cybernetics*, pp. 2369-74, 1997.

[7] Swain, P., Thompson, C., and Campbell, P., "The Charlotte™ Intra-Vehicular Robot," N95-23703, 1995.

[8] Baumann, R. and Clavel, R., "Haptic Interface for Virtual Reality Based Minimally Invasive Surgery Simulation," *Proceedings IEEE Int. Conf. on Robotics and Automation*, pp. 381-6, 1998.

[9] Edmond, C.V. et. al., "ENT Endoscopic Surgical Training Simulator," *Proceedings of Medicine Meets Virtual Reality*, pp.518-28, 1997.

[10] Ward, J.W. et. al., "The Development of an Arthroscopic Surgical Simulator with Haptic Feedback," *Future Generation Computer Systems*, vol.14, no.3-4, pp. 243-51, 1998.

[11] Adams, R.J., Moreyra, M.R., Hannaford, B., "Excalibur - A Three Axis Force Display," *Proc. ASME International Mechanical Engineering Congress and Exhibition*, Nashville, TN, 1999.

[12] Adams, R.J. and Hannaford, B., "Stable Haptic Interaction with Virtual Environments," *IEEE Transactions on Robotics and Automation*, vol. 15, No. 3, pp. 465-474, 1999.

[13] Adams, R.J., "Stable Haptic Interaction with Virtual Environments," Ph.D. Thesis, Department of Electrical Engineering, University of Washington, 1999.

[14] Pitrella, F.D., and Kruger, W., "Design and Validation of Matching Tests to Form Equal Groups for Tracking Experiments," *Ergonomics*, vol. 26, no. 9, pp. 833-845, 1983.

[15] Cobb, George W., *Design and Analysis of Experiments*, New York, NY: Springer, 1997.

[16] Philbin, D.A, Ribarsky, W., Walker, N., and Hubbard, C.E., "Training in Virtual Environments: Analysis of Task Appropriateness*," Proceeding of the IEEE 1998 Virtual Reality Annual International Symposium*, pp. 210, 1998