Workload and Performance Analyses with Haptic and Visually-Guided Training in a Dynamic Motor Skill Task

Joel C Huegel and Marcia K O'Malley

Abstract This chapter presents the implementation of a progressive haptic guidance scheme for training in a dynamic motor skill acquisition task similar to some dynamic surgical tasks. The training is administered in a haptic and visual virtual environment. The results of the task training protocol concurrently compare the performance and workload of the proposed haptic guidance scheme to a similar visual guidance scheme and to virtual practice with no guidance. The human-user training protocol lasted eleven sessions over a two-month period. The computerized version of the NASA Task Load Index (TLX) was administered to all participants during each session, thereby providing subjective workload data across the entire protocol. The analysis of the experimental results demonstrate that only early in the protocol, the progressive haptic guidance group outperforms all other groups. The workload analysis suggests that participants using the proposed haptic scheme have a significantly lower mental load and report less frustration than the others. These findings can be transferred to other virtual training environments used for surgical task training.

1 Introduction

Haptically enabled virtual environment (VE) technologies are being used for skill training applications in such areas as medical procedures, sports skill acquisition, and rehabilitation. These technologies provide reliable data acquisition, analysis, feedback, and evaluation of motor skill task performance while simultaneously providing a comparatively low-cost and low-risk training platform. Virtual environments used for training intend to reduce risk, improve and accelerate learning over

---

Joel C Huegel
Tecnologico de Monterrey-Campus Guadalajara, Guadalajara, Mexico, e-mail: jhuegel@itesm.mx

Marcia K O'Malley
Rice University, Houston, TX e-mail: omalleym@rice.edu

traditional training schemes, and transfer what is learned in the simulation environment to the equivalent or targeted real world task [10]. Virtual training environments (VTEs) are implemented either to provide an environment for practice that is as similar as possible to the real task or to act as an assistant by augmenting the feedback in some way during training. Examples include simulators for surgery residents and head-up displays (HUDs) for pilots. Haptics can play an important role both in improving simulation fidelity and in providing augmentation during training. Despite the current use of VTEs, what is still being debated in the research community is the measurable benefit of haptically augmented VTEs. This paper compares the performance and workload of groups of participants utilizing a progressive haptic guidance scheme to utilizing similar visual and written guidance schemes as well as virtual practice in a motor skill task carried out in a long-term training protocol.

Short-term performance can be improved when haptic feedback is provided. In fact, previous studies have shown that the addition of haptic feedback to VEs can provide benefits over visual and auditory displays for performance enhancement, increasing dexterity and the sensation of realism and presence [3,4,12,15]. However, there exist only a limited number of published studies aimed at determining the long-term training efficacy and outcomes of VTEs that provide augmentation or guidance in the objective task. The results that do exist are inconclusive or contradictory. For this paper we define skill performance to be a measure of output or ability in the task being studied while skill training is the protocol designed to improve performance over a period of time. In prior work, Huegel et al., summarized the key issues in developing haptic guidance training schemes [7]. First, the task to be studied must be difficult enough to require multiple sessions to master. Second, the guidance must be removed progressively to avoid dependence, and third, the amount of guidance must be based on significant task measurements. Later Huegel et al. designed and implemented a haptic guidance scheme that demonstrated these three characteristics [8]. In this paper, we also consider the human-factors requirement that the haptic guidance should not cause an increase in the associated subjective workload.

Currently, very few published studies have investigated the workload effects of haptic guidance or even augmentation during long-term training. Kalawski et al. provided a top-down system engineering overview to understanding human factors surrounding virtual environments in general but did not specifically address haptics or augmentation [9]. The seminal work by Tan et al. on human factors in the design of haptic interfaces covers the psychophysical measurements of human kinematics and forces but not subjective measurements of workload [17]. Zhou et al. recently investigated spare cognitive capacity of surgeons training with haptic feedback in a laparoscopic procedure [19]. The participants were asked to solve two digit multiplications in their mind while performing the surgical skill on a simulator. Zhou found that the participants tended to pause to solve the math problem and simply took longer to complete the surgical task. When in the condition of cognitive loading, both novices and experts could perform the task faster with the haptic feedback. This study included haptic feedback, but did not include hapic guidance per se. Furthermore, Zhou was interested in cognitive capacity rather than the workload of the task itself. When Rosenberg introduced the concept of virtual fixtures he suggested that mental workload could be reduced but did not investigate it further [14]. For extensive discussions regarding performance and workload see work by P.A. Hancock [5]. Hancock has addressed such issues as the effects of control order, input device types, and augmented feedback [5]. While Hancock did extend the research to encompass both augmentation and training, to the authors' knowledge, there has not been an investigation of workload and haptic guidance during long-term training. An increased workload might be justified if there is a significant increase in performance, however a thorough investigation must be conducted to evaluate the cost benefit [9]. With these considerations in mind, we selected to use the NASA-TLX developed by Hart and Staveland for assessment of subjective workload [6]. Thus, in this paper, measurements of both motor skill performance and subjective workload are recorded and analyzed, thereby investigating both concurrently.

The paper is organized as follows: Section 2 presents the methods used including the apparatus and VE, task description, experiment design, guidance schemes to be implemented, participant description, and data analysis. Section 3 presents the results while Section 4 discusses the findings and contributions. Section 5 draws the conclusions of this experiment.

2 Methods

A training experiment was conducted over two months to concurrently compare the performance and subjective workload associated with the proposed haptic guidance scheme, a similar visual guidance scheme, a written guidance scheme (via printed instructions), and virtual practice. The participant training was performed in a VE dynamic task as shown in Fig. 1. Participants completed the computerized version of the NASA Task Load Index throughout the training protocol, thereby obtaining six workload scores from the eleven sessions for analysis and comparison across the four guidance modes.

2.1 Apparatus, Virtual Environment, and Task Description

The experimental setup is depicted in Fig. 1. The setup was comprised by a nineteen inch LCD video display and a two degree of freedom (DOF) force feedback joystick (Immersion IE2000). The chosen virtual environment was two point masses connected by a spring and damper in parallel as shown in the inset of Fig. 1 and previously documented and utilized by Huegel et al. [8]. All participants, regardless of group assignment, received visual and haptic feedback of the moving masses and targets via the LCD display and joystick.

The task chosen met the task-complexity criteria for as set forth by Todorov et al. for studying training effects in the presence of haptic feedback and augmentation [18]. Additionally, according to O'Malley et al., the reflection of the force feedback generated by the dynamic interactions necessitate accurate human control for success [13]. The task, illustrated in Fig. 1, was to manipulate the motion of the point mass m_1 via the 2-DOF haptic joystick, and thus indirectly control the mass m_2 through the system dynamics in order to hit as many of the diagonally placed targets as possible during each 20-second long trial.

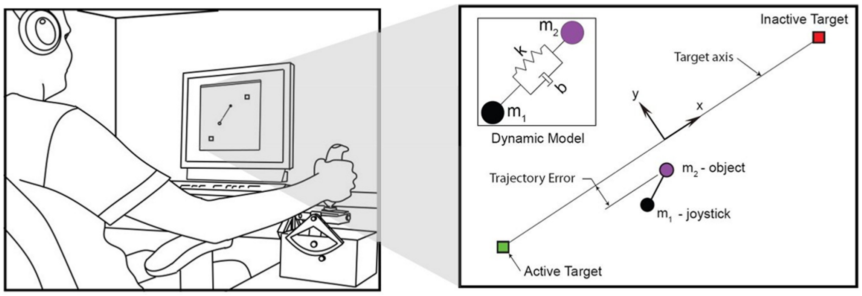


Fig. 1 A participant is sitting at the VTE (left). The VTE includes a visual feedback display and a haptic joystick for force feedback, both of which provide feedback of the system dynamics to all trainees regardless of guidance scheme. Target hitting task (right). The participant controls the location of m_1 in order to cause m_2 (object) to hit the desired target. Inset shows the virtual underactuated system (adapted from [8]).

2.2 Experiment Design and Guidance Schemes

In order to concurrently measure and compare the performance and workload effects of haptic and visual guidance to verbal guidance as well as virtual practice we designed an experiment consisting of one evaluation session, followed by nine training sessions, and one retention session for a total of eleven sessions as shown in Fig. 2. Our protocol, similar to the one implemented by Hancock, studies performance and workload effects [5]. The nine training sessions were spaced two to five days apart. The retention session was at least 30 days later. Each training session contained three subsessions as shown in Fig. 2. Each trial lasted 20 seconds for a total session duration of approximately nine minutes.

Subjective workload was measured using the NASA Task-Load Index (NASA-TLX), developed by Hart and Staveland [6] and later implemented in a computer based questionnaire. Prior to the experiment, each participant was provided with a description of the NASA-TLX. The computerized version of the NASA-TLX was integrated to the dynamic task testing such that as soon as the participant finished the pre-guidance and guidance subsessions, the first of the two-step procedure of the NASA-TLX appeared on-screen. In the first step participants rated their perceived workload demand on each of six scales: mental, physical, temporal, performance, effort, and frustration. Then, in the second step, the participants were asked to pairwise compare all six scales, thereby weighting them. Percentage scores were then computed.

Three types of haptics-enabled virtual training schemes have been proposed, thereby exploiting the capabilities of virtual training environments. One scheme is to first present the performance of an expert (human or robotic) to a trainee via visual and haptic feedback, and then allow the trainee to practice the task unassisted [16]. A second approach, *virtual fixtures*, requires the trainee to perform the task with enforced restrictions or reductions of the degrees of freedom of the task ([1, 2, 14]). A third approach, shared control, modifies the dynamics of the system so as to encourage the correct behavior from the trainee ([3, 4, 13]). We select a virtual wall design for the trajectory error guidance and a shared control design for the input frequency guidance. Both visual and haptic guidance scheme designs were replicated from prior work by Huegel et al. [8].

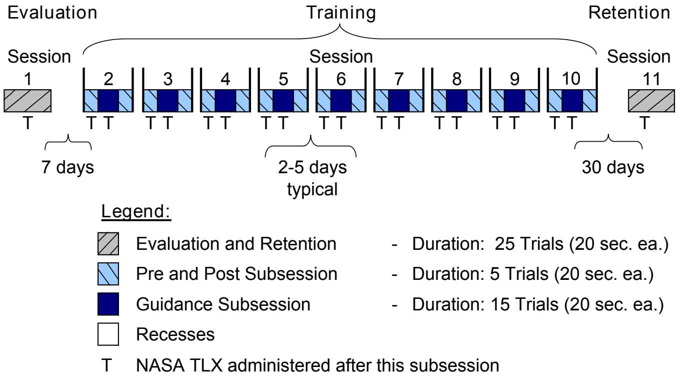


Fig. 2 Training experiment design: consists of eleven sessions. Only during guidance subsessions did the haptic and visual groups receive the corresponding progressively decaying guidance. Rest periods between sessions are indicated with braces. The NASA TLX was administered after the pre-guidance and guidance subsessions (adapted from [8]).

2.3 Participants and Data Analysis

The experiment involved 32 healthy participants, (7 female and 25 male; 30 right-handed and 2 left-handed; ages 18 to 51) primarily undergraduate students with no previous experience with haptic devices. A university IRB approved form was used to obtain informed consent. For all participants, values for the number of target hits (n_{hit}) for each subsession were determined by averaging the scores of the trials of that subsession (five trials for pre- and post-guidance subsessions and fifteen trials for the guidance subsession). Thus each of the 32 participants had three data points for each of the eleven sessions of training resulting in a total of 1056 observations.

The data were fitted with power functions. For the subjective workload analysis, each participant completed the computerized version of the NASA TLX after the first and second subsessions resulting in two data points per session averaged over the eight subjects per group for a total of 88 observations for each of the task load measures.

3 Results

We present the results and data analysis for the four guidance groups including comparisons of performance and workload. Figure 3 shows the performance of the four groups in hit count (n_{hit}) for both the guidance subsession and the post-guidance baseline subession. Pre-guidance subsession data is not included as it showed the least variation between groups. Each participants' scores from 5 trials (or 15 trials during guidance) are averaged to get one mean score per subsession. The data points plotted in Fig. 3 represent the mean of the subsession scores of the eight participants of a particular group with error bars indicating the standard error of the mean. The n_{hit} scores for all participants show increasing trends across all sessions as training progressed with saturation at approximately 23 hits. In order to visualize trends in performance across sessions that suggest learning, power functions were fitted to the data according to the following equation:

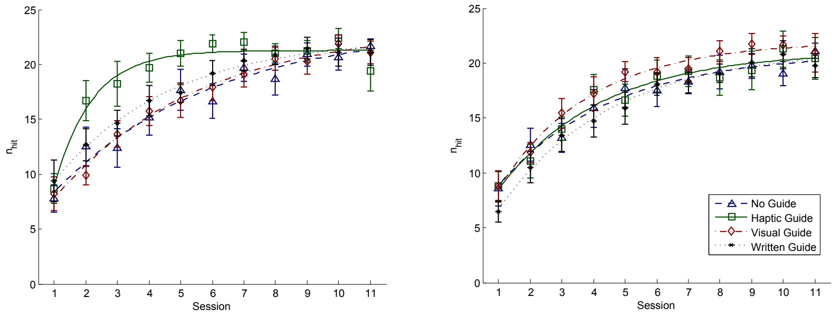
$$y = -ax^b + c \tag{1}$$

where a, b, and c are the parameters of the equation and have R^2 values in excess of 0.95. A summary of the curve fitting results, including estimated parameters and goodness of fit for each of the four groups of participants is shown in Table 1. During the guidance subsession, the haptic group reached saturation at a significantly faster rate (parameter b) than the other three groups. In other words, the haptic guidance did significantly assist in performing the task in the first seven sessions of training.

Table 1 Summary of the curve fitting procedures for the hit count data of each group

	Goodness of fit Guidance		**Goodness of fit Post-Guidance**	
Guidance Group	R^2	Fit Parameters	R^2	Fit Parameters
No-Guidance	0.96	$a = 31.7, b = 0.15, c = 23.8$	0.97	$a = -85.0, b = -0.06, c = 93.6$
Haptic Guidance	0.95	$a = -13.5, b = -1.33, c = 22.1$	0.95	$a = -56.6, b = -0.11, c = 65.0$
Visual Guidance	0.98	$a = 33.0, b = 0.15, c = 25.4$	0.97	$a = -34.3, b = -0.21, c = 42.7$
Written Guidance	0.98	$a = -85.0, b = -0.08, c = 91.4$	0.99	$a = -164, b = -0.03, c = 173.3$

The subjective workload tested via de NASA Task Load Index (TLX) also presents significant results. Figure 4 shows the overall workload computed across all sessions for the four groups. The visual guidance group reported the greatest workload significantly different than both the control and written guidance groups.



(a) Hit count During Guidance Sub-session (b) Hit count During Post-Guid. Sub-session

Fig. 3 Measure of Performance n_{hit} with increasing trends regardless of guidance mode. During guidance subsession the data shows a significantly faster rate of the haptic guidance group.

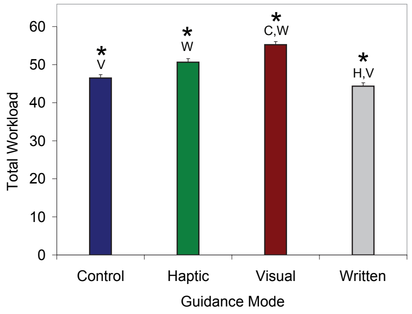
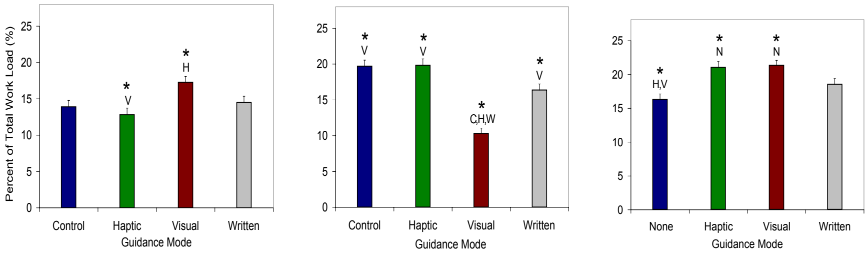


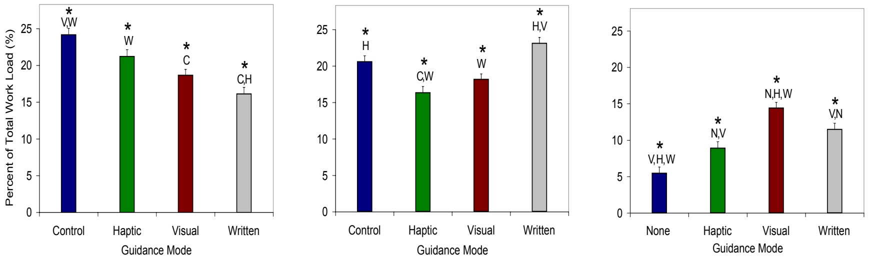
Fig. 4 Total subjective workload computed via the NASA-TLX for the four groups: control (C), haptic guidance (H), visual guidance (V), and written guidance (W). Significant differences are shown above the mean with the initials of the groups that are different. Error bars indicate standard error of the means.

The haptic guidance group also had a significantly higher score than the written group but not significantly different than the other two groups. Upon analyzing the six measures separately, more details emerge. Figure 5 shows individual subplots for each of the six measures. The error bars indicate standard error of the means of the eight participants per group. The haptic guidance group displayed significantly less mental workload than the visual group. Neither was significantly different from the control and written guidance groups. As expected, the haptic group reported significantly higher physical workload, likely due to the added forces to be contended with, while the visual group reported significantly less physical demand than the other three groups. In terms of the temporal workload, both haptic and visual groups reported greater workload than the control group. The last significant result is that the visual group reported a significantly greater frustration than any of the other three groups.

Analysis of variance on the six subjective workload measures indicated a significant main effect of guidance mode in both the pre-guidance and guidance subsessions but failed to indicate a main effect of session in any of the six measures. These results are summarized in Table 2.



(a) Mental workload index (b) Physical workload index (c) Temporal workload index



(d) Performance index (e) Effort workload index (f) Frustration workload index

Fig. 5 NASA Task Load Index. The six workload indices show the differences between the four groups of performers: Control (C), Haptic (H), Visual (V), and Written (W) guidance scheme. Error bars indicate standard errors. Significant differences are indicated with an asterisk (*) and the initial (C, H, V, or W) of the different groups. For bars not indicated with asterisks, the data failed to show significant differences.

Table 2 NASA TLX Subjective workload measures show significant main effects from the guidance mode. Asterisk (*) indicates $a = 0.05$ confidence interval.

TLX Scale	Pre-Guidance	Guidance
Mental	$0.024*$	$0.025*$
Physical	$< 0.0001*$	$< 0.0001*$
Temporal	0.25	$0.032*$
Performance	$< 0.0001*$	$0.0037*$
Effort	$0.002*$	$0.0002*$
Frustration	$< 0.0001*$	$< 0.0001*$

4 Discussion

This paper presents the results and analysis of both a performance measure (hit count) and a workload measure (NASA TLX) of a training protocol in a virtual dynamic task, thereby comparing three guidance modes to a control group. We have demonstrated that the protocol must have a significant duration so that training, rather than performance, can be evaluated. The dynamic task chosen was sufficiently difficult to demostrate performance improvements across multiple sessions as required by the Todorov criteria. The data indicates that the haptic guidance does outperform other schemes while guidance is active, but does not hold during the post-guidance session. This is an improvement, however, to prior work where an ad-hoc haptic guidance scheme never showed better performance than the control group [11]. With regard to the workload, the tradeoff between performance and workload must be evaluated. On the one hand, the addition of guidance schemes may be warranted if the workload is reduced, even if the performance is not increased. On the other hand, guidance schemes that do generate an improvement in performance, still may not be acceptable if the workload is unduly increased. Therefore, the haptic guidance schemes must demonstrate not only performance improvements but also prevent adverse changes to the workload.

5 Conclusions

This paper implemented and analyzed performance and workload of a novel progressive haptic guidance scheme for training in a dynamic motor skill acquisition task. The training was administered in a haptic and visual virtual environment. The results of the training protocol confirmed that haptic guidance accelerates performance improvements while the guidance is active. However, during post-guidance evaluations, the group that received the guidance does not significantly outperform either the visual guidance group or the control group. An additional group with written guidance (printed instructions) was included, but no significant differences were identified between this group and the control group. Throughout the training protocol the subjective workload was measured with the NASA-TLX. Significant differences in terms of guidance mode were identified both for total workload and for the six workload categories. The visual guidance group reported the highest overall workload. Moreover, this group reported significantly higher mental load and frustration while the haptic guidance group reported the lowest mental load, thereby suggesting that a well designed haptic guidance scheme, as the one implemented in this study, can be less obtrusive in terms of workload and still generate similar training results to both virtual practice and visual guidance schemes. The success of the progressive nature of the guidance suggests that while haptic guidance is beneficial early in a training protocol, it should be gradually removed to avoid potential dependence. These findings may be applied to an array of virtual environments including ones used for surgical task training, vehicle control, sports training, and rehabilitation among others.

Acknowledgements

This research was supported in part by a grant from the National Science Foundation (IIS-0448341). The authors also acknowledge support received from the Tecnolgico de Monterrey to complete the research reported in this chapter.

References

1. J. J. Abbott and A. M. Okamura. Stable forbidden-region virtual fixtures for bilateral telemanipulation. *Journal of Dynamic Systems, Measurement, and Control*, 128(1):53–64, 2006.
2. N. Bernstein. *The Coordination and Regulation of Movements*. Pergamon, London, 1967.
3. J. Emken and D. Reinkensmeyer. Robot-enhanced motor learning: accelerating internal model formation during locomotion by transient dynamic amplification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(1):33–39, March 2005.
4. P. Griffiths and B. Gillespie. Shared control between human and machine: using a haptic steering wheel and aid in land vehicle guidance. In *Proc. IEEE 12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS '04)*, pages 358–366, Chicago, IL, March 2004.
5. P. A. Hancock. Effects of control order, augmented feedback, input device and practice on tracking performance and perceived workload . *Ergonomics*, 39(9):1146–1162, 1996.
6. S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In P. Hancock and N. Meshkati, editors, *Human Mental Workload*, pages 139 – 183. Elsevier Science/North-Holland, Amsterdam, Netherlands, 1988.
7. J. C. Huegel, O. Celik, A. Israr, and M. K. O'Malley. Expertise-Based Performance Measures in a Training Virtual Environment. *Presence: Teleoperators and Virtual Environments*, 18(6):449–467, 2009.
8. J. C. Huegel and M. K. O'Malley. Progressive Haptic and Visual Guidance for Training in a Virtual Dynamic Task. In *Proc. IEEE Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems*, pages 343–350, Waltham, MA, 2010.
9. R. S. Kalawsky, S. T. Bee, and S. P. Nee. Human Factors Evaluation Techniques to Aid Understanding of Virtual Interfaces. *BT Technology Journal*, 17(1):128–141, January 1999.
10. J. Kozak, P. A. Hancock, E. Arthur, and S. Chrysler. Transfer of training from virtual reality. *Ergonomics*, 36(7):774–784, 1993.
11. Y. Li, V. Patoglu, and M. K. O'Malley. Negative efficacy of fixed gain error reducing shared control for training in virtual environments. *ACM Transactions on Applied Perception*, 6(1):21, January 2009.
12. M. K. O'Malley and A. Gupta. Passive and active assistance for human performance of a simulated underactuated dynamic task. In *Proc. IEEE 11th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS '03)*, pages 348–355, Los Angeles, CA, March 2003.
13. M. K. O'Malley, A. Gupta, M. Gen, and Y. Li. Shared Control in Haptic Systems for Performance Enhancement and Training. *Journal of Dynamic Systems, Measurement, and Control*, 128:75–86, 2006.
14. L. B. Rosenberg. The use of virtual fixtures to enhance operator performance in time delayed teleoperation. *Journal of Dynamic Systems Control*, 49:29–36, 1993.
15. E.-L. Sallnäs, K. Rassmus-Gröhn, and C. Sjöström. Supporting presence in collaborative environments by haptic force feedback. *ACM Transactions on Computer-Human Interaction*, 7(4):461–476, December 2000.
16. G. Srimathveeravalli and K. Thenkurussi. Motor skill training assistance using haptic attributes. In *Proc. IEEE First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, (WHC '05)*, pages 452–457, March 2005.
17. H. Tan, M. Srinivasan, B. Eberman, and B. Cheng. Human factors for the design of force-reflecting haptic interfaces. *ASME Dynamic Systems and Control*, 55(1):353–359, 1994.
18. E. Todorov, R. Shadmehr, and E. Bizzi. Augmented feedback presented in a virtual environment accelerates learning of a difficult motor task. *Journal of Motor Behavior*, 29(2):147–158, June 1997.
19. M. Zhou, D. Jones, S. Schwaitzberg, and C. Cao. Role of Haptic Feedback and Cognitive Load in Surgical Skill Acquisition. In *Proc. of the Human Factors and Ergonomics Society 51st Annual Meeting, 2007*, pages 631–635, Baltimore, MD, October 2007.