Joel C. Huegel*
Department of Mechatronic and Electronic Engineering
ITESM, Guadalajara Campus
Guadalajara, Mexico

Ozkan Celik
Ali Israr
Marcia K. O'Malley
Department of Mechanical Engineering and Materials Science
Rice University
Houston, Texas

Expertise-Based Performance Measures in a Virtual Training Environment

Abstract

This paper introduces and validates quantitative performance measures for a rhythmic target-hitting task. These performance measures are derived from a detailed analysis of human performance during a month-long training experiment where participants learned to operate a 2-DOF haptic interface in a virtual environment to execute a manual control task. The motivation for the analysis presented in this paper is to determine measures of participant performance that capture the key skills of the task. This analysis of performance indicates that two quantitative measures—trajectory error and input frequency—capture the key skills of the target-hitting task, as the results show a strong correlation between the performance measures and the task objective of maximizing target hits. The performance trends were further explored by grouping the participants based on expertise and examining trends during training in terms of these measures. In future work, these measures will be used as inputs to a haptic guidance scheme that adjusts its control gains based on a real-time assessment of human performance of the task. Such guidance schemes will be incorporated into virtual training environments for humans to develop manual skills for domains such as surgery, physical therapy, and sports.

1 Introduction

Virtual environment (VE) technology can provide reliable data acquisition, analysis, feedback, and evaluation for training of humans in motor skill tasks, while simultaneously providing a low-cost and low-risk training platform. The aim of any VE used for training is to reduce risk, improve and accelerate learning over traditional training schemes, and transfer what is learned in the simulation environment to the equivalent or targeted real world task. These virtual training environments (VTEs) are often designed either to provide a virtual practice (i.e., unaugmented) medium that matches the equivalent physical medium as closely as possible, or to behave as a virtual assistant to improve training effectiveness by providing augmented feedback or guidance during training.

While numerous practice-only (unaugmented) VTEs have been developed and analyzed, there exist only a limited number of published studies aimed at

Presence, *Vol. 18, No. 6, December 2009*, 449–467

*Correspondence to jhuegel@itesm.mx.

determining the efficacy of assistance-based VTEs. What results do exist are inconclusive or contradictory, highlighting the need to carefully design and implement the VTE assistance schemes. We noted in our own prior augmented VTE experiments that, despite several reformulations of the delivery of error reducing shared control, the performance of groups experiencing our haptic guidance in addition to the visual and haptic feedback of the task dynamics never exceeded that of groups that merely practiced the task with visual and haptic feedback in the absence of guidance (Li, Patoglu, & O'Malley, 2009). We hypothesize that there may be other skills required for our target-hitting task that are not addressed by the error-reducing haptic guidance algorithm. Therefore, in this paper, we analyze performance data obtained from a nonguidance (practice alone) VTE target-hitting task to uncover several skills that are required for successful task completion. We propose expertise-based performance measures for these task skills and analyze the correlations between three measures (trajectory error, input frequency, and smoothness ratio) and the stated task objective of hitting as many targets as possible in the allotted time.

Previous studies have shown that the addition of haptic feedback to VEs can provide benefits over visual and auditory displays for performance enhancement, reducing learning times, increasing dexterity, and increasing the sensation of realism and presence (O'Malley & Gupta, 2003; Sallnäs, Rassmus-Gröhn, & Sjöström, 2000; Griffiths & Gillespie, 2004; Jay, Stevens, Hubbold, & Glencross, 2008; and Emken & Reinkensmeyer, 2005). To exploit the capabilities of virtual environments with haptic feedback, various virtual training schemes have been proposed. One scheme is to first present the performance of an expert (human or robotic) to a trainee via visual and haptic feedback, then allow the trainee to practice the task unassisted (Henmi & Yoshikawa, 1998). A second approach requires the trainee to perform the task with enforced restrictions or reductions of the degrees of freedom of the task as proposed by Bernstein (1967) and more recently implemented as *virtual fixtures* (Rosenberg, 1993; and Abbott & Okamura, 2006). A third approach, termed *shared control* in the literature, serves to modify the dynamics of the system by imposing a control effort that elicits the desired behavior of the participant (O'Malley, Gupta, Gen, & Li, 2006; Griffiths & Gillespie, 2004; Emken & Reinkensmeyer). A comparative study of these last two approaches performed by Srimathveeravalli, Gourishankar, and Kesavadas (2007) showed slightly better performance from the shared control approach over the virtual fixture approach.

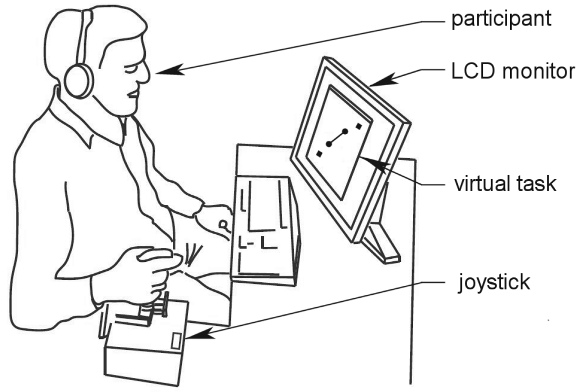
While these virtual training schemes have demonstrated effectiveness in enabling improved task performance, they have not yet conclusively demonstrated effectiveness in accelerating developmental progression (learning) or in increasing overall task performance after a period of training. Sutherland et al. (2006), for example, reviewed 30 studies utilizing simulation (or VTEs in some form) for surgical training. In all 30 studies, VTEs did not outperform traditional training schemes, and in fact VTEs only outperformed control groups who received no training at all. Similarly, for a simple pick and place assembly task with a cognitive component, Adams, Klowden, and Hannaford (2001) found no significant learning benefit from training in a virtual environment. Furthermore, in our manual target hitting task, Li et al. (2009) showed how a haptic guidance VTE designed in an ad hoc fashion resulted in negative efficacy when compared with the unassisted practice of the control group. In contrast, Morris, Tan, Barbagli, Chang, and Salisbury (2007) found more accurate recall of force profiles as a result of visual and haptic training than from visual or haptic training alone, but noted that the haptic feedback was novel for all participants. In another experiment, Feygin, Keehner, and Tendick (2002) compared visual and haptic feedback in the performance of a 3D path following training task. They found that while visual training was significantly better for teaching the trajectory shape, dynamic aspects were more effectively learned from haptic guidance. Feygin et al. qualified their findings by stating that the experiment was too short to arrive at conclusions about overall training outcomes. Liu, Cramer, and Reinkensmeyer (2006) demonstrated improved reproducibility of a novel 3D path with haptic and visual guidance, but gains were not retained a few trials after guidance was removed. These findings were supported in later explorations of the effect of haptic guidance on training of a steering task (Crespo & Reinkensmeyer, 2008). A common conclusion of Feygin's and Morris's work as well as our own prior work is that the best types of guidance are those that are tailored to present specific or key skills related to the task. Furthermore, according to Todorov, Shadmehr, and Bizzi (1997) and Adams et al. (2001), the value of virtual training environments (VTEs) will be realized when they are used for relatively complex tasks rather than for simple tasks.

A common approach to transfer to the trainee an understanding of the key skills of tasks in VTEs has been to "record" expert performance and then "play" this performance back to the trainee via the haptic channel. One classic example is the "virtual teacher" experiment that tested the ability of a command input preshaping technique to communicate the time-optimal strategy for a manual task (Gillespie, O'Modhrain, Tang, Pham, & Zaretzky, 1998). The results of the virtual teacher were inconclusive, yet Gillespie et al. argued that the result would have supported the approach if a primary skill of the task had been demonstrated. Another example is the haptic and visual playback of expert vertebral palpation by Williams, Srivastava, Conatser, and Howell (2004) where the trainee passively experiences the prerecorded positions of a tactile examination. In another case, Rissanen et al. (2007) proposed to use the skills of an expert explicitly in a VTE for the purpose of training force exertion profiles. All of these examples provided to the novice some form of assistance or guidance based on expert performance, yet the expertise was not explicitly defined. Furthermore, in the previous examples, the researchers did not identify the primary components or skills of the expert's performance for representation in their VTE, thereby assuming that regardless of the realism of VTE, it would communicate the key skills to the trainee. The identification of key skills required for the task has been noted by Todorov et al. (1997) as an important first step in developing successful guidance schemes.

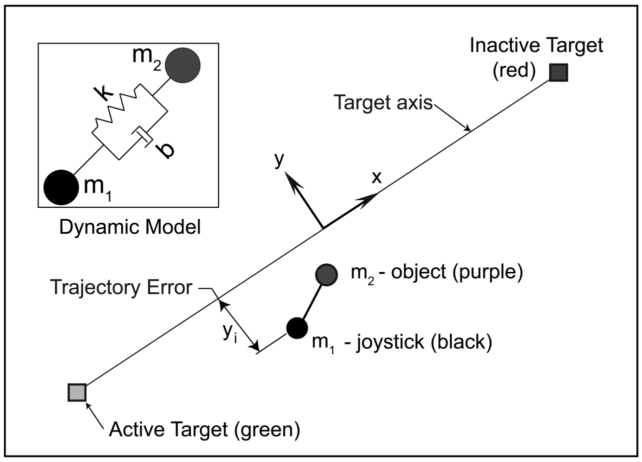
In an effort to determine the key skills necessary for success, some researchers have chosen to observe complex tasks in which there are clear and significant differences between high performing experts and inexperienced novices (Williams & Ericsson, 2005; Abernethy, Thomas, & Thomas, 1993). There has been a preference to study training domains that are closely related to equivalent real world tasks such as surgery, sports, music, and aviation. In practice, expertise is understood to mean exceptional levels of performance in the task of interest. In the surgical domain, Rosen, Hannaford, Richards, and Sinanan (2001) analyzed expert and novice surgeon performance during a typical laparoscopic procedure, finding significant differences between the groups in 14 interaction types. In a survey of surgical simulation for training, Gallagher et al. (2005) insisted on the need to clearly define and categorize expert performance for the purpose of establishing proficiency criteria to evaluate surgery trainees objectively regardless of the simulation used. Thus, the criteria for objectively categorizing expertise may be as important as the degree of realism of the VTE. In fact, Tzafestas, Birbas, Koumpouros, and Christopoulos (2008) state that any haptic surgical simulator must be assessed in two ways: not only as a training tool but also as a skill assessment tool. O'Toole et al. (1999) provided evidence that the performance of two groups, experts with more than 1,000 procedures performed, and novices with no experience, could be differentiated using simulator metrics. Other fields require similar objective measures of motor skill performance such as flight training, sports, and even rehabilitation (Lintern, Roscoe, Koonce, & Segal, 1990; Abernethy, Farrow, & Berry, 2003; Celik et al., 2008).

Recognizing the need to base haptic guidance on key skills required for the task at hand, we analyze the performance of 17 participants in a month-long training protocol for a rhythmic target-hitting task. The data were collected as part of a prior study reported by Li et al. (2009) where we compared two haptic guidance schemes to nonguidance (practice alone) in order to determine the efficacy of an error-reducing shared controller (ERSC). The fixed-gain ERSC scheme showed negative efficacy compared to nonguidance while the strategy (or minimum guidance dosage) ERSC scheme showed no significant difference compared to nonguidance. These results call into question the validity of error-reduction as the sole basis of our haptic guidance paradigm for the target-hitting task. Our previous findings motivate our current analysis to identify the key skills of the target-hitting task in order to design a more effective guidance scheme. In this paper, we identify two key skills required to successfully execute the target-hitting task and define performance measures for the skills. Then we explore the correlations between the measures and the task objective of maximizing target hits. We then investigate whether experts (defined in terms of hit count) are statistically significantly better at the task over time when compared to novice and intermediate performers using the new performance measures. We propose that the identification of the key skills and the employment of related performance measures in progressive guidance control schemes will accelerate and improve training outcomes in VTEs.

This paper is organized as follows. The first part of Section 2 revisits the methods used in the prior experiment including the apparatus, task details, and experimental procedure. The latter part of Section 2 defines which nonguidance (practice alone) data from the prior experiment is to be included in our study. Section 2 also describes the data analysis and statistical analysis that we conducted. The results are presented in Section 3. Section 4 gives a discussion of the findings and contributions of this paper. Section 5 draws the conclusions of this analysis.



(a) Experimental setup



(b) Schematic of Task and Dynamic model

Figure 1. *(a) Experimental setup includes a force feedback joystick and the graphical interface. (b) Schematic shows the target hitting task and its coordinate frame as well as the modeled 4-DOF second order dynamic system.*

2 Methods

In our prior work, we compared the performance of haptic guidance in the form of an error-reducing shared control (ERSC) scheme to nonguidance (practice alone; Li et al., 2009). In contrast, here we reanalyze the nonguidance data from the same experiment in order to more fully understand how participants are achieving their level of performance. Performance and motion data were extracted and analyzed firstly to gain insight into the strategies adopted by participants who were adept at the task; and secondly to identify and measure key skills required for task success. The following sections review the relevant details of the experiment previously reported, the data from which we reanalyze in this work.

2.1 Apparatus and Task Details

A haptic interface and computer monitor were used to present a virtual target-hitting task to the participants. The experimental setup shown in Figure 1(a) included a force feedback joystick (Immersion IE2000) and a 19 in LCD display with a 60 Hz graphics software loop rate for visual feedback. The haptic control loop ran at 1 kHz on a 2 GHz Pentium computer while position and velocity data were captured and stored at 20 Hz. During the task, the typical wrist torque exerted by the joystick motors was less than 1 Nm.

The dynamic second order manual task was modeled as two point masses, m_1 and m_2, connected by a spring and damper in parallel as shown in the inset of Figure 1(b). This two-mass system had four degrees of freedom (DOF), namely the planar motion of each of the point masses. It was, therefore, underactuated since the only control inputs were the planar motion of m_1. The task, illustrated in Figure 1(b), was to manipulate the motion of the point mass (m_1) via the 2-DOF haptic joystick, and thus indirectly, through the system dynamics, to control the movements of the mass (m_2) in order to hit as many of the diagonally placed targets as possible during each 20 s trial. Once a target was hit, the current target became inactive and the opposite target became active and so forth (see Figure 1[b]).

In order to increase task complexity, three system parameter sets were presented to all participants throughout all trials. Each parameter set includes a specific mass of m_2, spring stiffness, and damping to provide unique resonant frequencies (f_r), namely 1.00 Hz, 0.709 Hz, and 0.490 Hz. There was no information about f_r provided to the participants, hence they had to identify the changes based on the behavior of the virtual system (displayed via both the visual and haptic channels). There exists an upper bound to the target hit count score (n_{hit}) due both to the time limit for each trial (20 s) and to the dynamics of the two-mass system. As the excitation frequency exceeds the natural frequency determined by the mass and stiffness of the virtual system, the amplitude of m_2 is attenuated relative to the motion of the joystick m_1, thereby requiring larger displacements by the participant. The maximum performance observed in the experiment was approximately 35 target hits per trial. Participant data analyzed in this paper are well-distributed across the range of achievable scores.

The two mass system is well suited for an experimental study of human performance enhancement and training with haptic assistance because haptic feedback, generated by the dynamic interactions of the two masses, is

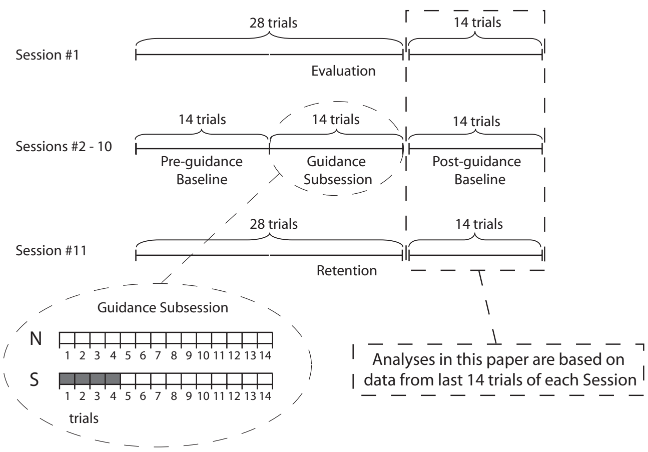


Figure 2. *Schematic representation of the experiment procedure shows one evaluation session, nine training sessions, and one retention session. Each training session contains three subsessions of 14 trials each: pre-guidance baseline, guidance, and post-guidance baseline. The N group received no guidance throughout all subsessions. In Li's experiment (Li et al., 2009), the S group received guidance for only 4 of the 14 trials of the guidance session. This study does not include any data from the guidance subsession; rather, it includes only the nonguidance data from both the N and S groups in the 14 post-guidance baseline trials. Figure adapted from Li et al.*

necessary for the human to accurately control the motion of the system (O'Malley et al., 2006). Additionally, the control of this system is sufficiently complex to require training (Todorov et al., 1997). In contrast, tasks chosen by Yokokohji, Hollis, Kanade, Henmi, and Yoshikawa (1996) and Adams et al. (2001) as the basis for testing virtual environment training were found to be too simple to draw conclusions regarding the efficacy of the virtual training environment.

2.2 Experimental Procedure

A training experiment was designed and conducted utilizing the target-hitting task to acquire movement and performance data for subsequent analysis. The experiment was composed of 11 sessions, including an initial evaluation session, nine training sessions, and one retention session as illustrated in Figure 2. The training sessions were spaced 48–120 hr apart and the retention session was one month after the last training session. Within each training session, there were three subsessions of 14 trials per session. The participants were allowed breaks of up to 5 min between subsessions to avoid fatigue. They were given the specific objective of hitting as many targets as possible in each 20 s trial, but no additional instructions were provided. A target hit was registered whenever the center position of m_2 was detected to be within 4 mm of the target center.

2.3 Participants

In this study, participants' data from the prior study (Li et al., 2009) were analyzed to determine key skills for the target-hitting task. In the experiment reported in Li et al. (2009), we examined the effect of different doses of haptic guidance provided to three groups of participants: No Guidance (N group)—no guidance throughout the entire experiment; Assisted (A group)—error-reducing haptic guidance during all trials of the guidance subsession of all of the training sessions; and Strategy (S group)—error-reducing haptic guidance provided during a fraction of the trials of the guidance subsessions of all of the training sessions.

Because in this work we seek to analyze nonguidance (practice alone) data that are unaffected by guidance, we include data from the eight participants who did not receive any form of haptic guidance but rather only practiced the task (N group). We also include in this study the data of the nine participants in the strategy group (S group) for whom the data failed to show significant differences from the nonguidance group (N group), suggesting that the guidance scheme had no effect on performance. For the strategy group (S group), the guidance was active for only four of the 42 daily trials. Moreover, this analysis includes data from only the last 14 trials of each session (which were without guidance for all 17 participants) as shown in Figure 2. The other guidance group in the Li et al. (2009) study (A group) received haptic guidance during all 14 trials of the guidance subsession and did exhibit significant differences from the nonguidance group (N group) even during the postguidance subsession. Therefore, the participants from this assistance group (A group) were not included in the analysis reported here to avoid confounding the results with the effects of guidance. Thus, for this analysis, the data from all 17 participants was without guidance. The 17 participants are subsequently subdivided into three expertise groups based on their performance in the first and last sessions: four experts, nine intermediates, and four novices for a total of 17 participants. The participants were all undergraduate students (ages 18 to 24 years old, five female, 12 male, two left-handed), with no previous experience with haptic devices. A university IRB approved form was used to obtain informed written consent from all participants.

2.4 Data Analysis

Our objective is to determine a set of performance measures that strongly correlate with the objective of the task. We used hit count as the objective measure of performance for the task throughout our previous study, since participants were instructed to maximize target hits in each 20 s trial. In order to compare performance regardless of the virtual system parameters, the total hit count per trial (count/Hz) is normalized by the following equation:

$$n_{\mathrm{hit}} = \frac{1}{f_{\mathrm{r}}} \times (\text{hit count}) \tag{1}$$

Visual inspection of the participants' data, presented in several ways, leads to the definition of three performance measures that capture the key skills for the target-hitting task: trajectory error, input frequency, and smoothness ratio.

2.4.1 Trajectory Error. An initial inspection of the data in the form of recorded participant movement traces reveals some qualitative aspects for differentiation of various behaviors demonstrated by the participants. Figure 3 shows representative position traces from training in the task. Some participants, such as Participant A, begin with erratic and slow motion, and continue to be erratic throughout training. Others, such as Participant B, begin erratically but, during training, learn to excite the system along the target axis. Still others, such as

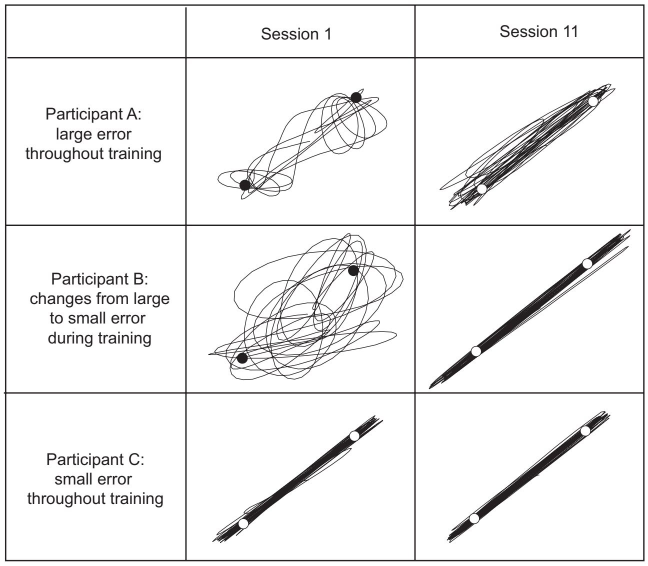


Figure 3. *Sample traces for three typical participants shows varying improvements and performance differences.*

Participant C, excite the system along the target axis from the very beginning of training. These data emphasize the need to follow the target axis to achieve a high number of hits, and were the motivation for the error reducing shared control (ERSC) algorithm. We previously reported that the ERSC algorithm uses a time-independent error measure (Li et al., 2009) instead of a time-dependent error measure as used in similar work by Gillespie et al. (1998) and Patton and Mussa-Ivaldi (2004).

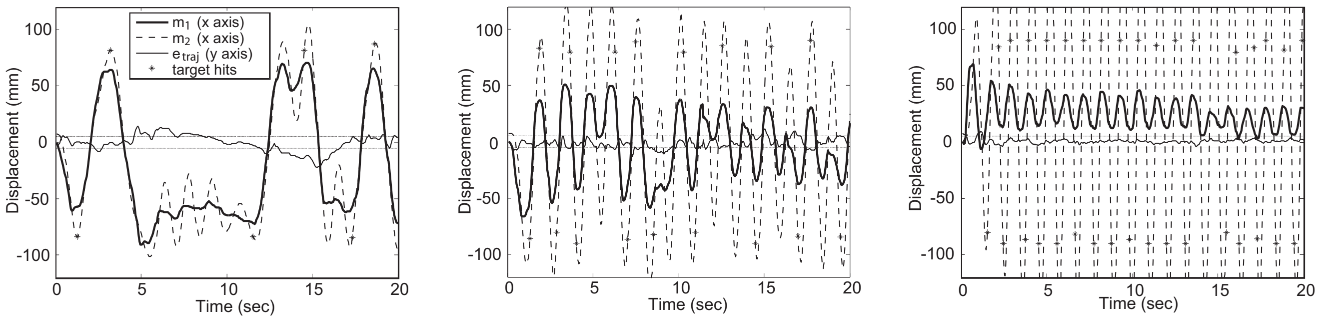
Based on an inspection of Figure 3 and the recognition of the need to minimize deviations from the target axis, we propose trajectory error as a performance measure for this task. We define trajectory error as the absolute magnitude of the deviation from the target axis (i.e., along the y axis, as shown in Figure 1[b]) of the joystick position (m_1) at each sampled instant summed for the entire trial (n = 400 samples). Trajectory error is expressed in units of millimeters. The target axis, shown in Figure 1(b), is the diagonal line passing through both targets and along the oriented x axis. Mathematically,

$$e_{\text{traj}} = \sum_{i=1}^{n} abs(y_i) \tag{2}$$

We choose to use the error of the joystick (m_1) rather than the manipulated mass (m_2) because we are analyzing the performance of the participant. Our prior analysis reported by Li, Patoglu, and O'Malley (2006) also used a trajectory error measure, but it was defined as the RMS deviation of the trajectory of the mass (m_2) from the target axis. Other researchers have also used error measures, including Celik et al. (2008) and Colombo et al. (2005) who correlated robotic based error measures of performance to the clinical stroke measures that therapists have used.

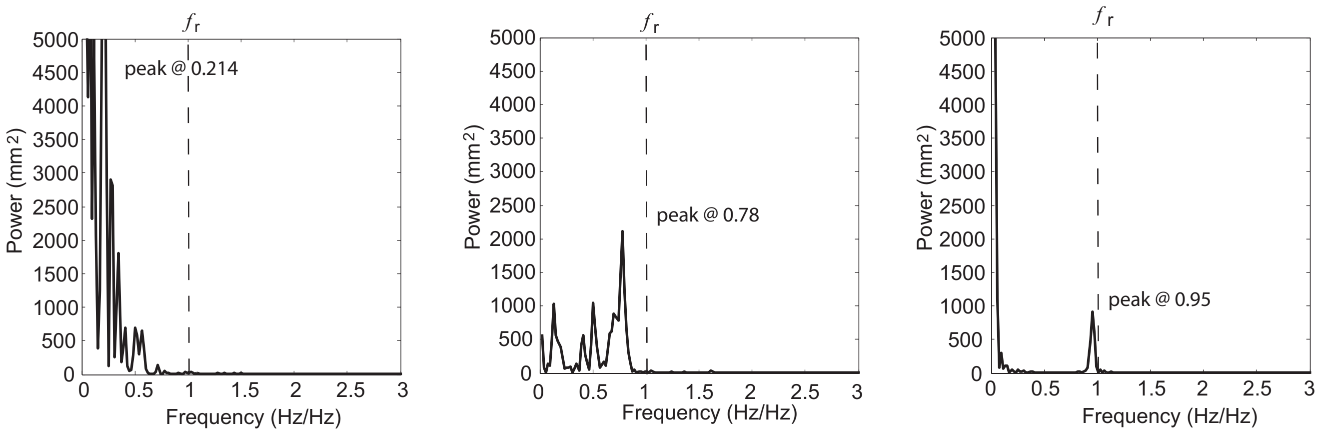
2.4.2 Input Frequency. Because prior analysis showed negative efficacy of the fixed-gain ERSC, we question the validity of error reduction as the sole basis for a successful guidance scheme (Li et al., 2009). Analysis of individual participant data for the target hitting task reveals the need for additional measures of performance that are time or state space dependent. Therefore, we examine representative time-series plots of the participants. Figure 4 shows traces of position versus time for trial 10 of session 4 (approximately midway through the training protocol) for three different participants. The trajectory error (e_{traj}) as previously discussed is represented by the area from the zero reference to the thin black line. The solid thick line is the position of the mass m_1 (system input) while the dashed line represents the position of m_2 (system output). Both are along the x axis (as shown in Figure 1[b]). Figure 4(a) illustrates the typical low performance of a participant who has yet to learn the task with high e_{traj} (22.9 mm) resulting in a low n_{hit} score (6 hits). The performance of another participant in Figure 4(b) shows the ability to maintain low e_{traj} (14.7 mm). However, the participant achieves only a moderate n_{hit} score (16 hits) due to the apparent inability to leverage the dynamics of the controlled system and excite the system near its resonant frequency. In contrast, the participant in Figure 4(c) shows good performance by being able to maintain low e_{traj} (5.55 mm) as well as to provide a consistent input excitation frequency of 95% of the resonant frequency, resulting in a high n_{hit} score (33 hits).

Based on the observation of the importance of input excitation frequency, we propose input frequency (f_{input})



(a) low performance (b) moderate performance (c) high performance

Figure 4. *Displacement time traces from Trial 10 of Session 4 for three typical participants. (a) The high e_{traj} and irregular input motion of a low performer. (b) The low e_{traj} but inconsistent input motion of a moderate performance example. (c) A high performer's low e_{traj} and consistent excitation.*



(a) low performance (b) moderate performance (c) high performance

Figure 5. *FFT position power spectra for the same three participants shown in Figure 4 again from Trial 10 during Session 4. (a) The erratic input spectrum of a low performer. (b) The fairly consistent but slower f_{input} of a moderate performance example. (c) The extremely consistent and low power of a high performer.*

as a measure of performance in a trial. In order to compute f_{input}, we take the Fast Fourier Transform (FFT) of the x axis position data of the joystick, m_1. The FFT power spectrum is a convenient way to determine the amplitude and frequency of the motion that is being applied to a system and was used by Huang, Gillespie, and Kuo (2007) in a similar task to quantify performance. Figure 5 shows the frequency spectra of the same three data sets (trial 10 of session 4) shown in Figure 4. Figure 5(a) shows a participant who is inconsistently exciting the system and whose data shows wide spectral variability. In contrast, Figure 5(b) shows the data of a participant who is exciting the system in a fairly consistent manner. Finally, Figure 5(c) shows a small but very clear spike indicating that this participant is exciting the system at only one frequency which is 95% of the resonant frequency of the virtual two-mass system. The challenge for the first participant is to increase the input frequency to the system resonant frequency. The participant must identify the resonant frequency and then provide consistent input motion commands near that frequency to the joystick m_1 in order to achieve a significant increase in n_{hit} score. For clarification, even though the FFT plot is called a "power spectrum," in this particular case it has units of mm^2. Additionally, because Li's experiment used three separate parameter sets (Li et al., 2009), our definition includes a normalizing coefficient. The equation for the second performance measure f_{input} is given in units of Hz/Hz (dimensionless) as follows:

$$f_{input} = \frac{1}{f_r} \times f(arg[max(\text{FFT})]) \tag{3}$$

Therefore, exciting the system at the resonant frequency will give a value of $f_{input} = 1(\text{Hz/Hz})$ regardless of the system parameter set.

2.4.3 Smoothness Ratio. After observing traces and time series plots of participants, we plotted velocity profiles to explore relationships between movement smoothness and expertise. Flash and Hogan (1985) showed that the tangential speed profile of the hand during point-to-point reaching movements of healthy participants can be well represented by an optimally smooth speed profile that minimizes the jerk, the time derivative of acceleration.

Later, the optimally smooth speed profile for a rhythmic movement was given by Hogan and Sternad (2007) as

$$v_{\text{mj-rhythmic}}(t) = A\left(5\,\frac{t}{d^2} - 10\,\frac{t^3}{d^4} + 5\,\frac{t^4}{d^5}\right) \tag{4}$$

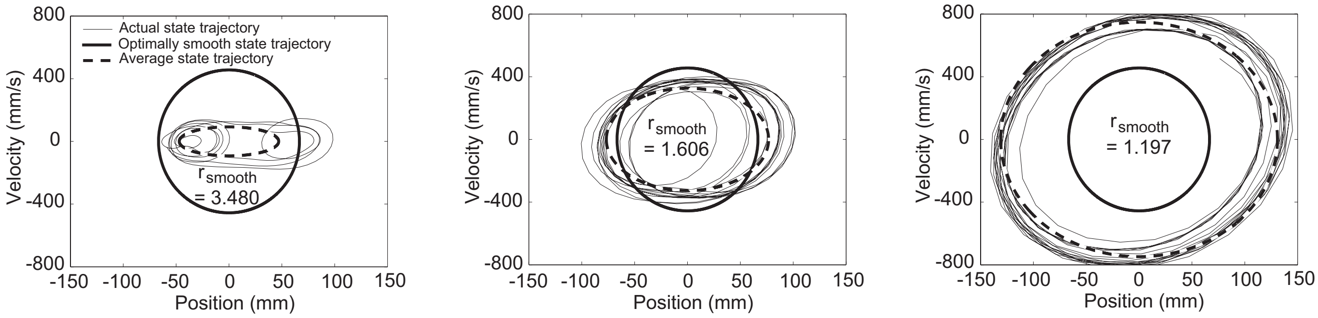
where A is the amplitude of the movement, d is the duration of the half-cycle movement, and t is time. The full-cycle smooth speed profile can be accurately approximated by a sinusoid with appropriate amplitude and frequency, and we have used this simplified version in this study. For our target-hitting task, the smoothness of the movement can be calculated for m_2 on the x axis (shown in Figure 1[a]), regardless of whether the speed profile of m_1 is smooth or not.

Dingwell, Mah, and Mussa-Ivaldi (2004) argued that the smoothness of movement property was carried over to the endpoint (or the object) while manipulating a flexible object, by using a "minimum object crackle" (crackle being the third time derivative of acceleration) cost function. In contrast, Svinin, Goncharenko, Zhi-Wei, and Hosoe (2006) argued that the smoothness of the hand was still the important factor, even when moving flexible objects. Svinin et al. used a "minimum hand jerk" cost function and carried the dynamic constraints imposed by the object onto the hand. They demonstrated that the profiles generated by using this cost function would hold for multi-mass objects, whereas speed profiles calculated by Dingwell's model would not be valid. In our task, we observed that object movement smoothness was more important than the hand movement smoothness in order to achieve high performance. For example, a participant could utilize a ballistic strategy that slings the second mass. This could generate a high hit count without requiring smooth movements of the hand. For this reason, we measured the smoothness of movement of the object as in Dingwell et al. Our work differs from both Svinin et al. and Dingwell et al. in that our task does not require either the participant hand or the object to stop at the target but rather pass through the target. In other words, there is no outer condition on the amplitude of the input or output. Therefore, our task does not have a unique optimal solution, and direct implications cannot be derived from the work by Dingwell et al. or Svinin et al.

We define a smoothness ratio measure (r_{smooth}) using the recorded state-space (position vs. velocity) trajectory of m_2 and an optimally smooth trajectory as shown in Figure 6. A rhythmic movement in this state-space appears as an ellipse. There are many ways to compare two ellipses, for example, the area between the two ellipses can be considered as an error. The nature of the task at hand, however, makes it possible to overshoot the targets, thereby causing a wider (greater amplitude) and taller (higher velocity) ellipse than the optimally smooth ellipse, while still being successful at the target hitting task. Hence, we opt for a shape-wise comparison. The smoothness ratio is defined as

$$r_{smooth} = \frac{a_{actual}/b_{actual}}{a_{nominal}/b_{nominal}} \tag{5}$$

where a_{actual} and b_{actual} represent the major and minor axes of the average ellipse calculated from the recorded



(a) low performance (b) moderate performance (c) high performance

Figure 6. *Actual, average, and optimally smooth state trajectories for the same three participants that were shown in Figure 4 and in Figure 5. All samples are again from Trial 10 of Session 4. (a) The state trajectory of a low performer that is inconsistent and nonsmooth. (b) The state trajectory of a moderate performer that is fairly consistent yet demonstrates a velocity that is lower than the optimal. (c) The state trajectory of a high performer that is consistent and demonstrates a velocity and displacement that is higher than the optimal velocity.*

data by considering the points of intersection with the axes (points of zero velocity and zero position). An initial portion (1.75 s) of each 20 s trial is trimmed from the data before calculating a_{actual} and b_{actual}. Then a_{nominal} and b_{nominal} are calculated from the optimally smooth rhythmic movement that has a duration equal to the inverse of the resonance frequency of the system (f_r). When the sinusoid approximation for the speed profile is used, the movement amplitude does not need to be specifically defined, as it gets canceled when calculating the ratio $a_{\mathrm{nominal}}/b_{\mathrm{nominal}}$. After the simplifications, this ratio becomes

$$\frac{a_{\mathrm{nominal}}}{b_{\mathrm{nominal}}} = \frac{1}{2\pi f_r}. \tag{6}$$

With this definition of r_{smooth}, the measure approaches unity as the state trajectory of m_2 approaches an undistorted but scaled version of the optimally smooth ellipse. As it can be easily deduced from Equations 5 and 6, r_{smooth} effectively becomes a measure of the actual average movement period normalized by the period of the movement at the resonant frequency of the system. Therefore, it is expected that r_{smooth} will be highly correlated with f_{input}. Figure 6 shows the state trajectories for the same trials (trial 10 of session 4) that were shown in Figure 4 and in Figure 5. The smoothness ratio, as previously discussed, is represented by the shape of the state trajectory. Figure 6(a) illustrates the typical low performance of a participant who has yet to learn the task with a distorted state trajectory demonstrated by a high r_{smooth} (3.480) and that has a low n_{hit} score (6 hits). The moderate performance of another participant in Figure 6(b) demonstrates the participant's ability to maintain a fairly low r_{smooth} (1.606). The low excitation velocity, however, produces only a moderate n_{hit} score (16 hits). In contrast, the participant in Figure 6(c) demonstrates an undistorted state trajectory, thereby maintaining a low r_{smooth} (1.197) score that results in a high n_{hit} score (33 hits).

2.5 Expertise Analysis

We seek to identify and measure the key skills of the manual target hitting task in order to improve the design of guidance schemes that can be conveyed in a virtual training environment (VTE). In order to identify the key skills, we investigate the differences between expert, intermediate, and novice performers in a quantitative and systematic way. In lieu of a standardized method to determine a participant's level of expertise, and recognizing the broad range of definitions for expertise in the literature (see the seminal work by Fitts, 1964 and more recent work by Dreyfus & Dreyfus, 1986), we use a statistical measure after

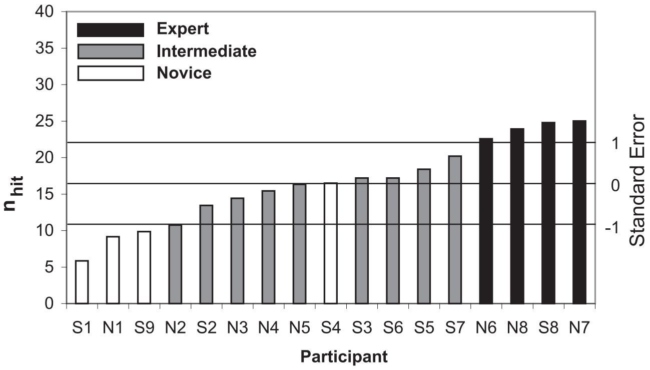


Figure 7. *Performance in the initial evaluation session (prior to training) of all participants shows designation of experts one standard error above the mean. Novices and intermediates are both below one standard error above the mean. Hit count is normalized by the natural frequency of the system which varies as presented in this section.*

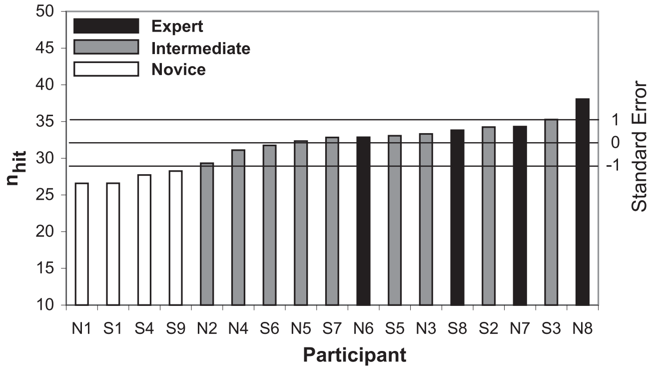


Figure 8. *Performance of all participants in the retention session shows the designation of novices who performed worse than one standard error below the mean. The intermediates, who performed worse than one standard error above the mean in the evaluation session, had performances that were intermixed with the experts in the retention session.*

the initial evaluation session to differentiate experts from intermediates and novices in the virtual practice VTE.

In order to determine the expertise of the participants, we analyze the number of target hits for each of the 17 participants in the initial evaluation session as shown in Figure 7. Any participant whose performance was greater than one standard error above the mean of all participants is deemed to be an expert (participant IDs: N6, N7, N8, and S8 as designated by Li et al., 2009). The remaining participants are considered novices or intermediates; those who performed worse than one standard error above the mean.

In a similar way to how we define the expert group, any participant with an n_{hit} score less than one standard error below the mean of all participants during the last training session is considered a novice (participant IDs: N1, S1, S4, and S9). Furthermore, those who perform worse than one standard error above the mean in the first session (to differentiate them from the experts) and better than one standard deviation below the mean in the last session (to differentiate them from the novices) are called intermediate performers. Figure 8 shows the distribution of performance for each participant, classified by their group assignment, at the end of the training protocol. Interestingly, the experts identified in Session 1 are not necessarily achieving the highest n_{hit} scores in Session 11 but are intermixed with the intermediate performers. The distribution of the 17 participants is four experts, nine intermediates, and four novices. These three groups (expert, intermediate, and novice) are used as the basis for comparison to validate our selection of new performance measures for the target-hitting task. Our group definitions are consistent with expertise groups defined in the literature as follows.

- **Expert** One who is able to perform the task well at the beginning of the training—also called masters, teachers, or autonomous performers in the literature (Williams & Ericsson, 2005; Abernethy, Neal, & Koning, 1994; Henmi & Yoshikawa, 1998; Gallagher et al., 2005; Nistor, Allen, Dutson, Faloutsos, & Carman, 2007; and Fitts & Posner, 1967).
- **Novice** One who performs the task in a superficial way, doing poorly at the outset and only marginally improving throughout training—also called beginners, students, or cognitive learners in the literature (Williams & Ericsson, 2005; Abernethy et al., 1994; Henmi & Yoshikawa, 1998; Gallagher et al., 2005; Nistor et al., 2007; and Fitts & Posner, 1967).

- **Intermediate** One who begins poorly but improves rapidly early in training until he or she is as good as, or better than, the expert—also called competent, proficient, or associative learners in the literature (Williams & Ericsson, 2005; Abernethy et al., 1994; Gallagher et al., 2005; and Fitts & Posner, 1967).

3 Results

For all participants, values for each session for both the objective measure of n_{hit} and the performance measures of e_{traj}, f_{input}, and r_{smooth} were determined by averaging the scores of the last 14 trials of each session. Thus, we have a total of 187 observations (17 participants and 11 sessions) for each measure. To compare participant performance by level of expertise, a performance group average score for each measure was determined from the participants' session scores to give one value per session per group. The data were fit with linear and exponential curves using MATLAB, and the best fit curves were determined from the R^2 values. Analysis of variance (ANOVA) was used to determine significance among groups.

3.1 Correlation of Performance Measures

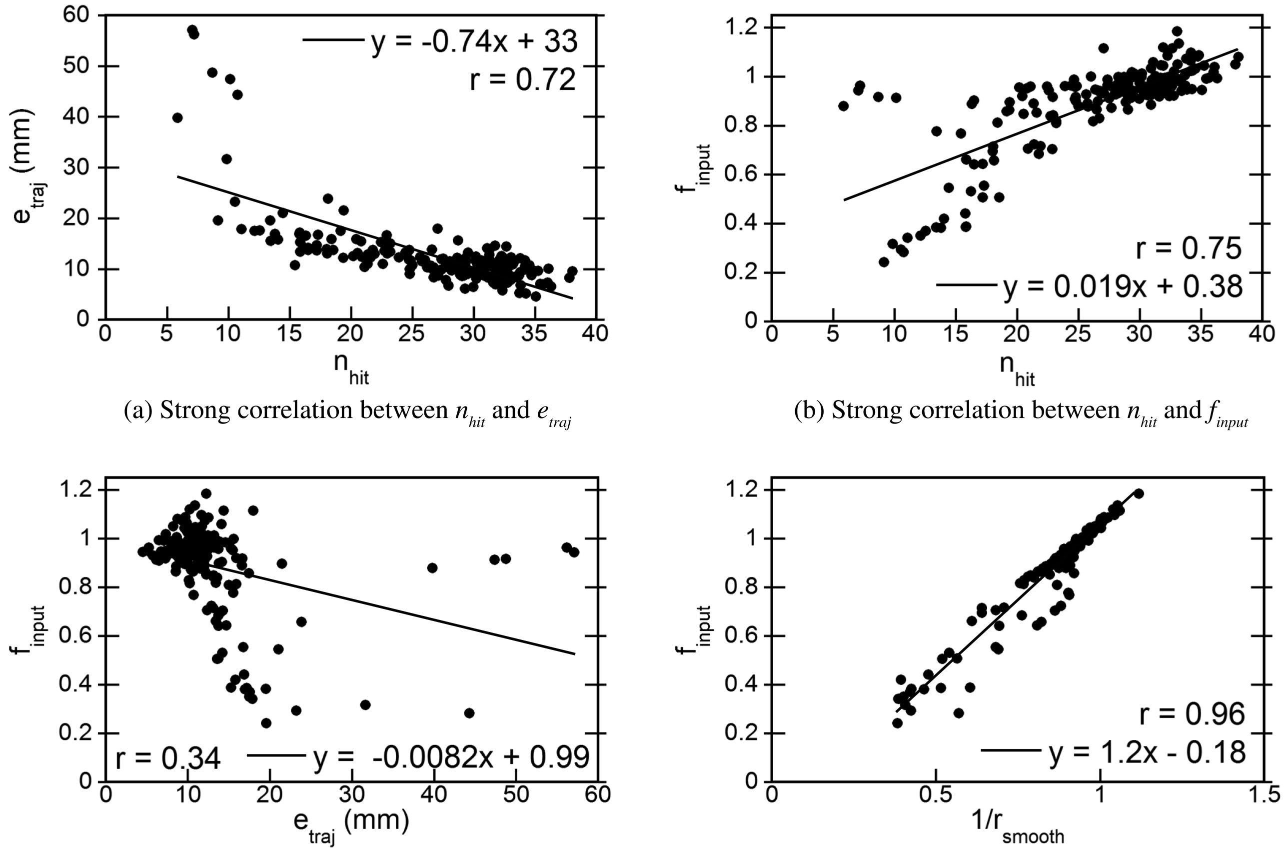
We first examined correlations between the task objective, n_{hit}, and the three performance measures introduced in the preceding section, namely, e_{traj}, f_{input}, and r_{smooth}. Figure 9 shows the relationships between the four measures in four panels. A straight line is regressed through the data points of each panel and the equation as well as the correlation coefficient of each fit are shown in each panel. First, there exists a strong correlation between n_{hit} and e_{traj}, $r(185) = -0.715$, $p < .01$, as well as between n_{hit} and f_{input}, $r(185) = +0.754$, $p < .01$. Second, the correlation between f_{input} and e_{traj} is significant ($p < .01$); however, the lower correlation coefficient, $r(185) = -0.336$, indicates that the two secondary measures are only loosely correlated by a straight line function. A two-factor multiple regression (f_{input} and e_{traj} as independent variables and n_{hit} as the dependent variable) indicates strong correlations, $r(184) = 0.90$, between the independent and dependent variables. Both predictors are significant ($p < .01$), but while f_{input} is directly correlated to n_{hit}, e_{traj} is inversely correlated to n_{hit}. These correlations suggest that both f_{input} and e_{traj} are important factors in predicting the normalized hit count, and thus are indicators of success in the completion of the primary task. Third and finally, the correlation between f_{input} and $1/r_{\text{smooth}}$ is very strong, $r(185) = -0.96$, $p < .01$, suggesting as expected that these two measures can be used interchangeably. Values for r_{smooth}, therefore, are not reported in the remainder of the results. An analysis of the correlations between these measures confirms that, in order to achieve a maximum number of target hits in our task, participants must excite the virtual system close to its resonant frequency (i.e., natural frequency) and, by extension, generate a smooth movement of mass m_2 while keeping the masses along the straight line joining the two targets.

3.2 Expertise Analysis

We analyzed the performance of all 17 participants in terms of each performance measure to verify expected trends during the training experiment.

Figure 10 shows the n_{hit} scores as a function of session for the three participant groups (experts, intermediates, and novices). Each data point is the n_{hit} average for a group at the corresponding session, with error bars indicating standard error of the mean. Straight line and exponential functions were fit to the data in order to visualize learning effects as a function of sessions. A summary of the curve fitting results, including estimated parameters and correlation coefficients from goodness of fit for each of the three groups of participants, are shown in Table 1.

The experts initially had the highest n_{hit} scores and improved slowly until reaching a saturation level of approximately 37 hits (parameter c of the exponential function, see Table 1). Intermediates began with significantly lower n_{hit} scores than the experts (parameter $c - a$, $p < .05$, confidence intervals [18.9, 25.0] for experts and [4.8, 11.4] for intermediates) and reached saturation at a faster (but nonsignificant) rate compared to the



(a) Strong correlation between n_{hit} and e_{traj}

(b) Strong correlation between n_{hit} and f_{input}

(c) Weak correlation between e_{traj} and f_{input}

(d) Strong correlation between $1/r_{smooth}$ and f_{input}

Figure 9. *Correlation plots between performance measures. (a, b) The strong correlations between both e_{traj} and f_{input} with respect to the objective measure of n_{hit}. (c) The loose correlation between e_{traj} and f_{input}, suggesting the measures are independent. (d) The correlation between r_{smooth} and f_{input} is very strong.*

experts (parameter b, confidence intervals [0.03, 0.33] for experts and [0.28, 0.45] for intermediates). The 95% confidence bound for the saturation level of the intermediates coincides with that of the experts (parameter c, $p < .05$, confidence intervals [31.9, 41.3] for experts and [32.0, 34.3] for intermediates) indicating that both groups reached the same performance level toward the end of the experiment. Additionally, the intermediates reached 90% of the saturation level slightly after the sixth session. The novices started with the lowest n_{hit} scores and improved linearly with significant slope (parameter a, $p < .05$), hence failing to reach saturation during the experiment.

The average e_{traj} and f_{input} are shown in Figure 11 and Figure 12, respectively (results are averaged over the 11 sessions of the protocol). The error bars show the standard error of the group mean. Figure 11 shows decreasing trends of the mean e_{traj} while Figure 12 shows increasing trends of the mean f_{input} as training progressed. Analysis of both the e_{traj} and f_{input} measures of performance by group showed similar trends to the performance by group in terms of n_{hit}. In other words, novices had the worst performance, experts showed the best performance, and the intermediates started out somewhere in the middle, yet achieved performance comparable to the experts at some point during training.

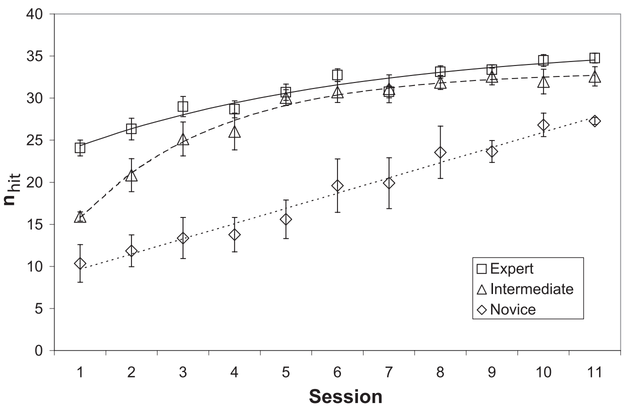


Figure 10. *Average n_{hit} as a function of session for the three groups of participants. Error bars indicate standard error of the mean.*

Straight line and exponential curves were fit to the data, with the details included in Table 1.

For e_{traj}, data for the expert and intermediate participants exhibited exponentially increasing trends, while the trends for novices were better characterized by a straight line function. For f_{input}, data for the experts and novices showed linearly increasing trends, while intermediates demonstrated exponential behavior.

Each performance measure was further analyzed using a two-way ANOVA in order to highlight significant effects of the between-subject factors: group and session. For all three performance measures, the main effects of group and session were significant. For n_{hit}, the effects of group and session were significant, group: $F(2, 154) = 180$, $p < .001$; session: $F(10, 154) = 41.4$, $p < .001$. For e_{traj}, the effects of both group and session were also significant, group: $F(2,154) = 23.5$, $p < .0001$; session: $F(10, 154) = 3.49$, $p < .0001$. Finally, for f_{input}, the effects of both group and session were also significant, group: $F(2, 154) = 51.3$, $p < .001$; session: $F(10, 154) = 7.97$, $p < .001$. The interaction effect of group and session was significant for n_{hit}, $F(20,154) = 2.03$, $p = .0086$, but was not significant for either e_{traj}, $F(20,154) = 1.36$, $p = .152$, or f_{input}, $F(20,154) = 1.30$, $p = .190$. The analysis indicates that the performance measures were significantly different between groups of varying expertise and that performance improved along sessions.

4 Discussion

The performance analysis described in this paper is motivated by a need to design haptic guidance controllers that convey key skills for a manual task in a virtual training environment (VTE) protocol. Previous attempts by our group to implement intuitively-designed haptic guidance, such as the error-reducing shared control (ERSC) used by Li et al. (2009) and O'Malley et al. (2006), have shown negative efficacy compared to non-guidance (practice alone). We hypothesize that the ERSC paradigm was ineffective because it failed to convey all of the key skills required to successfully execute the dynamic target-hitting task.

Participants' motion trajectories during their unassisted execution of the virtual target-hitting task were analyzed in order to identify skills that are required for successful task completion. Knowledge of key skills can be used in the design of guidance paradigms intended to improve the efficacy of training in virtual environments. The first key skill for our target-hitting task is the minimization of the trajectory error. The second skill is related to the excitation frequency of the system input. We focused our analysis of performance on the measured motion of m_1 in our underactuated dynamic system, which corresponds to the motion of the human via the joystick, since we were interested in directly assessing the participants' movements and performance of the task. Our trajectory error measure was based on the motion of m_1 relative to the target axis. Previously, we analyzed the error of the output of the second order dynamic system (Li et al., 2009). Such an analysis of performance based on the trajectory is important for tracking tasks such as those studied by Feygin et al. (2002). Because of the dynamics of the system in our target-hitting task, the motion of the output (m_2) is dynamically coupled to the motion of the input (m_1), and therefore similar decreasing trends are noted in the trajectory error measure over the course of training. For the input frequency measure described in this work, again we base our calculations on the motion of m_1, which directly corresponds to the motion of the input joystick and human participant. Others have focused on input frequency as we have (e.g., Israr, Kapson, Patoglu,

Table 1. *Summary of the Curve Fitting Procedures for the Performance Measure Data of Each Group*

		Goodness of fit				
Participant group	Measure	DOF	Function type	Function expression	R^2	Parameters
Expert	n_{hit}	8	Exponential	$-ae^{-bx} + c$	0.95	$a = 14.6$, $b = 0.18$, $c = 36.6$
	e_{traj}	8	Exponential	$ae^{-bx} + c$	0.65	$a = 3.60$, $b = 0.28$, $c = 8.18$
	f_{input}	9	Straight line	$ax + b$	0.68	$a = 0.006$, $b = 0.93$
Intermediate	n_{hit}	8	Exponential	$-ae^{-bx} + c$	0.99	$a = 25$, $b = 0.37$, $c = 33.15$
	e_{traj}	8	Exponential	$ae^{-bx} + c$	0.94	$a = 16.4$, $b = 0.74$, $c = 10.6$
	f_{input}	8	Exponential	$-ae^{-bx} + c$	0.99	$a = 0.53$, $b = 0.42$, $c = 1.03$
Novice	n_{hit}	9	Straight line	$ax + b$	0.98	$a = 1.81$, $b = 7.84$
	e_{traj}	9	Straight line	$ax + b$	0.88	$a = -1.96$, $b = 30.0$
	f_{input}	9	Straight line	$ax + b$	0.96	$a = 0.04$, $b = 0.47$

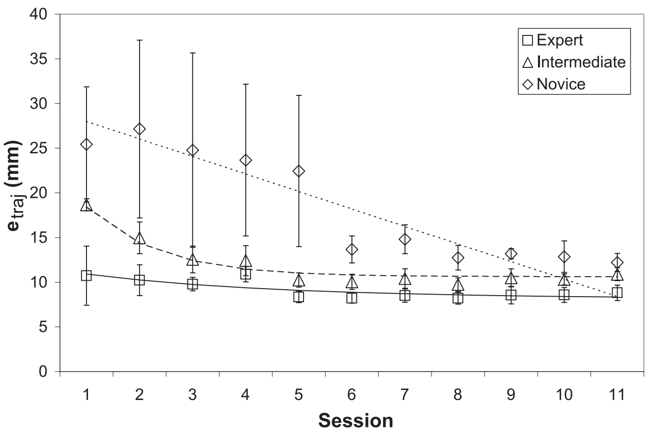


Figure 11. *Average e_{traj} as a function of 11 sessions for three groups of participants. Error bars indicate the standard error of the mean. Expert and intermediate data are best fit by exponential functions while novice data are best fit with a straight line function.*

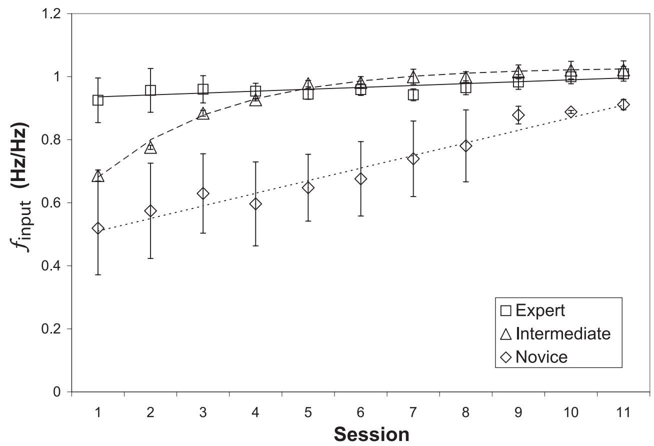


Figure 12. *Average f_{input} as a function of sessions for three groups of participants (error bars indicate standard error of the mean). Intermediate data are best fit with an exponential function; experts and novices are best fit by straight line functions.*

& O'Malley, 2009; and Huang, Gillespie, & Kuo, 2007). Conversely, some groups have approached the measure of rhythmic task performance by analyzing the smoothness of the system output. We have shown that when the optimally smooth state space trajectory is defined based on the resonant frequency of the system, a shape-wise comparison between the actual trajectory and the optimal trajectory results in a comparison of average movement frequency with resonant frequency. Hence, for rhythmic tasks, movement smoothness and frequency measures are inherently closely related. In fact, we have shown that they are highly correlated.

Performance measures that capture key skills for the target-hitting task were introduced, namely e_{traj}, f_{input}, and r_{smooth}. These measures were shown to correlate well with the objective measure of n_{hit}. In addition to

f_{input}, we conclude that the r_{smooth} measure, which describes the smoothness of a movement, is a measure of the consistency and correctness of the input or the output frequency for a rhythmic task using an underactuated linear system. Thus, either measure (f_{input} or r_{smooth}) could be used successfully to determine performance of the frequency skill related to this task. Either measure might have certain benefits depending on access to the input and output state variables, computation in real time or off-line, and types of disturbances in the system. After verifying that the results reported in terms of f_{input} and $1/r_{\text{smooth}}$ were very strongly correlated and almost identical, we chose to report our results only in terms of input frequency. Like f_{input}, e_{traj} is desirable because it also is not task-specific like the objective measure of n_{hit}. Additionally, while both f_{input} and e_{traj} correlate strongly with n_{hit}, they are poorly correlated with each other, suggesting their independence.

Finally, to investigate the learning effects of our target-hitting task, we studied performance (measured in terms of hit count, trajectory error, and input frequency) versus training session for our 17 participants. We categorized participants into three groups: expert, intermediate, and novice based on their performance in the objective task performance measure n_{hit} at the beginning and end of training. We investigated whether experts (defined in terms of hit count) are significantly better at the task over time when compared to intermediate and novice participants, even when using the new performance measures. Our analysis found statistically significant differences in performance between all three expertise-based groups.

We have identified the key skills necessary to achieve high numbers of target hits for the virtual target-hitting task, and we have defined two independent performance measures (trajectory error and input frequency) that correlate strongly with target hit count. Our intent in this work was to determine a set of performance measures that could be used to tune a haptic guidance scheme. Such performance measures should be necessary (independent from each other) and sufficient (complete as a set) to capture the required skills. For this task, we determined that two performance measures were necessary to capture the skills necessary to achieve high numbers of target hits in each trial. The correlations of both measures to the number of target hits suggest that trajectory error and input frequency measures could be successfully employed as gains for a progressively decreasing guidance controller that demonstrates the key skills to the trainee, providing more assistance to participants whose performance is poor, and less assistance to participants whose performance is good. Because guidance schemes based on minimizing error and tuning the excitation frequency would be based on the analysis of key skills required for the target-hitting task, it is expected that such schemes would outperform prior approaches such as fixed-gain ERSC or virtual fixtures (Li et al., 2009).

5 Conclusion

We propose that, employed in VTEs, haptic guidance paradigms must be based on measurements of the key skills that are critical to successful task completion. The performance of participants completing a virtual target-hitting task was analyzed to determine the key skills necessary for success, measured by the number of target hits during a trial. Two key skills of the virtual target-hitting task, namely minimization of trajectory error and excitation of the virtual dynamic system near resonance, were determined. We defined a performance measure for each skill and found a strong correlation with the objective hit count measure when comparing data from 17 participants of varying skill levels. The measures e_{traj} and f_{input} have high correlation to the objective measure of n_{hit}, yet have low correlation to each other, suggesting independence. A third measure, smoothness ratio (r_{smooth}), was evaluated, discussed, and shown to be equivalent to input frequency.

After defining the three performance measures, the participants were grouped based on their hit count scores into three performance-based groups (experts, intermediates, and novices). Learning effects in terms of each of the performance measures across the training sessions showed that improvements in trajectory error and input frequency indeed correspond to improved target hit count scores when examined by performance group. Future guidance schemes, aimed at enhancing the effectiveness of VTEs, should incorporate mechanisms for measuring and emphasizing these skills to the participants.

Acknowledgments

The authors wish to thank Yanfang Li for her valuable contribution to this work. This research was supported in part by a grant from the National Science Foundation (IIS-0448341). Joel Huegel's work at Rice was funded in part by a grant from the Instituto Tecnologico y de Estudios Superiores de Monterrey (ITESM) in Guadalajara, Mexico.

References

Abbott, J. J., & Okamura, A. M. (2006). Stable forbidden-region virtual fixtures for bilateral telemanipulation. *Journal of Dynamic Systems, Measurement, and Control, 128*(1), 53–64.

Abernethy, B., Farrow, D., & Berry, J. (2003). Constraints and issues in the development of a general theory of expert perceptual-motor performance. In J. Starkes & K. A. Ericsson (Eds.), *Expert performance in sports: Advances in research on sport expertise* (pp. 349–369). Champaign, IL: Human Kinetics.

Abernethy, B., Neal, R., & Koning, P. (1994). Visual-perceptual and cognitive difference between expert, intermediate, and novice snooker players. *Applied Cognitive Psychology, 8*(3), 185–211.

Abernethy, B., Thomas, K. T., & Thomas, J. T. (1993). Strategies for improving understanding of motor expertise (or mistakes we have made and things we have learned!). In J. Starkes & F. Allard (Eds.), *Cognitive issues in motor expertise* (pp. 317–356). Amsterdam, the Netherlands: Elsevier Science/North-Holland.

Adams, R. J., Klowden, D., & Hannaford, B. (2001). Virtual training for a manual assembly task. *Haptics-e Journal, IEEE Robotics & Automation Society, 2*(2), 1–7.

Bernstein, N. (1967). *The coordination and regulation of movements*. London: Pergamon.

Celik, O., O'Malley, M. K., Boake, C., Levin, H., Fischer, S., & Reistetter, T. (2008). Comparison of robotic and clinical motor function improvement measures for sub-acute stroke patients. In *Proceedings of the IEEE International Conference on Robotics and Automation* (*ICRA '08*), 2477–2482.

Colombo, R., Pisano, F., Micera, S., Mazzone, A., Delconte, C., Carrozza, M. C., et al. (2005). Robotic techniques for upper limb evaluation and rehabilitation of stroke patients. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 13*(3), 311–324.

Crespo, L. M., & Reinkensmeyer, D. J. (2008). Haptic guidance can enhance motor learning of a steering task. *Journal of Motor Behavior, 40*(6), 545–557.

Dingwell, J. B., Mah, C. D., & Mussa-Ivaldi, F. A. (2004). Experimentally confirmed mathematical model for human control of a non-rigid object. *Journal of Neurophysiology, 91*(3), 1158–1170.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. New York: The Free Press.

Emken, J., & Reinkensmeyer, D. (2005). 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.

Feygin, D., Keehner, M., & Tendick, F. (2002). Haptic guidance: Experimental evaluation of a haptic training method for a perceptual motor skill. In *Proceedings of the IEEE 10th Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems* (*HAPTICS'02*), 40–47.

Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 243–285). New York: Academic Press.

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Belmont, CA: Brooks.

Flash, T., & Hogan, N. (1985). The coordination of arm movements: An experimentally confirmed mathematical model. *Journal of Neuroscience, 5*(7), 1688–1703.

Gallagher, A. G., Ritter, E. M., Champion, H., Higgins, G., Fried, M. P., Moses, G., et al. (2005). Virtual reality simulation for the operating room: Proficiency-based training as a paradigm shift in surgical skills training. *Annals of Surgery, 241*(2), 364–372.

Gillespie, R. B., O'Modhrain, M., Tang, P., Pham, C., & Zaretzky, D. (1998). The virtual teacher. In *Proceedings of the ASME Dynamic Systems and Control Division,* 171–178.

Griffiths, P., & Gillespie, B. (2004). Shared control between human and machine: Using a haptic steering wheel to aid in land vehicle guidance. In *Proceedings of the IEEE 12th International Symposium on Haptic Interfaces for Virtual Envi-*

ronment and Teleoperator Systems (*HAPTICS '04*), 358–366.

Henmi, K., & Yoshikawa, T. (1998). Virtual lesson and its application to virtual calligraphy system. In *Proceedings of the IEEE International Conference on Robotics and Automation* (*ICRA 1998*), 1275–1280.

Hogan, N., & Sternad, D. (2007). On rhythmic and discrete movements: Reflections, definitions and implications for motor control. *Experimental Brain Research, 181*(1), 13–30.

Huang, F. C., Gillespie, R. B., & Kuo, A. D. (2007). Visual and haptic feedback contribute to tuning and online control during object manipulation. *Journal of Motor Behavior, 39*(3), 179–193.

Israr, A., Kapson, H., Patoglu, V., & O'Malley, M. K. (2009). Effects of magnitude and phase cues on human motor adaptation. In *Proceedings of the IEEE Third Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems* (*WHC '09*), 344–349.

Jay, C., Stevens, R., Hubbold, R., & Glencross, M. (2008). Using haptic cues to aid nonvisual structure recognition. *ACM Transactions on Applied Perception, 2,* 1–14.

Li, Y., Patoglu, V., & O'Malley, M. K. (2006). Shared control for training in virtual environments: Learning through demonstration? In *Proceedings of EuroHaptics,* 93–99.

Li, Y., Patoglu, V., & O'Malley, M. K. (2009). Negative efficacy of fixed gain error reducing shared control for training in virtual environments. *ACM Transactions on Applied Perception, 6*(1), 21.

Lintern, G., Roscoe, S. N., Koonce, J. M., & Segal, L. D. (1990). Transfer of landing skills in beginning flight training. *Human Factors, 32*(3), 319–327.

Liu, J., Cramer, S., & Reinkensmeyer, D. (2006). Learning to perform a new movement with robotic assistance: Comparison of haptic guidance and visual demonstration. *Journal of Neuroengineering and Rehabilitation, 31,* 20.

Morris, D., Tan, H., Barbagli, F., Chang, T., & Salisbury, K. (2007). Haptic training enhances force skill learning. In *Proceedings of the IEEE Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems* (*WHC '07*), 21–26.

Nistor, V., Allen, B., Dutson, E., Faloutsos, P., & Carman, G. P. (2007). Immersive training and mentoring for laparoscopic surgery. In *Nanosensors, Microsensors and Biosensors and Systems 2007, Proceeding of SPIE,* vol. 6528, 1–11.

O'Malley, M. K., & Gupta, A. (2003). Passive and active assistance for human performance of a simulated underactuated dynamic task. In *Proceedings of the IEEE 11th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (*HAPTICS '03*), 348–355.

O'Malley, M. K., Gupta, A., Gen, M., & Li, Y. (2006). Shared control in haptic systems for performance enhancement and training. *Journal of Dynamic Systems, Measurement, and Control, 128*(1), 75–85.

O'Toole, R. V., Playter, R. R., Krummel, T. M., Blank, W. C., Cornelius, N. H., Roberts, W. R., et al. (1999). Measuring and developing suturing technique with a virtual reality surgical simulator. *Journal of the American College of Surgeons, 189*(1), 114–127.

Patton, J., & Mussa-Ivaldi, F. (2004). Robot-assisted adaptive training: Custom force fields for teaching movement patterns. *IEEE Transactions on Biomedical Engineering, 51*(4), 636–646.

Rissanen, M. J., Kuroda, Y., Nakao, M., Kume, N., Kuroda, T., & Yoshihara, H. (2007). Toward visualization of skill in VR: Adaptive real-time guidance for learning force exertion through the shaping strategy. In *Proceedings of the IEEE Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems* (*WHC '07*), 324–329.

Rosen, J., Hannaford, B., Richards, C. G., & Sinanan, M. N. (2001). Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. *IEEE Transactions on Biomedical Engineering, 48*(5), 579–591.

Rosenberg, L. B. (1993). The use of virtual fixtures to enhance operator performance in time delayed teleoperation. *Journal of Dynamic Systems Control, 49*(1), 29–36.

Sallnäs, E.-L., Rassmus-Gröhn, K., & Sjöström, C. (2000). Supporting presence in collaborative environments by haptic force feedback. *ACM Transactions on Computer-Human Interaction, 7*(4), 461–476.

Srimathveeravalli, G., Gourishankar, V., & Kesavadas, T. (2007). Comparative study: Virtual fixtures and shared control for rehabilitation of fine motor skills. In *Proceedings of the IEEE Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems* (*WHC '07*), 304–309.

Sutherland, L. M., Middleton, P. F., Anthony, A., Hamdorf, J., Cregan, P., Scott, D., et al. (2006). Surgical simulation: A systematic review. *Annals of Surgery, 243*(3), 291–300.

Svinin, M., Goncharenko, I., Zhi-Wei, L., & Hosoe, S. (2006). Reaching movements in dynamic environments:

How do we move flexible objects? *IEEE Transactions on Robotics, 22*(4), 724–739.

Todorov, E., Shadmehr, R., & Bizzi, E. (1997). Augmented feedback presented in a virtual environment accelerates learning of a difficult motor task. *Journal of Motor Behavior, 29*(2), 147–158.

Tzafestas, C., Birbas, K., Koumpouros, Y., & Christopoulos, D. (2008). Pilot evaluation study of a virtual paracentesis simulator for skill training and assessment: The beneficial effect of haptic display. *Presence: Teleoperators and Virtual Environments, 17*(2), 212–229.

Williams, A., & Ericsson, K. (2005). Perceptual-cognitive expertise in sport: Some considerations when applying the expert performance approach. *Human Movement Science, 24*(3), 283–307.

Williams, R., Srivastava, M., Conatser, R., & Howell, J. (2004). Implementation and evaluation of a haptic playback system. *Haptics-e Journal, IEEE Robotics & Automation Society, 3*(3), 1–6.

Yokokohji, Y., Hollis, R. L., Kanade, T., Henmi, K., & Yoshikawa, T. (1996). Toward machine mediated training of motor skills—Skill transfer from human to human via virtual environment. In *Proceedings of the IEEE 5th International Workshop on Robot and Human Communication*, 32–37.