

# Generic and Systematic Evaluation of Haptic Interfaces Based on Testbeds

Evren Samur, Fei Wang, Ulrich Spaelter, and Hannes Bleuler

**Abstract**—The purpose of evaluation procedures is to achieve both qualitative and quantitative statements on haptic rendering realism and performance. Since a haptic interface provides an interaction between a user and a virtual environment, fidelity of a haptic interface directly affects the performance. To our knowledge, a standard, generic and reusable validation method which comprehensively addresses all the attributes of haptic feedback has not been realized yet. Despite the large number of human factor studies, only few of them have been proposed as well for haptic interface performance measurements. For this reason, we review validation procedures for haptic rendering and propose an evaluation method based on testbeds to obtain a systematic haptic interface assessment. We integrated the approaches of human factor studies into the testbeds to obtain a simple and yet complete measure of human-machine interaction performance. The testbeds were tested on a haptic interface, the IHP of Xitact SA, and performance results are presented. In the testbeds, performance metrics for generic haptic interaction tasks are expressed in terms of information transfer (bits) and sensory thresholds which are indeed device specific benchmark metrics. Thus, the suitability of a haptic interface for a defined task can be verified, device comparisons become possible and the obtained information can be used to identify possible improvements.

## I. INTRODUCTION

A haptic interface is a motorized and instrumented device that allows a human user to touch and manipulate objects within a virtual environment (VE) [1]. The haptic interface ensures haptic interactions between the user and VE in a haptic rendering process as shown in Figure 1.a. Therefore, the quality of a haptic device and controller design directly affects realism, presence and immersion of a VE.

There is a definite need for a norm for the evaluation of haptic applications as haptic technology is playing an increasingly important role in entertainment, like computer games, in steer-by-wire systems of next-generation cars, and in medicine as surgery assist devices or as interfaces in surgery simulators. Haptic interface evaluation is necessary in order to make both quantitative and qualitative statement on rendering realism, performance and enhancement. *Performance of a haptic interface*, in a broader sense, can be defined as its ability to render a wide range of haptic stimuli. Yet, to define performance measures is more challenging since purely technical device evaluation (i.e., bode-diagram, bandwidth) are not straightforward to interpret the results in terms of *perceived rendering quality* (i.e., “feel”). Hence, the

aim of this study is to apply human performance estimations to haptic interactions in order to have a systematic and complete evaluation method for haptic interfaces. Although there are a variety of task specific evaluation approaches for haptic interfaces, it has not yet been possible to define a standard for meaningful device comparison and assessment. An evaluation procedure for haptic interfaces should link device performance measures to the limits of human perception in order to obtain device-specific limits. Besides, it needs to be simple enough to be easily applied while taking all important attributes of haptic interaction into account. These necessities direct the attention to usability evaluations. *Usability evaluation* is defined as the assessment of a specific application’s user interface, an interaction technique, or an input/output device, for the purpose of determining its actual or probable usability [2]. Although usability studies have been extensively applied to traditional graphical user interfaces, there are only few systematic evaluation processes for haptic interfaces.

In this paper, we first review and classify the state of the art of haptic rendering evaluation techniques. Based on these resources, we apply Bowman and Hodges’ testbed evaluation approach [3] to haptic interactions and synthesize a set of evaluation testbeds and benchmark metrics for haptic interfaces. We describe the methodology of eight testbeds and demonstrate the experimental results on a force-feedback device, the IHP of Xitact SA [4]. Finally, we discuss the possible future extensions of our research.

### A. State of the art

A closer investigation of studies on the evaluation of haptic rendering shows that type of evaluations, aim of methods and performance metrics vary considerably in these studies. We have therefore categorized the evaluation methods in the literature that have been applied to haptic interactions including VE, control, device as well as the human operator (see Fig. 1). Some of these methods employ only algorithm validation and comparison based on rendering realism [5], [6], whereas some others studied control design and evaluation for haptic interfaces [7], [8], [9]. Several projects [10], [11] evaluated particular haptic devices based on technical performance metrics given by Hayward and Astley [12]. There are many human factor studies to assess the benefits of haptic feedback on sensory-motor control tasks. Peg-in-hole [13], [14], tapping [15], [16], targeting [17], haptic training [18], joint tasks in a shared VE [19] and object recognition [20], [21], [22] tests are the most frequently performed experiments in these studies.

This research has been supported in part by the NCCR Co-Me of the Swiss National Science Foundation

E. Samur, F. Wang, U. Spaelter and H. Bleuler are with Laboratory of Robotic Systems, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland (e-mail: evren.samur@epfl.ch)

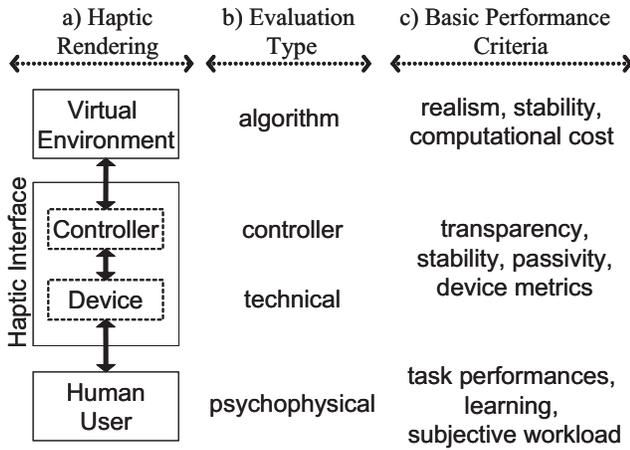


Fig. 1. Classification of haptic rendering evaluation techniques.

In spite of the large number of human factor studies, only few of the tests have been used to measure the performance of a haptic interface rather than the haptic feedback itself. Wall and Harwin [15] employed a tapping test in conjunction with Fitts' law [23] in order to establish a measure of human performance in a simple target selection task. They showed that the providing force feedback significantly reduced subjects' movement times. In another study [24], they measured the performance of their high bandwidth device in a perceptual context of roughness [21] in order to fully evaluate its contribution to the haptic system. They demonstrated that different haptic interfaces have different performance characteristics in rendering the surface roughness. Harders et al. [25] performed 3D peg-in-hole tests to compare three different haptic devices. Rendering hard virtual walls [8] has been the most mentioned benchmark topic in evaluating the performance of haptic interfaces. Guerraz et al. [26] suggested to use physical data from a haptic device to evaluate haptic user interfaces. Kappers et al. [27] performed haptic identification experiments using quadric surfaces and showed that both shape index, a quantity describing the shape, and curvedness had significant effect on haptic shape identification. Based on this research, Kirkpatrick and Douglas [28] used shape recognition as an evaluation method for complete haptic system. Their protocol can be used as a benchmark task to evaluate new haptic interface designs but it does not comprise all haptic interactions. Moreover, Tan [20] applied the absolute identification paradigm to sphere size identification for human performance estimations. Results were expressed in bits of information transfer and showed that humans could correctly identify at most 3 to 4 sphere sizes (corresponding to 2 bits) ranging from 10 to 80 mm in radius using the PHANToM (SensAble Technologies Inc.). This conclusion is also consistent with the results of manual length identification with physical objects given in literature, thus 2 bits of information transfer ( $IT$ ) can be used as the threshold of identification performance of human for device evaluation. Murray et al. [29] used this information transfer concept to evaluate their wearable vibrotactile glove.

## II. TESTBED EVALUATION METHODOLOGY

The main goal of the testbed evaluation approach is to find generic performance characteristics. This yields general and complete results which can be applied to any VE application using the tasks studied within a testbed.

### A. Categorization of haptic interaction tasks

The first step towards a complete testbed is to gain an intuitive understanding of the generic interaction tasks and current techniques available for the tasks [3]. The initial evaluation experiences, which can be extracted from human factor studies in the literature, result in a taxonomy, independent outside factors/variables, and performance measures. Outside factors consist of task, environment, user and system characteristics which may individually affect user performance. Performance measures, on the other hand, include both quantifiable metrics and subjective performance values.

Kirkpatrick and Douglas [28] suggested a taxonomy of haptic modes. They divided the taxonomy as motor control and perception, which is shown in Fig. 2.a. This highest level distinguishing is also consistent with haptic interaction categories of human sensory-motor system. Jandura and Srinivasan [30] divided haptic interactions into exploration and manipulation which are dominated by sensory and motor system, respectively.

Based on this initial categorization into motor control and perception, next step is defining generic interaction tasks. Bowman and Hodges [3] propose that many interactions within a VE may be divided into three general tasks as *travel* (movement of one's viewpoint), *selection* (act of choosing a virtual object) and *manipulation* (task of setting the position and orientation of a selected object) (Fig. 2.b). They further brake down these tasks into subtasks and included the haptic feedback into their taxonomy. These generic interaction tasks

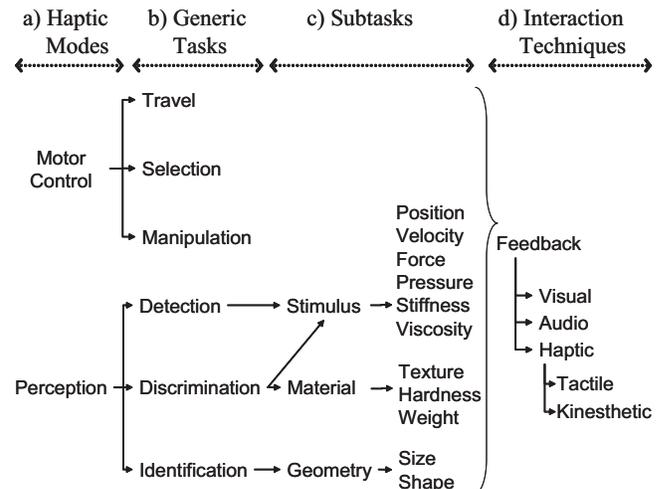


Fig. 2. Taxonomy of Haptic Interaction Tasks. a) the classification of haptic systems [28] according to the aim of a task, b) the generic interaction tasks related to haptics, c) the subtasks for perception mode (the subtasks for motor control mode can be found in [3]), d) the feedback interaction techniques

correspond to the motor control mode of haptics and do not represent the perception mode directly.

The three general methods in psychophysics, *detection*, *discrimination*, and *identification*, are used to determine human thresholds of perception. They may also be interpreted as the generic interaction tasks of object exploration dominated by the perception mode of haptics (Fig. 2.b). Detection and discrimination involve the measurement of sensory thresholds of perception of a stimulus (e.g., position, velocity, force, pressure, stiffness, and viscosity) as shown in Fig. 2.c. An *absolute threshold*, the minimum amount of stimulation required for a human to detect a stimulus, is determined by a detection experiment. On the other hand, a discrimination test results in a *differential threshold* which is the smallest difference in a stimulus detectable by a human. The differential threshold is also known as the just-noticeable-difference (JND) or difference threshold. The third generic task in perception mode is identification, which involves human ability to categorize stimuli without providing explicit references [20]. Psychophysical studies on human haptic identification of objects show that material properties (e.g., texture, hardness, and weight) compensate haptic object recognition in case of limited access to object structural cues (shape and size) but are not the primary basis for haptic object identification [31]. In other words, objects can be recognized by geometrical cues <sup>1</sup>. For this reason, identification of geometric dimensions and discrimination of material properties should be examined separately in order to properly evaluate the effectiveness of haptic interfaces on object recognition (Fig. 2.c).

After defining generic interaction tasks, corresponding interaction techniques can be determined. *Interaction techniques* combine natural human capabilities (particularly communication, motor, cognitive, and perceptual skills) with computer I/O devices [33]. The feedback techniques connect human sensory and motor system to haptic devices and include not only haptic but also visual and audio types (Fig. 2.d). The haptic feedback can be divided into two interaction technique components as kinesthetic (or force) and tactile feedback since they are sensed by different receptors in the human body and, therefore, have different effects on the sensory-motor control. As a result, the main independent variable should be feedback technique including two components of haptic feedback in the testbed evaluation of haptic interfaces.

### B. Performance metrics

Basic quantitative performance metrics for task analysis are time and accuracy which are easy to measure. However, these two metrics may not be enough to quantitatively assess and compare the effectiveness of a haptic interface. Therefore, a simple measure characterizing the human-machine interaction is required. Fitts [23] applied the concepts of information theory to the human motor control system and

provided an index of performance (*IP*) which expresses the results of a movement as a performance rate in *bit/s*. There are some studies in literature [13], [15], [16] showing that Fitts' law is an effective quantitative method for evaluating input/output devices including haptic interfaces. Similar to Fitts, Tan [20] proposed that information transfer between a human and a device which is expressed in *bits* is a possible way to make direct performance comparisons for different haptic interfaces. Consequently, the measures of information transfer are better than the measures of accuracy and time in evaluation of haptic interfaces.

The resolution limits and other metrics of a haptic device significantly influence the results of human threshold experiments. Weisenberger et al. [21] suggested that if two devices had different fidelities, then it was possible that differences in sensory thresholds might be attributable to the device rather than to the user's perceptual system. Therefore, careful repetition of described human factor studies leads to basic quantitative metrics for evaluation of haptic interfaces. For instance, sensory threshold values from detection and discrimination experiments for each kind of stimulus can be assessed to evaluate a particular haptic interface by considering the results from other devices and manual threshold experiments with real objects.

The resolution and the range of a system can also be combined and expressed in terms of *dynamic range* which is often given in decibels [34]. Dynamic range is the ratio of a specified maximum level of a parameter, such as force, to the minimum detectable value of that parameter.

## III. IMPLEMENTATION AND METHODS

Based on the taxonomy of haptic interaction (Fig. 2) and corresponding psychophysical experiments in the literature, eight testbeds were implemented to evaluate effectiveness of a haptic interface. Four examples representing the VE of testbeds are shown in Fig. 3.

### A. Travel and selection

Fitts' tapping task was chosen for travel and selection testbed which involves selecting targets by touching them and measuring the movement time between tapping the targets. The movement time (*MT*) equation of Fitts' law is given as Shannon formulation [35]

$$MT = a + b \log_2(A/W + 1) \quad (1)$$

where *a* and *b* are constants determined from experiments by curve fits, *A* is the distance of movement, and *W* is the width of target, which corresponds to required accuracy. The logarithmic term is called the index of difficulty (*ID*) which is in bits and the reciprocal of *b* is called the index of performance (*IP*) which expresses the results as a performance rate (bit/s). The parameters *a* and *b* are system dependent, thus can be used as benchmark metrics for performance comparisons of different haptic devices.

In this testbed, users are asked to tap alternately two virtual plates which are 50 mm long. The plates are generated on the floor of the workspace. In order to examine a wide

<sup>1</sup>In the text, geometrical and force cues are used in a correlated way. Otherwise, shape features are identified by force cues [32].

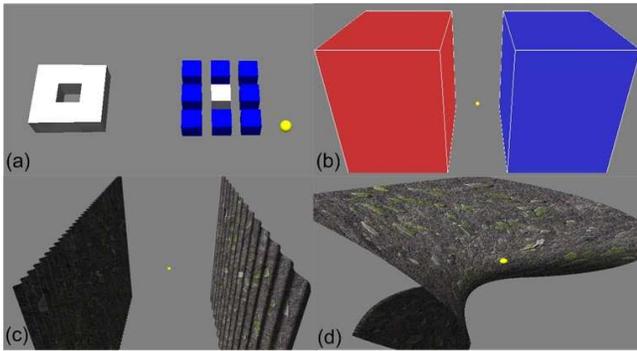


Fig. 3. Virtual experimental setup: a) Selection and manipulation b) Force discrimination c) Texture discrimination d) Shape identification

information range (i.e.,  $ID =$  from 1.0 to 6.0 bits), seven combinations of amplitude and target size are implemented. For this purpose, three different widths ( $W = 4, 8$  and  $16$  mm) and three center to center distances ( $A = 16, 64$  and  $256$  mm) between the target plates are studied. Two unnecessary combinations (i.e.,  $W, A = 4, 16$  and  $16, 256$ ) are not included to the experimental trials. Force feedback with a visual cue is provided to indicate taps in each case. The order of the 7 trials is randomly assigned. Each trial lasts 15 sec. and is followed by a 10 sec. rest period. The number of taps are recorded.

### B. Selection and manipulation

The peg-in-hole test was implemented for selection and manipulation testbed because it requires accuracy and precision to complete the manipulation task. In this testbed, the user selects the center object (i.e. peg) from a group of objects and places it within a target area (i.e. hole) as shown in Fig. 3.a. Similar to the tapping task, the  $MT$  formulation given in Eqn.1 is used where  $W$  stands for precision, the difference between the size of hole and peg (i.e.,  $W = H - P$ ).

In the experiments, a cube with an edge length ( $P$ ) of 10 mm is modeled as the peg. In order to obtain an information range ( $ID$ ) from 3.0 to 8.0 bits, six combinations of distance and precision are implemented. Four levels of precision ( $W = 1, 2, 4$  and  $8$  mm) and two levels of center to center distances ( $A = 64$  and  $256$  mm) between the peg and hole are studied. Two unnecessary combinations (i.e.,  $W, A = 1, 64$  and  $8, 256$ ) are not included to the experimental trials. Force feedback with a visual cue is provided to indicate selection. The density of the group is fixed to nine objects and not considered as an experimental variable. Each trial ends after three successful inserts and there is a 10 sec. rest period between each trial. Time is recorded in this testbed.

### C. Detection

The quantitative performance metrics of a haptic interface as absolute thresholds are basically position, velocity, force, pressure, stiffness, and viscosity (Fig. 2.c). In this testbed, only the force detection experiment is examined that reveals an absolute threshold for force.

The mostly preferred standard experimental method for measuring absolute and differential thresholds is the *method of constant stimuli* in which equally spaced stimuli are presented in a random order. This method reduces expectation errors, which result from human prediction of certain scenarios, occurring in case of presenting stimuli in ascending or descending order. In this testbed, force stimuli ranging from 0.1 to 0.6 N with 0.1 N increments are fully crossed with the three axis ( $X, Y$  and  $Z$ ) and two directions (+ and -). Each trial lasts 10 sec. and a force stimulus with different magnitude and direction is randomly presented to the user after 4 sec. At the end of each trial, users response whether the stimulus is detectable or not.

### D. Discrimination

Differential threshold values can be determined for a specific haptic interface by performing a discrimination experiment. In order to measure differential thresholds using the method of constant stimuli; two stimuli, one of which being the reference, are presented simultaneously and subjects indicate whether the other stimulus is perceived differently than the reference stimulus or not. Measured differential thresholds can also be represented by *Weber fractions* as  $\Delta I/I$  that is the ratio of the differential threshold ( $\Delta I$ ) to the magnitude of the reference stimulus ( $I$ ). Jones and Hunter [36] provided Weber fractions for human perception of position (0.08), movement (0.08), force (0.15), stiffness (0.23), and viscosity (0.34). Three discrimination testbeds revealing force, spatial period and stiffness thresholds were designed considering these Weber values.

1) *Force discrimination*: Three factors that must be controlled in this discrimination experiment are reference stimulus ( $I$ ), Weber fraction ( $\Delta I/I$ ) and direction of the force stimulus. In order to treat these three factors at the same time with less trials, a  $6 \times 6$  Latin square design is implemented. The experimental design is shown in Table I. Equally spaced values of  $I$  (from 1 to 6 N) are given in rows and columns represent the six values of  $\Delta I/I$ . The directions ( $+X, -X, +Y, -Y, +Z$  and  $-Z$ ) are assigned randomly within rows and columns with each direction appearing once in every row and every column. A total of 36 combinations are presented to users in a random order. Two virtual force regions, one of which is being the reference, are presented side by side in the middle of the workspace (see Fig. 3.b). Two forces have different magnitudes but the same directions. Users are asked to explore both regions and determine whether the other force stimulus is different than the reference stimulus or not.

2) *Texture discrimination*: The easiest way to judge the texture rendering capability of haptic interfaces is to use different periodic gratings and try to distinguish them. Two textures are modeled by sinusoidal gratings and located on the side walls of the virtual environment facing each other (see Fig. 3.c). The same Latin square design as shown in Table I is implemented in which three independent variables are the amplitude and spatial period of gratings and the Weber fractions for period. The amplitude is varied from

TABLE I  
THE 6 × 6 LATIN SQUARE DESIGN.

Reference, $I(N)$	Weber Fractions, $\Delta I/I$					
	0.1	0.2	0.3	0.4	0.5	0.6
1	+X	-X	+Y	-Y	+Z	-Z
2	+Y	-Y	+Z	-Z	+X	-X
3	-Z	+X	-X	+Y	-Y	+Z
4	-Y	+Z	-Z	+X	-X	+Y
5	-X	+Y	-Y	+Z	-Z	+X
6	+Z	-Z	+X	-X	+Y	-Y

1 to 6 mm with 1 mm increments and the spatial period is varied from 2 to 12 mm with 2 mm increments. The weber fractions take the values of 0.1, 0.2, 0.4, 0.6, 0.8 and 1.0. The orientation of the gratings is kept constant. Users are asked to explore both textures, which have different periods but the same amplitudes, and determine whether they can distinguish the difference between the periods of gratings or not. The amplitude of the gratings is varied for different pairs of textures.

3) *Hardness discrimination*: A differential threshold for stiffness can be determined by performing a testbed experiment for hardness discrimination. Two virtual surfaces with different stiffness characteristics are used in this testbed experiment. The stiffness values of surfaces varying from 0.3 to 1.8 N/mm with 0.3 N/mm steps are fully crossed with the same Weber fractions in Table I. Users are asked to correctly discriminate different hard virtual walls which are presented on the side walls of the workspace facing each other.

#### E. Identification

Identification of geometric properties can be divided into shape and size identification. Although these two subtasks require different experiments, similar quantitative performance metrics can be used for evaluation. Since human can correctly identify 4 sphere sizes ranging from 10 to 80 mm in radius [20], 2 bits of information transfer ( $IT$ ) can be used as the threshold of identification performance of humans for device evaluation. In other words, humans can correctly identify up to 4 stimulus categories ( $k = 2^{IT}$ ) which should be supported by each haptic interface.

In the absolute identification paradigm [20], given  $k$  stimuli,  $S_i$ , corresponds to  $k$  responses,  $R_j$ , where  $1 \leq i, j \leq k$ . The maximum likelihood estimate of information transfer,  $IT_{est}$ , for a particular response pair ( $S_i, R_j$ ) is given as

$$IT_{est} = \sum_{j=1}^k \sum_{i=1}^k \frac{n_{ij}}{n} \log_2 \left( \frac{n_{ij} \cdot n}{n_i \cdot n_j} \right) \quad (2)$$

where  $n$  is the total number of trials,  $n_{ij}$  is the number of joint event ( $S_i, R_j$ ) occurs, and  $n_{i,j} = \sum_{i,j=1}^k n_{ij}$ .

1) *Size identification*: Size identification testbed is based on the experiment conducted by Tan [20]. It involves haptically identifying different sizes of objects which are presented in a random order. A virtual sphere is displayed in the middle of the workspace without a visual cue. Four different sizes (radius,  $R = 10, 30, 60$  and  $80$  mm) are modeled in order to obtain 2 bits information range. At each trial, users

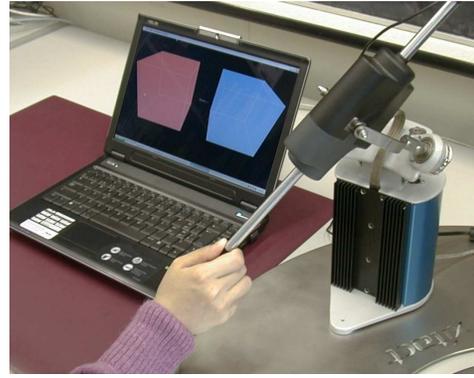


Fig. 4. Experimental setup

haptically explore the sphere and indicate the size by pressing the corresponding key on the keyboard. To avoid over-correct results, the total number of trials is set to 80 ( $n = 5k^2$ ).

2) *Shape identification*: The method proposed by Kirkpatrick and Douglas [28] is modified and used as shape identification testbed that requires the user to identify one of the quadric shapes. These shapes are constructed from two orthogonal parabolas and distinguished from one to another by the shape index  $S$ , a quantity describing the shape [27]. Four quadratic shapes are modeled: concave and convex cylindrical paraboloids ( $S = -0.5$  and  $0.5$  respectively) and two asymmetric saddles ( $S = -0.25$  and  $0.25$ ). One of the quadratic shapes ( $S = -0.5$ ) is shown in Fig. 3.d. They are presented in the center of the workspace without a visual cue. Similar experimental procedure as in the size identification test is performed. Curvedness, location and orientation of the shapes are not altered.

## IV. EXPERIMENTS AND RESULTS

### A. Experimental setup

The hardware components of the system include a LCD screen to display the virtual environment, and a force-feedback device (the IHP of Xitact SA [4]) to simulate haptic interactions (see Fig. 4). The xitact IHP is a 4-DOF haptic device designed for the simulation of a minimally invasive surgical procedure. The experiments were implemented on a standard laptop with Intel's Core 2 Duo 2.0 GHz CPU. A dual-thread software package was developed to implement the testbeds. Haptic and visual loop were updated at a rate of 500 and 100 Hz, respectively. The coordinate system was set in a way that  $X$ -axis points right and  $Y$ -axis points up.

### B. Participants and procedure

Five graduate students (one woman, four men), whose age ranges from 26 to 30, participated in the experiments. The procedure consisted of a separate training phase and a testing phase for each testbed. In the training phase, participants were told the procedure and they visually and haptically explored every stimuli in the testbed with the Xitact IHP. Once they explored the whole testbed in the training phase, the testing phase started. The overall experimental procedure lasted around 90 minutes.

TABLE II  
QUANTITATIVE RESULTS FOR THE XITACT IHP

Testbeds	Performance Metrics	Value	Units
Travel	$IP$ ( $1/b$ )	1.52	$bit/s$
	Intercept ( $a$ )	0.03	$s$
Manipulation	$IP$ ( $1/b$ )	2.5	$bit/s$
	Intercept ( $a$ )	2.0	$s$
Detection			
Force	Abs. Threshold	X: 0.6 Y: 0.3 Z: 0.5	$N$
	Dynamic Range	X: 19 Y: 36 Z: 18	$dB$
Discrimination			
	Weber fraction		
Force		X: 0.3 Y: 0.2 Z: 0.4	
	Texture	Z: 0.6	
Hardness		X: 0.4	
Identification			
	Info. Transfer ( $IT$ )		
Size		1.36	$bits$
Shape		1.12	$bits$

### C. Results and discussion

Quantitative performance results for each testbed are given in Table II. For the travel & selection and selection & manipulation testbeds, the index of performance ( $IP$ ) and intercept ( $a$ ), which are system dependent parameters, were determined by curve fits using Eqn. 1 as shown in Fig. 5. Small  $a$  and large  $IP$  values correspond to better performance and transparency (e.g., low friction and inertia). For instance, a comparison between our results and the results of a similar test performed using the PHANToM [15] show that the PHANToM enables faster movements ( $a = 0.028$ ,  $IP = 3.18$ ) than the Xitact IHP ( $a = 0.03$ ,  $IP = 1.52$ ). Lower inertia and cable transmission of the PHANToM provide slightly improved haptic transparency than the Xitact IHP which is actuated by a friction drive.

A linear statistical model was chosen for the analysis of detection testbed results. An analysis of variance (ANOVA) with the factors of stimulus level (six), direction (six), user (five) and interaction between stimulus and direction was performed to check the validity of the model. Only the stimulus,  $F(5, 140) = 7.84, p < .001$ , and direction,  $F(5, 140) = 19.50, p < .001$ , were statistically significant. The minimum detectable force stimulus for each direction was calculated based on the coefficients of the linear model. As shown in Table II, absolute force thresholds for  $X, Y, Z$  are 0.6, 0.3, and 0.5 N, respectively. These values can be interpreted as the minimum force that can be rendered by the Xitact IHP. Dynamic range, ratio of the maximum continuous force ( $X : 5.6, Y : 20, Z : 4$  N) to the minimum detectable force, for  $X, Y, Z$  are 19, 36 and 18 dB, respectively.

Differential threshold values expressed in terms of Weber fractions are given in Table II. An ANOVA for the force discrimination experiment with the factors of stimulus level (six), direction (six), weber fractions (six), user (five) and interaction between stimulus, direction and weber shows that the only significant factor is the interaction term,  $F(20, 140) = 2.27, p < .005$ . Careful investigation of the results reveals that since some of the experimental conditions exceed the maximum force, it was not possible to distinguish the difference for those stimulus levels. However, considering

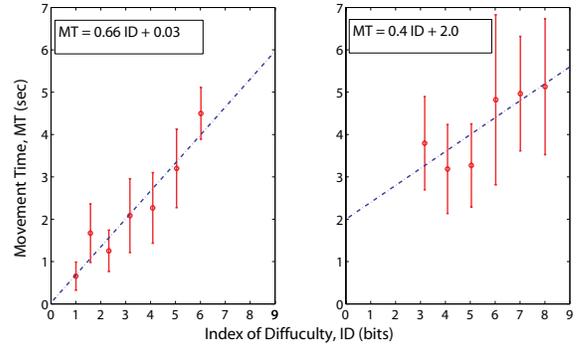


Fig. 5. Movement Time vs Index of Difficulty for the travel&selection (left) and selection&manipulation (right) testbeds. The data points represent the mean values of five subjects and the vertical bars correspond to the standard deviations over subjects. Dash lines represent linear curve fittings to the data points.

this fact, force weber fractions for  $X, Y, Z$  were found to be 0.3, 0.2, and 0.4, respectively. A similar ANOVA was performed for the texture discrimination testbed and the weber fraction level was significant,  $F(5, 140) = 19.48, p < .001$ . Neither amplitude nor spatial period level was significant. The weber fraction of the distinguishable spatial period is 0.6. The ANOVA results of the hardness discrimination experiment show that the effect of the weber fraction is the most significant,  $F(5, 140) = 4.89, p < .001$ , which yields a weber fraction of 0.4 for hardness discrimination. Comparing these results with Weber fractions for human perception of force (0.15) and stiffness (0.23) [36], we can conclude that the differences may be due to the device resolution limits.

The information transfer ( $IT$ ) of identification performance for size and shape was calculated using Eqn. 2. The estimated information transmitted in size and shape identification testbeds are 1.36 and 1.12 bits, respectively. These values are less than 2 bits of  $IT$ , the threshold identification performance of human, which shows that the performance is degraded by the use of the haptic interface.

## V. CONCLUSIONS

Evaluation procedures for haptic rendering must take into account all components of haptic interaction that are the VE, the haptic interface and the user behavior. In this paper, we proposed a series of evaluation testbeds derived from psychophysical experiments that examine haptic interactions. The proposed testbeds can be used to evaluate haptic interfaces in a generic context to judge on the quality of designs. Based on the taxonomy of the haptics mode [28], generic haptic interaction tasks for perception (detection, discrimination, and identification) have been derived. These testbed modules can also be seen as a complementary work to Bowman and Hodges' [3] testbeds on travel, selection, and manipulation. The thermal mode of haptic interaction is not included into the taxonomy. In the testbeds, performance results are measured in terms of information transfer expressed in  $bits$  and sensory (*absolute* and *differential*) thresholds which can be used as benchmark metrics to evaluate and

compare different haptic interfaces. The results of testbeds were demonstrated on a commercial force-feedback device and benchmark metrics for this device were presented.

During the testbed experiments, other performance metrics such as force and velocity profile and used workspace can also be collected for interface evaluations. Quantitative metrics are suitable for validation, yet, they do not completely represent the usability of the interface. Thus, qualitative metrics based on users' experiences during testbeds should be collected. User fatigue, discomfort and lack of immersion are strongly influenced by improper use of haptic feedback.

One of the potential limiting factors of generic evaluations is the assumption of a "perfect" VE. Results of the experiments depend on the rendering technique. However, in order to make a comparison between different interfaces, a standard VE system should be used. For this reason, in the future we will develop an open source library available to the haptic community in order to apply designed testbeds to other types of haptic interfaces. Another limiting factor is that in order to evaluate how well an application's haptic interface supports correspondent user tasks, application-specific evaluation methods should be conducted. For instance, the Xitact IHP, which is designed for laparoscopy, should be evaluated in terms of complete system and operation. However, the results of this kind of methods is only useful for similar applications, while, evaluations conducted in a generic context can be applied to many types of haptic interfaces.

## VI. ACKNOWLEDGMENTS

The authors would like to thank Lindo Duratti, Dr. Thomas Moix and Dr. Roger Gassert for their technical support and valuable comments.

## REFERENCES

- [1] R. B. Gillespie. Haptic interface to virtual environments. In *Robotics and Automation Handbook*. Tom Kurfess (ed.), CRC Press, 2004.
- [2] D. A. Bowman, J. L. Gabbard, and D. Hix. A survey of usability evaluation in virtual environments: classification and comparison of methods. *Presence: Teleoper. Virtual Environ.*, 11(4):404–424, 2002.
- [3] D. A. Bowman and L. F. Hodges. Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments. *Journal of Visual Languages and Computing*, 10(1):37–53, 1999.
- [4] Xitact SA. <http://www.xitact.com/products/lap/>.
- [5] E. Ruffaldi, D. Morris, T. Edmunds, F. Barbagli, and D. K. Pai. Standardized evaluation of haptic rendering systems. In *Proc. of IEEE Haptic Symposium*, pages 225–232, 2006.
- [6] P. Leskovsky, T. Cooke, M. Ernst, and M. Harders. Using multidimensional scaling to quantify the fidelity of haptic rendering of deformable objects. In *Proc. of Eurohaptics*, pages 289–295, 2006.
- [7] J. E. Colgate and J. M. Brown. Factors affecting the Z-width of a haptic display. In *IEEE Int. Conf. Robotics and Automation*, pages 3205–3210, 1994.
- [8] D. A. Lawrence, L. Y. Pao, A. M. Dougherty, M. A. Salada, and Y. Pavlou. Rate-hardness: A new performance metric for haptic interfaces. *IEEE Trans. Robotics and Automation*, 16(4):357–371, 2000.
- [9] B. Hannaford and J. Ryu. Time domain passivity control of haptic interfaces. *IEEE Trans. Robotics and Automation*, 18(1):1–10, 2002.
- [10] R. Ellis, O. Ismaeil, and M. Lipsett. Design and evaluation of a high-performance haptic interface. *Robotica*, 1996.
- [11] M. C. Cavusoglu, D. Feygin, and F. Tendick. A critical study of the mechanical and electrical properties of the phantom haptic interface and improvements for high-performance control. *Presence: Teleoper. Virtual Environ.*, 11(6):555–568, 2002.
- [12] V. Hayward and O. Astley. Performance measures for haptic interfaces. *Robotics Research: The 7th International Symposium*, pages 195–207, 1996.
- [13] B. Hannaford, L. Wood, D. McAfee, and H. Zak. Performance evaluation of a six axis generalized force reflecting teleoperator. *IEEE Transactions on Systems, Man, and Cybernetics*, 21:620–633, 1991.
- [14] B.J. Unger, A. Nicolaidis, P.J. Berkelman, A. Thompson, R.L. Klatzky, and R.L. Hollis. Comparison of 3-d haptic peg-in-hole tasks in real and virtual environments. In *IEEE/RSJ, IROS: 1751-1756*, 2001.
- [15] S.A. Wall and W.S. Harwin. Quantification of the effects of haptic feedback during a motor skills task in a simulated environment. In *Proc. of the 2nd PHANTOM Users Research Symposium*, pages 61–69, 2000.
- [16] K. Chun, B. Verplank, F. Barbagli, and K. Salisbury. Evaluating haptics and 3d stereo displays using fitts' law. In *Proc. of The 3rd IEEE Workshop on HAVE*, pages 53–58, 2004.
- [17] I. Oakley, M. R. McGee, S. A. Brewster, and P. D. Gray. Putting the feel in 'look and feel'. *CHI: 415-422*, 2000.
- [18] C.A. Avizzano, J. Solis, A. Frisoli, and M. Bergamasco. Motor learning skill experiments using haptic interface capabilities. In *IEEE, RO-MAN: 198-203*, 2002.
- [19] C. Basdogan, C. Ho, M. A. Srinivasan, and M. Slater. An experimental study on the role of touch in shared virtual environments. *ACM Transactions on Computer-Human Interaction*, 7(4):443–460, 2000.
- [20] H. Tan. Identification of sphere size using the phantom: Towards a set of building blocks for rendering haptic environment. In *ASME Annual Meeting:197-203*, 1997.
- [21] J. Weisenberger, M. Kreier, and M. Rinker. Judging the orientation of sinusoidal and square-wave virtual gratings presented via 2-dof and 3-dof haptic interfaces, 2000.
- [22] M. O'Malley and M. Goldfarb. The effect of force saturation on the haptic perception of detail. *IEEE/ASME Trans. on Mechatronics*, 7:280–288, 2002.
- [23] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Exp. Psychology*, 47:381–391, 1954.
- [24] S.A. Wall and W. Harwin. A high bandwidth interface for haptic human computer interaction. *Mechatronics*, 11:371–387, 2001.
- [25] M. Harders, A. Barlit, K. Akahane, M. Sato, and G. Székely. Comparing 6dof haptic interfaces for application in 3d assembly tasks. In *Proc. of EuroHaptics'06*, 2006.
- [26] A. Guerraz, C. Loscos, and H. Ritter Widenfeld. How to use physical parameters coming from the haptic device itself to enhance the evaluation of haptic benefits in user interface? In *Proc. of Eurohaptics*, 2003.
- [27] A.M. Kappers, J.J. Koenderink, and I. Lichtenegger. Haptic identification of curved surfaces. *Perception and Psychophysics*, 56 (1):53–61, 1994.
- [28] A. E. Kirkpatrick and S. A. Douglas. Application-based evaluation of haptic interfaces. In *Proc. of the 10th Haptic Symposium*, page 32, 2002.
- [29] A. M. Murray, R. L. Klatzky, and P. K. Khosla. Psychophysical characterization and testbed validation of a wearable vibrotactile glove for telemanipulation. *Presence: Teleoper. Virtual Environ.*, 12(2):156–182, 2003.
- [30] L. Jandura and M. A. Srinivasan. Experiments on human performance in torque discrimination and control. *Dynamic Systems and Control, ASME*, 55-1:369–375, 1994.
- [31] R.L. Klatzky, J. M. Loomis, S. J. Lederman, H. Wake, and N. Fujita. Haptic identification of objects and their depictions. *Perception and Psychophysics*, 54 (2):170–178, 1993.
- [32] G. Robles-De-La-Torre and V. Hayward. Force can overcome object geometry in the perception of shape through active touch. *Nature*, 412:445–448, 2001.
- [33] Matthew Turk and George Robertson. Perceptual user interfaces. *Communications of the ACM*, 43(3):32–34, 2000.
- [34] T. Moix. Mechatronic elements and haptic rendering for computer-assisted minimally invasive surgery training. *PhD Thesis EPFL*, No. 3306, 2005.
- [35] I. S. MacKenzie. Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, 7:91–139, 1992.
- [36] L. A. Jones and I. W. Hunter. A perceptual analysis of viscosity. *Exp. Brain Res.*, 94:343–351, 1993.