Towards a Standard on Evaluation of Tactile/Haptic Interactions

Ian Sinclair[1], Jim Carter[2], Sebastian Kassner[3], Jan van Erp[4],
Gerhard Weber[5], Linda Elliott[6], and Ian Andrew[7]

[1] MPB Technologies Inc., Montréal, Canada
ian.sinclair@mpbc.ca
[2] University of Saskatchewan, Saskatoon, Canada
carter@cs.usask.ca
[3] Technische Universität Darmstadt, Darmstadt, Germany
s.kassner@emk.tu-darmstadt.de
[4] TNO Human Factors, Soesterberg, Netherlands
jan.vanerp@tno.nl
[5] Technische Universität Dresden, Dresden, Germany
gerhard.weber@tu-dresden.de
[6] U.S. Army Research Lab, Fort Benning, USA
linda.r.elliott@us.army.mil
[7] HF Engineer, United Kingdom
andyand@talktalk.net

Abstract. Tactile and haptic interaction is becoming increasingly important; ergonomic standards can ensure that systems are designed with sufficient concern for ergonomics and interoperability. ISO (through working group TC159/SC4/WG9) is developing international standards in this subject area, dual-tracked as both ISO and CEN standards. A framework and guidelines for tactile/haptic interactions have recently been published as ISO 9241-910 and ISO 9241-920 respectively. We describe the main concepts and definitions in support of a new standard that describes how to evaluate tactile/haptic interactions and how to link this evaluation to previous standards. The new standard addresses three major aspects of the evaluation of a tactile/haptic system– the validation of system requirements, the verification that the system meets the requirements, and the overall usability of the system. Several measurement and analysis techniques are discussed, such as the calculation of scores for the determination of effectiveness. Tactile/haptic measurements have to be repeatable, and as an example we discuss how an appropriate model of the interaction with a virtual wall can be formed and used in evaluating a device.

Keywords: Guidelines, haptics, human computer interaction, standards, evaluation.

1 Introduction

Ergonomic standards go beyond providing consistency and interoperability. They help enhance usability in a number of ways, including improving effectiveness and

P. Isokoski and J. Springare (Eds.): EuroHaptics 2012, Part I, LNCS 7282, pp. 528–539, 2012.

avoiding errors, improving performance, and enhancing the comfort and well-being of users. Ergonomic standards provide a basis for analysis, design, evaluation, procurement, and even for arbitrating issues of international trade. Therefore, an ISO expert group has been working on standards documents for haptic interaction since 2005. ISO TC159/SC4/WG9 has reported on its progress at several conferences [11, 9, 10, 2] and published its first standard ISO 9241-920 Guidance on tactile and haptic interactions [6] in 2009; this was followed by a second standard, ISO 9241-910 Framework for tactile and haptic interaction [7] published in 2011.

As of 2012, the following countries are actively participating in WG9: Canada, USA, UK, The Netherlands, Sweden, Germany, South Korea, and Japan. Drafts produced by WG9 undergo a thorough review process, including rounds of commenting and voting on the drafts by National Technical Advisory Groups.

ISO TC159/SC4/WG9 is currently focusing on the evaluation of tactile/haptic interactions, to be published as ISO 9241-940. This paper presents a preliminary view into this future standard and invites participation in its evolution.

2 Framework for Evaluation

ISO TC159/SC4/WG9 recognizes that there are three major aspects of evaluation: validation, verification and usability, which relate the user, the requirements and the system under consideration.

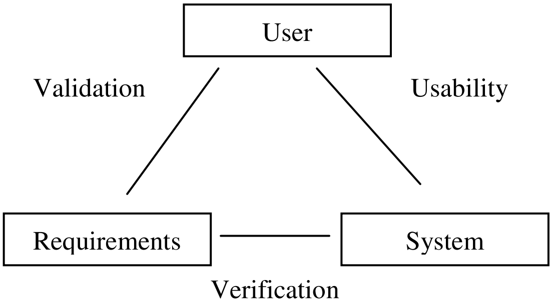

Fig. 1. User, requirements, system triangle

2.1 Validation

Validation evaluates the accuracy and completeness of the requirements' specifications. Boehm [1] states that validation involves answering the question: "Are we building the right system?" According to ISO/IEC-15288 [5], "This process performs a comparative assessment and confirms that the stakeholders' requirements are correctly defined."

Where criteria can be established from existing sources (e.g. ISO 9241-910), they should be specified in the requirements. Where a criterion cannot be established, usability testing (Section 2.3) becomes more important than validation.

Validation during development is generally performed by means of requirement reviews. Validation can also be performed on the set of user requirements that are used as input to an acquisition process (e.g. in a request for proposal). The quality of a validation is highly dependent on the quality of the communications between the user and the developer, especially on the clarity of the requirements documentation. Additional sources of pre-validated haptic requirements can come from International Standards such as ISO 9241-920 [6].

2.2 Verification

Verification tests the accuracy and completeness of the system and its operations. Boehm [1] states that verification involves answering the question: "Are we building the system right?" According to ISO/IEC 15288 [5], "The purpose of the Verification Process is to confirm that the specified design requirements are fulfilled by the system."

Verification requires specific criteria which can be measured by an appropriate technique. If specific criteria are not available, usability testing is an essential alternative to verification.

Verification during the development process is generally performed by testing that the system meets the requirements. Verification during acquisition is often limited to comparing user requirements to published specifications of a system, trusting that the published specifications have been properly verified. The quality of the verification is dependent on identifying the appropriate measures and measurement techniques. ISO 9241-940 will be suggesting measures and measurement techniques appropriate for haptic interactions.

2.3 Usability

Usability tests how well a user can operate or use a system. ISO 9241-11 [3] defines usability as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". It defines effectiveness as "the accuracy and completeness with which users achieve specified goals"; efficiency as "the resources expended in relation to the accuracy and completeness with which users achieve goals"; and satisfaction as "positive attitudes to the use of the product and freedom from discomfort in using it". It also defines context of use as, "users, tasks, equipment (hardware, software and materials), and the physical and social environments in which a product is used".

Usability testing answers the question, "Is the system right for the users and their tasks within the context of use?" Usability tests can be carried out during both development and system acquisition.

3 Evaluating Tactile/Haptic Interactions

The possible configurations for tactile/haptic interactions are large in number, and the possible tests that could be devised to evaluate such interactions are correspondingly numerous. To help make sense of the field, we list several types of interaction that can be evaluated in Table 1. Possible users range from experienced practitioners to the general public. The lists are not exhaustive, but form a representative set from which examples can be drawn and for which test procedures may be defined.

Table 1. Examples of interactions

Type	Auxiliary Display	Example	Test
Reality Simulation	Visual Display	Driving simulation	Lane change test
Virtual Display	Visual Display	Displaying earth map on a simulated globe	Tracing mountain ranges and valleys
Visual Control Panel	Visual Display	Selection button icons with haptic effects	Make selections in rapid sequence
Haptic Control Panel	Audible tones	Selection button icons with haptic effects but no visual counterpart	Navigate space to understand and make selections
Haptic Control Space	None	Controls for radio, air conditioning, seat positions while driving	Navigate control space to understand and make selections
Real Space Sensor	None	Cane with proximity sensors and haptic indications in the grip	Ease of navigating in a maze, especially by visually impaired users
Tactile Active Scanning	None	Braille reader with user-directed line pacing	Repeat verbally out loud a selection of Braille text
Tactile Passive Warning	None	Cell phone with silent ringer "tones" for selected callers	Recognize individual callers while distracted on other tasks

4 Validation

The validation of requirements for a tactile/haptic interaction begins with a clear statement of the goal of the interaction. Each requirement is then held up against this goal statement in order to judge its relevance to the goal.

For example, the goal could be to allow an artist to paint a scene in a virtual world. Detailed requirements can be laid out – camel hair brush to be held in either hand, the pressure of the brush on the virtual canvas to be reflected back to the user, simulating both water color and oil painting techniques. These should be translated to technical specifications – the required dynamic range and resolution of the brush pressure, for example.

The process can be aided by the construction of one or more operating scenarios. Such a scenario could reveal additional requirements – the speed of the brush stroke, the selection of colors from an on-screen pallet, for example. If the initial set of requirements had omitted a requirement such as brush stroke speed, then this can be inserted during the validation process.

If technical specifications are missing from requirements, then additional tests may be called for. For example, the range of brush pressures could be ascertained by having the artist paint a real canvas that is lying on a weigh scale. Several artists may be called upon to repeat the test, thereby accessing a range of opinions and thus extending the versatility of the planned system.

During the validation process, it will likely be necessary to prioritize the requirements. Limits can be encountered in both budget and available technology, leaving some desirable options to a future system. These may provide useful goals in the further development of tactile/haptic systems.

5 Verification

Tactile/haptic devices are generally constructed on electromechanical principles, independently of the tactile/haptic scenario software used in the interaction. Tactile devices are primarily directed at skin stimulation. They may use mechanical, thermal, chemical or electrical stimulation, although most rely on mechanical stimulation. Haptic devices are primarily directed at the kinesthetic senses, cues in the human body related to the sensing of joint angles, limb position and muscle tension. ISO 9241-910 presents general requirements for such devices, while Annexes A and B to that standard give examples of tactile and haptic devices.

5.1 Requirements for Devices

Whether a device is developed in a university laboratory or offered for sale, it should have a set of specifications attached to its performance. This allows comparison of devices. Meaningful comparison is possible only if the same measurement technique has been used to measure the parameter of interest.

As a principle of system engineering, each requirement should be measurable and testable. There is no point in setting a requirement that cannot be tested.

5.2 What Should Be Measured?

Lab measurements can be exacting and tedious, depending on the quantity to be measured. If a developer wants to characterize the device completely, then every clause in the specification should be either noted or measured. Table 2 shows examples of tactile/haptic system characteristics.

Table 2. Typical tactile/haptic system characteristics

Ergonomic	Temporal & General	Force & Torque	Environmental
Device-body interface	Bandwidth	Peak force & torque	Mobility
Degrees of freedom	System latency	Max continuous force & torque	Size
Motion range	Device latency	Min displayable force & torque	Weight
Working position	Maximum stiffness	Resolution of force & torque	Ease of installation
Limb support	Reliability	Dynamic range of force & torque	Ease of maintenance
	Fidelity	Peak acceleration	Thermal safety
	Modifiable to a task	Static friction	Electrical safety
	Adaptable to a task	Free space motion resistance	Mechanical safety
		Inertia	Acoustic noise

5.3 Measurement Resolution

The measurement of many attributes of haptic devices can be subtle. As a rule, the equipment used to measure an attribute should be about ten times higher in resolution than the resolution of the measurement that a researcher or user expects to make. At the same time, measurements of high resolution are not always required to determine if a device is suitable. It may suffice to measure the workspace to the nearest centimeter, rather than to sub-millimeter precision.

5.4 Context of Measurement

In order to allow for the widest possible scope of device evaluation, the draft standard presents verification techniques in two sections – one for the examiner who does not have measurement equipment at his or her disposal, and another for the laboratory equipped to make a variety of measurements.

6 Verification Example

6.1 Background

To illustrate the principle of verification, consider the example of a motorized haptic system. As a rule of thumb, when the maximum stiffness of the system is exceeded, unintended vibration will occur. Rather than making a passive presentation of touch, the system introduces energy into the interaction that reveals it to be a poor simulation of what it is trying to represent.

A convenient test for maximum stiffness is to present the device with a virtual wall with adjustable spring constant and no damping. The maximum stiffness is the highest spring constant of the wall that can be explored by the device without generating vibration.

The handle must be gripped by the user in the same manner that it would be gripped during normal operation. If held tightly, the measured stiffness will be slightly higher than its actual value in normal operation; if held loosely, the measured stiffness will be slightly lower than it would be in normal operation.

The following implicitly assumes a kinesthetic device.

6.2 Equipment

Consider a haptic simulation consisting of a virtual wall with a variable amount of springiness - the surface of the wall is modeled by a spring governed by Hooke's law. The haptic device is connected to a small virtual sphere, so that the sphere moves synchronously with the movement of the haptic device. The wall has a return force F proportional to the depth of penetration d by the virtual probe into the wall, where k is the constant of proportionality (also known as the spring constant).

$$F = -k \cdot d \tag{1}$$

The computer software should be set up so as to

1. give the user control over a point virtual probe by means of the haptic device under test;
2. locate a virtual wall, typically the local x, y or z Cartesian reference plane;
3. allow the user to place the virtual probe onto the surface of the wall, but to pull it back at will;
4. allow the user to increase or decrease the spring constant of the wall.

6.3 Test Procedure

The maximum spring stiffness that does not cause vibration is found by a bracketing technique

1. Set the wall spring stiffness to any value near the expected maximum stiffness.
2. If vibration occurs, reduce the spring constant.
3. If vibration does not occur, increase the spring constant
4. Continue increasing and decreasing the spring constant, bracketing ever closer the maximum stiffness where vibration does not occur.
5. When the smallest iteration crosses the line between vibration and no vibration, the maximum stiffness where vibration does not occur has been located.

A convenient way to vary the stiffness is to use a slide bar. When the mouse is used to select the slide bar, the roller on top of the mouse will move the slide bar, thus changing the stiffness. The user can vary the stiffness with one hand, sight unseen, while focusing attention on the device while it is controlled with the other hand.

7 Evaluating Usability

7.1 General

Data should be quantified so that a decision can be made as to the required qualities of the interaction. At the top level, the three components of usability are tested – effectiveness, efficiency and user satisfaction.

The purpose of validation in this context is to determine if a tactile/haptic interaction is usable for the purpose for which it was designed.

In order to validate the interaction, the examiner should ensure that goals have been specified to define the intention of the interaction. In many cases, the goals are set most concretely by positing a specific situation in which the interaction is to be used. For example, a hand-held touch screen has a number pad that gives a specific frequency of vibration when each number is pressed. A specific situation may be to dial a ten-digit phone number using the touch pad.

7.2 Test

A repeatable test procedure is then constructed from the situation. The procedure should include some means of measuring at least three components of usability set by ISO 9241-11 [3]:

Effectiveness measures can include:

- the success of each attempt to reach the goal
- percentage of goals achieved
- percentage of users successfully completing the task
- average accuracy of completed tasks

Efficiency measures can include:

- time to complete a task
- time taken on first attempt
- time spent on correcting errors

Satisfaction measures can include:

- frequency of complaints
- rate of voluntary use
- user rated ease of use

During a typical evaluation of usability, a representative set of users will place themselves in the test situation. They will follow the test procedure in an attempt to reach the goal by means of the tactile/haptic interaction.

7.3 Data Collection

The analyst collects the data on the three components of usability, plus optional auxiliary data on specific aspects of the interaction. The data is analyzed using common statistical procedures and the results are reported. A useful means of reporting is the common industry format for usability test reports, ISO 25062 [8].

For instance, in a dialing test with the hand-held touch screen, the success of dialing phone numbers and the time taken is measured by a monitoring computer. The experience of the user is assessed by a questionnaire, in which the user rates his impressions on a number of bipolar scales. In practice, higher quality data can be collected by comparing two or more means of interaction. Typically, a small number of variables will be altered between trials, so that the effect of the variables can be compared and a conclusion drawn as to which one is the best (or better) way of achieving the goal.

Additionally, trials may be run with and without vibrations in the touch pad of the previous examples. A variation may be tested with just a haptic 'click' under the character “5" in the centre of the number cluster, to see if the user is better able to centre the fingers while dialing the phone number.

7.4 Effectiveness

Some effectiveness measures include:

- reading speed
- speed of identifying an icon
- targeting speed
- moving speed
- reaction time

Tests will arrange for a score of effectiveness. A score of success p may be derived directly from the achieved result. If the goal of each run is yes/no success, then the number of successes n in m tries may be assessed. The score will then be $p = n/m$.

A number of users should test the tactile/haptic interaction, thereby reducing the possibility of individual bias. The exact number depends on the desired degree of certainty required for the test. Users should be selected as randomly as possible from the typical user group.

7.5 Efficiency

In a typical tactile/haptic interaction, the time taken to achieve the goal is measurable. Efficiency may then be conveniently calculated from the ratio of the individual scores used to determine effectiveness divided by the time taken to achieve each score.

The basic efficiency would be the mean of these scores. More useful is the decision as to which of two test runs is more efficient, and by how much.

7.6 Satisfaction

User satisfaction with an interaction may be assessed by any one of twelve methods presented in ISO 16982 [4], Usability methods for human-centered design. Methods vary from observation of users and questionnaires to expert opinion without direct user involvement. As set out in that standard, the choice depends on many factors – whether one is acquiring, designing or operating a system; whether the task to be performed is simple or complex, and whether the task is well known to the general population or relatively obscure.

In the present considerations of evaluation, we shall assume the use of a questionnaire. The questions thus presented can serve for a user-filled questionnaire, but also as the basis of an interview or a check sheet for the opinion of a subject expert.

We shall also assume the use of ordinal scales to rate user satisfaction, with the neutral opinion in the middle. The possible answers to a question such as "This interaction was easy to use" may range, for example, from "Strongly agree" to "strongly disagree".

8 Example of Evaluating Usability

8.1 Background

We consider the evaluation of the usability of a simulation of a surgical procedure. Reality simulation such as this may involve several interaction modalities – 6-DOF haptic feedback, stereo vision and high quality audio. The interaction may be intended for exploration, training or entertainment. In the case of surgery, the interaction would be one of precision, so fidelity and convincing immersion are important features.

We suppose that the user has experience with the actual scenario, and that he would judge the simulation against his experience with that scenario. As a case example, we may further suppose that the scenario is a surgical procedure involving the removal of a tumor from a brain. A skull section has been removed, and the surgeon is using a cauterizer, a bipolar coagulator and a suction tool.

The goal of the simulation is a realistic rendering of the visual scene and the feel of a surgical instrument as the surgeon wields it during a surgical procedure.

8.2 Test

The test will consist of parallel scans across the surface of the meninges (the membrane that covers the brain). The tool is a pen-like cauterizer which we shall refer to as a stylus for convenience and generality. The surgeon holds it as one would hold a pencil.

Fifty traces are made, each of approximately the same length. The stylus rides the surface of the meninges as the tool is moved by the surgeon. A very light force is maintained on the membrane. Success for each scan is measured by the ability of the surgeon to maintain a constant and correct force on the stylus while executing a scan of constant curvature. The centre of the position curve will typically be at the elbow of the surgeon, as he executes adduct-abduct motion of the elbow while keeping the wrist locked.

The raw data measured would be the downward force and the position of the tip of the stylus. The derived data could be any of a number of possibilities:

- the standard deviation of the force
- the distance along the arc during which the force exceeds a maximum or a minimum value
- the difference between the average value of the force and the target value
- the standard deviation of the radial position of the trace
- the distance along the arc during which the radial position deviated from the ideal radius by more than a certain value
- the distance along the arc during which the radial position deviated from the ideal parallel distance from the preceding trace

Considering the goal of the test (to make parallel lines at a certain pressure), we shall choose success to be a complete trace made within force limits and within radial distance limits of the preceding trace.

8.3 Analysis

Effectiveness will be measured by the percentage of the fifty traces that are successful. Efficiency will be measured by the success rate divided by the time taken to complete the fifty traces.

Satisfaction will be measured by the bipolar response to the statement, “Is this a realistic simulation of the actual operating scenario?"

The questionnaire could be expanded to include evaluating statements that solicit more fine-grained responses (*strongly agree, agree, neutral, disagree,* or *strongly disagree*).

“Does the stylus feel correct in my hand?"

“Does the stylus have the correct inertia?"

“Were there erroneous forces that I had to resist as I drew the traces?"

“Has the visual rendition been like an actual surgical site?"

“Has the feel been exactly matched to the visual rendition of the operating scene?"

8.4 Ramifications

The evaluation could be used

- to assess if a surgical simulator is satisfactory for training surgeons.
- to compare different brands of surgical simulator.

- to compare different haptic devices for use in a surgical simulator.
- to compare different orientations of the same haptic device.
- to compare different haptic and visual models of the brain surface.

9 Getting Involved

TC159/SC4/WG9 is continuously working to ensure that all guidelines are technically correct and feasible. You can get involved as an expert member of TC159/SC4/WG9, actively developing drafts of the planned work items. Even as a casual advisor, the members of WG9 are very interested in your opinions on tactile/ haptic-related terms and definitions, and hearing about your experience with measures for haptic devices or human performance.

References

1. Boehm, B.W.: Software Engineering Economics. Prentice-Hall, Inc., Upper Saddle River (1981)
2. Carter, J., van Erp, J.: Ergonomics of Tactile and Haptic Interactions. In: Proceedings of the Human Factors and Ergonomics Society 2006, pp. 674–688 (2006)
3. ISO: ISO 9241-11, Ergonomic requirements for office work with visual display terminals (VDTs) Guidance on usability. International Organization for Standardization (1998)
4. ISO: ISO/TR 16982: 2002 Ergonomics of human-system interaction – Usability methods supporting human-centered design. International Organization for Standardization (2005)
5. ISO: ISO/IEC 15288, Systems and software engineering – System life cycle processes. International Organization for Standardization (2008)
6. ISO: 9241-920 Ergonomics of human-system interaction – Part 920 Guidance on tactile and haptic interactions. International Organization for Standardization (2009)
7. ISO: 9241-910 Ergonomics of human-system interaction – Part 910 Framework on tactile and haptic interaction. International Organization for Standardization (2011)
8. ISO: ISO/IEC 25062: 2006 Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Common Industry Format (CIF) for usability test reports. International Organization for Standardization (2011)
9. van Erp, J., Carter, J., Andrew, I.: ISO's work on tactile and haptic interaction guidelines. In: Proceedings of Eurohaptics 2006, pp. 467–470 (2006)
10. van Erp, J.B.F., Kern, T.A.: ISO's Work on Guidance for Haptic and Tactile Interactions. In: Ferre, M. (ed.) EuroHaptics 2008. LNCS, vol. 5024, pp. 936–940. Springer, Heidelberg (2008)
11. van Erp, J.B.F., Kyung, K.-U., Kassner, S., Carter, J., Brewster, S., Weber, G., Andrew, I.: Setting the Standards for Haptic and Tactile Interactions: ISO's Work. In: Kappers, A.M.L., van Erp, J.B.F., Bergmann Tiest, W.M., van der Helm, F.C.T. (eds.) EuroHaptics 2010, Part II. LNCS, vol. 6192, pp. 353–358. Springer, Heidelberg (2010)