Adaptation of Ergonomic Criteria to Human-Virtual Environments Interactions

C. Bach & D. L. Scapin

INRIA, Domaine de Voluceau, 78153 Le Chesnay Cedex, France.

Cedric.Bach@inria.fr, Dominique.Scapin@inria.fr

Abstract: This paper introduces a new version of Ergonomic Criteria for HVEIs which have been tested for validity. An assignment experiment (Ergonomic Criteria *X* Usability Problems) has been conducted to ensure the appropriate understanding of the different criteria. The results of this study identify the criteria that are well understood and the ones that need further improvement.

Keywords: Ergonomic Criteria, Human-Virtual Environments Interactions, Ergonomic inspection methods, Experiment.

1 Introduction

Virtual Environments (VE) are developing fast and widely. Various contexts are now concerned, e.g., training, data visualisation, computer-aided design, tourism, art, games, etc. The cost of such systems, including interaction devices, is decreasing together with the improvement and availability of technology. However, just as it was the case for GUIs and for the Web, a large utilisation of such environments will depend on its usability (Williams and Harrison (2001).

One of the major problems is that, due to the relative youth of VEs, there is yet a rather small set of usability studies being published compared to other more classical computer-based environments. The lack of empirical studies relates also to several difficulties with testing VEs (Bowman et al., 2002): complexity of the physical environment; new role of the evaluator; few skilled users; few evaluation methods; variability and evolution of environments and devices. Nevertheless, the advice, principles, and recommendations are growing fast. Actually, there are already some efforts in gathering and presenting current ergonomics results (Gabbard and Hix, 1997; Kalawsky, 1999; Kaur, 1998; Stanney et al., 2000). However, there is a strong need to provide new user-centred methods for the design and evaluation of VEs (Durlach and Mavor, 1995).

This paper mentions briefly our approach for designing Ergonomic Criteria (**E.C.**) adapted to Human-Virtual Environments Interactions (**HVEIs**) and describes an experiment designed for testing the validity of those E.C.

2 Purpose and rationale

Among the numerous usability methods, this paper focuses on inspection methods based on empirically defined sets of dimensions. More specifically, the research consisted in assessing the potential adaptation of existing E.C. (Bastien and Scapin, 1993) tested thoroughly for GUIs (Scapin and Bastien, 1997) and the Web (Scapin et al., 2000). The initial step has been to gather all available ergonomics knowledge in the form of principles, advice, recommendations, experimental findings, then transformed into individual recommendations under a common format. Those 170 recommendations (see Bach et Scapin, 2003) were then assigned to the existing 18 elementary E.C. This worked for most recommendations. However one criteria was modified "*Significance of codes and behaviour*" and two new criteria were added "*Grouping-Distinguishing items by behaviour*" and "*Physical workload*", in order to take into account recommendations specific to HVEIs. Basically the specific recommendations relate mainly first to the behaviour of environments, agents, and avatars; and secondly the physical implications of certain types of HVEIs, that may, for instance, lead to motion sickness. The "new" 20 E.C. (see Table. 1, in which elementary criteria are in boldface) obtained were also ordered according to an inspection strategy (unlike the previous ones which were ordered

according to the number of recommendations underlying each criteria). The experiment used the same approach as in Scapin and Bastien (op. cit.) and consisted in the matching of concepts (the criteria) with their potential instances (usability problems). This step is a necessary step for assessing the E.C. as a classification tool or as a way to standardize design requirements or evaluation reports; further steps will test their value as a basis for ergonomic inspection.

1. **Compatibility**
2. Guidance
   2.1. **Legibility**
   2.2. **Prompting**
   2.3. Grouping / distinguishing items
      2.3.1. **Grouping/ distinguishing by Location**
      2.3.2. **Grouping/ distinguishing by Format**
      2.3.3. **Grouping/ distinguish. by Behavior**
   2.4. **Immediate Feed-Back**
3. Explicit Control
   3.1. **Explicit User Actions**
   3.2. **User Control**
4. **Significance of codes and behavior**
5. Workload
   5.1. **Physical Workload**
   5.2. Brevity
      5.2.1. **Minimal Actions**
      5.2.2. **Conciseness**
   5.3. **Information Density**
6. Adaptability
   6.1. **Users' Experience**
   6.2. **Flexibility**
7. **Consistency**
8. Error Management
   8.1 **Error Protection**
   8.2 **Quality of error messages**
   8.3 **Error Correction**

Table 1 : List of Ergonomic Criteria for HVEIs

3 Experiment

3.1 Method and Procedure

Ten subjects (6 male, 4 female) participated in the study. All have university degrees (4 to 7 years) and averaged 5.65 years of experience in software ergonomics (S.D. = 5.4). None of them had experience with HVEIs.

The experiment consisted of one individual session for each subject, through 3 phases:

In Phase 1, the subjects are invited to read a "E.C. for HVEIs" document. That document presents the list of 20 E.C., and then provides, for each criterion a definition, a justification, and examples of recommendations, all extracted from the Bach & Scapin (op. cit.) literature survey. That document remained available to the subjects throughout the whole experiment.

In Phase 2, the subjects are invited to assign to the E.C. a set of randomly presented usability design flaws. There are 40 examples of usability design flaws, each criterion being illustrated theoretically by 2 usability flaws. These examples have been either extracted from the literature or selected from existing VEs, or simulated. Each example is illustrated with an explanatory text (of course, with no reference to the criteria definition or examples wording) and in 10 cases, illustrated by a video. Each subject is asked for each example to choose the elementary criterion that best labels the design flaw.

In Phase 3, the subjects are presented with the same initial, individually assigned, randomly order of usability design flaws and is asked to confirm or not his/her initial elementary criteria assignment. This allowed the subjects to modify their judgement once they have looked at all the examples of design flaws. There is no time constraint on all phases of the experiment.

3.2 Data analysis and Results

Two types of data are considered in this first analysis: the time taken for each one of the sessions; and the criteria selection, by subject, for each one of the 40 usability design flaws.

Concerning performance time, the time dedicated to the reading of the "E.C. for HVEIs" document averages *22.42 mn (S.T. = 5)*. The time dedicated to the classification task averages *48.48 mn (S.T. = 9.2)*. The confirmation task time averages *23.12 mn (S.T. = 6.1)*. The overall task time for the whole assignment task (classification + confirmation) averages *72 mn (S.T. = 14.1)*. All these session times are fairly similar, no subject being outside *Mean ± 2 S.T.*

Concerning criteria selection, i.e., the assignment of the 20 pairs of usability design flaws to the 20 elementary criteria, global subjects performance is 68% of correct assignment (correct being coherent with the theoretical assignments) for the main criteria, and 59.5% for the elementary criteria. This global result is fairly encouraging, considering the short time dedicated to the discovery of the criteria definitions and the absence of experience of the subjects in HVEIs. In addition, detailed analysis of the confusion matrix (E.C. *X* Usability flaws) shows that some classification errors (i.e., not coherent

with the theoretical assignment) are related to the characteristics of the examples chosen, not always well understood by the subjects.

To sum up, the subjects assigned correctly 27 usability problems to the corresponding main criteria, all within a *Mean ± 2 S.T.* range; while they assign appropriately 24 usability problems to the corresponding elementary criteria, within a *Mean ± 2 S.T.* range.

In order to better understand the detailed assignments in the confusion matrix, a Cohen's Kappa (K) agreement statistic (see Bakerman and Gottman, 1986) was computed, which allows for identification of systematic confusions between elementary criteria. The global result, $K = .57$ is coherent with the mean frequencies obtained.

Analyzing further the results for the purpose of improving the E.C. descriptions lead to the classification of E.C. from their average frequency of identification together with their proportion of correct identifications (see Table 2).

	Mean Frequency of identification	
Proportion of Correct Identifications	≤ 2	> 2
0.66 ≤ p	Physical Workload	User Control
	Error Correction	
	Users' Experience	
	Explicit User Actions	
0.33 < p < 0.66	Conciseness	Flexibility
	Quality of error messages	Information Density
	Consistency	Minimal Actions
	Gr./Dist by Behavior	Legibility
	Error Protection	Immediate Feed-back
	Gr./Dist by Location	Compatibility
	Gr./Dist by Format	Prompting
0 < p < 0.33		Significance of codes and behavior

Table 2 :Classification of C.E. according to their frequency of identification and their proportion of correct identification.

Three categories of criteria have been distinguished: well defined criteria ($p \geq 0.66 \cap f \leq 2$); average defined criteria ($0.33 \leq p < 0.66 \cap f \leq 2$); and ill-defined criteria ($0 \leq p < 0.33 \cup f > 2$). This led to a decision table (Table 3) distinguishing the criteria that needed or not further improvement in their definition, justification or addition of counter-examples.

No modification	To be modified
Physical Workload	User Control
Error Correction	Flexibility
Users' Experience	Information Density
Explicit User Actions	Minimal Actions
Conciseness	Legibility
Quality of error messages	Immediate Feed-back
Consistency	Compatibility
Gr./Dist by Behavior	Prompting
Error Protection	Significance of codes and behavior
Gr./Dist by Location	
Gr./Dist by Format	

Table 3 : List of C.E. to be modified or not

Further analyses of the confusion matrix, for instance to distinguish variations in assignment within pairs of examples are under way. These analyses should lead to a precise strategy in assessing the detailed needs for modification.

4 Conclusion

The results from this study fall in the same range as previous studies (Bastien and Scapin, op. Cit.).

First of all, the E.C., by design, are built from existing knowledge (HVEIs recommendations available in the literature) which should, as of now, warranty their completeness.

Secondly, their assignment to usability problems is above average for main (68%) as well as for elementary criteria (59.5%). This is fairly satisfactory considering the short familiarization time with the "new" E.C. and the lack of subjects' experience with HVEIs.

Such results should improve as modifications are made on the definitions, justifications and more illustrative examples and counter-examples are added.

However, this study, which analysis remains to be completed, is only a first step with the goal of providing usability inspection methods for HVEIs.

First, it would be useful to define a coherent object model (i.e., a usable classification of HVEIs elements) to support an inspection strategy. This should also limit the difficulties encountered by the subjects to analyse thoroughly the complexity of virtual environments.

Secondly, further studies are needed to assess the "improved" E.C. as a basis for inspection.
This is currently under way. Three experiments are planned:
- one that will test the efficiency (i.e., the number and quality of usability problems diagnosed) of E.C. as a means to inspect HVEIs;
- another one that will study the intuitive evaluation of HVEIs, based solely on the experts' experience;
- a third experiment that will consist of actual users testing HVEIs. The rationale is to compare three ways the results in terms of role of the E.C. versus simple expertise and their coverage in terms of problem diagnosis versus empirical testing.

Finally, one should add the need for iteratively completing our recommendations data base, as empirical studies of HVEIs are progressing,

References

Bach, C. et Scapin, D.L. (2003) *Recommandations ergonomiques pour l'inspection d'environnements virtuels.* (Rapport de contrat). Projet EUREKA-COMEDIA, INRIA Rocquencourt, France, 2003.

Bakerman, R. and Gottman, J.M. (1986), *Observing interaction: an introduction to sequential analysis.* Cambridge: Cambridge University Press, 1986

Bastien, J. M. C. & Scapin, (1993), D. L. *Ergonomic criteria for the evaluation of human-computer interfaces.* (Technical report N° 156)., INRIA Rocquencourt, France, 1993.

Bowman, D.A., Gabbard, J.L., and Hix, D. (2002) *A Survey of Usability Evaluation in Virtual Environments: Classification and Comparaison of Methods.* Presence: Teleoperators and Virtual Environments. Vol. 11, n°4, 2002, pp. 404-424.

Durlach, B.N.I., and Mavor, A.S. (1995) *Virtual reality: Scientific and technological challenges.* Washington, DC: National Academy Press, 1995.

Gabbard, J.L., & Hix, D. 1. (1997), *A taxonomy of usability characteristics in virtual environments.* 1997. Avaible : http://csgrad.cs.vt.edu/~jgabbard/ve/taxonomy/

Kalawsky, R.S. (1999), *VRUSE–a computerized diagnostic tool: For usability evaluation of virtual/synthetic environment systems.* Applied Ergonomics, n°30,1999, pp. 11-25.

Kaur, K. (1998), *Designing virtual environments for usability.* Ph. D. Thesis, City University, London, 1998.

Scapin, D. L., and Bastien, J. M. C. (1997), Ergonomic criteria for evaluating the ergonomic quality of interactive systems. *Behaviour & Information Technology*, *16*, 220-231.

Scapin, D.L., Leulier, C., Bastien, C., Vanderdonckt, J., Mariage, C., Farenc, C., Palanque, P., and Bastide, R. (2000) *A Framework for 0rganizing Web Usability Guidelines.* 6th. Conference on Human Factors and the Web. June 19, 2000, Austin, Texas.

Stanney, K.M., Mollaghasemi, M., and Reeves, L. (2000), *Development of MAUVE, the multi-criteria assessment of usability for virtual environment system.* (Final Rep., Contract N°. N61339-99-C-0098). Orlando, FL: Naval Air Warfare Center Training Systems Division. 2000.

Williams, J.S. and Harrison M.D. (2001), *A toolset supported approach for designing and testing virtual environment interaction techniques.* International Journal of Human-Computer Studies, n°55, 2001, pp 145-165.