Obstacles and Perspectives for Evaluating Mixed Reality Systems Usability.

Cédric Bach

INRIA

B.P. 105 Domaine de Voluceau

78153, Le Chesnay, France

+33 1 39 63 51 09

cedric.bach@inria.fr

Dominique L. Scapin

INRIA

B.P. 105 Domaine de Voluceau

78153, Le Chesnay, France

+33 1 39 63 55 07

dominique.scapin@inria.fr

ABSTRACT

The goal of this paper is to survey the main issues with the ergonomic evaluation of MRS (Mixed Reality Systems) and to stimulate discussions for future research. A first point concerns definitions and specificities of MRS within the « reality / virtuality » continuum, and the incorporation of user issues. Another point concerns the combinatory character of the ergonomic knowledge to be applied to MRS entities (reality, « GUIs », « VR », and « MR » specific). A major issue concerns ergonomic evaluation methods, their current state, their advantages and drawbacks, particularly for user testing. Finally, the discussion points at various items which may be part of a future research agenda, such as the need for more usability data, for generic and well controlled experiments; for common testing platforms, for shared recommendations data bases, for design and assessment of inspection methods, for common task taxonomies and for common models of MR entities and situations, etc.; all of this possibly leading to increased knowledge based on shared benchmarks.

Keywords

Mixed reality, virtual environments, ergonomics, usability, evaluation methods, user testing, inspection methods, ergonomic recommendations, research agenda, user-centered approaches, interaction models, task models.

INTRODUCTION

Virtual Environments (VE) are being developed fast and widely, in various contexts (e.g., training, data visualisation, computer-aided design, tourism, art, games, etc.).

Mixed Reality Systems (MRS) follow the same path. For both types of environments, just as it was the case for GUIs and for the Web, a large utilisation will depend on its usability.

Facing the question of establishing a research agenda for contributing to more usable MRS, this paper attempts to draw a limited picture of the current state of knowledge on

mixed reality systems, from a user-centred perspective, and to discuss potential research avenues.

First, one needs to discuss definitions : MRS are viewed as a subset of VE, with an ergonomic perspective.

Secondly, through a brief account of VE and MRS studies, we highlight a few items explaining the lack of current knowledge on the usability of such systems, mainly related to their novelty.

Then, the paper mentions briefly a number of evaluation methods that can be applied to MRS, and focuses on the specific methodological challenges with user testing.

Finally, the discussion offers several items of interest for a common research agenda in the area of MRS ergonomics.

DEFINITIONS and SCOPE

For MRS, there is no fully agreed-upon definition, so far, but a set of features, e.g., as the ones mentioned in the Call For Papers for this (IUI / CADUI-associated) Workshop: *" ... integration of the physical and digital worlds in a smooth and usable way. This fusion involves the design and development of "mixed reality systems", including augmented reality, augmented virtuality, augmented video, and tangible systems ..."*. We definitely agree that: *"The diversity of terms used to denote these systems is evidence both of the amount of research activity in the field and the lack of a common conceptual framework for that activity"*. Obviously, one of the goals of this Workshop should be to progress on definitions and on a conceptual framework.

For VE, there are numerous definitions. Most often, they are techno-centred. A definition that we use for VE is the one for VR (Virtual Reality) from Loeffler & Anderson [14] : *"Virtual Reality is a three-dimensional, computer-generated, simulated environment that is rendered in real time according to the behaviour of the user."*

However, this definition should to be completed on two points:

- VE can be multi-users (for the evaluation of collaborative VE, see for instance [25])

- VE can be described along a « Real-Mixed-Virtual » continuum, as described by Milgram [16].

That continuum includes in fact a large spectrum of 3D computer-based interactive situations. On that spectrum, a number of terms used in the literature can be located, such

as Virtual Reality, Mixed Reality, Augmented Reality, Desktop or Immersive Environments, Augmented Virtuality, etc… it is even possible to incorporate within that continuum 3D CAD or VRML objects displayed on « classical » computers, which makes them belong to VR, even though very loosely.

The first point here is a matter of joining the techno-centred and the user-centred perspectives. The second point deals with the combinatory nature of MRS ergonomic knowledge (real, « GUI », « VR », and « MR » specific).

People might think that there is an enormous difference between a 3D object to be interacted with in a regular GUI environment, and that same object presented in a CAVE, using a 3D mouse. That difference is certainly there, but only from a technical point of view, not necessarily from a user's point of view.

With an ergonomic perspective, these two situations must be considered through their capacity to support users in achieving specified task goals, with efficiency, effectiveness and satisfaction. In some cases, an immersive situation can be beneficial (e.g., learning driving / control operations on a high-speed train, which involves physical simulation of the access to the driving instruments), while it may be less appropriate in others (e.g., learning specific procedures for that same train, which involve more cognitive sequence learning steps) where regular GUIs, or even paper instructions may be more appropriate.

In other words, depending on the task at hand, the usefulness of the support required may be different : in kinetics learning, immersion may make sense ; while in cognitive learning, it may not be as helpful.

The same point can be made for MRS which are constituted of both real and virtual elements within the same 3D interactive environment. As they can be located roughly at the midpoint of the Milgram continuum, they can be considered as a particular form of VE. However, their specificity is to stage a form of fusion between real and virtual worlds.

Even if MRS can be divided into Augmented Reality and Augmented Virtuality [17] depending on the exact physical or computer-based nature of the tasks and interactions involved, they relate to combinations of knowledge from various domains of specialization of ergonomics :

- Real objects, in the area of physical ergonomics, concerned with human anatomical, anthropometric, physiological and biomechanical characteristics (postures, materials handling, repetitive movements, musculoskeletal disorders, workplace layout, safety and health); in the area of ergonomics of every-day products and consumer products; in some cases, in the area of cognitive ergonomics and organizational ergonomics.

- Classical interfaces (GUIs) mainly in the area of cognitive ergonomics, concerned with mental processes, such as perception, memory, reasoning, (e.g., mental workload, decision-making, skilled performance, information presentation, commands and controls, etc.).

- Virtual environments are concerned as well by cognitive ergonomics, but also by physical ergonomics (e.g., in terms of VE behaviour, presence, cyber sickness, etc).

- Fusion of the previous elements, which corresponds to one specificity of MRS: the appropriate correspondence between the various real or virtual elements constituting MRS.

  This problem, also called « continuity », characterizes the perceptual and cognitive fluidity, from the users' point of view, between the real world and the virtual world. According to Nigay et al. [17], "*The perceptual continuity is verified if the user directly and smoothly perceives the different representations of a given concept. Cognitive continuity is verified if the cognitive processes that are involved in the interpretation of the different perceived representations lead to a unique interpretation of the concept resulting from the combination of the interpreted perceived representations.*".

  Such a concept is a challenge for MRS design that requires optimal matching between various features in terms of presentation, coding, meaning, and behavior.

  From an ergonomic point of view, one could state that it is not only an issue with interface consistency (e.g., providing information presentation as well as user initiated interactions with the same characteristics across operations and across applications), but also a complex compatibility issue, i.e., matching the various MRS individual elements with their referent, but also matching all related MRS elements which each other, depending on the tasks and context.

ERGONOMIC KNOWLEDGE ON MRS

Previous history of HCI concerning ergonomics knowledge is probably repeating itself : MRS novelty and lack of ergonomic knowledge.

With new technologies, usability concerns start with general questions, hypotheses, debates ; then questions are sorted and put to the test through experiments ; theories are offered ; data on usability grows rapidly as the technology becomes widely available ; data is then organized, made available through manuals, guides ; finally, style guides, architectures, design and evaluation methods are offered, tested, compared.

This has been proven true for GUIS, the Web, ... it should be the same for VE and MRS. However, there may be ways to accelerate the process by doing user testing early so that usability knowledge is gained rapidly, rather than having simply the technology perfected without user concern.

Also, as it has been verified that, when moving from GUIs to the Web, one should not « re-invent the wheel », but apply as much as possible sound ergonomic knowledge

that can be transferred to novel environments (information organisation, consistency, level of feedback, etc.).

Implementation and usage of VE and MRS is indeed very recent as it started in the early nineties and only grows at a fast pace since three or four years.

That novelty explains partly why the currently available ergonomic knowledge is relatively scarce, compared to the number of issues that need to be tackled in such complex environments. Such environments are relatively unstable in their implementation and usage ; lots of technical problems still need to be solved.

However, some actions can be envisioned to better incorporate a user-centred approach :

- The recommendation above to not « re-invent the wheel », is particularly true for MRS as their components are mixed, therefore, ergonomics knowledge needs to be gained for the novel, specific MRS issues (such as « continuity »), but currently available ergonomics knowledge should be applied to its non-specific elements (e.g., real objects, 2D displays, etc.). The amount of knowledge (empirical results, recommendations, standards, etc.) in ergonomics about real objects, « classical » interfaces is of course enormous, and needs to be carefully investigated in terms of applicability depending on context and tasks. Also, concerning ergonomic evaluation, a number of already available methods should be looked at in terms of their applicability and needs for adaptation to the specifics of MRS.

- Another recommendation above was to incorporate ergonomic concerns early. That is usually done through user testing (with or without hypotheses, on one or several contexts, on one system, or through the comparison of several systems, etc.). That is where a number of methodological problems arise.
  In order to identify the obstacles and perspectives in methodological terms, the next two sections deal first with ergonomics evaluation methods and secondly with MRS-specific issues when conducting user testing.

ERGONOMIC EVALUATION METHODS

In the context of MRS, which usability evaluation methods are available, and how can such environments be appropriately evaluated?

In the literature, there is no usability method yet specifically designed for MRS, except notations such as AZUR++, a notation for describing, and reasoning about the design of mobile mixed systems [6].

There has been already some research on how to guide VE design with usability considerations (e.g., in [26]), to consolidate usability dimensions and to design inspection methods, but much work is needed to extend their scope to MRS, to assess them and to compare them.

- Methods based on heuristics or recommendations are difficult to carry out due to the limited amount of recommendations available, to the fact that « classical » ergonomic recommendations have not yet been applied or extended to VE. In addition, such recommendations are not presented in a common unitary format, but often as experimental or « best practice » results ; see for instance Gabbard [7] [8]; Kaur [12]; Stanney[21] ; generally such documents are organized according to major categories of VE objects. Also, based on an extensive literature review, a set of 170 recommendations dedicated to VE have been extracted under a generic format and organized according to usability dimensions [1].

- Along those lines, with the goal of defining a structured ergonomic inspection method, an adaptation of the Ergonomic Criteria (E. C.) has been proposed and assessed in terms of intrinsic validity [2]. First of all the application of the E. C. inspection method requires expertise in VE and training on the 20 dimensions covered by these E. C.. Also, the method needs further evaluation for VE and extensions to MRS.

A method dedicated to VR, based on the *cognitive walkthrough* method, has also been designed by Sutcliffe [22]. This method is based on a theory of interaction (Norman [18]). This walkthrough analysis method uses three models (first on goal-oriented task actions, second on exploration and navigation in virtual worlds, third on interaction in response to system initiative) derived from the theory. Each stage of the model is associated with generic design properties. The evaluation method consists of a checklist of questions using the properties and following the steps of the method. That method could possibly be extended to MRS.

There are also methods, mainly for VE, based on recommendations, using computer-based support, such as :

- I-DOVE (Interactive tool for development of Virtual Environments) [11]. The goal of this prototype, currently being developed as a large-scale web-based application based on several sets of recommendations, is to offer context specific guidance for VE development, and alternative ways of searching and browsing, distinguishing user categories. The initial prototype was based on users interviews and was later evaluated by expert evaluation.

- Another tool is MAUVE [21]; also developed as a website, it incorporates design guidelines according to several VE categories such as navigation, object manipulation, input, output and so on (based on Gabbard's [8] taxonomy). A multi-criteria usability matrix is the support for organizing and retrieving recommendations. The evaluation process is supported in two steps: "traditional" heuristics stage, and prioritization of usability attributes. That capacity of tailoring the evaluation may be interesting for evaluators with

different points of view or organizational goals. The tool has not been evaluated yet.

- The last tool, which we know of, is a hypertext-based prototype developed by Kaur [13]. This tool present 45 generic design properties that specify the necessary support from the system for "successful" VE interfaces. Like the previous one, no usability evaluation plan was integrated during the development of the prototype.

There is also the question of adapting, at least partly, to MRS some of the « classical » ergonomics methods. Such methods are too numerous to be all discussed here (see for instance, [3], [4], [9]). However, one can mention three categories of methods that are general enough in their approach to be good candidates for MRS : questionnaires/ interviews, inspection methods and user testing. A number of questions must be solved in order to apply them to MRS.

- Questionnaire and Interviews allow gathering subjective data, often quite important to evaluate visual appeal, preferences, aesthetics, missing functionalities, and also very useful as a means to compare or cross-reference performance data. Such methods are certainly interesting candidates for being applied to MRS, providing specific lead questions for interviews and questionnaire items are tailored and validated for such environments. For questionnaires, Kalawsky [10] has designed VRUSE for measuring usability of a VR application in terms of users' attitude and perception. The 100 questionnaire items are organized under 10 usability factors: functionality, input, output, user guidance, consistency, flexibility, simulation fidelity, error correction, presence and overall system usability. This questionnaire has been tested in terms of reliability (Cronbach's alpha > 0.9).

- Inspection methods are also good candidates for supporting MRS evaluations. The problem there is the need for more data (particularly recommen-dations) and more data organization in order to cover the range of ergonomic problems related to MRS. Issues regarding recommendations identifi-cation and structuring into dimensions need to be looked at carefully; the history of HCI has shown so far that such dimensions can be established and efficiently contribute to evaluation of GUIs, the web, currently with VR ... but specificities of MR need to be taken into consideration (e.g., task compatibility, devices consistency for visualiza-tion, documents compatibility, innovative help systems, etc.).

- User Testing has been the major method in ergonomics and will probably remain as important for MRS. However, in order to apply that method to MRS, a number of methodological problems need to be tackled, including of course, the problems already experienced with a few cases of user testing of VE systems (see next section).

METHODOLOGICAL DIFFICULTIES WITH USER TESTING

User testing is certainly the preferred method to be used, particularly in order to alleviate the current lack of available usability data. However, many methodological problems require solutions.

First of all, let us use a metaphor both historical and aero-nautical. Looking back over a century ago, let us imagine that the following question was asked at that time: « which flying machine constitutes the best way to move rapidly in the air from one point to another : the blimp or the air-planes ? ». In those times, the airplanes trials were just starting ; it would have been difficult to find users (i.e., pilots) able to fly (usually the few pilots were the airplanes designers themselves) ; the underlying technology was only emerging, and often planes had numerous technical prob-lems or simply just crashed. If at that time one would have conducted comparative performance testing, there is no doubt that the blimp would have won over the airplanes ! However, by now, everyone would agree that airplanes are better that blimps for long distance transport of passengers.

Making the parallel with HCI technology, « classical HCI » would be our blimps and MRS would be our planes. In several ways, MRS are at the same point as planes in early 20[th.] century :

- There are very few experts that can operate them.

- The characteristics of tasks that can be performed in such environments are still quite vague.

- Learning by trying is still the rule.

- There are many problems to solve in order to « fly » those environments : on the technical aspects (e.g., computer graphics) ; on the interaction aspects (e.g., devices and modalities) ; and on their usability.

That state of affairs can explain why often results are disappointing [19] when classical graphics environments are compared to VE or MRS.

Before conducting user tests or to compare interactive situa-tions, it is best to :

- First make sure that the environments are already sufficiently well-designed so that well-known us-ability deficiencies are removed. This is a sound precaution helping to focus user testing on the real « new » usability problems and the test compari-sons on the real usability hypotheses, rather that obscuring the picture with unwanted usability problems, both for the user and for the experimen-tal data analyst. This can be achieved by careful assessment of the design, for instance through ap-plying available inspection methods and sets of recommendations.

- To alleviate as much as possible the various limitations of user testing due to the specificities

of VE and MRS. Some of those limitations are described in the next three paragraphs.

Limitations related to the physical environment
One of the major differences between MRS and traditional interfaces concerns their physical environment. MRS require a more sophisticated environment : the users rarely just sit in front of their computer ; they move from one place to another, they talk, they move parts of their body in order to interact.

This raises several problems in evaluation situations in which it is useful to prepare the experiments in order to avoid some disturbances, for instance, if the MRS application area in an office, it is necessary to limit the interaction zones in order to avoid collision with the chairs, tables, or cables ; other constraints can complicate the situations, for instance, interaction devices can be an obstacle to data collection : use of stereoscopy does not allow collecting data on video or on a monoscopic monitor, in which case, the evaluator must access directly to the stereoscopic data using a tracker or some parallel application software so that the evaluator does not become another user (even though a passive user) in the scene, which would certainly become a major bias in terms of « presence » [5].

MRS can also be multi-users and require a large experimentation space or they can be used outdoors, which makes difficult or even impossible the use of current usability laboratories, which are located in limited spaces to facilitate data extraction.

MRS using video projection can very easily augment the room temperature in the usability laboratories, which can become an important, often underestimated bias in the experiments.

Difficulties in the set up of user testing
First of all, the complexity of interactive situations with MRS may necessitate more resources that usual user testing : several evaluators may be needed to be able to extract the interesting data (e.g., checking on performance, on various modalities, on various media, etc.). Also, the use of video (several ones, simultaneously on various types of events) may be mandatory, as user behavioural sequences are more complex to extract and describe.

In addition to the complexity of software programming for such environments [23], setting up experiments may also be more complex as it requires more technical specialists to calibrate, tailor several types of technologies and various software supporting the complexity of MRS.

In case of devices or system breakdown, it may be more difficult and more time consuming to restart the devices or system, which may jeopardize the outcome and measures in the user testing. Therefore, it is even more important for MRS to be as stable as possible for the duration of the evaluation experiments.

Another obstacle which is specific to MRS (and to all novel applications) is that users may not know at all the way the situation works, which means extra time and caution in learning experimental requirements, task goals,

and ways to operate the various parts of the environment. This also leads to questions on how best to describe the tested situation without coaching too much the subjects if one wants to study their "intuitive" performance or preferences.

Along the same line, an additional difficulty is simply that it may be difficult to explain the complexities of MRS, considering that written instruction may not be sufficient beforehand, and impossible during the test when several entry or display devices are used together (the user cannot be using an eye-tracker, a data glove, watch a large display, and at the same time walk through a leaflet of task instructions). In addition, when needed, where to make some help system available ?

In some cases, as current technology is unable to support fully and consistently novel interaction paradigms, there is no way to test those new ideas unless using « Wizard of Oz » techniques which require trained specialists and specific, carefully balanced experimental design.

In other cases, the techniques used for interacting and those used for gathering subjects data obviously conflict ! For instance, the use of thinking aloud cannot work well when MRS use (as they may often do) voice recognition as an input mechanism !

Limitations related to the subjects
A first problem is that the application of MRS is not always directed by application needs, but by the design of new interaction paradigms, which makes difficult the specification of precise and accurate task and user requirements. It is therefore quite difficult to generalize because the users profile is often ill-defined ; sometimes, it is even the technical people that developed the MRS who are tested !

Another difficulty is that, at the current stage, it still impossible yet to practically distinguish, as it is usually done with "classical" HCI, the subjects in terms of experience with MRS (e.g., novices vs. experts). That holds true as well for the skills of human factors specialists in charge of the evaluations !

Experimental design may also encounter difficulties in the number of subjects needed to cover the many potential variables involved in the MRS. For instance, if one wants « simply » to compare various combinations of interactive devices associated to an MRS, such as 3 devices for each one of 3 user channels MRS (e.g., voice, gesture, eye-gaze), one needs 27 different testing situations (i.e., 9 combinations x 3 channels) in order to have all subjects participate to all possible combinations, which can become quite exhausting for the subjects and therefore potentially inconclusive for the experiment, due to fatigue, learning curve, etc.

And if one decides to set up the experimental design with non-repetitive measures, and associating different subjects' groups to each situation, for instance at a minimum of 5 subjects per cell, then the number of subjects required (135, i.e., 27 x 5) becomes fast very large and consequently quite complex to recruit and manage, not mentioning the

additional difficulty of making sure that the subjects' population is homogeneous.

Other limitations can be identified for the subjects such as the consequences of « cyber sickness ». Such limitations are both ethical (concerning the decision to run the user testing) and practical (concerning the post-experimental arrangements).

In terms of ethics, one can wonder if it is legitimate to expose subjects to physiological problems that can be serious, such as fainting fit or ataxia. In some countries, such as France, a law covers experimental situations of that nature and requires the experimental setting to incorporate the presence of certified physicians.

Concerning the management of subjects potential disorders after the experiments, protocols are available, e.g., Stanney [20], to make sure that a secure environment is provided to diagnose and to avoid potential incidents or accidents. Of course, it is obvious that, in the event of « cyber sickness », performance measures are completely biased, unless it is the goal of the experiment to test such disorders! In any case, it is mandatory to have ways, in an experiment, to diagnose early manifestations of such sickness, in order to limit its physiological consequences.

DISCUSSION ET PERSPECTIVES

This paper has presented a number ergonomic issues related to the evaluation of VE and extensions to MRS.

Among these, it may be fruitful for the future usability of MRS to investigate and to coordinate a number of research avenues.

In terms of ergonomic knowledge, a common effort should be pursued to ensure that MRS are, at all stages of design, evaluated for usability; that such data is sound and generic; and that, whenever possible, that knowledge is made available as recommendations and shared for further design and evaluation.

In terms of methods, several points can be made:

- For user testing, it would be useful to study the design of user testing protocols in order to alleviate the various biases identified.

- It is particularly important to make sure that user testing of MRS concerns well identified usability questions rather than just testing the environment as it is. Caution should be exerted in order to "clean-up" MRS from well-known problems before testing real ergonomic problems on new usability interaction paradigms.

- In addition, organizing some kind of efficient communication within the MRS research community may help to cross-reference and increase common knowledge on usability, e.g., from the various user tests performed in different laboratories. This would certainly help in the generalization of usability results and lead to commonly agreed generic recommendations.
  Using common testing platforms and testing protocols would help to compare and share results.

Also, this would facilitate the design and share of common recommendations databases. Regarding that, it would also be useful to share some mechanisms allowing the usable organization, storage, and retrieval of such sets of recommendations, using dedicated software (e.g., multiple guidelines bases management tools comparable to MetroWeb [15]).

- Under the assumption on the combinatory character of the ergonomic knowledge to be applied to MRS (real, « classic » & « web » computer-based, « VR », and « MR specific »), an effort should be made to distinguish the problems and to apply solutions from existing knowledge, for instance, on physical parameters of the MRS situations (e.g., noise, lighting, temperature, workload, perception, cognition, etc.), using knowledge related to physiology, psychophysiology, anthropometrics, etc., and of course software ergonomics. In that area, a number of results on GUIs, Web, etc., are obviously applicable (e.g., information presentation in 2D graphics, navigation on the internet, etc.), at least partly, to MRS situations.
  Also, issues related to social psychology and organizational ergonomics may apply, as such systems may concern organizational overall activities, cooperative work, virtual organizations, etc. In terms of measurements, one may also want to look at other measures than performance, such as preferences, or even, levels of addiction (e.g., in games), etc.

- Inspection methods are good candidates as complementary methods to user testing (for instance, just before user testing). Ergonomic Criteria (op. cit.) have been proposed to VE as a basis for ergonomic inspection, but it has been necessary not only to modify one criterion ("*Significance of codes and behaviour*") and to add two new ones ("*Grouping-Distinguishing items by behaviour*" and "*Physical workload*", but also to adapt their definitions, justifications, to add more illustrative examples and counter-examples, in order to take into account recommendations specific to VE. These criteria are also been currently tested in terms of compared efficiency (i.e., the number and quality of usability problems diagnosed) towards expert evaluation and user testing.
  These Ergonomic Criteria could be candidate for further adaptation to the specifics of MRS, providing the empirical data on such environments become available on a large scale. One can already consider that more criteria might be needed simply due to the fact that MRS concentrate many ergonomics issues from the Reality to the Virtuality; not mentioning the specific issue of "continuity".

In terms of MRS-specific questions, a large place should be considered for the issue of « continuity ». That issue,

which is certainly a major usability property [24] needs to be further investigated in order to better define the concept, its components, and distinguish the concept of « continuity » from other dimensions related to guidance, the structuring of information, the compatibility with the tasks or common practice, the consistency of information presentation, of modalities, of procedures, etc.

Another issue that seems to carry new types of problems is the design of help systems ; how should they be organized, which type of support should they use, etc.

In terms of models, it would be interesting to look at and to coordinate the characteristics of the models used for MRS (e.g., task models, interaction models, formal models, architectures, etc.), and to identify, when useful, the potential communication mechanisms between these models ... this has been a recurrent issue for GUIs (also in terms of system lifecycle development processes); it should not be different for MRS.

Other issues are worth discussing as well, such as the need for application and task models, for agreed upon extended task taxonomies, for instance, in order to be able to compare evaluation results, to generalize from one application domain to another one.

Partly linked to that issue is the need for shareable classifications of MR elements, of common MR objects models (and a shared vocabulary), not only to facilitate the analysis of MRS, but also both to compare results within the scientific community and to facilitate the design of strategies for inspections methods and heuristics.

Discussions on these various items, together with other issues talked about during the Workshop on "*Exploring the design and engineering of Mixed Reality Systems*", should lead to some insight for future research and contribute to increased knowledge on shared benchmarks. There is certainly lots of work at hand for this workshop, but also for the following ones!

REFERENCES

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

2. Bach C. & Scapin D. L. Adaptation of Ergonomic Criteria to Human-Virtual Environments Interactions. in *Interact'03*. IOS Press. 2003. pp. 880-883.

3. Bastien, J. M. C., Scapin, D. L. Les méthodes ergonomiques : de l'analyse à la conception et à l'évaluation. *Traité d'ergonomie, P. Falzon (Ed.), Masson.* 2003

4. Bastien, J. M. C., et Scapin, D. L. Évaluation des systèmes d'information et Critères Ergonomiques. *In Systèmes d'information et Interactions homme-machine, C. Kolski (Ed.), Hermès*. 2001.

5. Bowman, D.A., Gabbard, J.L., and Hix, D. 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.

6. Dubois, E., Gray, P.D., Nigay, L., ASUR++: a Design Notation for Mobile Mixed Systems. *Interacting With Computers, Special Issue on Mobile HCI*, Paterno, F. (ed), 2003.

7. Gabbard, J.L. *Researching Usability Design and Evaluation Guidelines for Augmented Reality (AR) Systems.* 2001. Available: http://www.sv.vt.edu/classes/ESM4714/Student_Proj/class00/gabbard/

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

9. ISO/ TS 16982. Ergonomics of human-system interaction – Usability methods supporting human-centred design. 2000

10. Kalawsky, R., VRUSE – a computerised diagnostic tool : for usability evaluation of virtual/synthetic environment systems. *Applied Egonomics*, Elsevier (ed) n°30, 1999, pp. 11-25.

11. Karampelas P., Grammenos D., Mourouzis A. & Stephanadis C. Towards I-dove, an interactive support tool for building and using virtual environments with guidelines. *Proceedings of HCI, 22-27 june 2003, Crete, Greece*, vol. 3, pp. 1411-1415.

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

13. Kaur, K. *Designing Usable Virtual Environments.* 1997. Demo available*:*
    http://www.soi.city.ac.uk/~dj524/demtool/frame.htm

14. Loeffler, C.E., and Anderson, T. The Virtual Reality Casebook. *New York: Van Nostrand Reinhold*, 1994.

15. MetroWeb: Available: http://www.isys.ucl.ac.be/bchi/research/metroweb.htm

16. Milgram, P., Takemura, H., Utsumi, A. and Kishino, F. in *SPIE 94*, Vol. 2351, Telemanipulator and Telepresence Technologies, 1994, pp. 282.

17. Nigay, L., Dubois, E., Renevier, P., Pasqualetti, L. and Troccaz, J. Mixed Systems: Combining Physical and Digital Worlds. *Proceedings of HCI, 22-27 june 2003, Crete, Greece*, vol. 1, pp. 1203-1207.

18. Norman, D. A., Cognitive engineering. *In D. Norman and S. Draper (eds), User centered system design: New perspectives on Human Computer Interaction (Hillsdale NJ: LEA)*, 1986, pp. 31-62.

19. Porcher Nedel, L., Dal Sasso Freitas, C. M., Jacon Jacob, L. and Soares Pimentas M. Testing the Use of Egocentric Interactive Techniques in Immersive Virtual Environments. *In Proceedings of INTERACT'03* (Zurich, September 2003), IOS Press, pp. 471-478.

20. Stanney, K. M., Kennedy, R. S. and Kingdon, K. Virtual Environment Usage Protocols. in *Handbook of Virtual Environments*, LEA Publishers, 2002, pp721-730.

21. 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.

22. Sutcliffe, A. G., and Kaur, K. D. Evaluating the usability of virtual reality user interfaces. *Behaviour & Information Technology*, vol. 19, n°6. 2000. pp. 415-426.

23. Träskbäck, M., Koskinen, T.and Nieminen, M. User-Centred Evaluation Criteria for Mixed Reality Authoring Application. *Proceedings of HCI, 22-27 june 2003, Crete, Greece*, vol. 3, pp. 1263-1267.

24. Trevisan, D., Vanderdonckt, J. and Macq, B. Continuity as a Usability Property. *Proceedings of HCI, 22-27 june 2003, Crete, Greece*, vol. 3, pp. 1268-1272.

25. Tromp, J. G., Steed, A. & Wilson, J. R. Systematic Usability Evaluation and Design Issues for Collaborative Virtual Environments. *Presence*, Vol. 12, n°3, june 2003. pp. 241-267.

26. Wilson, J. R., Eastgate, R. M., & D'Cruz, M. Structured Development of Virtual Environments. in *Handbook of Virtual Environments*, LEA Publishers, 2002, pp. 353-378.