



Science Arts & Métiers (SAM)

is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: <https://sam.ensam.eu>
Handle ID: <http://hdl.handle.net/10985/12778>

To cite this version :

Ahlem ASSILA, Jeremy PLOUZEAU, Aïda ERFANIAN, Yaoping HU, Frédéric MERIENNE -
Defining an Indicator for Navigation Performance Measurement in VE Based on ISO/IEC15939 -
In: 4th International Conference on Augmented Reality, Virtual Reality, and Computer Graphics,
Italie, 2017-06-12 - Lecture Notes in Computer Science - 2017

Any correspondence concerning this service should be sent to the repository

Administrator : scienceouverte@ensam.eu



Defining an Indicator for Navigation Performance Measurement in VE based on ISO/IEC15939

Ahlem Assila¹, Jeremy Plouzeau¹, Frédéric Merienne¹, Aïda Erfanian²,
Yaoping Hu²

¹ LE2I, Arts et Métiers, CNRS, Univ. Bourgogne Franche-Comté,
71100 Chalon-sur-Saône, France
{ahlem.assila, jeremy.plouzeau, frederic.merienne@ensam.eu}

² Dept. of Electrical and Computer Engineering, Schulich School of Engineering,
University of Calgary
{aerfania, huy}@ucalgary.ca

Abstract. Navigation is a key factor for immersion and exploration in virtual environment (VE). Nevertheless, measuring navigation performance is not an easy task, especially when analyzing and interpreting heterogeneous results of the measures used. To that end, we propose, in this paper, a new indicator for measuring navigation performance in VE based on ISO/IEC 15939 standard. It allows effective integration of heterogeneous results by retaining its raw values. Also, it provides a new method that offers a comprehensive graphical visualization of the data for interpreting the results. The experimental study had shown the feasibility of this indicator and its contribution to statistical results.

Keywords: Virtual Environment (VE); measure; performance; navigation; evaluation.

1 Introduction

Virtual reality is a technology that allows users to immerse in a virtual environment and enable them to navigate in order to explore it and to interact with its objects [1] [2]. This technology is realized, usually, via 3D real-time computer graphics and advanced display devices (e.g. head mounted displays (HMDs) or Caves) [3]. Generally, a virtual environment (VE) can be infinite and can afford numerous possibilities. To explore it, navigation interfaces and metaphors are necessary [2]. Undoubtedly, navigation in the virtual environment differs from the real world. It presents a main common task between virtual reality applications whatever if its goals are focused or not on moving [2].

As in all fields of application, evaluation is very crucial phase after designing a virtual reality application. Navigation performance is one of criteria to be considered when evaluating VE [4]. In literature, there are several measures proposed for making navigation performance evaluation (e.g. Task completion time [2], percentage of errors of path [5], navigation time [6], etc.). The choice between these measures differs from one study to another. It can depend, usually, on the application to be

evaluated. For examples, in the two studies [7] [8], authors have included only the measure of task completion time to evaluate navigation methods. In the study proposed by [9], authors have used mainly task completion time with the measure of error per task. However, analyzes and results reported are presented separately. Otherwise, Bliss, Tidwell and Guest [6] have been focused on the two measures of navigation time and wrong turns. Working in the same direction, we have been interested, in our previous work [10], to measure navigation performance when evaluating a guidance method in a VE. Thus, we have based on the two measures of trajectory precision and the completion time of navigation. We have firstly started by analyzing results separately which conduct us to some confusing results. These difficulties led us to think about aggregating both measures in one quantifiable value by dividing trajectory precision with the completion time of navigation. Although this formula is standard, we have found different other difficulties, mainly, in the analysis and interpretation of the huge amount of results in order to get a global conclusion. Also, the aggregation of the data into an overall score may ignore some important results that may be very useful in the analysis.

To that end, we propose in this paper another way for measuring navigation performance in VE by adapting indicator concepts defined by the ISO/IEC 15939. This standard allows defining a measure called “indicator” which combines heterogeneous data and specifies the procedure for interpreting the results. Our contribution bears on defining an indicator for navigation performance in VE, based on ISO/IEC 15939, with keeping the raw data. It includes not only the procedure for interpreting the results but also offers a clear graphical visualization to facilitate interpretation.

This article is organized as follows. Section 2 presents the related works about measurement concepts and the existed measures used in VE. Section 3 describes the motivation of our research and the proposed indicator of navigation performance, defined for evaluating VE based on ISO/IEC 15939. Section 4 presents the experimental study by applying the proposed indicator for evaluating a VE application. Section 5 introduces significant discussions about threats of validity of our study. Finally, section 6 draws some conclusions and outlines some future works about more research proposals.

2 Related Works

2.1 Measurement concepts based on ISO standards

Before using measure, we should think firstly about its definition and their main concepts. From literature and standards, several methodologies for measures definition have been proposed (e.g. Software Quality Measure [11], Practical Software Measurement [12], Systems and software engineering — Measurement process [13]). Based on standards, we distinguished heterogeneity in terminologies adopted. Some standards (such as ISO/IEC 14598 [14] and ISO 9126 [15]) refer to the term "Metric", whereas other standards such as ISO/IEC 25022 [16] adopt the term

"Measure". The distinction between the different terminologies of the measurements has been well explored in the field of software engineering. Indeed, the term "Metric" has been published in the two sets of ISO/IEC 14598 (1999-2001) and ISO/IEC 9126 (2001-2004) standards that correspond to the first international consensus on terminology for quality characteristics for the evaluation of the software product. Based on these two standards, a metric is defined as follows: *"a measurement scale and the method used for measurement"*.

In the new sets of standards, a new generation of quality standards for the software product has been proposed. It named "Systems and software Quality Requirements and Evaluation" (SQuaRE) [17]. This series of standards has replaced the two sets of ISO/IEC 14598 and ISO/IEC 9126 standards with ensuring uniformity and coherence between its terminologies. Nevertheless, SQuaRE avoided the use of the term "Metric" adopted in ISO/IEC 9126 and ISO/IEC 14598 and replaced it with the term "measure" in accordance with ISO/IEC 15939 (2007) [18]. This term has been adopted in the SQuaRE series of standards (such as (ISO/IEC 25000, [19], ISO/IEC 25021, [20], ISO/IEC 25022, [16])). The ISO/IEC 15939 (2007) has defined a measure, in general, as *"a variable to which a value is assigned as the result of measurement"* [13]. This standard distinguished it compared to the measurement term which is defined as *"a set of operations having the object of determining a value of a measure"*.

Based on the ISO/IEC 15939, we distinguished three kinds of measures as follows: base measure, derived measure and indicator. While, a base measure is defined as *"a measure defined in terms of an attribute and the method for quantifying it"*, a derived measure is defined as *"a function of two or more values of base measures"* [13]. The third kind concerned indicator is defined for responding to specific goals to be achieved called Information needs. Based on the ISO/IEC 15939, this concept is defined as *"a measure providing estimation or evaluation of specified attributes"* [13]. These attributes are quantified using objective or subjective methods. Generally, an indicator is structured around a measurement information model defined as *"a structure linking the needs to the relevant entities to the attributes of concern"* [13]. This model illustrates quantification and conversion process of the relevant attributes to indicators using measures that provide a basis for decision making [13]. Further, the definition of indicator requires the specification of an analysis model that responds to the information needs to be achieved. Following the ISO/IEC 15939 standard this model is describes as *"an algorithm or calculation combining one or more base and/or derived measures with associated decision criteria"* [13]. In turn, these decision criteria are used as a tool for ensuring the interpretation stage of indicator results. It consists on *"numerical thresholds or targets used to determine the need for action or further investigation or to describe the level of confidence in a given result"*.

2.2 Using indicators based on ISO/IEC 15939 standard

Since the specification of this measurement standard, the use of indicator concept has been more explored in the software engineering field. One advantage of use it is the fact that it can be considered as the most recent standard which provides a comprehensive measurement construction process (i.e. the measurement information model) ranging from the specification of its attributes to the establishment of

indicators that meet the specific requirements of stakeholders and their information needs [13]. In literature, several works have been focused on defining indicators which bear on the quality evaluation of software process (e.g. [21] [22] [23]). Other works have been used indicators to perform quality evaluation of software products and systems (such as, [24] [25]). In these works, authors have proposed some interpretations for supporting the analysis of indicators results.

Another prime specificity provided by indicator is its capability to combine various heterogeneous base and/or derived measures together considering a predefined information need. Based on ISO/IEC 15939, this combination can be made through the analysis model following two ways. The first is via mathematical calculations in order to have an overall score.

The second way deals with the definition of an algorithm or a model which specifies the measures associated of the indicator without effective aggregation of data and by keeping the raw data. In this same direction and based on ISO/IEC 15939, a more recent proposal about the specification of usability indicators in the field of Human-Computer Interaction has been proposed [26] [27]. These indicators mainly allow evaluating the quality of user interfaces of interactive systems, applied into traffic supervision field, with integrating effectively measures results extracted from both subjective and objective evaluation tools. Their main purpose is to prove how these indicators can support evaluators in the detection of usability problems considering effectively both the two sides of the evaluation (subjective and objective). In fact, this proposal describes how to define an indicator that performs effective integration with complementary and consistent manner [26] [27]. Furthermore, it proposed a strong basis for interpreting indicator results and supporting the decision-making in the evaluation via predefined decision criteria. In turns, it includes a set of recommendations for enhancing more the quality of user interfaces.

In this approach, the specification of a usability indicator involves three main steps [26]. The first is to identify the information need which bears on the main objective using the indicator. The second step concerns the specification of the base and derived measures associated to the indicator, with taking into account its intended purpose. The third step concerns the specification of the indicator via the specification of an analysis model that responds to the specified information need. This analysis model is defined by identifying the measures concerned and their associated decision criteria. Firstly, considering the objective of integration, the indicator is instantiated as a combination of measures for subjective and objective evaluation. The association of these measures was defined on the basis of a mapping model which combines the measures in a complementary and coherent way with respect to the same criterion to be evaluated. Secondly, for ensuring interpretation of indicator results, the decision criteria have been defined based on a set of simple rules that take into account each chosen analysis model. A rule consists in interpreting the values related to the combined measures of the indicator on the basis of predefined thresholds for each measure. The definition of a threshold has been used as a useful means of data analysis. This threshold allows indicating the acceptable results, which can be either above or below it [26] [27].

In addition, an indicator is represented graphically on the basis of the scatter plots (2D, 3D) graphs to effectively and concretely illustrate the integration between the combined measures. Fig. 1 illustrates an example of a hypothetical indicator

(instantiated as a pair of two measures M1 and M2) with the defined rules for the decision criteria (see [26]).

In this work, as presented in the Fig. 1, three levels of quality have been identified as follow [26]:

- Very bad (when all indicator measures have unacceptable/bad results);
- Bad (when at least one of the measures has an unacceptable/bad result);
- Good (in the case where all measures of the indicator have acceptable/good results).

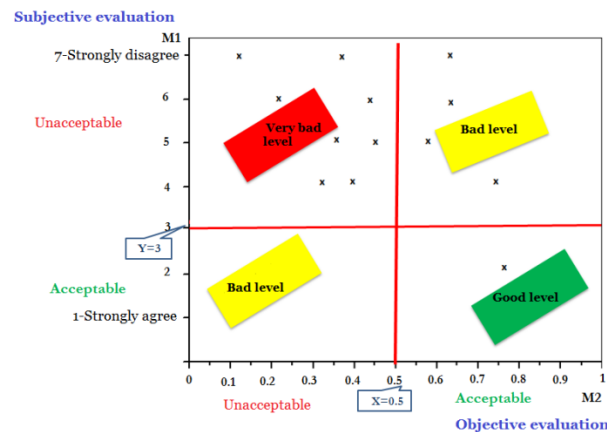


Fig. 1. Example of a hypothetical indicator with its decision criteria [26].

Considering the difficult phase of the selection of measures thresholds, this approach emphasizes the importance of involvement of a user expert in the evaluation during the experiments. His role is to select thresholds based, mainly, on the context of use of the evaluation [26].

2.3 Using measures for evaluating VE

For making evaluation of any designed interactive system, regardless of its field of application, the use of measures is very essential. In the Human-Computer Interaction (HCI) field, the evaluation of user interfaces covers, generally, a set of quality factor (such as usability) that in turns depend on criteria to be evaluated (for example usability can be measured in terms of effectiveness, efficiency, etc. [28]). These criteria can be evaluated using measures.

As in HCI domain, in virtual reality, the evaluation depends on the parameters/criteria to be evaluated (such as the sense of presence, navigation performance, cyber-sickness, etc.) [29]. Thus, the use of measures to make evaluation in virtual environment is very significant. In literature, several measures have been proposed. Those allow ensuring subjective or objective evaluation respectively with or without involving user perceptions [26]. For example, many authors have proposed presence questionnaires as a measure for performing subjective evaluation of the sense of presence in order to take an overall score ([30] [31]). Others have been focused to

define objective measures (such as heartbeat, electrodermal processes, reactions of the eyes, etc. [29]) for assessing sense of presence.

As virtual reality allows measuring any parameters in the simulation, there is an infinity of parameters/criteria that can be measured. Thus, navigation performance in VE can be performed using several measures (such as position, trajectory, orientation, task completion time [2] [10]). Nevertheless, the selection of these measures depends, primarily, on the application/simulator/tasks to be evaluated.

As introduced before, regarding our main purpose, we present in the next section the proposed indicator for navigation performance measurement in VE. It mainly allows integrating effectively heterogeneous measures based on a predefined information need, providing interpretation process and a clear visualization in order to better support the evaluation.

3 Measuring Navigation Performance in VE based on ISO/IEC 15939 standard

As explained in the section 2.2, Assila, Oliveira and Ezzedine [26] have proposed, recently, a generic process to be applied for constructing indicators based on ISO/IEC 15939 standard. Those are aimed to evaluate interactive systems based on predefined quality criteria with the integration of heterogeneous measures. This approach has been applied in the field of Human-computer Interaction and it has proved its feasibility. Believing the strength of this proposal and considering our goal, we propose to apply this process for defining a navigation performance indicator used in evaluating VE. This section comprises two parts. In the first sub-section, we introduce our motivation for specifying the proposed indicator. In the second section, we describe the measurement information model proposed for evaluating navigation in VE.

3.1 Motivation

For our knowledge, in the field of virtual reality the use of indicator concepts defined by ISO/IEC 15939 standard is not yet proposed. In this paper, we argue that the exploitation of this concept is very promising, regarding, firstly, that it is a standard characterized by its applicability and adoption in industry [26] [32]. Secondly, its capability to define a complete procedure for specifying, collecting and interpreting measures that respond to our goal [13]. Thus, we decided to propose a first indicator for evaluating navigation performance in virtual environment.

It contributes to:

- Integrate effectively heterogeneous results of the associated measures of indicator with a single way.
- Offer a simple way and quick understanding of results through the use of visual metaphors (in our case the scatter chart).

- Support evaluators in the analysis of their heterogeneous results and provides a basis for making interpretation via the decision criteria proposed.
- Allow creating matrix of choices for decision support.
- Allow detecting and fixing the causes of problems.
- Provide a comprehensive and visual conclusion to better manage the navigation performance in virtual environments.

3.2 Proposed indicator of navigation performance for evaluating VE

By following the process of defining indicators proposed by Assila, Oliveira and Ezzedine [26], we start first by identifying information need using the proposed indicator. As we explained before, our main goal for defining this indicator is to evaluate navigation performance in a virtual environment by integrating heterogeneous data. This indicator can be used in two ways, for assessing one pattern in VE or to compare several patterns together.

The second step consists on identifying base and/or derived measures to be used when specifying the indicator. To that end, we started firstly by specifying the entities and their attributes which are the most suitable with our information need. In our case, we identified the pattern to be evaluated as the entity from which we considered the path as the main attribute.

Subsequently, considering both our goal and the attribute to be considered we have focused on some existing measures, defined in literature [10], for measuring navigation performance in virtual environment. For base measures, we have considered: Length of trajectory followed, Total length of the trajectory, Time of arrival and Time of departures. As for the derived measures, we have selected the Trajectory precision and the Path time measures.

The third step consists on specifying the indicator itself. We called it as “*Navigation performance indicator in VE*”. Considering our information need, we enounced this indicator as a pair of the two following derived measures: the trajectory precision and the path time.

The measure of Trajectory precision allows measuring the precision for following the right trajectory. It is calculated as follows [10]: **Trajectory precision = Length of trajectory followed / Total length of the trajectory.**

For the Path time which measures the completion time of navigation, we adopted this formula [10]: **Path time = Time of arrival – Time of departures**

Since the visualization of the measures data generated by the indicator is a primordial phase to illustrate the effective integration of the results, we decided to represent our indicator by a scatter chart.

For interpreting indicator results, we have adopted the same principle proposed by [26] based on the definition of decision criteria. Those bear on specifying a set of rules associated to both combined measures. Each rule consists on checking if measures results are acceptable or not based on predefined thresholds. As proposed by the ISO/IEC 15939 standard, the use of thresholds is a useful technique for analyzing results. In turn, the selection of these thresholds should be made by a user expert considering, mainly, the context of use of the evaluation [26]. Thus, in our case, we decided to involve a user expert in the evaluation of VE when selecting thresholds of

measures. He /her should take into account the following aspects of the context of use according to the ISO 9241-11 standard which are: users, tasks, equipment (hardware, software and materials), and the physical and social environments [28]. In addition, following the approach of [26], the user expert should consider in our instance users' experiences with virtual reality technology and also their experiences with the system to be evaluated.

We present in Table 1 the detailed description of the measurement information model for our proposed indicator. As illustrated in this Table (see Decision criteria), we have proposed four rules for interpreting indicator results as follows:

- **Rule 1:** If (Path time is Acceptable) & (Trajectory precision is Acceptable)
=> Good level of navigation performance: No problem
- **Rule 2:** If (Path time is Unacceptable) & (Trajectory precision is Acceptable)
=> Bad level of navigation performance caused by Path time.
- **Rule 3:** If (Path time is Acceptable) & (Trajectory precision is Unacceptable)
=> Bad level of navigation performance caused by trajectory precision.
- **Rule 4:** If (Path time is Unacceptable) & (Trajectory precision is Unacceptable)
=> Very bad level of navigation performance caused by the two measures

Table 1. Specification of the measurement information model for the navigation performance indicator

Information need	Evaluate the performance of navigation in VE for one or more patterns.																	
Indicator	The navigation performance indicator in VE																	
Analysis model	The model is enunciated as a pair of the two measures (<i>Path time of a user</i> and <i>the trajectory precision by a user</i>) represented by a scatter chart.																	
Decision criteria	<p>Interpretation of each measure</p> <ul style="list-style-type: none"> For the <i>Path Time</i> measure named T: The smaller the value the better the performance. $0 \leq T \leq N$ where N is the maximum value obtained. The results are divided into two subranges according to the threshold X specified by an expert: <ul style="list-style-type: none"> Acceptable: for values $\in [0 ; X[$ Unacceptable: for values $\in [X ; N]$ For the <i>trajectory precision by a user</i> measure named P: The higher the value the better the performance. $0 \leq P \leq M$ where M is the maximum value obtained. The results are divided into two subranges according to the threshold Y specified by an expert: <ul style="list-style-type: none"> Unacceptable: for values $\in [0 ; Y[$ Acceptable: for values $\in [Y ; M]$ <p>Note: The user expert specifies thresholds X and Y taking into account the context of use of the evaluation and users' experiences.</p> <p>Decision criteria of the indicator</p> <table border="1"> <thead> <tr> <th colspan="2"></th><th colspan="2"><i>Path time</i></th></tr> <tr> <th colspan="2"></th><th>$[0 ; X[: \textit{Acceptable}$</th><th>$[X ; N] : \textit{Unacceptable}$</th></tr> </thead> <tbody> <tr> <td rowspan="2"><i>Trajectory precision</i></td><td>$[Y ; M] : \textit{Acceptable}$</td><td>Good level of performance No problem.</td><td>Bad level of performance caused by Path time.</td></tr> <tr> <td>$[0 ; Y[: \textit{Unacceptable}$</td><td>Bad level of performance caused by trajectory precision.</td><td>Very bad level of performance caused by the two measures.</td></tr> </tbody> </table>					<i>Path time</i>				$[0 ; X[: \textit{Acceptable}$	$[X ; N] : \textit{Unacceptable}$	<i>Trajectory precision</i>	$[Y ; M] : \textit{Acceptable}$	Good level of performance No problem.	Bad level of performance caused by Path time.	$[0 ; Y[: \textit{Unacceptable}$	Bad level of performance caused by trajectory precision.	Very bad level of performance caused by the two measures.
		<i>Path time</i>																
		$[0 ; X[: \textit{Acceptable}$	$[X ; N] : \textit{Unacceptable}$															
<i>Trajectory precision</i>	$[Y ; M] : \textit{Acceptable}$	Good level of performance No problem.	Bad level of performance caused by Path time.															
	$[0 ; Y[: \textit{Unacceptable}$	Bad level of performance caused by trajectory precision.	Very bad level of performance caused by the two measures.															
Derived measures	<i>Trajectory precision</i> : measures the rate of the well-traced trajectory.																	
Measurement Function	<i>Trajectory precision</i> (P) = Length of trajectory followed / Total length of the trajectory. $0 \leq P \leq 100\%$																	
Base measures	<i>Path time</i> (T): measures the total time between departures and arrivals. $0 \leq T \leq 1000s$																	
Measurement method	<i>Path time</i> (T) = Time of arrival – Time of departures																	

In other words, through our indicator, we have identified three levels of performance described as follows: good, bad and very bad, represented by three colors respectively: green, yellow and red.

During the analysis phase, the indicator results are represented by a scatter chart. Those will be distributed on the four areas identified by the indicator rules and will be grouped into clusters. Two cases are represented here. If the indicator is used to evaluate the navigation performance of one pattern then the conclusion can be derived directly from the Table of Decision Criteria (Table 1). Otherwise, if the indicator is used to compare different patterns, the conclusion can be established through the comparison of patterns clusters.

At the end of the evaluation, a complete and visual conclusion will be established. As a result, evaluators' decisions will be made based mainly on the evaluated case study.

4 Experimental Study

In this section, we are focused, mainly, to apply our proposed indicator of navigation performance in the guidance navigation evaluation of a VE in order to check its feasibility and its contribution compared to the statistical results obtained during our previous work [10]. To that end, we have taken the same experiment that we have already realized.

As described in [10], a VE has been created. It is composed by a grassy plain where the user could navigate using Razer Hydra game pads. As illustrated in this figure, a pair of shoes was displayed at the user's feet. These virtual shoes were following the user when he moved through the VE. In addition, in auditory feedback of walking steps gives the user the illusion of walking. Further, he could view both plain and shoes during his/her promenading.

When navigating, user had to follow an invisible path of navigation being guided by two different vibration patterns. Those have been called pushing and compass patterns [10]. The pushing pattern is composed by two different signals, one to tell the user to go forward and the second to tell him to turn. The compass pattern indicates the direction to follow by constant vibrations around ankles [10]. The path was configured as a sinusoidal curve with a period of 12 meters and 6 different amplitudes of 0.0 m, 0.3 m, 0.6 m, 1.2m, 2.4 m and 4.8 m. The order of the different amplitude was randomized. The only constant was the first part and the last which was the straight line. The path ran each amplitude for two periods, corresponding 24 meter (see Fig. 2).

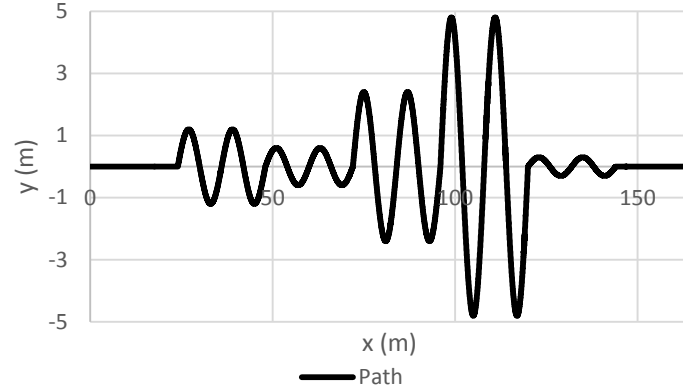


Fig. 2. Example of a generated path

In the following, we present the experimental protocol, the indicator results obtained and we end up by a conclusion.

4.1 Experimental protocol

The experiment study was conducted under an ethics clearance following the Canadian Tri-Council Ethics Guidelines. It has been based on the same context of use composed by the following components:

- **Users:** during experiment, users 'sample was involved 8 participants (5 females and 3 males), with an average age of 30 years (aged 25 to 44). There have different profiles, with no experience either in virtual reality technology or research.
- **Environment:** the experimentations were took place in a research room at the University of Calgary (Canada).
- **Equipment:** Each participant stood stationary in front of a 24" flat widescreen, which displayed the VE to minimize cybersickness for the participant. As depicted in Fig. 3, the screen was located in an arm distance from the participant and inclined at about 40° with respect to the participant. The head of the participant titled naturally forward, allowing his/her gaze line perpendicular to the surface of the screen. The height of the screen was adjustable to ensure this perpendicularity following the height of the participant. As mentioned above, participant used Razer Hydra game pads to navigate in the VE. Two Vibrotac bracelets were anchored around the ankles of the participant. The motor under the control box of each bracelet was located exterior-laterally just above the ankle joint. To eliminate the interference of any footwear, the participant was barefooted with socks during the experiment [10].
- **Tasks & experimental scenario:** As we have described in our previous work [10], the scenario followed during the experiment included six parts. The first was to perform pretests for checking the participant's color vision, handedness,

footedness and their sensibility to the vibrations. Thus, participants with color vision or vibration sensibility troubles were asked to stop the experiment. Subsequently, participants were invited to perform three learning and two testing sessions [10]. During the first learning session, the participant focused on learning navigation in the VE using only the game pads. Then he/she was to learn one of the two patterns for guiding navigation (pushing or compass pattern). Afterward, the participant underwent the first testing session; he/she had to navigate in the VE by following firstly the guidance given by the learned pattern from the bracelets. Subsequently, he/she had to perform the third learning session in order to learn the second pattern of guidance from the bracelets. Therefore, he/she executed the second testing session which consisted on navigating in the VE based on this learned pattern. After each testing session, participants were asked to answer a questionnaire about the simulation [10]. For information, each participant undertook the two patterns of guidance in a random order. Between any two sessions, there was a 10-minute break for each participant. Each participant was involved in the experiment for about 1.5 and 2 hours, including both pre-test and experimental blocks. Regarding the captured data, the actual trajectory and path time of navigation were logged, for each participant, in real time during the experiment via our VE.



Fig. 3. The experimentation equipment

4.2 Results obtained & conclusion

After experimentation, we were generated in our previous work all participants results for the Path time measure and the trajectory precision measure [10]. Fig. 4 illustrated the path time obtained by each participant using both patterns. Further, the results of

trajectory precision measure using the two patterns, were be obtained by dividing the length of trajectory followed with the total length of the trajectory (see Fig. 5). In our previous work, when analyzing results separately with paired t-test (Fig. 4 and Fig.5), we have noted a heterogeneity especially for results of trajectory precision measure (see Fig.5). This made the analysis difficult although the results of path time measure have showed that the participants have spent more time in navigation using the pushing pattern.

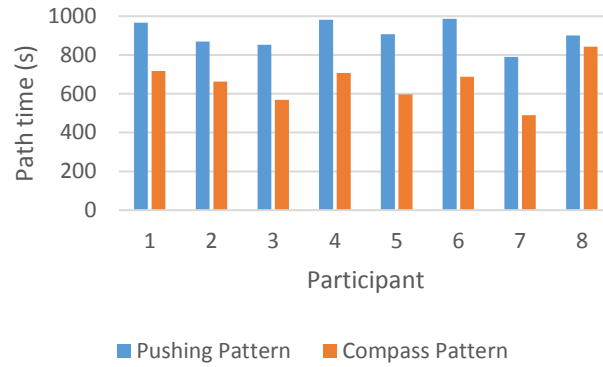


Fig. 4. Path time results [10]

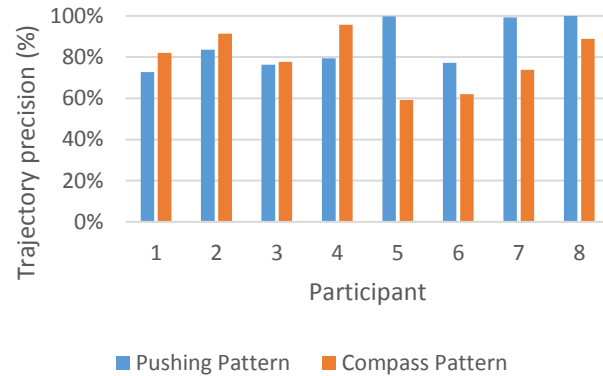


Fig. 5. Trajectory precision results [10]

Subsequently, we have calculated the performance index by aggregating the two measures by dividing the trajectory precision with the path time.

$$P = \frac{\text{Trajectory precision}}{\text{Path time}}$$

As illustrated in Fig. 6, the heterogeneity of results obtained by each participant made the interpretation task more complicated and difficult for reaching an effective conclusion. In addition to that, we have proceeded to calculate the mean performance index for the two patterns. Results obtained have been shown that the mean

performance index of compass pattern is higher than those of pushing pattern. The paired t-test returned a p-value of $p = 0.022$, lower than the significant level $\alpha = 0.05$. Observations of the results and the statistical analysis allow then to conclude that the compass pattern is more efficient to guide user through a VE. Nevertheless, these results still insufficient and qualitative analyses can be improved by applying our proposed indicator for navigation performance.

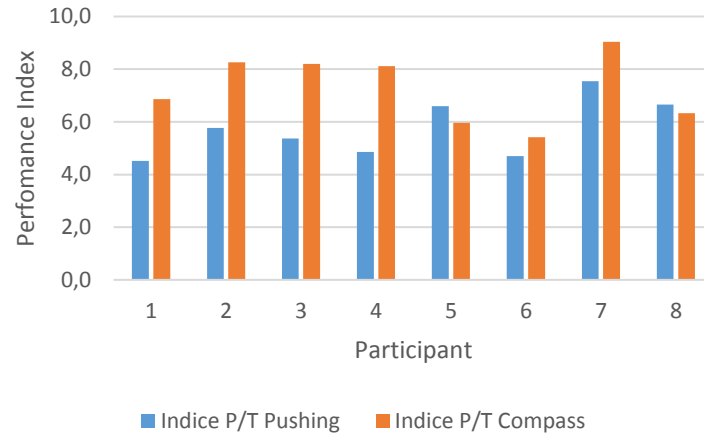


Fig. 6. Statistical results for performance index obtained in our previous work [10]

According to the results from previous article presentation, the three figures 4, 5 and 6 are needed to analyses data and give some early conclusions. Fig.4 shown that in every case, participants used more time to complete the path with the Pushing pattern than with the Compass pattern. Fig.5 indicated that some participants were better in precision with the Pushing pattern while others did better with the Compass pattern. Finally, Fig.6 allows concluding that the loss of precision with the Compass pattern is compensated by the path time. Users are better with the compass pattern.

By applying our proposed indicator for navigation performance, which integrates effectively the measures of the trajectory precision and the path time, we found the same conclusion with one representation and clear analyses. Fig.7 shows the equivalent precision which is above the threshold for most participants. It also shows that participant used less time to complete the path with the Compass pattern. We have selected the two thresholds (850s for the path time and 75% for the trajectory precision) based mainly on the context of use of our experiement and users' experiences as described above. Further, indicator results indicate that the majority of users results using Compass pattern have good levels of performance compared to those using Pushing pattern. Moreover, thresholds, barycentre and standard deviation as seen in Fig.7 allow to show that compass pattern increase performances as results are mainly located in the good level space division.

In summary, although that statistical analysis is needed to conclude that the Compass pattern is better in guiding user, we argue that our indicator shows clearly and simply the different impacts of guiding patterns with keeping the raw data. Therefore, this indicator provides a feasible method for measuring navigation performance with an effective integration of heterogeneous results.

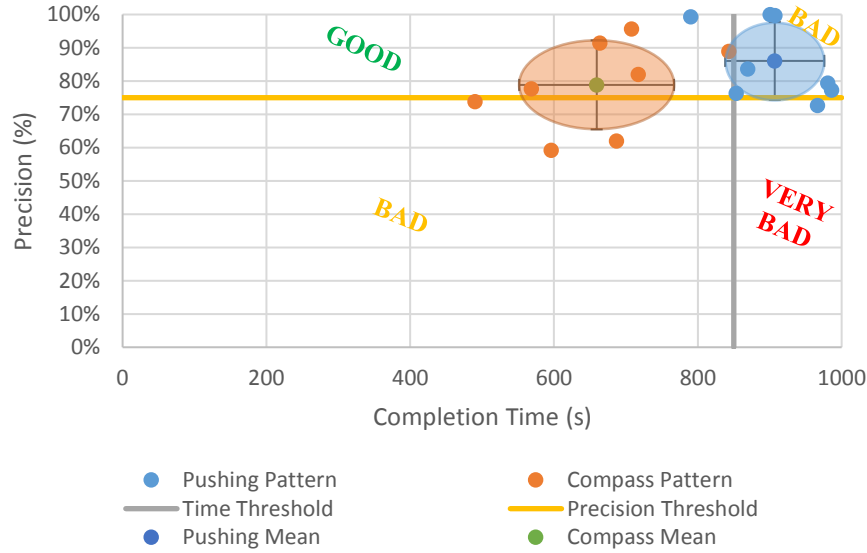


Fig. 7. Mean results of the indicator for the two patterns

5 Discussion

During this study, we have examined the four threats of validity proposed by Wohlin [33]: construction validity, conclusion validity, internal validity and external validity. The threats to construction validity illustrate the relationship between theory and observation; whether the treatment adequately reflects the cause and whether the outcome adequately reflects the effects [33]; and in our case, whether, the considered measures adequately reflect the navigation performance in VE. To minimize this risk, we relied on the well Known and referenced measures dedicated for measuring navigation performance in VE. Nevertheless, its calculation methods have largely depended on our designed VE.

The threats of conclusion validity are those that affect the ability to draw the right conclusion about the relationship between treatment and the outcomes of our study (effective integration of heterogeneous results to facilitate interpretation) [33]. To reduce this threat, we required the consideration of the context of use when selecting thresholds of measures. However, we cannot control this bias and therefore we have accepted this risk.

The threats associated with internal validity draw influences that may affect the independent variables with respect to the causal relationship between the treatment and the result obtained [33]. In our case, those concern, particularly, the participants involved in the experiment. To minimize this risk, different participants have been invited, by convenience, without retaining specific profiles based mainly on the purpose of our VE.

Lastly, the threats of external validity are conditions that limit our ability to generalize the results of our experiment outside the scope of our study [33]. To minimize this threat, our indicator can be applied in the evaluation of navigation performance of other VE with taking into consideration the tools used for navigation. Therefore, precision calculation will depend on the designed VE.

6 Conclusion & Perspectives

This study shows an innovative way to analyze data from virtual reality experiment. We demonstrated the use of indicators in order to evaluate and compare two modalities. Here the indicator we used concern the navigation performance. Using this method allows to observe and analyze numerous data in an easy way.

Nevertheless, virtual reality is ruled by several parameters which are often needed to be evaluated, such as cyber-sickness, sense of presence, cognitive load, and more. These parameters depend on several measures and the method exposed in this paper could be used to evaluate them. Thus, it would be a tool to describe a virtual reality application through cyber-sickness indicator, sense of presence indicator and more. To go further, it would be possible to combine all data in one indicator or at least some indicator together to evaluate the virtual reality experience. Using this kind of indicator could leads to decision making in real-time to adapt the simulation parameter to increase the user's virtual reality experience.

Acknowledgments.

This research work was supported by the region of Burgundy Franche-Compté.

References

1. Louka M.N.: Augmented and Virtual Reality Research in Halden 1998–2008. In: Skjerve A., Bye A. (eds) *Simulator-based Human Factors Studies Across 25 Years*. Springer, London (2010).
2. Pouzeau, J., Paillot, D., Chardonnet, J. R., Merienne, F.: Effect of proprioceptive vibrations on simulator sickness during navigation task in virtual environment. In: *International Conference on Artificial Reality and Telexistence Eurographics Symposium on Virtual Environments*, pp. 10–28, Japon (2015).
3. LaViola, J.J.: A Discussion of Cybersickness in Virtual Environments. *SIGCHI Bulletin*, vol. 32, pp.47–56 (2000).
4. Walker, B.N., Lindsay, J.: Navigation performance in a virtual environment with bonephones. In: *Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display*. pp. 260–263. Limerick, Ireland (2005).
5. Terziman, L., Marcal, M., Emily, M., Multon, F.: Shake-Your-head: Revisiting Walking-in-Place for Desktop Virtual Reality. In: *Proceedings of ACM VRST2010*, pp. 27–34. ACM, Hong Kong (2010).
6. Bliss, J.P., Tidwell, P.D., Guest, M.A.: The effectiveness of virtual reality for administering spatial navigation training for firefighters. *Presence*. 6 (1), 73--86 (1997).

7. Kopper, R., NI, T., Bowman, D., Pinho, M. S.: Design and Evaluation of Navigation Techniques for Multiscale Virtual Environments. In: Proceedings of IEEE Virtual Reality conference (VR'06). pp. 175-182, IEEE Computer Society Washington, DC, USA (2006).
8. Ardouin, J., Lécuyer, A., Marchal, M., Marchand, E.: Navigating in Virtual Environments with 360° Omnidirectional Rendering. In: IEEE Symposium on 3D User Interfaces, pp. 95—98. Orlando, FL, USA (2013).
9. Zielasko, D., Horn, S., Freitag, S., Weyers, B., Kuhlen, T.W.: Evaluation of hands-free HMD-based navigation techniques for immersive data analysis. 2016 IEEE Symposium on 3D User Interfaces (3DUI). pp. 113--119, Greenville, South Carolina, USA (2016).
10. Plouzeau, J., Erfanian, A., Chiu, C., Merienne, F., Hu, Y.: Navigation in virtual environments: design and comparison of two anklet vibration patterns for guidance. 2016 IEEE Symposium on 3D User Interfaces (3DUI). pp. 263--264, Greenville, South Carolina, USA (2016).
11. Suetendael, N. V., Elwell, D.: Software quality metrics. Atlantic City International Airport, NJ: Federal Aviation Administration Technical Center (1991).
12. McGarry, J., Card, D., Jones, C., Layman, B., Clark, E., Dean, J., Hall, F.: Practical software measurement: Objective information for decision makers. Boston, MA: Addison-Wesley (2002).
13. ISO/IEC 15939.: International Organization for Standardization/International Electrotechnical Commission: Systems and software engineering— Measurement process (ISO /IEC 15939: 2007(E)). Geneva, Switzerland: Author (2007).
14. ISO/IEC 14598-1.: Information technology -- Software product evaluation -- Part 1: General overview. International Organization for Standardization, Geneva, Switzerland (1999).
15. ISO/IEC 9126-1.: Software engineering - Product quality - Part 1: Quality model. International Organization for Standardization, Geneva, Switzerland (2001).
16. ISO/IEC 25022.: Systems and software engineering - Systems and software quality requirements and evaluation (SQuaRE) - Measurement of quality in use. International Organization for Standardization, Geneva, Switzerland (2012).
17. ISO/IEC 25010.: Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models. International Organization for Standardization, Geneva, Switzerland (2011).
18. Abran, A., Al-Qutaish, R., Cuadrado-Gallego, J.: Analysis of the ISO 9126 on Software Product Quality Evaluation from the Metrology and ISO 15939 Perspectives. WSEAS Transactions on Computers. 5(11), 2778--2786, World Scientific & Engineering Academy and Society (2006).
19. ISO/IEC 25000. : Ingénierie des systèmes et du logiciel -- Exigences de qualité des systèmes et du logiciel et évaluation (SQuaRE) -- Guide de SQuaRE. International Organization for Standardization, Geneva, Switzerland (2014).
20. ISO/IEC 25021.: Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Quality measure elements. International Organization for Standardization, Geneva, Switzerland (2012).
21. Antolić, Ž.: An example of using key performance indicators for software development process efficiency evaluation. R&D Center Ericsson Nikola Tesla, 6 (2008).
22. Feyh, M., Petersen, K.: Lean software development measures and indicators—A systematic mapping study. In: Lean Enterprise Software and Systems Proceedings—4th International Conference. pp.32--47, Galway, Ireland (2013).
23. Monteiro, L., Oliveira, K.: Defining a catalog of indicators to support process performance analysis. Journal of Software Maintenance and Evolution: Research and Practice, 23 (6), 395—422 (2011).
24. Moraga, M. A., Calero, C., Bertoa, M. F.: Improving interpretation of component-based systems quality through visualization techniques. IET Software, 4 (1), 79–90 (2010).

25. Staron, M., Meding, W., Hansson, J., Höglund, C., Niesel, K., Bergmann, V.: Dashboards for continuous monitoring of quality for software product under development. In: I. Mistrik, R. Bahsoon, P. Eeles, R. Roshandel, & M. Stal (eds.), *Relating System Quality and Software Architecture*. pp. 209--229, Waltham, MA: Morgan Kaufmann Publisher (2014).
26. Assila, A., Oliveira, K., Ezzedine, H.: Integration of Subjective and Objective Usability Evaluation based on ISO/IEC 15939: a Case Study for Traffic Supervision Systems. *International Journal of Human-Computer Interaction*. 32, 12, 931--955 (2016).
27. Assila, A., Oliveira, K., Ezzedine, H.: An Environment for integrating subjective and objective Usability findings based on Measures. *IEEE 10th International Conference on Research Challenges in Information Science*. IEEE, pp. 645--656, Grenoble, France (2016).
28. ISO 9241-11.: ISO 9241-11, Ergonomic requirements for office work with visual display terminals (VDT) s- Part 11 Guidance on usability (1998).
29. J. Van Baren., W. Ijsselstein.: Compendium of presence measures. Deliverable 5 for OmniPres project: Measuring presence: A guide to current measurement approaches. (2004).
30. Witmer, B. G., Singer, M. J.: Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence*. 7(13), pp. 225—240 (1998).
31. Gerhard, M., Moore, D., Hobbs, D.: Continuous presence in collaborative virtual environments: Towards the evaluation of a hybrid avatar-agent model for user representation. In: A. de Antonio, R. Aylett, D. Ballin (eds.): *Proceedings of the International Conference on Intelligent Virtual Agents*. pp. 137—153. Madrid, Spain (2001).
32. Staron, M., Meding, W., Nilsson, C.: Framework for developing measurement systems and its industrial evaluation. *Information and Software Technology*. 51 (4), 721–737 (2009).
33. Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslén, A.: *Experimentation in Software Engineering*. Springer-Verlag Berlin Heidelberg (2012).