Understanding VR Software Testing Needs from Stakeholders' Points of View

[‡]Stevão A. Andrade, [†]Alvaro Joffre U. Quevedo, [‡]Fatima L. S. Nunes, [‡]Márcio E. Delamaro
[‡]Universidade de São Paulo (ICMC, EACH), Brasil
[†]Ontario Tech University Oshawa - ON, Canada
{stevao,delamaro}@icmc.usp.br, fatima.nunes@usp.br, alvaro.quevedo@ontariotechu.ca

***Abstract*—Software testing is a critical activity to ensure that software complies with its specification and guarantee for delivering high quality products. However, there are a number of challenges to be faced as new software domains are created. Although there are some works in the literature that explore the application of software testing practices specific to the context of Virtual Reality (VR), these works tend to reflect only the researcher's point of view, leaving aside the players involved in the VR development process, contributing to a gap that makes it difficult to popularize such practices. Aiming to reduce this gap, this paper presents an analysis under the perception of stakeholders involved in the process of developing and using VR applications. We conducted a survey with 88 respondents, comprising different roles, interviewed during a virtual reality scientific conference and later complemented by electronic dissemination, reaching five different profiles. The results showed that there is a major concern among all the interviewed profiles regarding the impact of faults in VR applications. In addition, we investigated these results in both software and hardware aspects. The results related to software pointed to a greater concern of the players regarding visual aspects and non-functional requirements. As for hardware-related aspects, the biggest concern is related to limitations in mechanisms that are responsible for providing input data to applications.**

***Index Terms*—software testing, virtual reality, validation**

I. INTRODUCTION

The production of quality software has become an incessant pursuit in the software development industry as software and technology have gained more and more space in people's daily lives, and it is responsible for a wide range of tasks in the most diverse types of application domains [1]. A software development cycle consists of a series of phases, ranging from engineering, design, implementation, verification and validation requirements to software maintenance [2]. To be successful, software development teams also need to take into account aspects such as user satisfaction, in addition to all the functionalities of the software itself [3]. Similar to the software development phase, we believe that the software testing phase should also have this type of concern.

Similarly, technologies such as virtual reality (VR), which allows development of systems with advanced resources, such as images, sounds, videos and three-dimensional (3D) real-time interaction, also follow a similar development model, in addition to having their own standards [4]. However, due to their unique characteristics, it is clear that, unlike other type of software domains, systematized software testing techniques and criteria in the literature cannot be precisely applied to their context [5].

Although there are some works in the literature that explore the application of software testing practices specific to the context of VR [6]–[8], studies point out that there is still a difficulty in bridging the gap in knowledge transfer between academia and practitioners [5].

We believe that part of this problem is related to the fact that the works tend to mostly explore the researcher's point of view, leaving aside the players involved in the context of VR development, who will be mostly responsible for using the knowledge produced. Thus, in order to fill this gap, stakeholders' points of view need to be understood regarding their real needs. This could allow testing approaches to be targeted at what these groups think is really important, making it possible to prioritize efforts in faults that, ideally, should not manifest themselves in the final product of a VR application.

The aim of this study is to obtain a view of the interest groups, in addition to investigating the knowledge of these groups regarding specific types of faults in VR applications. As example of points of interest we can cite: which types of faults are more critical, which are less relevant, and how much each one affects the quality of the final product. To do so, the stakeholders' opinions were collected from conducting a survey, in which the participants were asked to answer a set of questions about what types of faults in VR applications contribute to a negative experience, among other aspects.

The remainder of the paper is organized as follows. Section II briefly presents some related works, in Section III we discuss the survey design of this study. Results are presented in Section IV. A discussion about the research questions that guided the survey is presented in Section V. Section VI analyzes the threats to validity. Finally, we draw conclusions and provide some directions for future work in Section VII.

II. RELATED WORKS

Güleç et. at. [9] presented a survey in order to investigate whether it is necessary to develop new software development methodology for VR projects. After analyzing a vast collection of papers, they concluded that it is useful to develop a new software development process for VR projects in order to reduces the risk of errors since many studies show that have indicated that the reuse of existing processes can result in failure when applied to multidisciplinary fields, which is the case with VR applications.

Our survey target on exactly this gap, however it explores a more specific aspect within the context of software engineering, which is the adoption of software testing practices.

Complementarily, Noghabaei et. al. [10] conducted a study to understand the industry trends and limitations in adopting VR technologies. They investigated the utilization growth of VR technologies by measuring the potential cost and time savings for VR development in order to improve communication and visualization among different stakeholders. The results show that despite industry experts foresee a strong growth in the use of VR technologies over the next 5 to 10 years, there are some inherent limitations in adopting new VR technologies, such as lack of budget, upper management's lack of understanding of these technologies, design teams' lack of knowledge.

Our work is motivated by this perception that there is still a misinformation in the roles involved in the context of development and use of VR applications and tries, to provide some insights by identifying the perceptions and limitations of stakeholders so that future work can create ways to mitigate part of the challenges existing in the context of software testing for VR applications.

III. Survey Design

The survey was planned following the guidelines proposed by Linaaker et al. [11] to design effective surveys for the software engineering area. Therefore, the process used while conducting the survey was broken down into seven major steps, represented in Figure 1 (white boxes represent the steps followed to apply the survey, and blue boxes identify the phases related to data extraction and interpretation) and discussed in the subsections below.

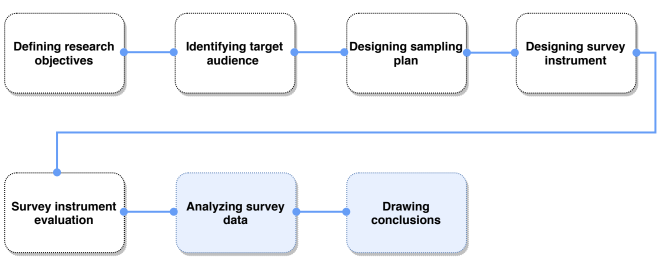

Figure 1: Process adopted in the survey conduction

A. *Defining research objectives*

The aim of the survey was formulated using the Goal, Question, Metric (GQM) approach [12] and intends to understand stakeholders' points of view regarding the application of software testing practices in VR applications.

In addition to understanding the opinion regarding the need for testing practices, the survey also attempts to understand stakeholders' major concerns about types of faults in VR applications that contribute to poor experience. We perform this assessment by analyzing which characteristics are considered as most critical, and therefore could be prioritized during the test phase of VR applications.

Specifically, we studied the following research questions:

- $\mathbf{RQ_1}$: What is the perception of stakeholders regarding the need to prioritize testing efforts in specific characteristics in VR applications?
- $\mathbf{RQ_2}$: What are the stakeholders' perceptions regarding which hardware fault characteristics most contributes to a negative/poor experience in VR applications?
- $\mathbf{RQ_3}$: What are the main types of faults that stakeholders are aware in VR applications?

B. *Identifying target audience*

The survey was initially planned and designed to reach any role involved in the development and use of VR experiences.

C. *Designing sampling plan*

To reach the target audience, the survey was conducted during the *21st edition of the Symposium on Virtual and Augmented Reality (SVR)*, which is the first conference on Virtual Reality and Augmented Reality in Brazil, held annually by the Brazilian Computer Society (SBC in Portuguese).

We decided to carry out the survey during this event due to the fact that SVR brings together researchers, students and professionals from various academic fields, industrial and commercial areas interested in the advances and applications of VR.

In addition to that, we also made an online version of the survey available in order to reach a general audience that reflects the perceptions from the point of view of users of the technology.

D. *Designing survey instrument*

We reviewed the objectives proposed in the first stage of the process and defined the way that the data provided by the participants would be collected. The web questionnaire was divided into four sections. An electronic version with all the collected responses is provided [13].

The first section contains a presentation of the survey in which the objectives are described, as well as the target audience and clarifications regarding the importance of participation. The second section aims to characterize the respondents' role and understand their perceptions regarding the use of VR technologies. The purpose of the third section was to understand the critical factors that could difficult the ability of using and enjoying VR applications. Finally, the objective of the fourth section was to understand the knowledge of stakeholders concerning the various types of faults that exist in VR applications, as well as to understand whether stakeholders judge them as critical.

E. *Survey instrument evaluation*

Designing a survey instrument is an iterative activity that involves an evaluation by carrying out a pilot run to introduce improvements in the survey instrument [11]. Therefore, to avoid ambiguously and poorly-worded questions, we pilot-tested our questionnaire with a small group of 3 researchers in order to collect some feedback and understand if the questions were clear and if the material provided enough information to carry out the survey.

F. Analyzing survey data

Most questions were designed to fit a model classified as closed questions. Generally closed-ended questions provide a fixed list of response choices or categories and ask respondents to select one or more options as their answer.

This model collects quantitative data, and can therefore allow us to make a better analysis through charts and quantitative methods.

G. Drawing conclusions

After compiling the results, the respondent data was analyzed and the results are presented in the next section. To avoid over confidence in the results, they were evaluated and reviewed individually from a critical viewpoint by all the authors of the paper and are discussed through charts and tables.

IV. Results

The survey was answered by 88 stakeholders with several roles in total, as detailed in the next subsections.

A. Characterization of participants

In order to understand the role of the participants regarding how they classify themselves within the context of virtual reality, the first question of the survey was *"What is your role when using VR?"*

The pie chart shown in Figure 2 depicts this information. Having 47% of the responses, the *user* role was the one that obtained the largest number of respondents, followed by *students* with 20%, *teachers* with 16%, *designers* with 9%, and finally *developers* with 8%.

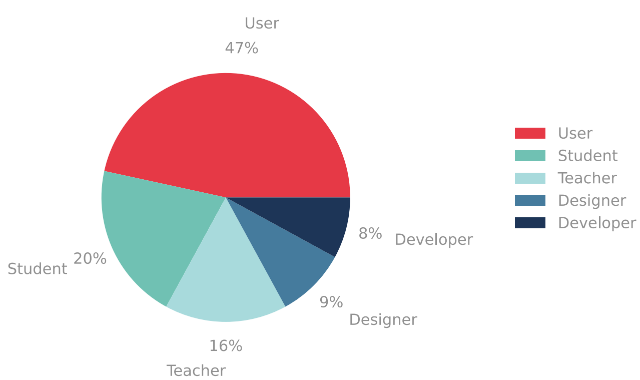

Figure 2: Self-expressed VR role

The fact that the survey was initially carried out during a scientific conference related to virtual reality and co-located with the *Conference on Graphics, Patterns and Images* (SIBGRAPI) and *Brazilian Symposium on Games and Digital Entertainment* (SBGames) meant that we were able to obtain a wide range of different roles, which is a positive factor for the purpose of this study as it enables us to have different views

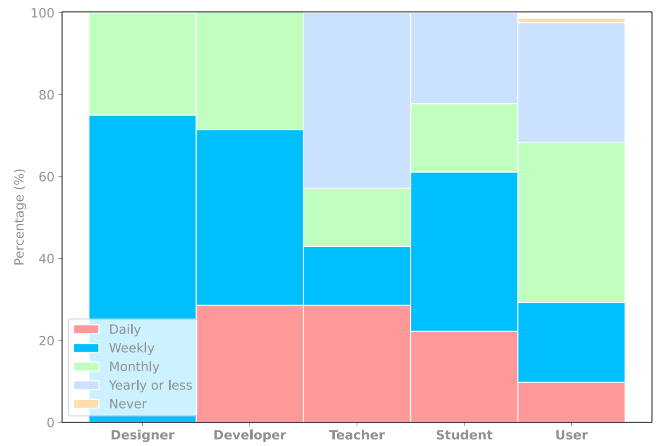

Figure 3: Comparison on the responder's role and frequency of use

about the investigated topic that can help reduce the bias of the conclusions obtained.

After identifying the respondents' role, our aim was to find out how often they use applications that include some sort of virtual reality resources. The rationale behind this question helps us understand respondents' behavior, especially as the role of each respondent is linked to the frequency of use. Therefore, we asked *"How often do you use virtual reality applications?"*.

The results demonstrate that there is no consensus regarding the frequency of use of this type of application, although in general the tendency to use it is higher as the combination of results presented in daily and weekly use represent over 45% of participants. However, there is still a large portion that makes use of it infrequently, which fits in the quadrant annually or less, which corresponds to 25% of respondents. Capturing the frequency of VR usage of the stakeholders is important because it allows us to establish some interesting associations. For example, between the role and the habit of use (Figure 3).

When verifying the association between the role of the respondents and the frequency of use of VR applications, it can be observed that there is a trend in the usage growth of this type of technology, regardless of the role analyzed as most of the public points out usage that mostly varies on a daily and monthly basis.

The increase in the use of this type of technology is a trend because, as pointed out by Boas et. al. [14], the use of technologies related to VR is increasingly and it may affect every field in which computers are involved.

Stakeholders tend to have many characteristics in common, yet there is a great discrepancy as to the purposes for which they use VR applications. Thus, we also asked respondents the following question about their habits of use so that we could characterize the major areas of use: *"What type of VR applications do you use most?"*.

Analyzing the Figure 4, it can be observed that the main use of VR applications is for entertainment, in which games were the most common answer among respondents (66% of responses). In addition to games, applications related to media (11%), simulation (11%), productivity (6%), modeling (3%) and therapy (2%) were also pointed out.

Milgram [15] proposed the idea of *Reality-Virtuality Continuum* in which reality is situated on one side and virtual environments entirely generated by computer or VR are located on the opposite side. Moving from the reality side to the VR side, users can experience aspects of augmented reality and mixed reality, until reaching total immersion in completely virtual environments. This vast spectrum of possibilities allowed the production of new experiences that increasingly explore the interaction between virtual environments and users, making virtual environments increasingly popular.

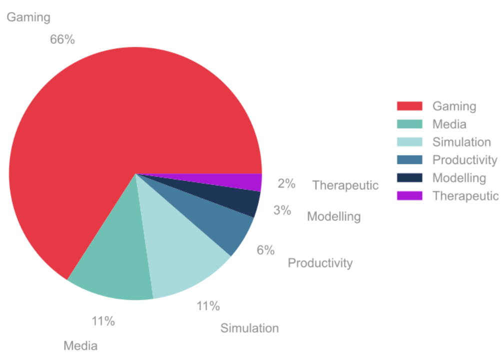

Figure 4: Main areas of VR applications use

Thus, aware that VR is a technology that is increasingly present in people's daily lives, to conclude the step of characterizing the respondents' role, we decided to ask them about their opinion concerning how much malfunction or bugs can contribute to a negative experience when using VR applications. To do so, we measured the respondents' opinions using a *Likert scale* [16] for the following statement: *"Malfunctions or bugs contribute to a negative experience using VR applications"*.

The results presented in the *Likert scale* shown in Figure 5 enable us to understand the perceived importance that respondents expect regarding the quality of the experience when using a VR application. Up to 70% of the respondents (69 in total) agree or strongly agree that malfunction or bugs contribute to a negative experience in VR applications.

Implicitly, these results point out that developers should pay attention to aspects of the quality of the developed product as this is a factor that, from the stakeholder's point of view, can have a strong impact on the experience during use.

Thus, we move on to the second part of our questionnaire, whose main objective was to understand the respondents' points of view regarding the fault aspects that are mostly annoying in VR applications.

B. Software feature concerns

Our main objective in this analysis was to measure the perception of the stakeholders regarding the frustration regarding some specific characteristics that make up VR applications.

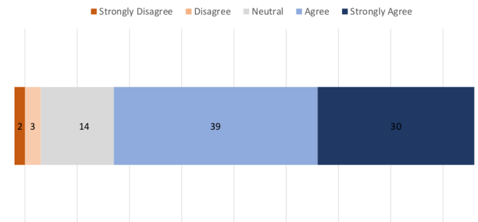

Figure 5: Stakeholder's perceptions (in absolute numbers) regarding bugs in VR application experiences

To do so, we split the major concepts that constitute a VR application into features, based on existing architectural classification for VR [17] and characteristics observed in 3D game development models [18]. The categorization and problems that are connected to it include:

- Visual aspects: are related to faults that affect the VR graphics in general. They can range from small flickering to serious texture issues.
- Audio aspects: these can range from volume drop differences to major audio failures.
- Design aspects: when a scene is poorly constructed, which might end up with problems such as holes in the ground or invisible walls, etc.
- Artificial intelligence aspects: mostly related to actions performed by actors in the scene. They can generally lead to problems such as anomalous behaviors or *pathfinding* algorithm issues.
- Physics aspects: despite being normally handled by the engines, problems in this category may include problems with breakables objects and general dynamic behaviors.
- Stability aspects: these include characteristics related to problems such as non-intentional freezes, crashes, and loading.
- Performance aspects: they refer to the speed with which the hardware processes the code and include problems related to the scene frame rate, load time and minimum device requirements.
- Networking aspects: these are specifically related to server-client connectivity and occur mostly in applications that make use of this type of feature.

After clarifying to the respondents about the concept for each of the characteristics described above, they were asked to rate, on a substantial severity scale how much the faults related to each of the concepts negatively impact the experiences when using VR applications: *"Which of the following affects usability, annoys or frustrates your experience using VR applications?"*

The results presented in Figure 6 show that there is no absolute consensus in the respondents' opinion. The scale represented on the left side of the graph defines the degree of importance for each of the major concepts evaluated. Value 1 represents a *Very Low* perception, 2 a *Low* perception, 3 *Neutral* perception, 4 *High* perception, and, finally, 5 a *Very High* perception.

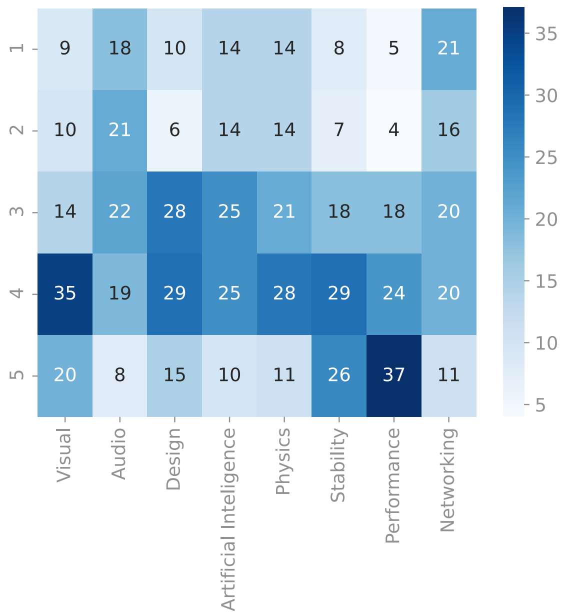

Figure 6: Heatmap of usability/usage aspects that annoy or frustrate the experience in VR apps

It can be observed that, in general, there is a similar degree of relevance for all the aspects evaluated, as the *Low* and *Very Low* perceptions were the least marked in almost all categories.

The category that presented the least perception of importance was related to *Networking* resources. A possible explanation for this result may be the fact that not all types of VR applications make use of networking resources in its working mechanics. Therefore, a respondent who uses this type of application may judge that this is not an important feature to be considered.

On the other hand, aspects such as *Design*, *Visual*, *Physics* and *Performance* were the ones that received the most votes as features that can cause great frustration or discomfort when presenting problems when using VR applications. Unlike networking features, these are features are present in most VR applications, therefore they are considered fundamental for the user to have an adequate experience.

Regarding *Visual* aspects, the results are in line with other studies that assess the impact and appeal that visual aspects have in VR applications. Marchiori et al. [19] demonstrated in an experiment, by checking vital signs of a group of users, that visual aspects are essential factors that contribute to the user of a VR application to feel really immersed and to enjoy an experience as expected. Thus, the perception of how problems of this nature are reflected in the user experience is a key factor for the success of an application.

Normally, a *Physics* simulation is responsible for providing the dynamic behaviors and collision detection to virtual objects in virtual environments in order to simulate a real world. Thus, it plays a vital role both to ensure that the application works correctly and in the user's perception which is related to the degree of user satisfaction when interacting with the application.

Finally, disorders related to poor *Performance* while running a VR application can cause cybersickness, such as nausea, disorientation, headaches, sweating, and eye strain [20]. Cybersickness symptoms might cause discomfort and consequently disrupt the experience of the user. Thus, in order to provide a pleasant experience in a VR application, performance, which is part of the list of non-functional requirements of the application, is a factor that needs to be taken in consideration.

C. Hardware feature concerns

VR often relies on specialized hardware and similarly to software aspects, its interaction with an application is also subject to faults. The main devices that can comprise an experience in VR can include a head-mounted display, which connects to a personal computer for a powerful immersive VR experience. Other hardware peripherals such as haptic, motion controllers, mouse and joysticks are often used to allow real-time interaction in addition to other types of peripherals to handle tracking, capture gesture or voice [21].

The next point investigated in the survey was related to the experience of using VR devices, assessing how much faults related to such devices can affect the user experience: *To what extent do you agree that the following IMPEDE your ability to properly use and enjoy VR experiences?* .

Specifically, we delimit seven distinct aspects related to the capabilities of hardware that are commonly used in VR applications [21]:

- Technical capabilities of the application **(A1)**: this type of capacity is related to aspects concerning components and sensors that make up the hardware used in VR experiments. For example, camera-based tracking, gyroscopes, accelerometers, and even magnetometers that are responsible for functions such as object orientation, head tracking and rotational tracking.
- Nausea when using the application **(A2)**: although there is no consensus regarding the definitive factor for these symptoms, there are various technical aspects of VR that can induce nausea sickness, such as mismatched motion, field of view, motion parallax, and viewing angle. Additionally, the amount of time spent on virtual reality can exacerbate symptoms.
- Fidelity of the virtual world **(A3)**: this deals with aspects that specify what the "visual display" should show the user. It covers perception of the distance of objects from the user's eyes and is also related to the perception of object scale and how we perceive motion which are pillars that make up characteristics of sensation and perception.
- Presence in the real world **(A4)**: although it is not a feature that is essentially related to the VR experience, the fact that this specific type of application, in some cases, provides a higher degree of immersion which excludes the user's perception of aspects related to the real world is a matter of concern about who is there when the user is experiencing a fully immersive application.

- Interacting with real world objects **(A5)**: VR literature points out that it is still not possible to reconstruct a virtual environment that is fully identical to the real world. Although it is possible to build photorealistic environments, interactions on the other hand are tricky depending on the level of dexterity required. An example is the replication of perception (such as weight and physics) of real objects. Thus, there are studies that evaluate how it is possible to interact within virtual environments together with objects arranged in the real world and how it can impact the user experience.
- Interacting with real world peripherals **(A6)**: this aspect is intrinsically linked to the communication of VR applications with hardware devices (keyboard, mouse, motion controllers) used to provide inputs and how these inputs are interpreted by the applications.
- Providing input to the virtual world **(A7)**: this is a complementary aspect to that of Interacting with real world peripherals, but deals with other types of devices, such as position trackers, gesture or voice captors.

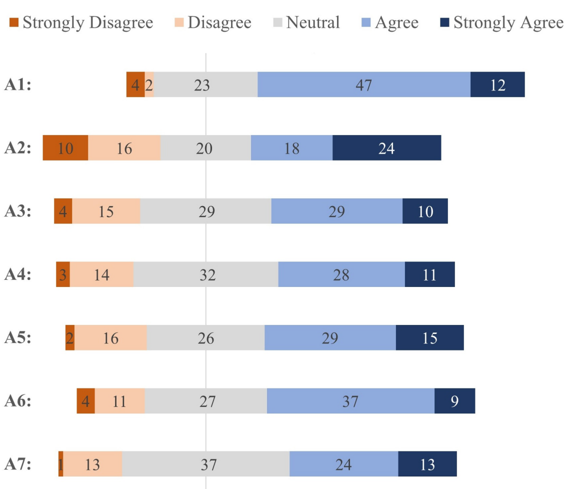

Figure 7: Hardware usage aspects that hampers or frustrate the experience in VR apps

The *Likert scale* was used to rate the compliance among the respondents concerning agreeing or disagreeing with the statement related to the hardware aspects that impede the ability to properly use and enjoy VR experiences. Figure 7 reproduces the perception of stakeholders in relation to the seven evaluated aspects. In general, it can be observed that there is a high tendency towards neutral, tending to a perception of agreement that the evaluated aspects can actually prevent or negatively impact the experience in VR applications.

Although the vast majority of aspects presented a similar trend, it is interesting to observe that the **A1** (technical capabilities of the application) aspect showed the highest margin of agreement, that is, most respondents agreed that when such aspects do not work properly, it tends to impede the the use of VR applications negatively impacting the user experience.

There is a range of factors that can impact sensations of cybersickness and as presented by Rebenitsch and Owen [22], theoretical foundations to unify the results are still lacking. Moving to the software testing field, approaches that explore the result of adoptions from different fields of view can be an alternative to control and partially decrease the effects of cybersickness, as well as limit the degrees of freedom in the navigation mechanisms. Another possibility is to investigate the variation of speed when users navigate a scene, because although there is no absolute consensus, this is a characteristic that is constantly related to cybersickness [23].

From the point of view of fidelity to the real world, a balance is needed between aspects such as processing capacity and development time as more realistic experiences normally require more complexity, and consequently a higher production time. The act of producing more complex applications also directly impacts the cost of applying the software test as the cost of the test activity is directly linked to the complexity of the software to be tested [24].

Regarding the presence in the real world, it is a factor that can sometimes be left out as it deals with characteristics that are inverse to immersion, which is one of the pillars of virtual reality. However, with the experimentation of new possibilities of interaction, there is also an increase in the risks of accidents involving the use of VR devices [25], thus it is a factor that cannot be disregarded. Aspects related to interaction with real objects in conjunction with VR experiences can increase the degree of complexity of the application, but must be balanced, as it can increase risks of causing accidents as the application must be prepared to identify real objects and how to react to them.

Finally, the ability to provide inputs in real time and how the VR application reacts to this is a key factor, because together with immersion, real-time interaction is one of the main characteristics of VR and must be taken into account when evaluating aspects related to quality.

Although interaction and providing input techniques have been widely used in VR applications, they still exhibit some limitations and can be the source of faults. Some of the limitations are related to response time and visual feedback issues [26]. Therefore, software testing techniques can exploit this space in order to ensure that applications can deliver solutions that present precise techniques to select objects in 3D scenes, which guarantee appropriate visual feedback, among others.

In the following section, we focused on cataloging the knowledge of stakeholders regarding the most common types of faults that could occur in VR applications.

D. *Fault perception*

The different types of devices and technologies used to provide the feeling of immersion in VR experiences can directly influence the universe of possible faults that can occur in an VR application. As applications reach a degree of maturity and development, capable of creating increasingly realistic environments, there is also a tendency to increase complexity. Consequently, the need for quality is one aspects that must be addressed.

Driven by those thoughts, we decided to investigate the stakeholders' knowledge regarding the main types of faults that may occur in the universe of VR applications. The main aim of this stage is to know from stakeholders how much they recognize about typical types of faults that can occur in VR applications. To do so, we categorized the main types of faults and asked the stakeholders if they were aware of that particular type of fault and if it was considered as critical.

We divided the most common faults into groups (visual, audio, design, artificial intelligence, physics, stability, performance and networking), using the same classification that was adopted in the questions investigated in subsection IV-B. Based on that, we asked the respondents: *"Regarding the types of faults listed below, check all the types you know and think are critical"*.

As in the previous sections, we provided a document to the respondents in order to clarify and assist them in the process of understanding the meaning of each of the categories and the type of the fault characteristics presented in the survey. Table I presents a compilation of the results for this question. We divided the responses between the classes of characteristics evaluated, using the following pattern: the column "Fault characteristics" represents the type of fault specified, the column "Respondents" identifies the total number of respondents who were aware of the existence of that fault and judge it to be critical and, finally, the column "%" presents the percentage according to the total number of respondents.

Regarding faults related to audio aspects, it can be observed that between 30% up to 50% of the respondents pointed out some characteristics that can be judged as critical. Faults related to audio aspects are usually linked to problems when creating the content, which are often only noticed when the content is used. Thus, the problems normally appear when implementing actions that should trigger an audio as a consequence, besides other possible problems are related to checking whether the sounds are in sync with the animation it is correlated.

The fault characteristic that had the highest number of votes in the survey was related to *distortions* in the audio. Problems related to distortion can be linked to the microphone used in the audio reproduction, to the volume control of the application (when the quality of the media is relatively low), or even a misinterpretation of the problem, which may originate in cable connections or volume buttons.

Artificial intelligence is an extensive area of study that requires investigation by itself. However, since the use of AI mechanics has become increasingly ubiquitous in all areas of technology, we chose to analyze two specific aspects in the VR context. The two characteristics evaluated involve mainly *pathfinding* situations (i.e. problems related to navigating from source to destination in paths modeled in the scene) and problems in the scene behaviors (when a given user action does not result in the reaction expected by the scene).

About 75.3% of respondents pointed out that the inability to continue a movement (*Stuck*) is considered a serious fault. In fact, this is a decisive factor, especially in VR applications that have a strong dependence on this aspect so that the flow of the scene can happen according to the user's interactions with the application (commonly present in games and training applications).

Table I: Respondents' responses to the most critical types of faults

Fault characteristics	**Respondents**	**%**
Audio		
Audio drops	31 / 88	35.2%
Skiping	36 / 88	40.9%
Distortion	45 / 88	51.1%
Missing sound fx	38 / 88	43.3%
Volume too high/low	30 / 88	34.1%
Artificial Intelligence		
Does not move	39 / 88	45.9%
Stuck (unable to move)	64 / 88	75.3%
Scene Design		
Stuck spot	26 / 88	29.5%
Stick Spot	16 / 88	18.2%
Scene hole	37 / 88	42%
Invisible obstacles	47 / 88	53.4%
Missing geometry	43 / 88	48.9%
Networking		
Lag	66 / 88	76.7%
Scoring errors	31 / 88	36%
Invisible players	14 / 88	16.3%
Cannot connect / Drop connection	42 / 88	48.8%
Performance		
Low frame rate	59 / 88	67.8%
Higher loading time	35 / 88	40.2%
High minimum requirements	20 / 88	23%
Physics		
Object do not break	31 / 88	35.2%
Objects floating abnormally	35 / 88	39.8%
Problems interacting with objects	58 / 88	65.9%
Unrealistic gravity	36 / 88	40.9%
Impossible to pile objects	14 / 88	15.9%
Stability		
Freeze	58 / 88	65.9%
Crash	56 / 88	63.6%
Cannot load the app	40 / 88	45.5%
Unresponsive	37 / 88	42%
Visual		
Clipping	34 / 88	39.1%
Z-Fighting	18 / 88	20.7%
Screen tearing	38 / 88	43.7%
Missing textures	51 / 88	58.6%
Vsible artifacts	28 / 88	32.2%

Scene design faults are linked to aspects such as projection errors. When some things in a scene do not have a coherent semantics, those faults include a user being stuck in some space of the scene (stuck spot), or when there is not the fluidity needed to move in some space within the scene (stick spot), in addition to scene projection problems in general (which include the categories of scene hole, invisible obstacle and missing geometry).

For this specific category of faults, the results were mixed. Few users indicated low criticality for elements such as stick spot (18.2%), problems related to the projection of the scene received greater attention, ranging from 42% to 53.4% of the respondents. The question of visual aspects having a greater weight for this characteristic of faults reinforces the perception that was observed in subsection IV-B, in which a large part of the respondents pointed out visual aspects as being factors that can negatively impact the users' experience.

Some aspects related to networking may not be directly linked to the application, since applications normally expect connectivity to perform some specific type of task. However, understanding which points the stakeholders judge as critical can be an interesting factor so that applications can provide mechanisms to become more resilient when dealing with these type of faults.

The main characteristic of faults pointed out by stakeholders was the presence of lag (76.6%). The term *lag* is associated with the loss of packets over a connection, or the excessive use of the bandwidth. Packet loss can be a problem related to the internet service provider, but the excessive use of bandwidth can be correlated with the way the application communicates with a server in scenarios where an application needs to consume some service to work properly.

Performance issues are usually associated with hardware but they can also hide other problems, such as the lack of optimization of the application's computational process. Both facts impact the points investigated in the survey category.

The most remembered factor in the survey was problem related to a *low frame rate* (67.8%). When designing a VR scene, developers usually set a minimum frame rate. Not being able to achieve this frame rate in a scene implies that there is much information in the scene and the hardware responsible for reproducing it is not able to meet the demand, or there are processing bottlenecks and it is necessary to perform profiling tasks to analysis and understand how the functions that compose the application are performing in order to optimize those that have a higher computational cost.

Regarding fault characteristics related to Physics, there is a great challenge, since currently most VR experiences exploit engines that provide to the developer all the features that involve a simulation of objects that move, collide and interact under various physical forces, such as gravity. Thus, there must be a concern to use a reliable engine and focus the testing aspects from an *integration testing* [27] point of view focusing on how the application communicates and makes use of the features provided by the engine.

The characteristics revealed in our survey point to possible faults that encompass all aspects of physics in VR applications, since it is a very important subsystem, producing impacts on both usability and animation in scenes. Among the characteristics observed, the most impacting according to stakeholders was that of *problem interacting with objects*, emphasized by 65.9% of respondents. Problems related to interaction with objects may also be related to a peripheral that supports motion capture, such as gloves, or how the communication between models and the objects is being carried out. A clear understanding of how the physical interaction model behaves is the key to dealing with these possible type of faults.

The concept of stability explored in the survey refers to the predictability of the application, i.e., whether it is able to behave or not according to the developers' intentions. Therefore, in order to assess how much the stakeholders are aware of the flaws involved in this category, we evaluated characteristics such as *freezes*, *crash* and *loading* problems.

All the characteristics achieved a high degree of attention from respondents, ranging from 42% up to 65.9%, with the two most evaluated faults were related to *crash* and *freeze* (respectively 63.6% and 65.9%). Crash is considered as extremely serious faults, as it can affect the execution of the application. It is usually linked with the triggering of an exception that was not properly handled by the application and that ends up reaching the most superficial layer of the application, causing an unexpected shutdown, while freeze can be linked to the loss of input information or how the application handles the inputs made by the user and reflects on its behavior in the scene.

Finally, we verify the impact of faults related to possible visual aspects. Visual faults may not necessarily be critical for the application to function properly, but as VR applications have a high appeal in faithfully reproducing aspects of the real world in a virtual world, observing possible visual problems becomes a task that can determine the success of a VR application. Therefore, we have listed five common types of faults that can be linked to visual aspects and that can avoid the risks exposed above.

The two characteristics most mentioned by stakeholders were *screen tearing* and *missing texture*, with 43.7% and 58.6%, respectively. Screen tearing is a visual fault that occurs when there is an inability to project a frame on the screen correctly, usually this is a fault that can be related to performance problems (this result shows a correlation in the respondents' perception, since faults related to performance also received high attention from respondents). Missing texture are problems related to the way a program deals with the application of textures on objects that make up a scene, which can result in the appearance of flat surfaces or simply placeholders.

V. Discussion

The results of subsections IV-A and IV-B provide us some evidence about the answer to the first research question (RQ_1) defined by this study that investigated the perception of stakeholders regarding the need to prioritize software testing efforts in specific characteristics of VR applications. In general, the stakeholders understand the importance of the need for software testing practices in VR applications and provided feedback about visual aspects, physics and performance as the most critical type of faults.

Although VR is still a relatively new frontier, researchers are developing new testing strategies. One of the major concerns pointed by the stakeholders is related to performance issues of VR application. The review conducted by Chandra et al. presented a list of aspects related to the performance of VR applications [28] that could be addressed.

To deal with visual aspects, the main strategy used is to work with user specifications and exploit the experience of the testing team. The process is normally driven by the product's scope requirements. This allows the testing team to understand the potential scenarios for user engagement. Thus, the results of this work can provide one more aspect to be taken into account during this process.

However, considering that the development time and, consequently, the time that can be applied to the software testing activity, is limited, it is necessary to understand critical points that can be used to prioritize activities on a scale of importance.

The second research question (RQ_2) raised by this study focused on understanding the perception of stakeholders regarding fault characteristics in VR hardware aspects. The results presented in the subsection IV-C indicated relevance for all aspects evaluated, with a special emphasis on features linked to sensors that are responsible for complementing it as a VR application when measuring movement and direction in space and devices connected to data entry tasks.

There are physical consequences to wearing VR devices, such as headaches, motion sickness, and eye strain. A big challenge is to find a way to minimize user fatigue and discomfort while using an application as much as possible. To deal with this type of problem, a great effort is made in accessibility tests, aiming to ensure that there are no accessibility problems for users with visual, physical, hearing or cognitive disabilities. Since VR devices and applications are still in early stages, there are currently no safety solutions for developers to follow.

The fact that even small types of faults are important in VR applications led us to seek to understand the view of stakeholders on the various types of faults that can affect such an application. The results indicated in subsection IV-D provide an answer to the last research question (RQ_3) of our study that sought to understand *"What are the main types of faults that stakeholders are aware of in VR?"* We understood that not all stakeholders are aware of the different types of faults that can affect VR applications, which shows a risk, because, by not considering these risks, it can cause them to be overlooked involuntarily.

However, the results of the RQ_3 allowed us to better understand how stakeholders assess specific types of faults in relation to their criticality, in addition to discussing how each of the most pointed ones is triggered in the context of VR applications.

This information can be an auxiliary factor in the context of the software testing point of view, as this data can help the decision-making process with regards to prioritizing the validation of specific points of an application, allowing for balance strategies, focusing on equalizing the needs of the project, taking into account both the point of view of a quality team and that of the rest of the team, including end users.

The results obtained with the development of the survey are in line with studies that investigate the need for specific software development practices in VR context [9] besides it also point out directions for the development of specific software testing research applied to the virtual reality context, evidencing a series of opportunities that involve the development of research that encompasses joint aspects of software engineering and virtual reality.

VI. Threats to Validity

This work is subject to the threats to the validity reported for this type of studies, thus the results must be interpreted carefully. We discuss the validity concerns based on classification proposed by Wohlin et al. [29].

Threats to internal validity are related to the diversity and representativeness of the samples collected. To avoid this threat we initially distributed the survey in a scientific conference of VR specialists, enabling us to obtain responses from a wide range of profiles that comprise the VR area and later we complemented the study distributing it online in order to reach an audience of technology users. Thus, converging to results that can take into account all the profiles involved in the context of this type of applications.

Construct validity addresses problems about how well whatever is intended to measure has actually been measured. Regarding this, all the data presented in this study correspond to the aggregated assessment of all responses. We also seek to evaluate the questions about different types of graphs and discuss what possible consequences of the results presented. To boost confidence, the data were analyzed not only in tables and graphs, but we also made extra analyses available in the survey repository*.

In order to guarantee the reliability of the survey (conclusion validity), we carefully followed the survey guidelines process proposed by Linaakeer et al. [11]. As well as providing all the resources and data available in the study repository to enable the third party data review, all the results were interpreted by all authors to avoid interpretation errors.

To avoid external validity, we reinforce that the survey reflects the stakeholders' interpretation of importance and necessity of software testing in VR applications. The answers may not necessarily represent an accurate picture of reality and, therefore, may present some degree of subjectivity. Aspects such as experience, as well as the degree of contact with VR technology may have influenced the responses. To avoid this threat, we tried to interview as many stakeholders as possible and seek to avoid data correlation analyses to avoid an interpretive bias from the researchers' point of view.

VII. Conclusions and Future Works

The main characteristics and types of faults used to understand the perception of stakeholders were extracted from books and articles related to faults [18] and architecture of VR applications [17]. The study made it possible to understand the opinion of a group of five different roles, with different habits and focuses on using VR applications. The results showed that all roles agree that the presence of faults in VR applications can lead to a negative experience in their use, implying the need for attention of software testing practices.

*https://github.com/stevao-andrade/vr_survey

Among the seven main characteristics of a VR application, visual aspects and non-functional requirements, such as stability and performance, were identified as the aspects that most negatively impact the users' experience. We also observed hardware characteristics. The seven points analyzed explore mainly important aspects for the use of VR applications, highlighting points related to the sensors, due to the fact that it is an essential part necessary to provide the sensation of immersion to the user and interaction with other devices that allow the user to provide inputs for applications.

Regarding perception related to faults, in general, stakeholders were aware of the most common types of faults for each of the eight categories evaluated. The main faults pointed out were related to artificial intelligence, visual aspects and characteristics linked to non-functional requirements. This result proved to be in agreement with the general perception assessed at the beginning of the research.

The results of this work enable us to draw conclusions from different points of view: looking from the perspective of the quality assurance team responsible for the development of VR applications, the results allow us to understand how to balance the activities to reach both the demand of the project development and the expectations of the players involved in the development and use of VR applications. From the point of view of a researcher, it allows enthusiasts to develop new techniques and test approaches to direct their efforts to aspects considered important by the players involved in the universe of a VR application, enabling future research with less bias from the point of view of the researcher. The results are also more aligned with the real needs of the field of application.

As future work, we intend to extend the evaluation of this research, collecting a greater number of opinions from stakeholders, so that we can have a more accurate view of the needs of each one, through an individual analysis of each profile.

Acknowledgements

Stevão A. Andrade research was funded by *São Paulo Research Foundation* (FAPESP, Portuguese: Fundação de Amparo a Pesquisa do Estado de São Paulo), process number 2017/19492-1. The authors also thank the support of FAPESP process number 2019/06937-0, *Brazilian National Council for Scientific and Technological Development* (CNPq, Portuguese: Conselho Nacional de Desenvolvimento Científico e Tecnológico) process number 308615/2018-2, *National Institute of Science and Technology Medicine Assisted by Scientific Computing* (INCT MACC) process number 157535/2017-7, and *Natural Sciences and Engineering Research Council of Canada* (NSERC) Discovery grant RGPIN-2018-0591.

References

[1] O. Sievi-Korte, S. Beecham, and I. Richardson, "Challenges and recommended practices for software architecting in global software development," *Information and Software Technology*, vol. 106, pp. 234–253, 2019.

[2] R. Pressman and B. Maxim, *Software Engineering: A Practitioner's Approach (Eighth Edition, English Abridgement)*. Ji xie gong ye chu ban she, 2016.

[3] J. Scholtz and E. Morse, "Using consumer demands to bridge the gap between software engineering and usability engineering," *Software Process: Improvement and Practice*, vol. 8, no. 2, pp. 89–98, 2003.

[4] IEEE Standards Association, "P2048.4 - standard for virtual reality and augmented reality: Person identity," 2020, available from: https://standards.ieee.org/project/2048_4.html.

[5] S. A. Andrade, F. L. S. Nunes, and M. E. Delamaro, "Towards the systematic testing of virtual reality programs," in *2019 21st Symposium on Virtual and Augmented Reality (SVR)*, 2019, pp. 196–205.

[6] X. Y. Zhao, L. M. Xu, and H. Li, "Software testing applications based on a virtual reality system," *Journal of Electronic Science and Technology*, vol. 5, pp. 120–124, 2007.

[7] A. Bezerra, M. Delamaro, and F. Nunes, "Definition of test criteria based on the scene graph for vr applications," in *Virtual Reality (SVR), 2011 XIII Symposium on*, May 2011, pp. 56–65.

[8] A. C. S. Corrêa, F. L. S. Nunes, and M. E. Delamaro, "An automated functional testing approach for virtual reality applications," *Software Testing, Verification and Reliability*, vol. 28, no. 8, p. e1690, 2018.

[9] U. Güleç, M. Yilmaz, and V. Isler, "A literature survey: Is it necessary to develop a new software development methodology for virtual reality projects?" *J. UCS*, vol. 23, no. 8, pp. 725–754, 2017.

[10] M. Noghabaei, A. Heydarian, V. Balali, and K. Han, "A survey study to understand industry vision for virtual and augmented reality applications in design and construction," *arXiv preprint arXiv:2005.02795*, 2020.

[11] J. Linåker, S. M. Sulaman, R. Maiani de Mello, and M. Höst, "Guidelines for conducting surveys in software engineering," 2015.

[12] V. R. Basili, "Goal question metric paradigm," *Encyclopedia of Software Engineering*, pp. 528–532, 1994.

[13] "Understanding vr software testing needs from stakeholders' points of view," 2019, available from: https://forms.gle/Jmc5NoEHoz3Kz9ov5.

[14] Y. Boas, "Overview of virtual reality technologies," in *Interactive Multimedia Conference*, vol. 2013, 2013.

[15] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, "Augmented reality: A class of displays on the reality-virtuality continuum," in *Telemanipulator and telepresence technologies*, vol. 2351. International Society for Optics and Photonics, 1995, pp. 282–292.

[16] G. Albaum, "The likert scale revisited," *Market Research Society. Journal.*, vol. 39, no. 2, pp. 1–21, 1997.

[17] R. Capilla and M. Martínez, "Software architectures for designing virtual reality applications," in *European Workshop on Software Architecture*. Springer, 2004, pp. 135–147.

[18] L. Levy and J. Novak, *Game development essentials: Game QA & testing*. Cengage Learning, 2009.

[19] E. Marchiori, E. Niforatos, and L. Preto, "Analysis of users' heart rate data and self-reported perceptions to understand effective virtual reality characteristics," *Information Technology & Tourism*, vol. 18, no. 1-4, pp. 133–155, 2018.

[20] S. Davis, K. Nesbitt, and E. Nalivaiko, "A systematic review of cybersickness," in *Proceedings of the 2014 Conference on Interactive Entertainment*, 2014, pp. 1–9.

[21] S. M. LaValle, *Virtual Reality*. Cambridge University Press, 2019.

[22] L. Rebenitsch and C. Owen, "Review on cybersickness in applications and visual displays," *Virtual Reality*, vol. 20, no. 2, pp. 101–125, 2016.

[23] R. H. So, W. Lo, and A. T. Ho, "Effects of navigation speed on motion sickness caused by an immersive virtual environment," *Human factors*, vol. 43, no. 3, pp. 452–461, 2001.

[24] B. Keogh, "Between triple-a, indie, casual, and diy," *The Routledge companion to the cultural industries*, pp. 152–162, 2015.

[25] O. S. University, "Virtual reality, real injuries: How to reduce physical risk in vr," 2020, available from: www.sciencedaily.com/releases/2020/01/200108092448.htm.

[26] F. Argelaguet and C. Andujar, "A survey of 3d object selection techniques for virtual environments," *Computers & Graphics*, vol. 37, no. 3, pp. 121 – 136, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0097849312001793

[27] P. Ammann and J. Offutt, *Introduction to software testing*. Cambridge University Press, 2016.

[28] A. N. R. Chandra, F. El Jamiy, and H. Reza, "A review on usability and performance evaluation in virtual reality systems," in *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 2019, pp. 1107–1114.

[29] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in software engineering*. Springer Science & Business Media, 2012.