https://doi.org/10.2352/EI.2024.36.9.IQSP-267

Quality of Experience of Virtual Reality-Based Communication Applications

Mylene C.Q. Farias; Texas State University; San Marcos, TX

Abstract

In this paper, I present the proposal of a virtual reality subjective experiment to be performed at Texas State University, which is part of the VQEG-IMG test plan for the definition of a new recommendation for subjective assessment of eXtended Reality (XR) communications (work item ITU-T P.IXC). More specifically, I discuss the challenges of estimating the user quality of experience (QoE) for immersive applications and detail the VQEG-IMG test plan tasks for XR subjective QoE assessment. I also describe the experimental choices of the audio-visual experiment to be performed at Texas State University, which has the goal of comparing two possible scenarios for teleconference meetings: a virtual reality representation and a realistic representation.

Introduction

In the past decade, significant progress in imaging and computing technologies has unleashed the creation of more accurate representations of the physical world, enabling immersive and interactive experiences that more closely resemble reality [1]. As a consequence, immersive applications have attracted a lot of interdisciplinary attention, with the development of multimodal human-computer interactions that allow users to fully immerse themselves within realistic or virtual environments or to interact with real/virtual elements seamlessly blended with the physical world [2]. Immersive experiences can be found at different points along the "virtuality continuum," ranging from photorealistic settings, through mixed reality approaches, to completely virtual environments [3]. This continuum is illustrated in Figure 1.

In this scenario, high-speed connections and high-quality imaging systems are essential for the development of next-generation immersive experiences and applications in areas such as healthcare, education/training, arts & entertainment, remote work, marketing, and automotive. But the success of these emerging applications will depend on the quality of experience (QoE) they provide users [4]. As defined by Qualinet [5], QoE refers to the "degree of delight or annoyance of applications or services resulting from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the users personality and current state." For immersive technologies, immersive media experiences (IMEx) extend the concept of QoE by encompassing elements such as the sense of presence, immersion, and motion sickness, among others [6].

Both QoE and IMEx are shaped by three influencing factors (IF): the system, the context, and the human user. For immersive experiences, the system IF plays an important role, affecting immersion, presence, and interaction itself. This impact has been extensively studied in the literature, with several studies analyzing the effect of the design of the device itself on IMEx [7,8]. The context IF, which includes the user's environment, is often studied together with the system IF, since it is difficult to separate these two IFs. Finally, an important IF for immersive media are human factors, which include perceptual characteristics such as visual, audio, and spatial perception that make each user unique. It is well known that impairments may reduce perception and cause imbalances, resulting in uncomfortable symptoms and cybersickness. Subjective and instrumental assessments are commonly used to study human IFs in IMEx. In this context, an open area of research is the design of subjective and instrumental assessment methods to estimate the user immersive media experience.

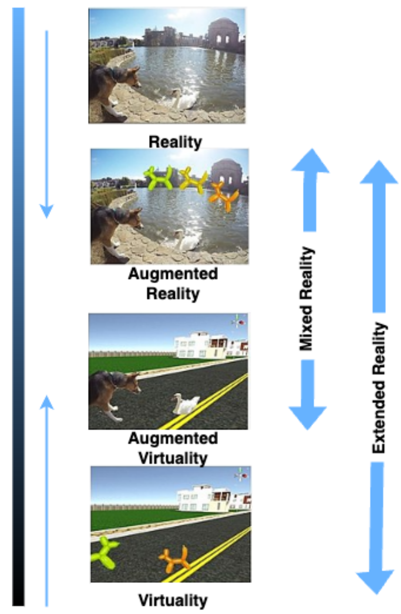

Figure 1: Extended reality scheme based on Paul Milgram's scheme [3]. Figure adapted from Xr4all, CC BY-SA 4.0, via Wikimedia Commons.

In this paper, I discuss the preparation studies and pilot tests performed with the goal of running a subjective experiment to analyze two possible scenarios for teleconference meetings: a realistic representation and a virtual reality representation. The complete experiments are being performed in the Summer of 2024 at Texas State Universtiy and are part of the VQEG-IMG[1] test plan[2] for the definition of a new recommendation for subjective assessment of eXtended Reality (XR) communications (work item ITU-T P.IXC). In this paper, I describe the VQEG-IMG test plan goals and the methodology of the experiments to be performed at Texas State University. I also discuss the pilot tests, the virtual environment, and experimental setup considered in this study.

The VQEG-IMG Test Plan

The purpose of the VQEG-IMG test protocol is to answer a particular research question: Do changes in a System Influencing Factor (independent variable) have an impact on a QoE Constituent (dependent variable) while users carry out a fundamental communication task within controlled circumstances? Therefore, two primary goals for the VQEG-IMG tests are:

1. To methodically regulate an independent variable and assess its influence on QoE; and
2. To evaluate entire black box systems without delving into individual variables.

As mentioned earlier, the test plan will result in the definition of a new recommendation for subjective assessment of eXtended Reality (XR) communications (work item ITU-T P.IXC). The Recommendation will encompass the following elements:

- Establish a uniform method for describing test conditions and variables.
- Normalize fundamental communication tasks (as outlined in the proposal below).
- Standardize a select set of measures applicable across various tests and systems.
- Provide suitable statistical analysis methods for describing measure outcomes and their implications.

The VQEG-IMG test plan includes test paradigms that estimate several aspects of QoE in XR communications. With this goal, the following measures are being considered:

- Task performance and conversation analysis, measured using objective measures with the goal of understanding the impact of system IFs. Examples of these measures include task success rate (e.g., number of guessed items per unit of time) and post-hoc analysis of conversations (e.g., turn taking, interruptions between speakers, requests for repetition, etc.).
- System performance questionnaires, like for example single-item or short questionnaires targeted to analyze the effect of a test condition in user quality (e.g. Absolute Category Rating, ACR, defined in ITU Recommendation P.913 [9]).
- Human factors/QoE constituents, such as simulator sickness, immersion, presence, among others. These measures explore the influence of the system in different QoE factors and depend on the specific research question.

To explore all these aspects, several types of tasks to be performed in the experiment have been envisioned:

- Audio-communication task: Survival Task;
- Visual-communication tasks: Charade / Physiotherapy;
- Object-based communication tasks: Block building; and
- Environment-based communication task: Treasure Hunt.

The Call for participation in VQEG-IMG test plan was released on May 12th, 2023. The goal was to look for laboratories to perform subjective assessment tests based on the protocol defined for the Recommendation, so that the method itself can be validated before proposing it to ITU-T. The deadline for proposals was on June 16, 2023. On June 29, 2023 the 12 proposals received were presented at VQEG Meeting at San Mateo, California, by VQEG-IMG chairs: Jesús Gutiérrez (Universidad Politécnica de Madrid, Spain) and Pablo Pérez (Nokia, Spain). The 12 proposals were from the following institutions, which included researchers from Spain, The Netherlands, Italy, Germany, Sweden, Poland, UK and US:

1. Universidad Politécnica de Madrid (Spain) - UPM
2. Nokia (Spain) - NOK
3. CWI (The Netherlands) - CWI
4. AGH University of Science and Technology (Poland) - AGH
5. University of Padova (Italy) - PAD
6. University of Roma 3 (Italy)- UR3
7. TU Ilmenau (Germany) - TUI
8. University of Surrey (UK) - UoS
9. University College London (UK) - UCL
10. Texas State University (USA) - TSU
11. Keysight Technologies & University of Malaga (Spain) – KUM
12. Lulea University of Technology (Sweden) - LUT

Table 1 shows the experimental tasks to be performed by each participant. The experiments will be performed during the 2024 year.

Audio-Visual Communication Tasks

In this section, we describe the audio-visual communication tasks, which are the target of our studies. The audio-visual experiments will be performed during the year of 2024. Texas State University is involved in this particular set of experiments, whose execution will be distributed over the year.

Survival Audio Task

The survival task, described in ITU-T Recommendation P.1301 [9], simulates a survival scenario where participants are given a number of items or objects. Survival tasks were originally devised to examine how decision-making groups perform under pressure. In the original version, participants are

[1] `https://vqeg.org/projects/immersive-media-group/`
[2] `https://www.vqeg.org/VQEGSharedFiles/MeetingFiles/2023_06_Sony_USA/VQEG_IMG_2023_125%20VQEG_IMG_WorkPlan_2023-06-v1.pdf`

Task	Labs
Audio Communication: Survival (conversation)	UPM, NOK, AGH, TUI, **TSU**, KUM
Visual Communication: Movement (body centered)	UPM, CWI, **TSU**, KUM, LUT
Object-based communication: Block Building (object centered)	UR3, UCL, LUT
Environment-based communication: Escape Room (environment oriented)	NOK, PAD

Table 1: List of Laboratories that answered the Call for participation in VQEG-IMG test plan.

asked to envision themselves in a survival scenario following an accident (like a plane or spacecraft crash). They are provided with a list of fifteen items that remain intact and undamaged post-landing. Initially, participants individually prioritize these items based on their perceived importance for the group survival. Then, they collaborate to establish a consensus ranking. Finally, participants reevaluate their rankings individually to compare them before and after the group discussion. This original version of the task typically lasts one to two hours.

To streamline the task, a modified version was proposed that consisted only of a group discussion aimed at selecting six essential objects for survival. Additionally, the original 15-item or 12-item list was subdivided into three 5-item or 4-item lists, one for each participant. This aims to prevent participants from being overwhelmed by a lengthy list (which might require repeated readings) and to ensure active participation from all members. Each 5-item or 4-item list was supplemented with photographs to aid participants in identifying uncommon objects and to expedite memory recall, thereby preventing prolonged pauses during the discussion. The task concludes once all participants agree upon and recapitulate the list of the six chosen objects. The Survival task offers the benefits of fostering natural turn-taking exchanges, accompanied by relatively concise written materials. ITU-T Recommendation P.1301 [9] details four survival scenarios: winter, at sea, on the moon, and in the desert.

In the proposed experiment, the simplified version will be used, with the task being divided into three stages:

1. First, each participant is allocated up to 5 minutes to get prepared, which includes analyzing the list of items, contemplating their significance, and formulating arguments for discussion.
2. Second, participants collaboratively construct an importance ranked list of these items during the discussion. In this stage, single measures for Quality of Experience (QoE) factors, such as ACR, are employed to provide a measure of IFs.
3. Participants complete final questionnaires to formalize their rankings.

These stages can be repeated for different scenarios and conditions.

Charade and Physiotherapy Visual Tasks

In the visual tasks, the aim is to have participants engage in a visually-driven task, where communication relies solely or predominantly on visual cues. Examples encompass employing sign language for conversation or leading a physical training regimen. Two subtasks are planned, in which the participant have different degrees of freedom (DoF) within the communication immersive space. The first task is the 3DoF Charade, while the second task it the 6DoF Physiotherapy.

The 3DoF Charade task involves participants playing the game Charade in a virtual environment. A charade game is a form of entertainment where participants act out a word or phrase without speaking, typically by using gestures, facial expressions, and body movements. The aim is for other players to guess the word or phrase correctly within a time limit. Charades often involve teams competing against each other, with one player from a team acting out the word or phrase while the rest of the team tries to guess it. It is a popular party game that encourages creativity, communication, and quick thinking. The implementation of this experiment will require at least two players, with one of the players acting out the words (Participant A) and the other trying to guess the words (participant B). Participant A will receive the list of words that Participant B must guess within a specified time frame. Participant A cannot offer verbal cues and must convey clues solely through physical gestures (mimicking). The game is finished when all words are guessed or when time expires. Success will be measured by the number of words guessed per unit of time.

The 6DoF Physiotherapy task involves a physiotherapy training session where physical movements are executed by participants. One of the participants acts as the instructor (Participant A) who will demonstrate a finite set of movements, while another that participant must accurately replicate the movement (Participant B). The session concludes when Participant A confirms Participant B has correctly executed all movements. Success is measured by the time taken for one move to be completed and by the number of moves completed per unit of time. It is worth noticing that since a 6 DoF virtual environment is being used, participants have unrestricted movement within the space. Also, in both cases, success is assessed by an objective metric: the frequency of task completions (words or movements) within a fixed time duration.

Visual Communication Setup for a Virtual Reality Environment

As mentioned earlier, the Intelligent Multimedia Processing for Advanced Computing Technologies (IMPACT) Laboratory in the Computer Science Department at Texas State University will be performing the 3DoF Charade task in a virtual reality environment. In this section, we describe the study performed to determine the adequate physical (hardware and software) setup and experimental methodology for the experiment. It is important to emphasize that the setup has to allow participants to mimic by using hand signs to convey the meaning of the words in the charade game.

Hardware and Software Setup

The experiment will be performed in a virtual reality environment where participants pick an avatar and are able to see other participants and part of their avatar's body. The charade task requires the use of Head Mounted Displays (HMDs) that provide 3 degrees-of-freedom (3 DoF) so that participants are able to move their upper body, including their heads and hands. More specifically, it is very important that, besides head tracking, HMDs also have hand tracking capabilities to allow participants to gesticulate and visualize the hands accurately. After analyzing some of the HMDs in the market, we decided to use Meta Quest2, which has eye, head, and hand-tracking capabilities and it is recognized by many VR meeting applications.

To avoid biases and lagging because of rendering, The VR environment should be a simple VR room where participants can move freely and see each other. In the original Call-for-Participation, Mozilla Hubs[3] was suggested as a VR environment for the experiment. The Mozilla Hubs platform was developed by Mozilla, the creators of the Firefox web browser. As depicted in Figure 2, it is a free platform that enables users to create and interact in VR spaces. It allows users to choose a pre-made avatar or to upload a customized one. Although the avatars can be full-body ones, including hands, after testing it with Quest2 we found out that that it did not recognize the HMD hand-tracking and, therefore, it could not be used for the charade task.

Figure 2: Mozilla Hubs Meeting application, available at `hubs.mozilla.com`.

Therefore, we looked at the current state-of-the-art of VR meeting applications to understand which ones could be used in our experiment. We selected and tested seven additional applications to experiment: Horizons, Spatial, RecRoom, AltSpaceVR, VRChat, Meeting VR, and GlueVR. All these applications are available for use for free. A detailed description of these applications was performed by Osborne *et al.* [10]. We tested them looking for the ones that accepted at least a half-body avatar representation with hands. Unfortunately, some applications do not support hand-tracking or have limited hand-tracking capabilities showing movements of only a few fingers of the hands (e.g. Spatial and RecRoom). After our tests, we chose Meta Horizons and VRChat as our VR meeting platforms.

[3] `https://hubs.mozilla.com/`

Horizons is a free virtual reality social platform developed by Meta (formerly known as Facebook). It was designed as a space where users can interact, socialize, play games, and create content in VR. Figure 3 shows examples of working virtual environments in Horizons. Notice that in these environments, which are often meant for work collaboration, users are represented by half-body with hands avatars. Similarly to other platforms, in Horizons, users can either pick a pre-made avatar, customize an avatar, or upload an avatar created using a specific software. The half-body with hands avatar in Horizons shows the hand movements as long as the hands are in front of the HMD visor. Therefore, while playing a charade game in Horizons, users can gesticulate with the movements of the hands and of each of the five fingers being displayed correctly. Notice from Figure 3 that the available avatars in Horizons look more realistic than the ones in Mozilla Hubs.

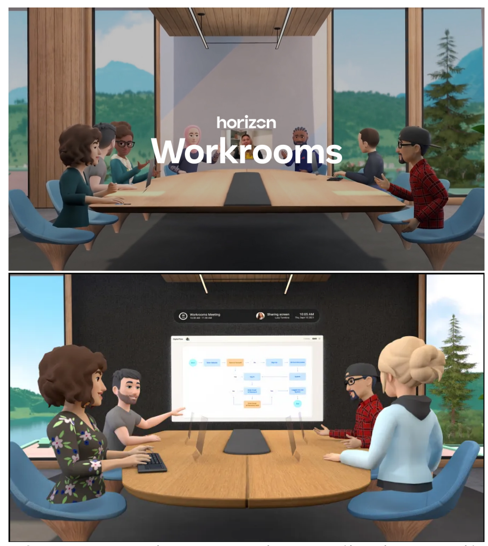

Figure 3: Meta Horizons Meeting application, available at `forwork.meta.com/horizon-workrooms`.

VRChat is a popular social platform that allows users to create, share, and experience virtual worlds and environments in virtual reality or via a desktop mode. VRChat gained popularity due to its user-generated content, diverse community, and the freedom it offers users to explore and create within virtual spaces. It is often used for socializing, gaming, role-playing, and even for educational or creative purposes by individuals and groups around the world. Figure 4 shows some examples of avatars and social rooms in VRChat. Notice that VRChat allows creating very distinct avatars, which are very popular among users. Most avatars look like character of video games or cartoons. VRChat supports the hand-tracking required and allows users to gesticulate and mimic. This allows it to be used as a virtual environment for playing the charade game.

Experimental Methodology

Before the experiment can be performed, a couple of experimental parameters have to be defined. In terms of the virtual

Figure 4: VRChat Meeting application, available at `hello.vrchat.com/`.

Figure 5: Virtual rooms used in pilot study: (top) VRChat and (bottom) Horizons.

environment, as mentioned earlier, a simple VR room is necessary to avoid biases and computational complexity. Figure 5 shows the two virtual environments used in the pilot study with Horizons and VRChat. In the programmed experiment, similar environments will be used and the participants will be placed in front of each other in the VR room. Also, to make the preparation time shorter, the participant will choose one avatar from a set of pre-installed and pre-selected avatars available at the time of the experiment.

The experiment will be performed at the Computer Science Research Lab at Texas State University. Participants will be from the university community. We will try to increase diversity by inviting not only computer science undergraduate and graduate students, but also staff, family members, and students from other areas. A baseline experiment with Microsoft Teams will also be performed for comparison purposes. The subjective experiment will be divided into 4 stages:

- **Pre-Questionnaires:** In this stage, participants will be asked to sign a Consent Form and fill a Pre-experiment Questionnaire. The consent form describes briefly the experiment and its risks, attesting that the collected data will only be used for research purposes. The Pre-experiment Questionnaire has several questions and is divided in different sections. The first section collects personal data from the participants, such as gender, education level, age, experience with VR, HMDs, and charade, mother language, etc. The second part of the questionnaire is a Simulator Sickness Questionnaire (SSQ) [11], which aims to measure sickness in the context of motion sickness in virtual reality.
- **Training:** This session consists of a small trial where the experimenter explains orally the charade game, detailing the task, how the words are communicated to them, what he/she is allowed to do, etc. In particular, participants are asked to try out the interface and gesticulate.
- **Main Session**: In the main session, the participants play the charade game. Only two participants (players) will be used. The participants will be able to stand to gesticulate. The participants will not know each other to avoid social biases or an unfair comparison. The words will be selected randomly from a list of pre-selected words. The words will be simple, given the fact that it is expected that many participants will not have English as his/her mother tongue.
- **Post-Questionnaires:** In this stage, participants will fill out the post-experiment questionnaire, which is also divided into several sections. The first section is a repetition of the SSQ questionnaire answered before the experiment. This allows determining if the experiment caused any motion sickness problems. In the second section, the participants will fill out the NASA task load index (TLX) questionnaire [12], which provides multi-dimensional ratings of overall workload based on a weighted average of six subscales: mental demands, physical demands, temporal demands, performance, effort, and frustration. In the third and last section, participants will answer a few questions about presence and social interaction.

Conclusions and Future Work

In this paper, I discussed the many challenges of estimating the user QoE for immersive applications. In particular, I detailed the VQEG-IMG test plan for XR subjective QoE assessment, describing the audio-visual experiment to be performed at Texas State University. The experiment consists of a charade game to be played by two participants in two different virtual reality applications. The Microsoft Teams interface will be used as a benchmark non-virtual reality application for comparisons purposes. The goal is to compare two different scenarios for teleconference meetings: a virtual reality representation and a realistic representation. The experiment will be performed throughout the summer of 2024.

References

[1] J. G. Apostolopoulos, P. A. Chou, B. Culbertson, T. Kalker, M. D. Trott, and S. Wee, "The road to immersive communication," *Proceedings of the IEEE*, vol. 100, no. 4, pp. 974–990, 2012.

[2] P. Pérez, E. Gonzalez-Sosa, J. Gutiérrez, and N. García, "Emerging immersive communication systems: overview, taxonomy, and good practices for qoe assessment," *Frontiers in Signal Processing*, vol. 2, p. 917684, 2022.

[3] 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. Spie, 1995, pp. 282–292.

[4] M.-A. Moinnereau, A. A. de Oliveira Jr, and T. H. Falk, "Immersive media experience: a survey of existing methods and tools for human influential factors assessment," *Quality and User Experience*, vol. 7, no. 1, p. 5, 2022.

[5] K. Brunnström, S. A. Beker, K. De Moor, A. Dooms, S. Egger, M.-N. Garcia, T. Hossfeld, S. Jumisko-Pyykkö, C. Keimel, M.-C. Larabi *et al.*, "Qualinet white paper on definitions of quality of experience," 2013.

[6] A. Perkis, C. Timmerer, S. Baraković, J. B. Husić, S. Bech, S. Bosse, J. Botev, K. Brunnström, L. Cruz, K. De Moor *et al.*, "Qualinet white paper on definitions of immersive media experience (imex)," *arXiv preprint arXiv:2007.07032*, 2020.

[7] H. Kharoub, M. Lataifeh, and N. Ahmed, "3d user interface design and usability for immersive vr," *Applied sciences*, vol. 9, no. 22, p. 4861, 2019.

[8] Y. M. Kim, I. Rhiu, M. Rhie, H. S. Choi, and M. H. Yun, "Current state of user experience evaluation in virtual reality: a systematic review from an ergonomic perspective," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 63, no. 1. Sage Publications Sage CA: Los Angeles, CA, 2019, pp. 1274–1275.

[9] International Telecommunication Union, "Subjective Quality Evaluation of Audio and Audiovisual Multiparty Telemeetings," International Telecommunication Union, ITU-T Recommendation P.1301, September 2012.

[10] A. Osborne, S. Fielder, J. Mcveigh-Schultz, T. Lang, M. Kreminski, G. Butler, J. V. Li, D. R. Sanchez, and K. Isbister, "Being social in vr meetings: A landscape analysis of current tools," in *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, 2023, pp. 1789–1809.

[11] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *The international journal of aviation psychology*, vol. 3, no. 3, pp. 203–220, 1993.

[12] S. G. Hart, "Nasa task load index (tlx)," 1986.

Author Biography

Mylene Farias received her B.Sc. degree in electrical engineering from Federal University of Pernambuco (UFPE), Brazil, in 1995 and her M.Sc. degree in electrical engineering from the State University of Campinas (UNICAMP), Brazil, in 1998. She received her Ph.D. in electrical and computer engineering from the University of California Santa Barbara (UCSB), USA, in 2004 for work in no-reference video quality metrics. Currently, she is an Associate Professor in the Department of Computer Science at Texas State University. Previously, she was an associate professor in the Department of Electrical Engineering at the University of Brasilia (UnB). To date, she has published 54 scientific journals and 110 peer-reviewed conference papers. She is currently an Associate Editor for IEEE Signal Processing Letters and SPIE Journal of Electronic Imaging, in addition to being an Area Editor for Elsevier Signal Processing Image Communication. Dr. Farias is a Senior Member of IEEE and the IEEE Signal Processing Society. She is also a member of ACM. She has served as a member and area chair in technical program committees of several conferences. Her current interests include video quality metrics, video processing, multimedia signal processing, immersive media, and visual attention.