Measuring Virtual Reality Experiences is more than just Video Quality

Hanan Alnizami, James Scovell, Jacqueline Ong, Philip Corriveau

What is VR

While Virtual Reality has drawn much attention and publicity the past few years, it is not a new concept in technology.

Virtual Reality (VR) allows the user to experience a completely digitized environment while attempting to disconnect the user from her/his real world. In his book, Virtual Reality, Howard Rheingold defines it as an experience in which a person is surrounded by a three-dimensional computer-generated representation, and is able to move around in the virtual world and see it from different angles, to reach into it, grab it, and reshape it [1]. While Virtual Reality has drawn much attention and publicity the past few years, it is not a new concept in technology. VR dates back to the 1960s when Ivan Sutherland pioneered the first head-mounted display at MIT [2], which was then a room-size V.R. machine, with an helmet so heavy that it had to be supported by a mechanical arm suspended from the ceiling [3]. Soon after, HMDs were adopted for military applications [4, 5]. Then on, the US Navy, the US Army, and NASA all invested in VR in hopes of building flight and combat simulators. The US Army deployed the Integrated Helmet and Display Sighting System (IHADSS) on the AH-64 Apache helicopter. Despite the monocular display, the IHADSS greatly contributed to the proliferation of all types of HMDs [6].

Since then, VR has expanded to various applications including automotive, medicine, education, and architecture [7, 8], offering invaluable information-sharing experiences across many applications such as gaming, entertainment, education, and commerce. This new and innovative way of interaction has enabled users to unique experiences such as telepresence [7, 9], and high interactivity [10], especially in virtual commerce experiences from the comfort of one's home.

The explosion of devices available for consumer consumption has been incredible and varies in the quality of implementation for a range of budgets. Regardless of which segment one is aimed at creating immersive experience in VR is not an easy endeavor, and assessing VR interactions holistically is a demanding and complex procedure. VR experiences represent a constellation of engineering metrics which, while can be challenging to simply evaluate independently, they interact together to make or break an experience. It is vital to have a good understanding of Holistic VR experiences and know how to properly assess them. Failure to provide well-tuned virtual content could cause undesired physiological implications on the user.

Virtual Reality Engineering/User Metrics

VR experiences are driven by multiple types of metrics such as visual performance, auditory cues, user interaction, and ergonomics. While a plethora of VR literature has been deployed on enhance computer graphics, display technologies, and input tracking among others, little to no literature has been found that focuses on VR experience usability evaluation and overall ergonomics.

Head Mounted Display Ergonomics

This is an essential pillar for providing an immersive VR experience. One of the significant limitations for these HMDs is their substantial weight, due to the attached computing and display hardware. HMD weight can markedly affect head balance, body posture and locomotion, which in turn can retard voluntary motion and action in response to visual stimuli. A heavier device weight can also increase the mismatch between visual, vestibular and proprioceptive cues, leading to motion sickness symptoms. Not only is weight important, but the distribution of weight around the HMD could play a role in users range of motion and overall experience, see Figure 1.

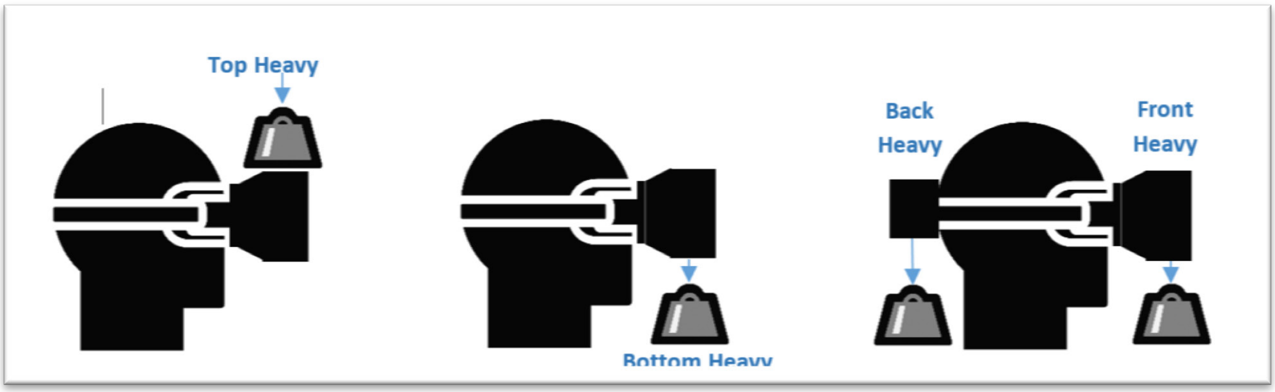


Figure 1. HMD weight and weight distributions can affect overall muscular and physical stress

HMD temperature has become an evident limitation that has been increasingly self-reported by users, especially gamers, who engage in lengthy VR sessions. Typically, HMD manufacturers seek HMD designs that offer good seal around the user's face in order to prevent light leakage that could produce blur and glare on the HMD screen, affecting user's experience. Latter design decisions and poor ventilation solutions within the HMD have caused the relatively small amount of air trapped within the face cup to increase in temperature and humidity.

Whether it is weight, weight balance, pressure, fit and finish, ambient temperature of the HMD's face cup, or overall hygiene, failure to provide good user design will guarantee breaking the VR experience.

Well-designed controller ergonomics are also crucial to allow for comfortable interaction with the virtual world. Controller weight, button design, finish, and hygiene are some of many ergonomic aspects that should be considered.

VR Visual Performance

An area where the Video Quality Experts Group has excelled for many applications and usages, is another critical component in providing a truly immersive experience. Negative experiences such as dropped frames or tearing can disrupt the user interactions breaking the perceived reality. Visual experiences afforded by HMDs are the best when the rendered visual images closely match other sensory inputs, such as motion, balance and proprioceptive feedback. This is especially important for situations where the user is moving around and actively interacting with the visual world.

There are many visual factors that can impact this aspect of VR such as resolution, refresh rate, flicker, field of view, pixels per degree, etc. While each individual variable can be isolated and evaluated to understand acceptable experience thresholds, to truly understand the overall visual experience eventually the variables must be combined to understand potential interactions. Unfortunately, due to the infantile nature of VR we must begin by isolating individual variables for evaluation.

There are some variables like resolution that can be manipulated using the developer settings built-in to some systems. Otherwise a test harness would need to be developed to manipulate each variable. Due to the “black box” nature of various VR systems it can be challenging to do comparisons between systems to isolate a single variable, which is another benefit of developing specific test harnesses that allow researchers to manipulate one variable at a time.

Refresh rate is another variable that can significantly impact the perceived visual performance. Some VR systems utilize a methodology called decoupled refresh rate, where content movement will update at a lower rate than the user head movement allowing for a lower compute cost thus reducing the potential for performance based issues. Though there is little information available on decoupled refresh rate to understand the potential impact to the user.

Common issues caused by impaired system performance include dropped frames or tearing. In an effort to combat this, system developers have developed various methodologies to alleviate these issues. These methodologies include blending of frames, projecting new frames based on user movement and the last rendered frame. These methods will help reduce performance related issues but there is a lack of understanding as to what magnitude of degradation would be too much to handle and what negative impact these methodologies may have. These attempts to combat system performance based issues is a benefit for the end user but an extra challenge for researchers attempting to evaluate various solutions.

While there are many variations, there are generally three ways to evaluating user perception for variables such as frame drops, tearing, refresh rate and others. First is to find natural variation between devices or systems. In the case of refresh rate for example, this is achieved by identifying: 1) various refresh rates for devices to be tested (i.e. 30, 60, and 90Hz); 2) an application that works across devices; 3) representative use cases that would exacerbate the impact of refresh rate. When running the study, participants are tasked with completing tasks and providing subjective experience ratings to understand if any differences exists among conditions. There are some potential issues with this methodology, such as finding natural variation. This methodology also assumes that all other variables are held constant as to prevent confounding factors, which is often unlikely.

The second way is developing a test harness that would allow for test variable manipulation while holding all other variables constant. The trick with developing test harnesses is manipulating variables in a way in which the variable is representative of how the variable would naturally occur. It is important to keep in mind that developing a test harnesses can be a time consuming and costly endeavor.

The third way to evaluating user perception is via expert assessments. Using experts in a study can be especially controversial as the sample size is extremely limited and can often be biased by their experience and knowledge in the space. Having unbiased and industry recognized evaluators becomes crucial if the data is to be accepted.

VR Audio

Audio is a component of VR that can easily be overlooked but is critical to many of the experiences that are of interest to most users. In VR, senses are muted to the real world and as such, the brain relies on stimuli presented through the display and speakers or headsets to accurately and comfortably orient the user in space. 3D positional audio for example has become critical not only for providing an immersive VR experience but for orienting the user in their virtual space. It is no longer good enough to know that something is going outside of your field of view. If a zombie is coming from behind you to your left, you need to have audio that can accurately convey this to the user.

Other Ecosystem Variables

There are other aspects of system performance that can impact the experience including latency and accuracy. These hold true for both the HMD and any controllers that may be used. Latency is a fairly straight forward concept to evaluate but is critical to ensure a positive experience. Excessive latency can cause a variety of issues from minor annoyance to extreme nausea. Accuracy includes a variety of variables that can all impact the perception of perceived reality. It is not as simple as a point and shoot. HMD’s and controllers need to be evaluated for directional accuracy, drift, static noise, and scaling errors. Minor variation in any one of these variables may not break the experience but a combination or excess of any one variable can have significant negative effects.

Evaluation Methodologies

Traditional methodologies frequently used by VQEG such as MOS were able to be applied to evaluate many of the variables discussed above such as resolution, refresh rate, thermals, frame drops, tearing and more.

While these base methodologies were applied for some variables, we have been forced to develop adapted methodologies, hardware and software solutions. For example, to understand the thermal impact of the HMD on a user, thermal and humidity sensors were attached to an Arduino board that was fixed to the HMD allowing for continuous real time data collection. This allowed for data to be easily mapped to user rating. The initial research focused on passive experiences but it will be important to also understand the thermal impact to a user during more strenuous experiences. A challenge for thermal testing was to accurately collect relative humidity data in a confined space such as the face cup of the HMD.

When evaluating various bit rate encodings, it is challenging to allow for real time application of the encoding method. Content has to be prerecorded for use in a study. The latter causes an issue if the user moves their head as content will not visually update by their movement. As a result, users are instructed not keep their head in a fixed position to avoid inducing extreme nausea.

Best Practices

When recruiting users to participate it is important to ask the proper screening question to ensure a safe and positive study session. General questions that con preclude users from participating include a history of susceptibility to motion sickness, heart conditions, and eyewear that could interfere with wearing an HMD.

When switching between test harness settings or applications it can be beneficial to ask users to close their eyes until the new content is available as the transition can be disorienting. As the users are immersed in the HMD and cannot see the outside world without removing the HMD every time, asking participants to speak aloud and have a proper training sessions with the rating scale can be extremely beneficial. It was not uncommon for users to forget the rating scale and needed to be reminded, so it is important to ensure the user understands the rating scale each time they give a rating.

If participants do report any eyestrain or nausea, it is important to have water available and ensure they do not immediately drive. Lastly, it is very important to frequently check in with users to ensure they are not suffering from any side effects and that they have the ability to quit at any time.

Conclusion

Visual quality is not enough, VR is much bigger than just the visual experience. The inability to present the user with accurate and consistent stimuli from any of the variables discussed above whether visual, auditory, interactive, or ergonomic, could affect users' neurophysiological responses and physical comfort such as dizziness, nausea, eye strain, and overall muscular and skeletal fatigue. Excessive weight can impede and limit behavioral responses and induce viewing discomfort due to poor HMD fit and will induce undesirable visual, muscular and cognitive symptoms.

Bibliography

[1] H. Rheingold, "Virtual reality: Exploring the brave new technologies", Simon & Schuster Adult Publishing Group, 1991.

[2] I. E. Sutherland, "A head-mounted three dimensional display", *Proc. of the AFIPS Fall Joint Computer Conference, part I*, pp. 757–764, San Francisco, USA, Dec. 1968.

[3] V. Heffernan, "Virtual reality fails its way to success", *The New York Times Magazine,* Nov. 2014. [Online]. Available: https://www.nytimes.com/2014/11/16/magazine/virtual-reality-fails-its-way-to-success.html?_r=0

[4] J. E. Melzer and K. Moffitt, "HMD design--putting the user first", in Head-mounted displays: Designing for the user, McGraw-Hill, pp. 1–16, 1997.

[5] C. E. Rash, "Helmet mounted displays: Design issues for rotary-wing aircraft", vol. 93. SPIE Press, 1999.

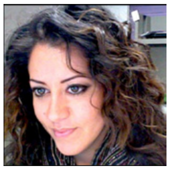
[6] C. E. Rash and J. S. Martin, "The impact of the US Army's AH-64 helmet mounted display on future aviation helmet design", 1988.

[7] F. P. Brooks, "What's real about virtual reality?", *IEEE Computer Graphics and Applications*, vol. 19, no. 6, pp. 16–27, Nov-Dec. 1999.

[8] C. Demiralp, C. D. Jackson, D. B. Karelitz, S. Zhang, and D. H. Laidlaw, "Cave and fishtank virtual-reality displays: A qualitative and quantitative comparison", *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 3, pp. 323–330, May-Jun. 2006.

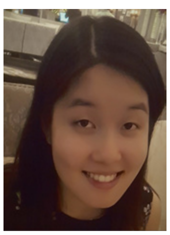
[9] M. Y. Hyun and R. M. O'Keefe, "Virtual destination image: Testing a telepresence model", *Journal of Business Research*, vol. 65, no. 1, pp. 29–35, Jan. 2012.

[10] A. Mollen and H. Wilson, "Engagement, telepresence and interactivity in online consumer experience: Reconciling scholastic and managerial perspectives", *Journal of Business Research*, vol. 63, no. 9–10, pp. 919–925, Sep-Oct. 2010.



Hanan Alnizami *is a Human Factors Engineer at Intel. She specializes in setting UX influenced engineering requirements that drive software, hardware, and platform specifications. She received her PhD in Human-Centered Computing from Clemson University in 2015.*

James Scovell *is a Human Factors Engineer working at Intel. He specializes in experimental design and predictive model development in areas such as system performance, visual quality, and human computer interaction.*



Jacqueline Ong *has been a user experience software engineer for two years at Intel, driving competitive analysis for virtual reality and machine learning use cases with Intel solutions.*

Philip J. Corriveau *is a Senior Principal Engineer and Director of End to End Competitive UX at Intel. He directs a team of human factors engineers conducting user experience research across Intel technologies, platforms and product lines. Philip is a founding member of and still participates in VQEG.*