

Virtually Usable: A Review of Virtual Reality Usability Evaluation Methods

By Dana Martens, MFA D+T Candidate 2017 at Parsons, The New School for Design
May 14, 2016

Executive Summary

Throughout the last decade, Virtual Reality (VR) based systems have found footholds in new and varied industries, powering a wide variety of interactions between humans and machines. As new technologies such as advanced graphics engines, wearable hardware, and innovative interface systems have been created, conceptions and applications of virtual reality have changed along with it, growing every more varied as different mixtures of the above technologies are combined to help designers realize and create their virtually imagined experiences.

Due to this variety, both in application and in technology, one of the most challenging aspects of designing successful virtual reality experiences from a human-centered design standpoint has been assessing the usability of the system throughout the ideation, prototyping and implementation phases in order to provide guidelines that designers and developers can use to fix usability issues and build better interactions. While there has been extensive research refining traditional usability evaluation techniques for graphic user interfaces (GUIs) to work for virtual reality, creating new evaluation techniques and re-categorizations of virtual experience, and combining these two methods to account for the variety of VR technologies and applications mentioned above, the methodologies have not been synthesized into content that can help designers understand what technique to use, when to use it, and what the costs and benefits are for the methodology.

In this paper, I seek to address this issue by reviewing some of the dominant academic contributions in the field and pulling out the insights and best practices developed throughout the research in order to:

- Define virtual reality and briefly review the major types of VR technology commercially available
 - Virtual reality desktops
 - Head mounted displays (HMDs)
 - Cave Automatic Virtual Environments (CAVEs)

- Categorize some of the major differences between VR and traditional user interfaces - such as GUIs – and how these differences lead to specific types of usability evaluation issues.
- Briefly review and compare successful methodologies for VR usability evaluation that have been proposed and tested in earlier studies
- Using the categories defined above, build a framework showing where and when these tested evaluation methodologies might be best implemented

My intention is to help frame the robust and informative, yet dense and overlapping, body of academic research into a more digestible roadmap or toolkit that helps those tasked building and evaluating virtual reality applications –often not human computer interaction (HCI) experts – to quickly and easily identify which of the prevailing evaluation methods might best meet their needs.

Finally, I believe that one severe gap in the body of VR usability evaluation research I reviewed, is that newer works (within the last 5 years) are severely lacking. Since much of the earlier research relies on evaluation tools and technological constraints at the time of the study, I believe that current and emerging technologies such as wearables, facial & feature recognition, and more accessible/easy-to-use VE creation tools & (e.g. Google Cardboard) distribution platforms have the potential to support these evaluation methods in new ways. At the end of this paper, I will briefly sketch out some of my ideas for how these technologies may be used, and outline some related areas for further inquiry and experimentation.

What is Virtual Reality Anyway & How Many Different Types Could There Possibly Be?

According to Wikipedia, Virtual Reality, also sometimes known as immersive multimedia or computer-simulated reality, is any technology that “replicates an environment, real or imagined, and simulates a user’s physical presence and environment to allow for user interaction. Virtual realities artificially create sensory experience, which can include sight, touch, hearing, and smell.” Fittingly, Oculus’s Chief Executive, Brendan Iribe described it “...essentially, a hack on the human sensory system” (Economist). Given this definition, one thing can immediately be ascertained about the possible uniqueness of all VR interactions. Virtual reality is at its heart, a replication of a “*real*” environment or reality, “the state of things as they actually exist, rather than as they may appear or might be imagined.”[3]

This means that we as users tend to experience VR through the lens of our own experience of *real* reality, and all of the assumptions about how things work in the *real* world that go along with it. *Understanding Virtual Reality: Interface, Application, and Design* defines four key elements for experiencing reality, and by extension virtual reality: a virtual world, immersion, sensory feedback (responding to user input), and interactivity (Sherman, William, Craig 6). A review of the available literature agrees that VREs are distinct in that they stem from the use of partially or fully immersive systems, seeking to establish one's presence in the environment (Bowman, Gabbard, Hix 405). Given these elements, experts often also assume that the role of a Virtual Environments (VEs), "an instance of a virtual world presented in an interactive medium such as virtual reality," (Sherman, William, Craig 6) is to represent the real world as faithfully and naturally as possible. While some VEs may represent unnatural worlds, e.g. virtual information spaces, unnatural scales and points of view – surveying the whole universe or being on the surface of a molecule - they still must conform to or purposely reject these expectations, helping users build a cognitive model of how they can experience and interact with the unnatural world they are in – as opposed to the real world (Sutcliffe, Gault 834).

As technology has improved exponentially, both in terms of electronics and raw computing power, VR systems have improved as well, overcoming technological challenges such as graphics limitations and processing speed, as well as incorporating new interfaces and hardware. In addition, these improvements have led to a decrease in cost that have made developing virtual reality experiences more accessible to a wider range of industries and applications. While a VR experience is truly made up of the wide collection of inputs and outputs that support the key elements described above, we can roughly classify them by one of the most important sensorial interfaces, visual feedback.

Desktop Virtual Reality

Desktop virtual reality is any program that simulates a real or imaginary world in 3D graphics on a screen, just like your desktop computer at home. This is the oldest form of VR given technical constraints, and depicts the virtual world in three dimensions through a 2D screen or single projection of a screen. The majority of VR usability research involves Desktop VR as much of the literature was written in the 90's and early 00's, when they constituted the majority of industry applications (Kaur, Maiden, Sutcliffe 1996).

Head Mounted Displays (HMDs)

Head mounted displays “visually represent immersive space by insulating human vision in the visual space.”(Bchung 2015). By harnessing stereoscopic (one image per eye) visual perspective, HMDs help illuminate the feeling of depth as user’s brains naturally simulate depth in the virtual world. Head mounted displays have become more popular and varied as the price of the hardware falls and more and more developers have access to units. Most people automatically think of HMDs when VR is mentioned.

CAVE Automatic Virtual Environments (CAVEs)

A CAVE is literally what it sounds like, a virtual environment where a person is fully immersed in the virtual experience without wearing any visual displays, the virtual reality is mapped onto the wall of the room (or CAVE) the person is within. Most CAVEs involve projected video or graphics on the walls, floor and maybe ceiling, as well as stereo sound to represent the three-dimensional nature of the space (Virtual Reality Society 2016).

In addition, VR systems can have several different types of user input allowing them to interact with the virtual world. While some applications are purely passive, the system has complete control the entire time and the user is just along for the ride – like a virtual guided tour – most require some sort of user interaction through an input device such as a gamepad, joystick, mouse and keyboard, wand, or haptic interface.

Why Is Evaluating VR Different from Evaluating a Traditional User Interface?

Given the key elements of a VR experience (a virtual world, immersion, sensory feedback, and interaction) combined with the variety of output and input devices that can be used to achieve them, “VE applications have interaction styles that are so radically different from ordinary user interfaces that well-proven methods that produce usable [Graphic User Interfaces (GUIs)] may be neither appropriate nor effective,”when trying evaluating them (Bowman, Gabbard, Hix 405).

While the challenges posed by the diversity of VR systems is indeed great, these *radical differences* have been one of the most well-documented areas of research throughout the predominant literature, added to by individual authors’ (often one of a small rotating cast of HCI experts)

experience in conducting real-world evaluations of the technologies. In addition, since the VR usability research field is still rather narrow and comparatively new, most studies have built off of one another, refining and evaluating previous work and summarizing developments and realizations before adding their own perspective. Two of the major players, Sutcliffe and Kaur nicely sum up the scope of early research (much of it by themselves) regarding these developments in the introduction to their piece “A Usability Evaluation Method for Virtual Reality User Interfaces” which you can refer to for a detailed overview of the predominant milestones and studies (Sutcliffe, Kaur 415).

However, looking broadly across the research, one developing trend is to organize these differences and related evaluation issues in categories that inform design interventions development teams can use to correct for issues throughout the process. The authors of one later piece, focused on developing a framework for comparing and classifying methods of usability evaluation for VE’s, translate the importance of this methodology. “By focusing on usability from the beginning, developers are more likely to avoid creating interaction techniques (ITs) that do not match appropriate user task requirements and to avoid producing standards and principles for VE user interface development that are nonsensical” (Bowman, Gabbard, Hix 404). By looking across the different categorizations and distinctions the research proposes, I have developed five categories of essential differences between VEs and traditional user interfaces, and the evaluation issues that arise from them. One important addition to this categorization is the inclusion of specific issues that would not be addressed in a typical usability evaluations, specifically related to the limitations of the available technology on the reality of the virtual experience and problems stemming from the mismatch.

Real World Environment & Technology

Perhaps the biggest difference between VEs and traditional user interfaces is the physical environment in which the system is used and which input and output devices are used to interact with it. As outlined above, there are three main types of visual output systems currently in use, and numerous types of input devices from the traditional gamepad to complex computer vision and biometric feedback interfaces. In addition, while most typical user interactions through GUIs don’t rely on the position of a user’s whole body in space, VE’s may require users to stand or sit, to move around a large space and/or use whole body motions. (Bowman, Gabbard, Hix 405)

This leads to several complications when performing usability evaluations.

- Most HMDs don't offer a see-through window into the real world and evaluators must ensure that users won't knock into anything, trip over cables, or move outside of the range of recording equipment and tracking devices.
- Similar problems arise in CAVE environments where physical walls and obstructions can be difficult to see because of the projected visualizations
- Most VE displays don't allow simultaneous viewers (user and evaluator) to see the same perspective, so additional equipment must be setup so the evaluator can see the same image as the user
- When multiple views are possible, it is hard for the evaluator to both watch what the user is seeing and the physical actions of the user at the same time. Similarly, recording both the user and the interface can be difficult since the VE users may be mobile, requiring an extremely wide shot or post-editing to sync different video streams
- Interactive features of the technology such as voice-recognition may inhibit common evaluation practices such as asking the users to "think aloud" or providing instructions, hints, or questions during the evaluation
- In cases involving collaborative VEs, where more than one user is participating simultaneously, often physically separated, several issues emerge: techniques for gathering information from each user (often remotely) must be developed, the unpredictability of network behavior may act as a confounding variable, and different users may be using different configurations of input and output devices in a variety of physical environments.

Presence Related Issues

Another major difference between VEs and traditional user interfaces is the importance of presence, or the feeling of actually existing with the VE. As stated earlier, one of the key elements of successful VEs is immersion, the user's sense that they are present in the virtual reality, rather than the real world around them. Presence is key to immersion, and often requires a delicate and nuanced combination of sensory experiences from one or more of the VEs output devices. For this reason, several challenges arise when attempting to evaluate users interactions with virtual reality interfaces, many of them stemming from the real-time impossibility of the evaluator to both exist within the same VE as the user, observing and guiding their digital representation, while also being present to observe and guide their interactions with the physical interface (Bowman, Gabbard, Hix 405).

- If physically observing the user's interactions, evaluators risk breaking the user's presence if sensed. This is especially true in CAVE environments, where it is difficult for evaluators to observe from certain angles, given they may be seen or interrupt the projected graphics.
- If physically observing the user's interactions, evaluators may also risk interrupting presence when issuing new tasks, asking for feedback, or providing guidance.
- Alternatively, if evaluators choose not to interact with users at all, they must ensure that the VE is complete enough to require no interruptions for technical issues. They also have to ensure that the user is provided (and remembers) clear and detailed instructions so that they have a full understanding of the procedure, interface, and tasks they are to accomplish when immersed.
- Virtual Environments are often complex and require multiple types of sensory inputs to create and maintain presence. As a result, several different (and not always compatible) types of hardware and software must work together, all recording and interpreting different values. This complexity can make it challenging for evaluators to use the technology to help record and compare quantitative information such as timing, task errors, and number of critical incidents through the available technology.
- Similarly, traditional user interfaces only require one discrete type of user input (like a gamepad) while VEs can often require multi-modal types of input that combine events, gestures, position, voice, and/or whole-body motions. It is much harder for evaluators to process these multiple input streams simultaneously while also logging an accurate account of the user's actions.
- Measuring a user's perception of presence within a VE is often a difficult task unto itself, relying on reports that must capture and analyze fundamentally subjective concepts such as the perceived realism of the virtual world. There have been several studies in which post-hoc questionnaires are employed for this purpose, and purportedly obtain reliable and consistent measurements (Witmer & Singer, 1998; Slater, 1999, Usoh et al 2000).

User Issues

The relatively new and constantly changing nature of VR interactions makes it particularly challenging to recruit, compare, and test users using typical evaluation methods. In traditional usability evaluations, an effort is usually made to ensure that test subjects are directly engaged from the target user population, or a similarly representative sample is obtained. For example, good evaluations take things like gender equity, age distribution, and ranges of experience with the tested technology or

similar technologies into account. However, since virtual reality is a relatively new and still somewhat industry-specific (e.g. gaming) technology, it often presents unique challenges in selecting representative user samples. In addition, the persisting physical effects of prolonged immersion in VEs can limit the type and length of testing methods that can be used (Bowman, Gabbard, Hix 407).

- Many potential applications of Virtual Reality are still being explored for the first time in several industries and creative domains to solve problems and drive user engagement. As such, initial attempts have no precedents for who their ideal users might be, meaning that target user populations are often not well defined. This can make it difficult for evaluators to generalize performance results.
- Since VR is such a new technology, very few people can be considered experts in using VEs. For this reason, obtaining a representative sample of both novice and expert users may be impossible, since often the only people who could be considered experts with the technology are the development team themselves.
- Given that most VR users will inherently be novices, evaluators can make no assumptions about their ability to use or understand their interface given real-world and cultural experience. This can limit the types of evaluation techniques that may be employed, or require them to be reframed at a lower cognitive and physical level.
- Similarly, the results of evaluations may be highly variable among individuals for this reason, requiring larger samples of user to obtain a clear picture of the performance, especially if statistically significant results are required.
- In addition, since research into VR usability evaluation is still rather new and flexible, ranging over a large array of I/O devices, the potential design space studied by VE interaction techniques is quite broad. For this reason, evaluators must often compare several different techniques and I/O devices, or combinations thereof within a single evaluation. This means that users must be equally flexible, adapting to a wide variety of situations and interface changes throughout testing.

Researchers are well aware of the potential for VEs to cause simulation or virtual reality sickness, especially when using HMDs or CAVE interfaces. Thought to stem from the experience of multiple cognizance (competing awareness of reality) and related (usually technically caused) incongruities in different types of sensory experience, these mismatches can cause nausea, headaches, and vomiting with prolonged exposure.

- Very little is known about acceptable exposure time to VEs, limiting the types of user-oriented evaluations that can be employed (eliminating those that are lengthy) or similarly modifying the evaluation technique to restrict the length and frequency of sessions, incorporating rest periods and contingencies for sickness and fatigue.
- Since the specific causes of simulator sickness are still unknown and since VE applications contain new and diverse mixtures of interaction techniques, it can be difficult to evaluate accurate usability measurements without also factoring in ratings of user comfort (which can be achieved through subjective questionnaire based methods, or through directly monitoring physiological factors).

Issues Related to the Nature of Virtual Environments

Perhaps the largest and most significant difference between VEs and traditional user interfaces is rooted in our very perception of reality itself, and related assumptions and expectations of what is possible in virtual realities as a result. As discussed earlier, there are four key elements for experiencing virtual (or real) reality: a virtual (or real) world, immersion, sensory feedback (in response to user input) and interactivity. As such, the successful translation of these elements requires interactions with virtual worlds that mirror the complexity of our interactions with the real physical world, rather than those needed to understand and manipulate a traditional user interface.

- In the real world, one of the most significant indicators of “realness” is physicality, touch, or haptic feedback from our environment. The majority of VEs, especially older Desktop VEs and HMD varieties requiring little user engagement (such as a VR movie watching app) provide no haptic feedback to users at all, certainly not a direct mapping of touch to objects in the virtual world. Instead, they rely on re-mappings of other types sensory feedback, most frequently visual, one example being collision detection algorithms to help prompt users what their presence (cursor, avatar) is currently selecting, and what is selectable from the environment. This can often cause issues for complex manipulation and physical tasks, as well as orienting oneself appropriately to objects in the virtual space.
- Since exploration is also a key aspect of our real world, it is also a key element of virtual worlds, and users expect to be able to navigate their VE in a logical way. However the aforementioned need for visual indications of user presence to stand in for haptic sensation forces developers to employ alternate feedback techniques, such as magic ‘snap-to’ effects so

that nearby objects jump into the user's "hand", or the ability to select through VE objects like walls. These mismatches can lead to usability issues when users' expectations of how the interaction should work is limited by the available I/O devices.

- A similar issue only concerns graphical output, the power of the available VE technology to render realistic graphics. Being unable to meet the memory and graphic requirements for this realism can cause overall perceptual issues for users such as 3D depth or perspective distortion, lack of visual clarity due to low resolution, environmental rendering lags that can cause disengagement or sickness, exacerbated by the technical affordance of the hardware (desktops, HMDs, CAVEs) the visuals are presented through (Sutcliffe, Gault, Maiden 101).
- How users are represented in a VE (a cursor, avatar, hand etc.) and how they control their self-representation in that space through available input devices is very different than in a standard digital interface. Less than perfect rendering of user's natural actions becomes a much bigger issue when presence is paramount, and more complex manipulation of the VE is required (Sutcliffe, Gault 834).

By understanding key differences stemming from the uniqueness and complexity of VEs and the limitations available technology still places on VR interaction design, we can begin to develop a framework for choosing the appropriate evaluation method to identify and address specific issues mentioned above.

How Might We... Decide which Evaluation Method to Use?

Extensive research has been undertaken examining if traditional usability evaluations work for VEs, and if not, what modifications are necessary to help them overcome the issues raised above. While several studies have reviewed the efficacy and results of these methods, it is more useful to understand when, where, and how these evaluation techniques are best used. Later research, specifically that by Bowman, Gabbard and Hix in 2002, started solving this issue by reviewing and collating VE evaluation methods to date, according to different key characteristics they share: involvement of representative users, context of the evaluation, and the type of results produced. They felt that their specific classification was useful because it "...structures the space of evaluation methods and provides a practical vocabulary for discussion methods in the research community", however conceded that it also had the shortcoming of not conveying when each method should be

used – they felt that this was too specific to the research/testing being done (Bowman, Gabbard, Hix 420).

Their methodology is an excellent place to begin as each of the three characteristics address the main concerns of evaluators, what the evaluation will cost (users vs no), what the impact of the evaluation will be (general or application specific), and how the results can be applied (quantitative vs qualitative) Starting with these classifications – as well as supporting research on specific applications of each -I will quickly review predominant evaluation methods and tag them according to these three characteristics. Since many of these methods are extensions of traditional evaluation tools, I will also mention which tool they are derived from, what the main changes were, and why researchers felt they made it more appropriate for studying usability specifically in VEs. Given the stated shortcoming of this method, I will also attempt to identify the main issues with each type of evaluation, according to the categories I outlined earlier, and as such, where in the development cycle it should potentially be utilized. Finally, as much of the research is quite old, I will provide my own assessment of if/how I think new technologies can overcome any of these issues and provide recommendations for future evaluators to try.

| Evaluation type | Require User? | Generic/App Specific | Type of Result |
|--------------------------------|----------------------|-----------------------------|-------------------------------------|
| Informal Summative | Yes | Generic | Qualitative |
| Formal summative | Yes | Generic | Quantitative |
| Formative (informal or Formal) | Yes | Specific | Both <i>Depends on Formality</i> |
| Post-Hoc Questionnaire | Yes | Either | Both |
| Interview/Demo | Yes | Specific | qualitative |
| <i>Performance Model</i> | <i>No</i> | <i>Either</i> | <i>Quantitative**</i> |
| Heuristic Evaluation | No | Either | Qualitative |
| Cognitive Walkthrough | No | Specific | Qualitative |

**Model Modified from Bowman, Gabbard, Hix “Usability Evaluation in Virtual Environments: Classification and Comparison of Methods” 423.*

***No evaluation yet exists of this category*

Formative (informal or formal) (Gabbard, Hix, Swan 1999)

Requires Representative Users, Application-Specific, Quantitative or Qualitative

Formative Evaluations are observational in nature and seek to empirically assess interaction techniques by iteratively placing users in task-based scenarios. This evaluation method both seeks to identify specific usability problems and the design's overall ability to support user exploration, task learning, and task performance. More informal formative evaluations often yield only qualitative results (critical incidents, user comments, general reactions) while more formal versions can additionally generate quantitative results (task timing, error count etc.).

Formative evaluations are used to assess and improve the usability of an evolving design so they are best employed throughout the prototyping stages, informing each iteration of the prototypes overall design and usability.

VE Specific Issues

Given that these evaluations require users, several issues outlined above are possible including breaks in presence if evaluators need to interact with users, to issues with extensive memorization of tasks before heading into the VE, if evaluators don't want to interact. In addition, the type of technology used will be new for users and precautions must be taken to limit sickness and or disorientation before the exposure effects are known. One of the main tools of formative evaluations is a cognitive walkthrough where an evaluator assesses user interfaces by stepping through a representative list of common tasks users would likely complete when using the system and related measures that assess the interface's ability to support each. This approach is often used when supporting exploratory learning and can especially help understanding system usability for first-time and infrequent users. However, the very nature of virtual worlds makes these types of evaluations very difficult. As in the real world, this means there are a broad variety of interactions available to assess. Traditional formative evaluations like cognitive walkthroughs require task/goal frameworks that help to collect quantitative (time on task, error rate) and qualitative (observation of confusion, talk-aloud) information. However the nature of VEs means that there are usually more stages of exploration that are required than just completing a task.

One way this might be solved is to extend cognitive walkthroughs into a VR specific model that explores more broad interactions than just task-completion. One piece of research suggests creating three stages of interaction: task-based activity, navigation/exploration (whether for completion of a task or serendipitous), and system initiative (the user's role is to interpret actions taken by the VE and

respond to them when initiative is returned, or interrupt/seize the initiative back if necessary) (Sutcliffe, Kaur 417).

Recommended New Technologies

Given the varied nature of interactions that must be captured during formative evaluations, any technology that can help to simultaneously capture both real-world and in-world actions can assist. Advances in VR can potentially allow one evaluator to exist fully in-world to observe a user's avatar or other virtual presence or an AI controlled camera can do the same for later analysis. In addition, biometric feedback can help to measure user's reaction to in-world stimuli if immersion is being measured.

Summative: Formal (Bowman, Johnson & Hodges, 1999) & Informal (Bowman & Hodges, 1999)

Requires Representative Users, Generic, Qualitative and/or Quantitative

Summative evaluations do what their name suggests, they “sum up” or statistically compare two or more different configurations of a user interface design, components of the design, or specific interaction techniques, by having representative users try out each version while evaluators collect quantitative and qualitative information. They can be informally applied, usually collecting just qualitative data, or more formally applied, also helping to collect quantitative data (time on task, error rate). They specifically differ from formative evaluations in that they must compare more than one design. Research involving users is crucial for VE perhaps more so than other types of interfaces as the technology is relatively new and varied, meaning expertise in VR is hard. For this reason, summative evaluations are particularly important as they help to compare specific I/O combinations and or interactions techniques.

Summative evaluations typically occur after user interface designs are complete, and are comparing specific differences between configurations. As such, they are most appropriate for late-stage prototypes in which general usability has been established, but specific interaction or interface questions persist.

VE Specific Issues

These types of evaluations can be difficult as they involve users who may not be familiar with VR or the technology involved, especially if requiring HMDs and complicated inputs. It is also difficult to gather a representative user base if novice versus expert users are required explore how easy the technology might be for first-time users of different skill levels. In informal variations where more exploratory questions might be asked versus tasks recorded, interruptions in presence can also occur.

In order for a formal summative evaluation to work successfully for VEs, a consistent set of user task scenarios must be borrowed and refined from formative evaluations (best if already conducted) to allow for quantitative comparison between complicated user interfaces. In many cases, the results of formative evaluations inform the design of the summative study by determining specific usability characteristics that can be evaluated and compared (Bowman, Gabbard, Hix 418).

Recommended New Technologies

Since summative evaluations rely on a comparison of different user interfaces, any technology that can help to capture user input and/or reactions can be helpful. In terms of capturing inputs, advances in eye tracking technology can be used to show where the user might be looking for HMDs and CAVEs. Emotion recognition software could also be used to track engagement and confusion during specific interactions. This would probably be easiest in a CAVE environment as a HMD would obstruct the user's face.

Post-Hoc Questionnaire (Darken, Sibert 1996) (Slater, Usoh, Steed 1995)

Requires representative users, generic or application-specific, quantitative or qualitative results

Post-Hoc (after the fact) questionnaires are written questions that collect demographic information as well as the opinions, views, and preferences of users after they have already participated in a usability evaluation session. They are good for collecting qualitative and subjective data and are often easier to implement and more consistent than interviews.

Given their nature, post-hoc questionnaires are typically used throughout the prototyping process but in conjunction with another evaluation method.

VE Specific Issues

There aren't too many specific issues with questionnaires aside from the fact that they typically collect very subjective responses. Since VR is such a new technology, responses could be extremely varied as to comfort levels, preferences for I/O configurations etc. In addition, specific feedback that can lead to design interventions may be hard to obtain if users are very inexperienced with the technology and care must be taken to frame questions very specifically to gather the best information possible.

Recommended New Technologies

There aren't too many new technologies that assist with survey-taking although I believe that Google Forms and other online tools can speed up the process and generate quick previews of quantitative information regarding user preference. However it could be interesting to do post-hoc questionnaires within the VE itself, given that immersion and presence are such a big factor. I think exploring if users answer differently when "in world" versus the real world could be an interesting research direction.

Interview/Demo (Bowman, Johnson & Hodges, 1999)

Requires users, Application-specific, Qualitative Results

During an interview, evaluators can collect similar subjective information as in a questionnaire, however they can often go more into detail. "Structured interviews" have a preset list of questions and responses they are seeking to gather while "Open-ended interviews" allow the evaluator to ask broad questions without a fixed set of answers. These are more exploratory in nature and can allow the interviewer to ask more spontaneous and specific questions that they think of as the evaluation progresses. Often demonstrations of a prototype are used as an aid to guide the conversation.

Like post-hoc questionnaires, interviews and demos can be used throughout the prototyping process, however they may be stronger than questionnaires for earlier prototypes if used in an exploratory manner, since VR is still relatively varied and user preference may also vary widely.

VE Specific Issues

Like questionnaires, interviews don't have too many VE specific application issues aside from those involving recruiting a representative sample of users. However it may be difficult for users to talk about their subjective experience of the VE as they might not have the appropriate experience or vocabulary to describe issues they have had. This is especially true if the questions involve immersion and presence, as that can be difficult to collect quantitatively at this time.

Recommended New Technologies

Similar to questionnaires I would recommend any technology that could be used to more intelligently monitor a user's sense of presence in a VE. If immersion and presence are at heart a measure of how much a user feels like they are in the VE instead of the real world, I believe technologies that can capture unconscious responses from the user in response to virtual stimuli might be a direction to explore. This could include anything from emotion recognition, eye-tracking, biometric feedback, head-tracking etc. The goal would be to see if the user is fully responding to in-world feedback versus real world feedback.

Heuristic Evaluation (Steed, Tromp 1998) (Gabbard, Hix, Swan 1999) (Sutcliffe, Gault 2004)

Does not require representative users, Generic or Application-specific, Qualitative Results

In a heuristic evaluation, several usability “experts” separately evaluate a design by applying a set of design standards or “heuristics” that are relevant. Results from the experts are then combined and ranked to prioritize the criticality and related importance of fixing each issue.

While there are several heuristics for traditional user interfaces, only one piece of research so far has attempted to define heuristics for virtual environments. “Heuristic evaluation of virtual reality applications” published in 2004 defines 12 distinct heuristics for VEs which are either direct translations or interpretations of Nielsen’s heuristics published in 1994 given the author's own previous research on VE specific usability evaluation issues (Sutcliffe, Gault 833).

1. Natural Engagement
2. Compatibility with the User’s Task & Domain
3. Natural Expression of Action
4. Close Coordination of Action and Representation
5. Realistic Feedback
6. Faithful Viewpoints
7. Navigation and Orientation Support
8. Clear Entry and Exit Points
9. Consistent Departures
10. Support for Learning
11. Clear Turn Taking
12. Sense of Presence

Heuristic evaluations do not require users and as such are more appropriate to employ early in the design process, as experts can help to identify and correct general usability issues before users are required.

VE Specific Issues

Since heuristic evaluations require “experts” they may be very hard to employ considering how new VR is and how few usability experts exist in this field. But more importantly, the applicability of the above heuristics will be very difficult without a detailed list of representative tasks. The authors who published the above heuristics recommend performing a technology audit before conducting the evaluation, stating “Since VEs are constructed to represent the real world, user tasks should ideally mirror real world actions; however, in practice limitations of technology mean that some compromises have to be accepted” (Sutcliffe, Gault 833). Since VEs are typically constructed to represent the real world, they will always be limited by how real the available I/O devices can make the experience feel and conversely how awkward any necessary sensorial remapping is. In performing the audit, evaluators should set a baseline for how the technology can reasonably be expected to perform and can use the specific issues defined earlier in the section “Issues Related to the Nature of Virtual Environments” to guide their thinking around the following categories:

- Operation of the User’s Presence
- Lack of Haptic Feedback
- Interactive Techniques
- Realistic Graphics

The above heuristics can then be used or discarded based on whether or not the audit has identified if they occur in the VE at all. For example, if the user is automatically logged in to the experience and can’t leave it, there is no need to assess whether or not there are clear entry/exit points.

Recommended New Technologies

Heuristic evaluations will most likely be the hardest to perform until the input and output devices for virtual reality begin to standardize, as they may be doing already with the set of HMDs recently released and paired input devices for each hand that roughly work the same way in terms of how real they are. Given this move towards standardization, one recommendation would be to perform a technology audit of the available HMDs there are, seeking to pull out general heuristics for this

specific visual output category. This could perhaps be repeated for CAVEs if it seemed they are gaining in popularity and implementation.

Conclusion

While VR feels like a very new technology, research has been performed regarding its applications for the last fifteen years, albeit often in specific commercial and military applications. Now that VR is becoming more affordable and available to consumers, usability will become even more important for designers, as target audiences become broader and expectations higher. Given this shift, guidelines and methodologies for designers will also become more important in the coming years to help guide best practices around comfort and experience, guidelines that still only exist in a small but comprehensive body of academic research. In this paper I began what I hope to be a larger exploration of how these ideas might be translated more concretely and simply for everyday designers to employ when needed. Given the areas for further study and new technologies I outlined above, I would like to continue this research in the future by testing these methods with actual VEs created by peers and classmates, trying out some of the new technologies I think would help and documenting the experience, and creating a visual model or flowchart that could show this process. At the end of the day, it will always be hard to evaluate virtual realities, just as it is difficult to evaluate the usability of our own realities, even when we are in full control of our sensory experience. But as the technology gets better and the virtual gets more real, we will all be able to create, test, and experience a more immersive and magical world on every level.

Works Cited

Bchung. "User Experience @ Pratt." *Usability Evaluation Practices for the HMD VR*. Pratt, 18 Nov. 2015. Web. 5 May 2016. [Link](#).

Bowman, D., Johnson, D., and Hodges, L. *Testbed Evaluation of VE Interaction Techniques. Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (1999): 26-33. Web. 10 May 2016.

Bowman, D. and Hodges, L. *Formalizing the Design, Evaluation, and Application of Interaction Techniques for Immersive Virtual Environments. The Journal of Visual Languages and Computing* 10:1 (1999): 37-53. Web. 10 May 2016.

Bowman, Doug A., Joseph L. Gabbard, and Deborah Hix. "A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods." *Presence: Teleoperators and Virtual Environments* 11.4 (2002): 404-24. Web. 4 May 2016.

Darken, R. P. and Sibert, J. L. *Wayfinding Strategies and Behaviors in Large Virtual Worlds*. In Proceedings of CHI '96: ACM conference on Human Factors in Computing Systems (1996): 142-149. Web. 8 May 2016.

"CAVE Fully Immersive Virtual Reality - Virtual Reality." *CAVE Fully Immersive Virtual Reality*. Virtual Reality Society, 9 Jan. 2016. Web. 5 May 2016. [Link](#).

Gabbard, J.I., D. Hix, and J.e. Swan. "User-centered Design and Evaluation of Virtual Environments." *IEEE Comput. Graph. Appl. IEEE Computer Graphics and Applications* 19.6 (1999): 51-59. Web. 10 May 2016.

"Grand Illusions." *The Economist*. The Economist Newspaper, 29 Apr. 2015. Web. 3 May 2016. [Link](#).

Kaur, K., Maiden, N.A.M., and Sutcliffe, A.g. *Design practice and usability problems with virtual environments*. In: Virtual Reality World '96 conference, Stuttgart, Proceedings (IDG Conferences).

"Reality." *Wikipedia*. Wikimedia Foundation, n.d. Web. 3 May 2016. [Link](#).

Sherman, William R., and Alan B. Craig. *Understanding Virtual Reality: Interface, Application, and Design*. Amsterdam: Morgan Kaufmann, 2003. Print.

Slater, M., Usoh, M., and Steed A.. *Taking Steps: The Influence of a Walking Metaphor on Presence in Virtual Reality*. *ACM Transactions on Computer Human Interaction* 2:3 (1995): 201- 219. Web. 4 May 2016.

Steed, A. and Tromp, J. *Experiences with the Evaluation of CVE Applications*. In Proceedings of Collaborative Virtual Environments (1998).

Sutcliffe, Alistair, and Brian Gault. "Heuristic Evaluation of Virtual Reality Applications." *Interacting with Computers* 16.4 (2004): 831-49. Web. 4 May 2016.

Sutcliffe, Alistair, Brian Gault, and Neil Maiden. "ISRE: Immersive Scenario-based Requirements Engineering with Virtual Prototypes." *Requirements Engineering Requirements Eng* 10.2 (2004): 95-111. Web. 10 May 2016.

Sutcliffe, A. G., and K. Deol Kaur. "Evaluating the Usability of Virtual Reality User Interfaces." *Behaviour & Information Technology* 19.6 (2000): 415-26. Web. 5 May 2016.

Usoh, Martin, Ernest Catena, Sima Arman, and Mel Slater. "Using Presence Questionnaires in Reality." *Presence: Teleoperators and Virtual Environments* 9.5 (2000): 497-503. Web. 7 May 2016.

"Virtual Reality." *Wikipedia*. Wikimedia Foundation, n.d. Web. 1 May 2016. [Link](#).

Witmer, B. G. and Singer, M. J. *Measuring Presence in Virtual Environments: A Presence Questionnaire*. PRESENCE: Teleoperators and Virtual Environments 7:3 (1998): 225-240. Web. 7 May 2016.