Latency and Cybersickness: Impact, Causes, and Measures. A Review

Jan-Philipp Stauffert*, Florian Niebling and Marc Erich Latoschik

Human-Computer Interaction (HCI) Group, Informatik, University of Würzburg, Würzburg, Germany

Latency is a key characteristic inherent to any computer system. Motion-to-Photon (MTP) latency describes the time between the movement of a tracked object and its corresponding movement rendered and depicted by computer-generated images on a graphical output screen. High MTP latency can cause a loss of performance in interactive graphics applications and, even worse, can provoke cybersickness in Virtual Reality (VR) applications. Here, cybersickness can degrade VR experiences or may render the experiences completely unusable. It can confound research findings of an otherwise sound experiment. Latency as a contributing factor to cybersickness needs to be properly understood. Its effects need to be analyzed, its sources need to be identified, good measurement methods need to be developed, and proper counter measures need to be developed in order to reduce potentially harmful impacts of latency on the usability and safety of VR systems. Research shows that latency can exhibit intricate timing patterns with various spiking and periodic behavior. These timing behaviors may vary, yet most are found to provoke cybersickness. Overall, latency can differ drastically between different systems interfering with generalization of measurement results. This review article describes the causes and effects of latency with regard to cybersickness. We report on different existing approaches to measure and report latency. Hence, the article provides readers with the knowledge to understand and report latency for their own applications, evaluations, and experiments. It should also help to measure, identify, and finally control and counteract latency and hence gain confidence into the soundness of empirical data collected by VR exposures. Low latency increases the usability and safety of VR systems.

Keywords: virtual reality, latency, cybersickness, jitter, simulator sickness

1. INTRODUCTION

Cybersickness is a severe problem for the usage and safety of VR technology. It hinders both the broader adoption of VR technology and its overall usability. Cybersickness is closely related to simulator sickness and motion sickness. Early research describes cybersickness as a motion sickness in virtual environments (McCauley and Sharkey, 1992). Cybersickness is usually defined by a set of specific adverse symptoms in combination with the use of certain technologies, such as disorientation, apathy, fatigue, dizziness, headache, increased salivation, dry mouth, difficulty focusing, eye strain, vomiting, stomach awareness, pallor, sweating, and postural instability (LaViola, 2000; Stone Ill, 2017; McHugh, 2019). These symptoms are shared with related definitions of sickness, even though their severity might vary. Stanney et al. (1997) argues that cybersickness is connected to more symptoms in the disorientation cluster of the Simulator Sickness Questionnaire.
(SSQ) (Kennedy et al., 1993) than simulator sickness. The disorientation cluster contains several symptoms which do not all carry the explicit meaning of disorientation. Gavagni et al. (2018) show that motion sickness and cybersickness show the same severity of symptoms in extreme cases. Bockelman and Lingum (2017) distinguish cybersickness from other definitions of sickness by its “cyber” source. We use the term cybersickness to describe sickness with the aforementioned symptoms induced by Virtual Reality or Augmented Reality applications that do not apply external forces on the user. External forces are motion platforms or other motor actuated methods that move a user without the user’s own effort. These VR or AR applications provide stimuli predominately by visual perception.

Chang et al. (2020) review experiments that measure cybersickness. They describe the frequency of use for different subjective measurements. Out of 76 experiments, 58 (≈ 76%) use the SSQ (Kennedy et al., 1993). Less often used questionnaires are the Fast Motion Sickness scale (FMS, 6 experiments ≈ 8%, Keshavarz and Hecht, 2011), a forced-choice question (5 experiments ≈ 6.5%, Chen et al., 2011), the Misery Scale (MISC, 4 experiments ≈ 5%, Bos et al., 2010), the Motion Sickness Assessment Questionnaire (MSAQ, 3 experiments ≈ 4%, Gianaros and Stern, 2010) and the Virtual Environment Performance Assessment Battery (VEPAB, 3 experiments ≈ 4%, Lampton et al., 1994). In contrast, Davis et al. (2014) state that the Pencacola Diagnostic Index (Graybiel et al., 1968) is the “most widely used measure in motion sickness studies” (Davis et al., 2014, p. 6). They state that another widely used questionnaire besides the SSQ is the Nausea Profile (Muth et al., 1996) and further list the Virtual Reality Symptom Questionnaire (Ames et al., 2005). Another questionnaire in use is the Motion Sickness Susceptibility Questionnaire (MSSQ) (Golding, 1998). Here again, it becomes apparent how close cybersickness is to simulator sickness and motion sickness, since questionnaires are often used for multiple sickness definitions. The listed questionnaires are in use for research on cybersickness, but care has to be taken to understand their different usage and purpose. Many, like the SSQ, report on the sickness experienced at the time of answering the questionnaire while others like the MSSQ ask for past experiences to gauge sickness susceptibility that can play into the experience. The Nausea Profile is a scale for measuring nausea due to any cause, not a motion sickness-specific scale, while the MSAQ of the same group targets motion sickness and describes subscales for further differentiating motion sickness aspects.

There are different explanations how cybersickness comes into being and there are multiple factors that influence cybersickness. Explanations for cybersickness often precede the term cybersickness itself. They were created for motion sickness or simulator sickness and then adopted for cybersickness. Rebenitsch and Owen (2016) and LaViola (2000) list and discuss the following theories for cybersickness: the sensory mismatch theory (Reason and Brand, 1975; Oman, 1990), the poison theory (Treibman, 1977), the postural instability theory (Riccio and Stoffregen, 1991) and the rest frame theory (Virre, 1996). Oman (1990) describe their sensory mismatch theory as possibly underlying multiple different sickness definitions such as motion sickness and simulator sickness. Bles et al. (1998) adapt this statement to describe postural stability as underlying multiple different sickness definitions.

Factors that evoke cybersickness are “rendering modes, visual display systems and application design” (Rebenitsch and Owen, 2016, p. 102) as well as hardware-specific factors. Rebenitsch and Owen (2016) describe the former factors but leave hardware-specific factors such as latency open to be discussed in other publications. This review focuses on latency contributions to cybersickness. There are other hardware-specific factors such as tracking accuracy (Chang et al., 2016) that are not covered in this review. Latency describes the processing time incurred by the computer system used for the VR application. VR needs complex hard- and software to deliver the desired experience. Each part in the system contributes to the overall latency by itself and by the effects of its interaction with other parts.

Latency as an inherent property of computer system processing is easily introduced into complex architectures, and as such is subject to many evaluations. There are different angles toward research on latency in virtual environments that mutually influence each other. Effects of latency on cybersickness are found, which necessitate research into measurements and control of latency. Experiments that simulate latency are performed that gather more insight into its effects on cybersickness and user performance. And not least of all, latency in experiments performed in virtual environments needs to be reported in research articles. This review is thus organized as follows: First, we discuss experiments that show that latency contributes to cybersickness. Then, we describe ways to measure latency, which is essential for development of applications with consistent latency behavior. We then show how measured latency is reported in research articles to illustrate latency patterns in experiments.

2. EFFECTS

Table 1 shows an overview of different studies that show that latency contributes to cybersickness. The researchers conducted experiments with latency as the independent variable and cybersickness as the dependent variable. Latency is manipulated to create conditions of different motion to photon latency in the employed systems. For each condition, cybersickness is measured to compare sickness values between the conditions. Researchers measure cybersickness with subjective questionnaires or objective physiological measurements. The most often used questionnaire for the listed papers is the SSQ with six out of 11 papers (Meehan et al., 2003; Moss et al., 2011; St. Pierre et al., 2015; Kinsella et al., 2016; Stauffert et al., 2018; Casereman et al., 2019). Physiological measurements are postural stability or postural sway, heart rate, sweating and galvanic skin response. We list postural stability separate from the other physiological measurements to distinguish the different cases of usage. Frank et al. (1988) list postural stability separate from other physiological measurements. Kawamura and Kijima (2016) only observe postural stability. Postural instability

---

Frontiers in Virtual Reality | www.frontiersin.org 2 November 2020 | Volume 1 | Article 582204
often correlates with visually-induced motion sickness (Riccio and Stoffregen, 1991) and some studies have found it to be predictive of visually-induced motion sickness (Arcioni et al., 2019). Meehan et al. (2003) and Stauffert et al. (2018) only use heart rate and galvanic skin response. Their physiological measurements showed an effect of increased latency on heart rate. Gavagni et al. (2018) argue that forehead sweating is the best physiological indicator for motion sickness which shows the same symptoms as cybersickness in extreme cases. Their rollercoaster experiment only finds minor or moderate effects for heart rate and galvanic skin response. While heart rate may not be the best indicator of latency induced cybersickness, it supports the research that evaluates cybersickness with the SSQ.

Most research for latency and cybersickness tests only the effect of static latency added (Frank et al., 1988; DiZio and Lackner, 2000; Meehan et al., 2003; Moss et al., 2011; Kawamura and Kijima, 2016; Caserman et al., 2019; Palmisano et al., 2019; Kim et al., 2020). They introduce a fixed delay into the system and test different such latencies against each other. This is based on the assumption that most observed latencies are close to one mean latency, for which one fixed added latency per condition is an approximation. This simple latency model consistently shows an increase of latency in the VR simulation, leading to increased cybersickness or a more disturbed stand equilibrium.

Movement itself is important to experience latency induced cybersickness (DiZio and Lackner, 2000). Although, Moss et al. (2011) found no influence of latency in an experiment with a lot of head movement. They report that the head movement itself evoked sickness. It may be that sickness from other sources was stronger than the latency induced sickness thereby masking it. Without movement, the user might not feel the discrepancy between real world and virtual world widened by latency. An increase of head movement can increase cybersickness (Palmisano et al., 2019; Kim et al., 2020). Studies often involve a search task that requires head movement.

Taking into account that latency in measurements often shows irregular spikes, Stauffert et al. (2018) showed that not only uniform but occasional latency spikes provoke cybersickness. St. Pierre et al. (2015) and Kinsella et al. (2016) show that periodic latency like measured by Wu et al. (2013) contributes to cybersickness. They describe latency as consisting of a time-invariant and a periodic part. Periodic latency is described to follow a sine wave. St. Pierre et al. (2015) argues that the sine’s amplitude has more influence than its frequency. Kinsella et al. (2016) observes the opposite.

### 3. MEASURING LATENCY

The contribution of latency to cybersickness necessitates controlling latency in every VR or AR application. High latency and especially latency spikes can often only be detected by measurements, which in turn provide researchers and other developers with indications if and where interventions are needed during the development process. Approaches to measure latency are numerous and distinguish themselves in the amount of instrumentation they need, and which kind of latency they measure. Most approaches measure motion to photon latency, which is the time between a movement of some tracked object, and the effect corresponding to this movement shown on a screen, conveyed by photons emitted from the screen. Different tracked objects can be used to signify movement in the measurement of motion to photon latency, such as Motion Controllers or Head-Mounted Displays (HMD). The employed screens may be computer monitors, mobile phone screens or AR/VR HMD screens. The motion to photon latency is also called end-to-end latency. Table 2 shows an overview of approaches.

Measurements need to compare the time difference between the motion of a tracked object and a resulting response on a screen. The observed motion can be the onset of a motion (Feldstein and Ellis, 2020), special characteristics during a motion such as the peak of acceleration (Friston and Steed, 2014), the end of a motion (Chang et al., 2016) or arrival at a predetermined position (He et al., 2000) or a predetermined motion (Di Luca, 2010). A predetermined motion is usually a sinusoidal movement of a pendulum (Steed, 2008) or the circular movement of a turntable (Swindells et al., 2000). A motion can also be the passing of time in the form of timestamps (Siellhorst et al., 2007; Billeter et al., 2016; Gruen et al., 2020).

The screen shows either a copy of the motion (Roberts et al., 2009) or an encoded version of it (Becher et al., 2018). The system needs to track the tracked object, integrate it into its simulation and show a generated image on the screen (Mine, 1993; Feldstein and Ellis, 2020). The necessary processing time leads to the image on the screen always being delayed in contrast to the original, real motion. Additional steps such as Remote Graphics Rendering (Kämäräinen et al., 2017), or using additional computers to process tracking information, leads to increased latency (Roberts et al., 2009).

A straight forward approach uses a camera to observe both the real and the virtual motion and compare the delay between their chosen motion aspect. The analysis can be done by hand (Liang et al., 1991) or automated (Friston and Steed, 2014). Tracking cameras trade spatial resolution for temporal resolution. High spatial resolution is needed to better capture the real motion, but high temporal resolution is needed to determine a high precision latency value. A way around the dilemma is to fit the mathematical function of the known movement to the tracking data (Steed, 2008). This reduces uncertainty due to restricted resolution.

Camera based measurements do not work well with HMDs, because the lenses distort the image in a way that necessitates them to be very close to the lens. This way, they cannot record the real tracked object any longer. These approaches usually use a computer monitor as the observed screen. Some researchers remove the HMD lenses (Feldstein and Ellis, 2020) or use additional lenses that reverse the distortion (Becher et al., 2018).

An alternative is to observe the real motion separately from its virtual counterpart. The obvious extension is to
use two synchronized cameras (Kijima and Miyajima, 2016b). More often, the real motion is observed by a photodiode that gets covered (Mine, 1993) or a rotary encoder (Seo et al., 2017) that reports the orientation of the platform that the tracked object is placed on. The screen is monitored by one (Pape et al., 2020) or more photodiodes (Becher et al., 2018; Stauffert et al., 2020a). A photodiode has a high temporal resolution but can only measure one brightness value per measurement. The application to measure needs to display its tracking information in brightness levels to use photodiodes.

The chosen method determines how many latency values are measured. Sine fitting (Steed, 2008; Teather et al., 2009; Zhao et al., 2017) and cross correlation (Di Luca, 2010; Kijima and Miyajima, 2016b; Feng et al., 2019) only report one latency for one measurement run. If the latency between an event and its reaction on the screen is measured, the number of latency measurements that can be reported depends on the approach. Some methods need to provoke an event and then wait for the result, before it is possible to measure again (Liang et al., 1991; He et al., 2000; Swindells et al., 2000; Miller and Bishop, 2002; Roberts et al., 2009; Friston and Steed, 2014; Raen and Kjellmo, 2015; Kämäräinen et al., 2017; Seo et al., 2017; Feldstein and Ellis, 2020; Pape et al., 2020). Some approaches allow to measure the latency for every frame shown on the screen (Sielhorst et al., 2007; Papadakis et al., 2011; Wu et al., 2013; Billeter et al., 2016; Kijima and Miyajima, 2016b; Becher et al., 2018; Gruen et al., 2020; Stauffert et al., 2020a). Some approaches that only measure the latency of an event are usable to measure continuously, while others are not. We distinguish methods in Table 2 depending on the reported usage.

4. DESCRIPTION

Looking at the approaches to measure latency, we see that latency is reported in different ways. The reported values are often not comparable, as different papers use different systems

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Research simulating latency that tested for a connection to cybersickness.</th>
</tr>
</thead>
<tbody>
<tr>
<td>System</td>
<td>Task</td>
</tr>
<tr>
<td>Frank et al., 1988</td>
<td>Driving simulator</td>
</tr>
<tr>
<td>Stauffert et al., 2018</td>
<td>HMD Vive</td>
</tr>
<tr>
<td>Kawamura and Kijima, 2016</td>
<td>HMD DK2</td>
</tr>
<tr>
<td>Caserman et al., 2019</td>
<td>HMD Vive Pro full bodytracking</td>
</tr>
<tr>
<td>Moss et al., 2011</td>
<td>HMD No HMD</td>
</tr>
<tr>
<td>Kinsella et al., 2016</td>
<td>HMD</td>
</tr>
<tr>
<td>St. Pierre et al., 2015</td>
<td>HMD</td>
</tr>
<tr>
<td>DiZio and Lackner, 2000</td>
<td>HMD</td>
</tr>
<tr>
<td>Meehan et al., 2003</td>
<td>HMD</td>
</tr>
<tr>
<td>Palmisano et al., 2019</td>
<td>HMD</td>
</tr>
<tr>
<td>Kim et al., 2020</td>
<td>HMD</td>
</tr>
</tbody>
</table>

Frontiers in Virtual Reality | www.frontiersin.org
November 2020 | Volume 1 | Article 582204
with varying complexity. A less complex system is expected to show lower and more deterministic latency than a more complex system. Newer hardware often has lower latency with reduced determinism (McKenney, 2008). Some papers report multiple measurements of different systems. Table 3 lists only a subset of the numbers reported in the respective research papers. Interested readers are referred to the original publications.

An observation is that latency is not a constant value. Latency is different with different devices (Mine, 1993), different software configurations (Friston and Steed, 2014) or different input methods (Kämäräinen et al., 2017). Different usage patterns such as a change of the movement direction can influence latency (He et al., 2000). Even small changes in the measurement setup can make a difference. Latency measured in the upper part of a screen can be lower than latency measured in the lower part, due to the scan out sequence (Papadakis et al., 2011). The problem with latency measurements is that they are often performed “under optimized and artificial conditions that may not represent latency conditions in realistic application-oriented scenarios” (Feldstein and Ellis, 2020).

### Table 2: Comparison of latency measurement approaches.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Motion</th>
<th>Photon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device</td>
<td>Capture</td>
<td>Device</td>
</tr>
<tr>
<td>Becher et al., 2018</td>
<td>HMD</td>
<td>Rotary encoder</td>
</tr>
<tr>
<td>Di Luca, 2010</td>
<td>Tracked object</td>
<td>Photodiode</td>
</tr>
<tr>
<td>Billeter et al., 2016</td>
<td>LED timestamp</td>
<td>Camera</td>
</tr>
<tr>
<td>Feldstein and Ellis, 2020</td>
<td>HMD</td>
<td>Camera</td>
</tr>
<tr>
<td>Feng et al., 2019</td>
<td>HMD</td>
<td>Camera</td>
</tr>
<tr>
<td>Friston and Steed, 2014</td>
<td>Mouse</td>
<td>Camera</td>
</tr>
<tr>
<td>Gruen et al., 2020</td>
<td>Sub millisecond clock</td>
<td>Camera</td>
</tr>
<tr>
<td>He et al., 2000</td>
<td>Wand</td>
<td>Camera</td>
</tr>
<tr>
<td>Kämäräinen et al., 2017</td>
<td>Touch</td>
<td>Switch</td>
</tr>
<tr>
<td>Kijima and Miyajima, 2016a</td>
<td>HMD</td>
<td>Camera</td>
</tr>
<tr>
<td>Kijima and Miyajima, 2016b</td>
<td>HMD</td>
<td>Camera</td>
</tr>
<tr>
<td>Liang et al., 1991</td>
<td>Pendulum</td>
<td>Camera</td>
</tr>
<tr>
<td>Miller and Bishop, 2002</td>
<td>HMD</td>
<td>CCD array</td>
</tr>
<tr>
<td>Mine, 1993</td>
<td>Pendulum</td>
<td>Photodiode</td>
</tr>
<tr>
<td>Papadakis et al., 2011</td>
<td>Tracked object</td>
<td>Rotary encoder</td>
</tr>
<tr>
<td>Pape et al., 2020</td>
<td>Rigid body</td>
<td>Switch</td>
</tr>
<tr>
<td>Raaen and Kjelmo, 2015</td>
<td>HMD</td>
<td>Photodiode</td>
</tr>
<tr>
<td>Roberts et al., 2009</td>
<td>Hand</td>
<td>Camera</td>
</tr>
<tr>
<td>Seo et al., 2017</td>
<td>HMD</td>
<td>Rotary encoder</td>
</tr>
<tr>
<td>Sielhorst et al., 2007</td>
<td>Timestamps</td>
<td>Camera</td>
</tr>
<tr>
<td>Stauffert et al., 2020a</td>
<td>Tracked object</td>
<td>Motor driver</td>
</tr>
<tr>
<td>Steed, 2008</td>
<td>Pendulum</td>
<td>Camera</td>
</tr>
<tr>
<td>Swindells et al., 2000</td>
<td>Turntable</td>
<td>Camera</td>
</tr>
<tr>
<td>Teather et al., 2009</td>
<td>Tracked object</td>
<td>Camera</td>
</tr>
<tr>
<td>Wu et al., 2013</td>
<td>Manually Moved Bar</td>
<td>Camera</td>
</tr>
<tr>
<td>Zhao et al., 2017</td>
<td>HMD</td>
<td>Potentiometer</td>
</tr>
</tbody>
</table>

Camera based measurements has a camera that observes both the real tracked object and its virtual counterpart. Photodiode based measurements read the encoded information off a screen with a photodiode. The observation of the real object is done with a different sensor. Motion to Photon latency measurements use different devices where the motion originates from and which kind of screen emits the photon. The methods column describe how often it is possible to measure latency.

Frontiers in Virtual Reality | www.frontiersin.org 5 November 2020 | Volume 1 | Article 582204
The variability is usually reported by a mean value at least. Standard deviation and minimum/maximum values provide more insight. Histograms can be used to show even more information about what latencies are to be expected. We want to focus on these visualization methods here as a basis to understand the connection between latency and cybersickness. The different ways to describe cybersickness are used in the different simulated latencies of the cybersickness experiments of Table 1.

The sparklines in Table 3 give an impression of the shape of latency. The data is stretched to take the maximum amount in x and y direction and only shows the x axis segment that contains data. Sparklines are supposed to only give a general idea of the shape (Tufte, 2001). Stauffert et al. (2016) and Stauffert et al. (2020a) use a logarithmic y axis. The other papers use a linear y axis. Every sparkline has the measured latency in x direction and its probability in y direction. We exclude Stauffert et al. (2020a) systems where there is artificial latency introduced, but include systems that have artificially high system load but mimic real world scenarios.

A key difference between representations given in publications is if they include rare outliers. Some researchers show no outliers (Wu et al., 2013; Pape et al., 2020) while others do (Sielhorst et al., 2007; Stauffert et al., 2016, 2020a). Latencies usually cluster around one or multiple values. Wu et al. (2013) system 2 and Stauffert et al. (2020a) system 1 show one cluster. Pape et al. (2020) and Sielhorst et al. (2007) system 1 and 3, Wu et al. (2013) system 1 and Stauffert et al. (2016) show two clusters. Sielhorst et al. (2007) system 2 shows 3 clusters and Stauffert et al. (2020a) system 2 shows 9 clusters, each indicated by higher probabilities surrounded by lower probabilities in the histogram.

Each cluster’s distribution appears to follow a normal distribution though Sielhorst et al. (2007) system 1, Stauffert et al. (2016) and Stauffert et al. (2020a) system 2 show a more skewed distribution with a longer tail toward larger latencies, resembling more a gamma distribution. Pape et al. (2020) proposes to describe the distribution with a gaussian mixture model, i.e., an imposition of multiple normal distributions. Stauffert et al. (2018) argue to use an empirical distribution derived from the measurements. Multiple clusters presumably originate from the interplay of two or more parts running in decoupled loops in the observed system. Feldstein and Ellis (2020) list processing stages such as simulation or rendering that contribute to the final latency pattern with their runtime and communication behavior.

Besides the general distribution, there may be temporal patterns. Stauffert et al. (2020a) found reoccurring latency spikes with a uniform interarrival time. Wu et al. (2013) found a sinusoidal latency pattern.

### 5. DISCUSSION

We have shown how latency is measured. The necessary instrumentation varies from simple observations of the VR equipment (Steed, 2008), to the need of specific software to run (Friston and Steed, 2014), to required modifications of the hardware (Stauffert et al., 2020a). The motion may be evoked manually (Wu et al., 2013) or with a pendulum (Mine, 1993) or a turntable (Chang et al., 2016). Latency is observed from one distant observer with one camera (He et al., 2000), multiple distant observers with synchronized cameras (Gruen et al., 2020)
or close observers that are attached to the moved device and the screen (Di Luca, 2010).

Most researchers that measure latency report a mean latency value with an optional standard deviation. Some report a minimum and maximum value in addition. More insight is provided by histograms and plots showing the temporal behavior (Wu et al., 2013). There is research into whether latency influences cybersickness. Most compare the effect of one latency condition with another condition that has a time invariant increased latency (Frank et al., 1988; DiZio and Lackner, 2000; Meehan et al., 2003; Moss et al., 2011; Kawamura and Kijima, 2016; Caserman et al., 2019; Palmisano et al., 2019; Kim et al., 2020). This is based on the most often reported mean latency. Latency jitter as described in latency histograms and periodic latency patterns are shown to also contribute to cybersickness (Stauffert et al., 2018). All approaches to report latency find a counterpart where latency is simulated and shown to influence cybersickness.

There is more research into latency for VR systems than for AR systems, mainly because the technology is often times easier to handle. Many AR systems are simulated with VR systems until AR technology makes the simulated features possible. While less researched, AR systems show similar problems (Sielhorst et al., 2007).

### 5.1. Limitations on Latency Comparability

There are many factors that can influence latency and the predictability. Kijima and Miyajima (2016a) show that HMD prediction and timewarp (van Waveren, 2016) make a difference. Asynchronous timewarp uses a shortcut to update the displayed image after it was rendered, which yields different values when measured to a system that looks at motion controller movement that is only updated in the simulation of the virtual world. A sequential scan-out process leads to the eyes getting the information at different points in time so it can make a difference which screen is taken for measurement (Papadakis et al., 2011). He et al. (2000) found different latency depending on the movement direction of the tracked object. Manufacturers optimize latency with prediction that may fail (Gach, 2019).

Latency reporting depends on the observed system. The values in Table 3 are not comparable to one another because some do not measure certain stages of computation or use other hardware. Even though the values are not comparable, they are often reported in a similar fashion with one mean value and a standard deviation.

Spatial jitter can be similar to latency jitter by offsetting tracking positions in an unexpected way. Some measurement methods can not distinguish between latency jitter and spatial jitter by their design. 2D pointing performance suffers with spatial jitter (Teather et al., 2009). Spatial jitter is likely to evoke cybersickness as well and may partially be described in the latency jitter studies already. Some measurement methods measuring related phenomena further complicates the comparison.

### 5.2. Latency Variability

VR and AR applications require substantial computational power to create virtual environments. Computer systems to provide the experience are optimized for performance rather than real-time, i.e., guaranteed response times (Mckenney, 2008). Some applications such as robotics and space exploration require such deterministic runtime behavior of software. Modern operating systems do not provide real-time capabilities and even the Linux PREEMPT_RT patches cannot provide reliable real-time runtimes (Mayer, 2020). Without a real-time operating system, there may be unforeseeable latency spikes that can harm VR experiences, even if latency was previously acceptable.

Researchers agree that “the delays vary substantially” (Kämäräinen et al., 2017) and often try to “illustrate the variations in latency of real systems” (Friston and Steed, 2014) by reporting more than one mean latency value. As a caveat, the “latency testing on isolated virtual reality systems under optimized and artificial conditions may not represent latency conditions in realistic application-oriented scenarios” (Feldstein and Ellis, 2020). Care must be taken to measure as close to the use case as possible to best represent the expected latencies. The best case would be to measure during exposure.

Rare latency outliers show latencies much larger than the average (Stauffert et al., 2020a). Networked applications often only look at the 95th, 99th, and 99.9th percentile (Vulimiri et al., 2013) to estimate response times. Teather et al. (2009) use the 95th percentile to describe their motion-to-photon latency measurements. Stauffert et al. (2018) provide a first study with latency spiking behavior including the top one percent but more research is needed to understand if regarding only the 95th or 99th percentile is sufficient. Some web applications found the need to include the remaining one percent of latencies in their analyses (Hsu, 2015).

Latency jitter can be reduced with prediction (Jung et al., 2000). Incorporating latency jitter in the prediction model increases the prediction performance (Tumanov et al., 2007). Prediction, however, introduces its own side effects such as over anticipation (Nancel et al., 2016).

### 5.3. Desirable Latency Values

How much latency is tolerable for a good VR experience? Carmack (2013) says that it should be below 50 ms to feel responsive and recommends less than 20 ms. Attig et al. (2017) look at HCI experiments without VR that report no impact on usability when latency is below 100 ms. Humans can detect visual variations at 500 Hz (Davis et al., 2015) and latency below 17 ms (Ellis et al., 1999, 2004; Adelstein et al., 2003). Although, Feldstein and Ellis (2020) indicate that perceivable latency does not necessarily cause cybersickness. Jerald (2010) measures a minimum latency threshold of 3.2 ms in one of the participants, but adds that the exact perceivable latency may depend on the virtual environment.

### 5.4. Need to Measure Latency

Measuring latency helps to become aware of bottlenecks in employed hard- and software (Swindells et al., 2000; Di Luca, 2010). Without measuring, those problems may never be detected and may influence an otherwise sound experiment. Many researchers, however, do not report latency. The 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)
saw 104 published papers. 85 papers conducted a user study in virtual reality. Only 6 reported the latency of the employed VR system. Although a reported mean latency strengthens trust that the systems performed as expected, latency jitter may still have occurred during experiments and may have impaired individual measurements.

Which approach to use depends on the application and possibilities of the researchers. A detailed analysis helps to judge the application’s performance but everything is better than not measuring at all. Every researcher should be able to do manual frame counting (He et al., 2000) as shown in Feldstein and Ellis (2020) that compare the results of different evaluators. Sine fitting (Steed, 2008) reduces imprecisions in the video analysis. Even though it is more involved than manual frame counting, software can help with this analysis (Stauffert et al., 2020b). Beyond these basic approaches, the choice of how to measure latency depends on the specific hard- and software used. Design your measurement system to fit your VR system guided by the approaches in Table 2. Research should strive toward measuring latency for every frame shown on the employed screen to assure validity of observations and to maximize insight. Measuring latency can hint at problems, latency values then have to be interpreted to find an intervention if need be.

6. CONCLUSION

Latency is one of the characteristics of a computer system that is often discussed to have a major impact on the system’s usability. Research shows that larger latencies and latency jitter can influence well-being in a negative way in the form of cybersickness. Yet little research of VR experiences check and report the latency behavior of their employed computer system. Only 7% of the papers published at the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) conducting user studies in virtual reality reported their motion to photon latency. Latency may introduce unwanted effects that are not obvious to the researchers and reviewers if a latency value is not reported.

Latency is not restricted to one value but changes over time and with the VR system usage pattern. More elaborated test setups are required to capture these dynamics. Research is only beginning to understand the implications of time-invariant latency. Even the occasional latency spike will contribute to cybersickness. Measuring latency is of importance to understand better the influence on cybersickness and to understand where latency might not be the main cause for cybersickness.

AUTHOR CONTRIBUTIONS

J-PS conducted the literature review and took the lead in writing the manuscript, he collectively discussed, and developed concepts to measure and report latency. FN worked on the manuscript and supervised the project. ML conceived the original idea, collectively discussed and developed concepts of own research on latency, and supervised the project. All authors provided critical feedback and helped shape the research, analysis, and manuscript.

REFERENCES


Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Stauffert, Niebling and Latoschik. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.